

Retours sur la *Sunbelt 2018* : réflexions et perspectives sur l'analyse des réseaux sociaux

Maxime Cornet¹
EHESS

Soumis le 01/10/2018, mis en ligne le 30/10/2018

Type de soumission

Actualité.

Titre anglais

The 2018 Sunbelt Conference: Reflections and Perspectives on Social Networks Analysis

Mots clés

Conférence *Sunbelt*

Keywords

Sunbelt Conference

Introduction

Alors que la 38^{ème} édition de la *Sunbelt conference* se termine sous le soleil estival néerlandais, il est temps de revenir sur ces six jours riches en discussions, échanges, et en développements méthodologiques. Cette année, la foule était au rendez-vous pour se réunir sous le dôme de la *St Martin's Cathedral*, et entre les murs des bâtiments historiques de l'Université d'Utrecht.

Si l'on doit retenir une chose des présentations et discussions entourant la conférence, c'est bien la diversité des intervenant.e.s, disciplines et approches représentées en ce mois de juin 2018 à Utrecht. De fait, l'analyse de réseaux sociaux apparaît comme éminemment pluridisciplinaire, et les thématiques abordées aussi variées que les disciplines qui les mobilisent. Des sessions méthodologiques et techniques présentant les dernières innovations statistiques et conceptuelles, aux questionnements liés à l'emploi de l'analyse des réseaux dans les champs de la recherche historique et archéologique, en passant par les sessions thématiques présentant les

¹ Auteur : maxime.cornet@zoho.com

approches variées autour d'un objet spécifique : les communautés en ligne, les réseaux de collaboration scientifiques, les réseaux d'entreprises, et bien d'autres. Les échanges particulièrement riches amenés par cette confrontation d'idées et de points de vue divers amène à mettre en perspective un certain nombre de concepts méthodologiques et épistémologiques fondamentaux au cœur de l'analyse des réseaux sociaux.

En effet, un certain nombre de questionnements récurrents égrenaient la conférence, et l'on pouvait les retrouver en fil rouge, sous des formes différentes dans nombre de sessions. Des questionnements profonds, mais essentiels, et sur lesquels tou.te.s les intervenant.e.s ont livrés leur point de vue. De la définition de l'acteur, à la mise en relation simultanés de dimensions multiples dans l'analyse. De la construction de l'espace à la prise en compte de la temporalité dans laquelle s'inscrivent les liens. De la mise en place du terrain, à l'usage des sources. Tous ces questionnements surviennent d'une manière ou d'une autre dès l'instant où l'on aborde un objet par l'analyse des réseaux sociaux, mais la confrontation de visions issues de disciplines et champs divers permet de les redéfinir et de les mettre en discussion de manière intéressante.

L'une de ces interrogations, en somme transversale à un bon nombre de ces problématiques, concerne les données manquantes, leur définition, et leur traitement. Il apparaît que le problème ne s'aborde pas de la même manière selon la nature du réseau, de la discipline et du terrain mobilisable.

L'origine des données et la nature des liens : mettre l'histoire en réseau

Si certains questionnements sont pluridisciplinaires, la manière dont ils sont abordés ne peut être détachée de la discipline qui les traite. Ainsi, à Utrecht, durant trois sessions consacrées spécifiquement aux *défis posés par l'analyse des réseaux historiques et archéologiques*, se sont posées les questions de l'analyse de la trace, de la temporalité dans laquelle s'inscrivent les réseaux, de la nature des acteurs, et surtout, de la nature des sources : par qui sont-elles rapportées, comment, et dans quel contexte ? Toutes ces questions sont primordiales quelle que soit la discipline lorsque l'on aborde la question des réseaux. Mais elles sont particulièrement mises en avant dans le cas de réseaux historique et archéologiques, d'autant que ces disciplines ont l'habitude de traiter des questions de contextualisation des sources et d'inscription temporelle et spatiale des événements. Les

sujets sont divers, et l'analyse des réseaux sociaux s'emploie pour traiter de sujets variés :

- l'émergence d'une économie de marché « moderne » à Naples sur fond d'unification Italienne pour Schisani, Vitale, Giordano et Primerano². Les auteurs usent ici d'une base de données multi sources, composée de tables inter-reliées décrivant les acteurs économiques et socio-politiques d'Italie du Sud lors de la « première mondialisation » : l'*IFESMEZ database*³. Les auteurs disposent ainsi d'un réseau de 753 firmes opérant à Naples au début du XIX^e siècle dont les liens sont attestés par des sources multiples. Ils peuvent ainsi mettre en évidence l'émergence d'un marché capitaliste à Naples au grès des crises économiques secouant l'Italie et l'Europe tout au long du XIX^e siècle. *Via* diverses mesures de centralité, ils mettent en avant la prévalence de l'homophilie jusqu'en 1880, puis l'effondrement des firmes très interconnectées et l'émergence de structures financières hétérogènes ;

- l'évolution du marché de l'emprunt dans une petite ville Allemande au début du XIX^e siècle pour Stark et Bixler⁴. Les auteurs usent de jeux de données longitudinaux, extraits de livres de prêts datant de la période, et sélectionnent cinq coupes espacées d'une année chacune et s'échelonnant autour de la réforme du prêt de 1825 en Westphalie. L'idée étant de prendre en compte la dimension dynamique des transformations observées, les auteurs usent des Modèles Stochastiques Orientés Acteurs (SAOM). Ils observent une baisse de l'effet de saint Matthieu au fur et à mesure de la modernisation du marché de l'emprunt : les liens interpersonnels jouent progressivement un moindre rôle, au fur et à mesure que le marché s'institutionnalise. On observe la transition d'un marché médiéval sur-encasté allant vers un marché plus moderne ;

- la répression de l'opposition politique dans la première république Autrichienne pour Petz, Wenninger, Tzaferis et Pfeffe⁵. Les auteurs étudient un réseau judiciaire constitué à partir des archives du tribunal pénal de Vienne. Ils observent l'évolution des condamnations en

² Schisani, Vitale, Giordano & Primerano, 2018, « Analysing structural changes in corporate networks over the Italian political unification. The case of the business Neapolitean area ».

³ <http://www.ifesmez.unina.it/pa/> (consulté le 28/09/2018).

⁴ Stark, Bixler, 2018, « Modeling the dynamics of a rural credit market in the 19th century ».

⁵ Petz, Wenninger, Tzaferis & Pfeffe, 2018, « Networks of political prosecution in the consolidated dictatorship of the first Republic of Austria in 1935 ».

fonction de l'appartenance politique des suspects, après une série de réformes judiciaires donnant à l'administration policière un pouvoir accru, et introduisant une nouvelle série d'interdits (illégalité des partis d'opposition, etc.). En définitive, l'appartenance politique des individus ne semble pas avoir d'influence sur leur condamnation, On observe en revanche un taux extrêmement important de condamnés parmi les prévenus.

- les relations entre la police montée Canadienne et ses auxiliaires Aborigènes pour Ayala⁶. L'auteur étudie la création du corps de la police montée, originellement mis en place pour contrôler les territoires sauvages de l'Ouest du Canada et les populations qui s'y trouvaient. Ces territoires ne disposant d'aucune infrastructure et étant mal cartographiés, la police montée s'entoure rapidement d'un nombre en rapide expansion d'auxiliaire Aborigènes. Les archives de la police montée, ainsi que les journaux tenus par ces officiers permettent de retracer partiellement les relations qui les unissent à leurs auxiliaires, les mariages qui surviennent, les jeux d'influence et de pouvoir, et ainsi, de construire une ébauche ethnographique.

Toutes ces présentations traitent d'objets variés, et usent de l'analyse des réseaux dans des objectifs différents : révéler des évolutions structurales dans les marchés financiers suite à la survenue d'un événement particulier pour Schisani et Bixler, mettre en avant les mécanismes de répressions suivant un coup d'état pour Petz, ou dresser un état des lieux des relations sociales entre colons et natifs pour Ayala. Il demeure que tous ces auteurs abordent à un moment ou à un autre la question des sources et de la construction du lien dans leur présentation. La question du manque de traces et du contexte de celles qui ont été conservées se pose à tout.e historien.ne, quel que soit l'angle d'approche retenu pour traiter son objet. Cependant, dans le cas d'une analyse de réseau sociaux, la mobilisation d'archives omettant, délibérément ou non (pour des raisons politiques, parce que l'évènement est jugé mineur, dû à la destruction ciblée ou non de documents, à la perte de documents thématiques, etc.) peut profondément affecter la nature et la structure du réseau. L'observation est encore plus apparente dans les cas de Petz et d'Ayala : les archives du tribunal sont-elles exactes ? Que se passe-t-il en amont de la comparution

⁶ Ayala, 2018, « Social networks as historical research tools: the case of the Canadian mounted police and its aboriginal auxiliaries (1874-1900).

devant le tribunal ? etc. La nature politique de la répression peut éventuellement s'exprimer dans un contexte autre que le cadre réduit du tribunal de Vienne. Le constat exprimé par Ayala est encore plus dur, comment mesurer les biais introduits dans ses données par le fait que les seules traces disponibles couvrant la période et l'espace étudié proviennent des archives et des mémoires de colons. Les descriptions des Aborigènes y sont normatives et les données les concernant sont souvent vagues. Cette observation soulève la question du traitement des données manquantes en analyse de réseau, manque d'autant plus criant que le terrain est lointain et l'information souvent parcellaire. Néanmoins, les observations soulevées au cours de cette session valent pour l'ensemble des disciplines mobilisant l'analyse de réseau. La contextualisation du matériel récolté et la conscience des limites des données (de quel point de vue sont-elles émises, et surtout qui sont ceux absent de l'analyse ?) est primordiale.

Le terrain et le réseau : des réponses variées selon les contraintes

Le traitement des données manquantes dans les réseaux ne préoccupent pas que les historien.ne.s ou les archéologues, qui ont par définition un accès limité au terrain de par le caractère limité des traces subsistantes de celui-ci, mais aussi l'ensemble des autres disciplines qui mobilisent ces techniques. Pour les disciplines qui disposent souvent d'un accès direct au terrain, comme la sociologie ou l'économie, une partie de la réponse vient sans aucun doute de la phase de collecte des données : comment minimiser les biais, inclure le plus d'individus concernés par l'objet dans l'étude, et avoir la meilleure idée possible de la position des individus que l'on ne peut pas interroger ? De nombreux intervenant.e.s se sont exprimé.e.s sur la question de la collecte des données de réseau :

- Stys⁷, dans son étude sur le leadership et la distribution des services en République Démocratique du Congo, met en place une association entre sociogramme, cartographie géopolitique, et données d'archives (recueils financiers et transactionnels, etc.). En interrogeant des individus issus de milieux divers (étrangers/natifs) évoluant au sein d'environnements variés : l'auteur procède à une collecte échelonnée en fonction de zones identifiées grâce au croisement d'un ensemble de cartes : zones

⁷ Stys, 2018, « Take me to your leader: Employing mixed methods and SNA in the study of public authority and governance in conflict-affected states ».

d'influence des différents groupes armés, territoire urbain/rural, ressources disponibles, etc. L'idée étant de saisir la distribution inégale de l'accès aux services, et le type d'entité qui fournit ceux-ci (gouvernement, groupes armés, communauté locale, etc.) en fonction du contexte.

- Hogan, Janulis, Philips, Melville, Mustanski, Contractor, Birkett⁸, s'interrogent eux sur les innovations possibles à apporter au sociogramme : à partir de données longitudinales issues du sous échantillon RADAR de la cohorte Américaine YMSM, ils tentent de comprendre les évolutions observées dans les réseaux collectés *via* les sociogrammes : qui sont les acteurs qui tendent à disparaître et dans quelle mesure peut-on corriger le biais introduit *via* les outils de mesures auto-rapportées ? Ils se rendent compte que les méthodes de collecte interactives (*OpenEddy* ou *Gensi*) sont plus efficaces que les sociogrammes papiers traditionnels : les participant.e.s voient leur réseau de dessiner au fur et à mesure qu'elles et ils ajoutent des alter au réseau. On observe de fait une limite de l'effet de fatigue de l'enquêté.e qui a moins tendance à sous déclarer ses liens dans les vagues avancées de l'enquête.

Ces deux approches sont loin d'être les seules innovations dans la collecte des données de réseau présentées à la conférence. On pourrait aussi citer, entre autres, Lizardo, Hachen, Chandler et Wang⁹, qui usent de données collectées *via* une application pour smartphones qui extrait automatiquement les journaux d'appels et l'historique des messages des enquêté.e.s, à but de comparaison avec les sociogrammes collectés de manière plus traditionnelle. L'aspect important à retenir de ces présentations est que la méthodologie d'enquête revêt une importance particulière, et que ces questions doivent être abordées en amont, si l'on veut éviter les biais de structure introduits par les nœuds ou les arrêtes manquantes.

Le traitement des données manquantes : innovations statistiques

Si des solutions existent pour tenter de minimiser les biais introduits lors de la collecte des données, il existe aussi un ensemble de techniques statistiques permettant de gérer *a posteriori* les données manquantes.

⁸ Hogan, Janulis, Philips, Melville, Mustanski, Contractor, Birkett, 2018, « Examining changes in alter nomination over time: results from a longitudinal cohort study ».

⁹ Lizardo, Hachen, Chandler et Wang, 2018, « Examining the empirical linkages between different kinds of ties in dynamic ego networks ».

Ainsi, une session était consacrée à ces questions, *Données manquantes : effets et traitements*.

Smith, Moody et Morgan¹⁰ tentent, sur un ensemble de réseaux simulés à partir de paramètres variables, de comprendre l'effet des données manquantes en fonction des caractéristiques du réseau. On sait que dans les réseaux auto-rapportés (constitués à partir de générateurs de noms, etc.), les acteurs les plus centraux ont une probabilité plus importante d'être manquants. En prenant en compte ce fait dans les simulations, il semble que les données manquantes ont un effet particulièrement important sur la transitivité et la centralité, et un effet moins important sur la densité. Il semble de plus qu'en règle générale (excepté pour les mesures dépendant de la réciprocité des liens), l'imputation donne de meilleurs résultats que le jeu de données partiel.

À partir de cette observation, quelques auteurs proposent des méthodes d'imputation. Krause, Huiser et Snijders¹¹ proposent par exemple, pour les données longitudinales, une méthode d'imputation basée sur les modèles stochastiques orientés acteurs (en effectuant une simulation de la trajectoire du réseau à partir des paramètres du SAOM).

L'approche statistique du problème est bien entendu, compte tenu de la discipline et des données mobilisées (qui sont bien souvent simulées) très structuraliste. Il reste que les observations soulevées sont intéressantes : les données manquantes ne surviennent pas de manière aléatoire dans le réseau, et l'effet de ces données sur la structure observée est à prendre en compte. L'imputation des données manquantes est aussi un angle intéressant à étudier. Il n'empêche que ces techniques montrent aussi leurs limites. L'approche présentée plus haut ne fonctionne, par exemple, qu'avec des données longitudinales. De plus, ces méthodes ne proposent une solution acceptable que du point de vue des métriques structurelles globales (elles ne permettent pas d'étude individuelle de la position de nœuds spécifiques), et surtout, elles ne résolvent pas (ou pas totalement) les biais introduits lors de la collecte des données : elles infèrent un résultat à partir des données disponibles.

On peut retenir de l'ensemble de ces présentations qu'il n'existe en définitive pas de solution universelle au problème des données manquantes en analyse des réseaux sociaux. Il est cependant, à mon sens, nécessaire

¹⁰ Smith, Moody, Morgan, 2018, « Network sampling coverage III: Imputation of missing network data under different network and missing data condition ».

¹¹ Krause, Huiser, Snijders, 2018, « Multiple Imputation for missing data in longitudinal network models ».

de l'avoir en tête à chaque étape du processus scientifique : lors de la construction de l'objet, pour bien définir les groupes concernés par l'étude, et le type de lien, lors de la collecte des données pour veiller à n'exclure aucun de ces groupes, lors de l'analyse des données, car il est toujours difficile de recueillir des jeux de données complets, et que nous l'avons vu, les données manquantes ont une influence qu'il convient de garder en tête sur les métriques décrivant la structure du réseau.

Plus largement, si l'on dépasse la thématique des données manquantes pour revenir au cadre plus vaste de la conférence, il apparaît que le dialogue pluridisciplinaire autour de l'analyse des réseaux sociaux amène un ensemble de questionnements forts sur les principes méthodologiques traditionnels de celle-ci. La diversité des champs d'application pousse à inventer de nouvelles manières de la mettre en œuvre, rendant cette 38^e édition de la *Sunbelt* particulièrement riche.