

Sur un court séjour en terrain disciplinaire étranger

Conférence MARAMI 2019, Dijon

Novembre 2019

Titre anglais

On a short stay on foreign disciplinary field. MARAMI Conference 2019, Dijon

Mots clés

MARAMI, Informatique, Communautés, Graphes dynamiques

Key words

MARAMI, Computer Science, Communities, Dynamic Graphs

Introduction

La précédente conférence MARAMI (Modèles & Analyse des Réseaux : Approches Mathématiques & Informatiques) à Avignon faisait la part belle aux sciences sociales, la dixième édition qui vient de se tenir à l'IUT de Bourgogne¹ était, à une poignée d'exceptions près, mono-disciplinaire et peu ouverte à d'autres disciplines que l'informatique. L'auteur de ces lignes, géographe de formation, a donc eu la chance de participer à sa première conférence hors sciences sociales.

Se rendre à une conférence d'une discipline autre que la sienne est sans doute l'un des moyens les moins coûteux - en énergie, pas nécessairement au niveau financier vu les tarifs de certaines conférences - pour avoir une idée globale du paysage actuel de la recherche dans la discipline en question. Le fait que cette édition se soit tenue en anglais et qu'elle ait attiré au-delà des seuls cercles francophones ajoutait une plus-value certaine à la conférence.

1. <https://iutdijon.u-bourgogne.fr/marami/>

Le lecteur ou la lectrice ne trouvera pas ici un résumé de l'ensemble des communications présentées, certaines sont accessibles en ligne sur les sites personnels des auteur.e.s (ou plus souvent leur git²), d'autres devraient donner lieu à des articles dans *Social Network Science* dans les mois à venir. Il s'agira plutôt ici de mettre en lumière les questions de recherche dominantes dans cette communauté et de pointer certaines particularités pouvant sembler exotiques lorsqu'on les observe depuis une autre discipline. Il importe de préciser que, les journées étant longues, les présentations (20 minutes questions comprises, les keynote speakers ayant droit à 40 minutes) s'enchaînent à un rythme soutenu, je n'ai pas nécessairement perçu l'intérêt et la richesse de l'ensemble des communications présentées lors de cette conférence.

1 Graphes dynamiques, graphes multicouches et communautés

Trois grandes questions de recherche m'ont semblé dominer cette conférence : l'étude des graphes « multicouches³ », l'étude des graphes dynamiques et la détection de communautés, une supplémentaire et découlant des deux précédentes étant la détection de communautés dans les graphes dynamiques. Il peut paraître surprenant que la détection de communautés soit encore un enjeu tant certaines méthodes et certains algorithmes semblent s'être imposés de façon parfois routinière ces dernières années (algorithme de Louvain couplé à la mesure de la modularité pour apprécier la qualité de la partition). Le fait que certains logiciels ne proposent que cette option explique sans doute, en tout cas dans les disciplines où les chercheurs et chercheuses n'écrivent pas leurs propres programmes (soit l'immense majorité en sciences sociales), cette apparente domination. Or, ce champ reste dynamique et différentes mesures alternatives permettant d'apprécier la qualité des partitions furent présentées.

En ce qui concerne les graphes dynamiques, plusieurs approches coexistent sans qu'il soit toujours possible de savoir si elles parviennent à dialoguer les unes avec les autres : *graph stream*, succession d'instantanés, variations continues avec fenêtres variables d'agrégation des données. L'un des champs de recherche sécant aux précédents concerne la génération de données permettant d'étudier ces graphes dynamiques. La rareté de jeux de données de grande taille permettant ces études a été fréquemment évoquée lors des communications. Les détails techniques m'ont parfois échappé, brièveté des présentations oblige, et il n'est pas certain que ces méthodes

2. Plateforme permettant de déposer et de partager des données, du code, des présentations, etc.

3. Les auteur.e.s utilisaient le terme *multilayer* qui désignait selon les cas un graphe multiplexe ou un graphe multi-niveaux.

soient implémentées dans un futur proche dans les logiciels généralement utilisés en SHS - évoquer R était le meilleur moyen de provoquer des grimaces chez mes collègues informaticien.ne.s...

2 Particularités et angles morts

2.1 Explorer, visualiser

Un des participants étrangers a confié lors d'une pause être surpris du faible nombre d'outils présentés lors de cette conférence, qu'il s'agisse d'outils d'exploration et/ou de visualisation. Seules trois présentations - dont celle du troisième *keynote speaker*, Jean-Daniel Fekete - étaient consacrées à cet aspect. Le prototype en ligne présenté par Stefan Bornhofen et Marten Düring et permettant l'exploration de graphes multiplexes et dynamiques semble prometteur⁴ mais cela reste un prototype en ligne de plus (peu de documentations, bugs, pérennité très aléatoire). La méthode proposée par Sergey Kirgizov et Éric Leclercq pour mettre en évidence l'évolution des communautés dans les graphes dynamiques est stimulante et semble-t-il facilement reproductible dans R (l'exception qui confirme la règle). Enfin, la présentation de Jean-Daniel Fekete était à la fois un plaidoyer pour l'exploration visuelle des données et une présentation du dernier outil développé au sein de l'équipe Aviz, le logiciel PaohVis permettant l'exploration d'hypergraphes dynamiques⁵.

2.2 Des données peu questionnées

À quelques exceptions près (une enquête sur la sociabilité des étudiant.e.s belges par Kristel Vignery, l'approvisionnement de Montpellier en tomates par Milo Monier *et al.*), les personnes présentes ne produisent pas leurs données. Ces données sont issues d'études antérieures (cabinet d'avocats d'affaire de Lazega, données biologiques, données *Socio-Patterns*), récupérées en ligne et « nettoyés » (données Twitter, données issues du *Land Matrix Project*) ou générées en fonction des objectifs de recherche. Lorsque la question posée se veut strictement computationnelle (trouver un algorithme performant, mesurer tel ou tel indicateur, tester telle méthode de partitionnement ou d'agrégation), ce manque de curiosité concernant le processus des productions des données pose peu de problèmes - cela ne correspond pas à nos habitudes en sciences sociales mais les questions de recherche sont autres. Par contre, lorsque certains auteurs prétendent tirer de leurs études une utilité sociale, ce manque de recul paraît plus préoccupant. Cela était sans doute dû au manque de temps (les scrupules méthodologiques

4. <https://blizaar-lab.recherche.cergy.eisti.fr/>

5. <https://www.aviz.fr/Research/paohvis>

liés aux données sont aussi peu vendeurs en conférence qu'en article) mais la présentation des travaux de Ronaldo Menezes (l'un des trois *keynote speakers* de la conférence) sur la géographie des délits manquait cruellement d'un regard critique sur les sources utilisées. Que la masse des données numériques pouvant être analysées soit en croissance constante est certain ; que cela permette de mieux comprendre le fonctionnement de nos sociétés ne l'est toujours pas.

2.3 Le réseau à tout prix ?

Toute personne ayant un jour eu l'occasion d'organiser un événement lié à l'analyse de réseaux le sait, le terme même de réseau est tellement polysémique que l'utiliser est le meilleur moyen de recevoir des propositions très hétérogènes qui, parfois, ne contiennent pas la moindre trace d'analyse de réseau (formalisation des données, méthodes, concepts, etc.). Il est presque rassurant de constater que la communauté informatique ne fait pas exception, plusieurs communications revendiquant l'analyse de réseaux sans jamais l'utiliser ni même en avoir besoin.

Ronaldo Menezes crée des réseaux pour étudier des distributions spatiales d'événements ponctuels (pourquoi pas mais l'intérêt n'est pas évident), Mohammed El Hassouni *et al.* créent des réseaux multicouches pour étudier des similarités entre films - alors que des mesures de similarité lexicale auraient donné un résultat équivalent - et je n'ai pas encore compris pourquoi et comment Roberto Interdonato et ses collègues mobilisaient la détection de communautés en partant d'images satellites relatives au couvert végétal...

Conclusion

L'analyse de réseaux, comme toutes les méthodes quantitatives, fournit à différentes communautés disciplinaires un langage, des méthodes, voire des outils en partie communs, ce qui est supposé permettre un dialogue entre ces communautés. Or ce dialogue reste l'exception plutôt que la règle et assister en tant que géographe à une conférence d'informatique permet de comprendre pourquoi. Si le vocabulaire est le même, les questions de recherche différent, tout comme l'attention aux données ou à leur collecte.

J'aimerais pouvoir écrire le contraire mais il ne m'est aujourd'hui pas plus possible de recommander MARAMI à un chercheur ou une chercheuse en sciences sociales que de recommander les rencontres Réseaux et Histoire à un chercheur ou une chercheuse en informatique ou en physique. Il est nécessaire de pouvoir présenter ses travaux à ses pairs et les deux conférences cités à l'instant le permettent ; créer les lieux de dialogue entre disciplines, et notamment entre sciences sociales et sciences formelles, reste me

semble-t-il un objectif à atteindre.