

Théorie des réseaux

Vincent Miele, Ingénieur de recherche CNRS, Laboratoire de Biométrie de Biologie Évolutive

« Network Science » : l'écosystème

L'analyse des réseaux a connu un essor important dans les années 2000, avec l'apparition d'une terminologie ad hoc, les « Complex Networks » et le « Network Science », portée principalement par la communauté des physiciens. Il s'agit d'étudier toute forme de données d'interactions modélisées sous la forme d'un réseau ou plus formellement sous la forme d'un graphe (sommets et arêtes) qui est l'objet mathématique associé.

Du côté des applications, on trouve des approches de modélisation par différents types de réseaux dans de nombreuses disciplines :

- biologie : régulation génique, similarités de gènes, métaboliques, interaction protéines-protéines, neuronaux, interaction chromosomique 3D
- écologie : trophiques, mutualistes, mouvements, génétiques
- sciences humaines et sociales : "face to face", relations sexuelles, appels téléphoniques, amitiés, réseaux sociaux sur le web

qui sont appuyées par la récolte généralisée de données massives (GPS, mobiles, séquençage, etc.).

Du côté méthodologique, la communauté « Informatique Mathématique » est active sur le sujet naturellement au travers de la théorie des graphes, aussi bien au niveau de la Recherche que de l'offre de formation :

- Comité de programme de MARAMI <http://lipn.univ-paris13.fr/marami2013/MARAMI13/Comites.html>
- Equipe Complex Network du LIP6 <http://www.complexnetworks.fr/>
- Groupe de travail du GdR IM <http://gtgraphes.labri.fr>
- INRIA Bamboo

Les physiciens s'intéressent aux systèmes complexes :

- Alain Barrat et son groupe <http://www.cxnets.org/>, Bruno Gonçalves
- l'IXXI et les séminaires réseaux/SHS <http://www.ixxi.fr>

Les statisticiens et probabilistes proposent de nouveaux développements mathématiques :

- le groupe SSB <http://ssbgroup.fr>
- les acteurs du mini-symposium de la SMAI (Laurent Decreasefond, Marc Lelarge, Viet Chi Tran)

Par ailleurs, les avancées méthodologiques sont souvent réalisées par des écologues, des bio-informaticiens (très nombreux), des sociologues ... qui sont au contact de la donnée et abordent le domaine par le prisme de la modélisation.

On note également le très fort succès de conférences internationales sur le sujet :

- NetSci <http://netsci2013.net> (300 soumissions et 400 participants en 2013, dont peu de français)
- CompleNet <http://complenet.org>

et l'apparition cette année de deux journaux dédiés, « Network Science » (Cambridge UP) et « Journal of Complex Network » (Oxford UP).

Les développements méthodologiques

De nombreux travaux dans le domaine consistent à analyser un réseau unique et fixe. Du point de vue applicatif, le réseau sert à extraire de l'information, par l'utilisation de techniques descriptives. Du point de vue théorique, le réseau consiste en une observation d'un processus (une réalisation d'un modèle pour les statisticiens ou d'une loi de connectivité universelle pour les physiciens) à décrire analytiquement. Cependant, le temps des développements méthodologiques pour l'analyse de réseaux statiques et uniques semble désormais révolu. Un saut conceptuel important se dessine, avec la complexification des problèmes étudiés (étude d'entités couplées, dynamiques ou en interaction) et de nouvelles directions de recherche apparaissent.

► Les analyses topologiques

De nombreux travaux portent sur la topologie des réseaux, avec des approches descriptives, modèles ou algorithmiques. Il s'agit d'obtenir la meilleure description de la connectivité d'un réseau. La recherche de sous structures connectées dites "communautés" a par exemple fait l'objet de développements inspirés par la physique et l'informatique (optimisation de la modularité par des approches itératives [Blondel]), l'algèbre (analyses spectrales [Van Mieghem]) ou la statistique (modèles à variables latentes [Daudin]). Du point de vue mathématique, les modèles de graphes aléatoires sont nécessaires non seulement comme éléments d'une analyse descriptive, mais également pour des approches analytiques ou par simulation (Monte-Carlo si le modèle est génératif) pour étudier la significativité d'une caractéristique observée dans un réseau. Ces modèles constituent la brique élémentaire de toute analyse *in silico* de processus autour des réseaux (voir après) et leur estimation peut constituer une véritable challenge computationnel pour des réseaux de grande taille [Zanghi,Vu,Chandrasekhar]. Par ailleurs, on note également des liens entre topologie des réseaux et topologie algébrique [Vergne]. Enfin, pour évaluer la robustesse ou la fragilité d'un réseau suite à la suppression de sommets ou d'arêtes, la percolation est étudiée, de façon analytique [Zif] ou par simulation. Cependant, les interactions modélisées par les réseaux sont de plus en plus considérées dans un contexte, ce qui se traduit mathématiquement par l'utilisation de covariables sur les sommets ou les arêtes à intégrer dans les approches décrites précédemment [Mariadassou]. De même, le contexte est également spatial dans nombre de réseaux dits "spatiaux" (en particulier écologiques) où les sommets sont des entités situées dans l'espace [Dray]. Ces éléments de contexte conduisent à complexifier les approches au delà de l'étude simple des connections.

1. V.D. Blondel, J-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics* (2008)
2. P.Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge University Press (2011)
3. J.-J., Daudin, F. Picard, S.,Robin, Mixture model for random graphs, *Statistics and Computing* (2008)
4. H. Zanghi, F. Picard, V.Miele and C. Ambroise, Strategies for online inference of model-based clustering in large and growing networks, *Annals of Applied Statistics* (2010)
5. D.Q. Vu, D.R. Hunter, M. Schweinberger, Model-based clustering of large networks, *Annals of applied statistics* (2013)
6. A.Chandrasekhar, M.O.Jackson, Tractable and Consistent Random Graph Models, arXiv:1210.7375 (2012)
7. Anais Vergne, Laurent Decreusefond, Philippe Martins, Reduction Algorithm for Simplicial Complexes, *Infocom* (2013)
8. R.Ziff, Getting the Jump on Explosive Percolation, *Science* (2013)
9. M. Mariadassou, S. Robin, C. Vacher, Uncovering Latent Structure in Valued Graphs: A Variational Approach. *Annals of Applied Statistics* (2010)
10. Dray S, Pélissier R, Couteron P, Fortin M-J, Legendre P, [...], Community ecology in the age of multivariate multiscale spatial analysis, *Ecological Monographs* (2012)

► Les phénomènes dynamiques sur les réseaux

Pour de nombreux travaux, le réseau n'est pas l'objet de l'étude mais le support de la diffusion d'un processus d'intérêt. Selon les cas, on parle de cascade, de diffusion ou de propagation de "failure", de rumeurs, d'épidémies, sur un réseau [Lelarge]. Au niveau mathématique, ces études s'appuient sur des modèles épidémiques par équations différentielles [Machens, Decreusefond], et nécessairement s'appuient sur les différents modèles de graphes aléatoires. On note également des travaux basés sur la théorie du contrôle qui décrivent la contrôlabilité de différents réseaux réels et modèles [Liu]. De nouveau, l'introduction de covariables induit un degré de complexité supplémentaire, lorsqu'il s'agit d'intégrer dans un modèle de diffusion les effets de la topologie et les effets de cofacteurs comme l'homophilie ou la proximité géographique [Centola]. Par ailleurs, des travaux modélisent la contagion complexe (par exemple, il faut être exposé à une rumeur venant de plusieurs amis avant de la croire) ou le comportement collectif dans les réseaux [Gleeson]. De nouvelles questions mathématiques restent ouvertes quand il s'agit de modéliser une diffusion non homogène sur le réseau ou de supposer un espace d'état multiple [Melnik] ou même continu, par opposition à l'état binaire ("0/1", "malade/sain" par exemple). Une récente direction consiste par ailleurs à considérer la diffusion dans le réseau comme un processus à mémoire (prendre en compte le sommet précédent pour déterminer le sommet suivant) [Rosvall], ce qui demandera de nouveaux développements mathématiques.

1. M. Lelarge. Diffusion and Cascading Behavior in Random Networks. *Games Econ. Behav.* (2012)
2. A. Machens, F. Gesualdo, C. Rizzo, A. E. Tozzi, A. Barrat, C. Cattuto, An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices, *BMC Infectious Diseases* (2013)
3. L. Decreusefond, J-S. Dhersin, P. Moyal, V.C. Tran Large graph limit for an SIR process in random network with heterogeneous connectivity, *Annals of Applied Probability* (2012)
4. D. Centola, The Spread of Behavior in an Online Social Network Experiment, *Science* (2010)
5. Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási Controllability of complex networks, *Nature* (2011).
6. J.P. Gleeson, D.Cellai, J.Onnela, M.A.Porter and F.Reed-Tsochas, A Simple Generative Model of Collective Online Behaviour, arXiv:1305.7440 (2013)
7. S. Melnik, J. A. Ward, J. P. Gleeson, and M. A. Porter, Multi-stage complex contagions, *Chaos* (2013)
8. M. Rosvall, A. Esquivel, A. Lancichinetti, J.West, R. Lambiotte, Networks with Memory, arXiv: 1305.4807

► Les réseaux temporels, dynamiques

La nécessité de modéliser l'apparition ou la disparition d'arêtes (ou plus rarement de sommets) conduit à l'étude de réseaux dits temporels, évoluant ou dynamiques [Holme]. Les notions abordées pour les réseaux statiques sont alors à redéfinir dans ce cadre [Wang]. De nouveau, le rôle de covariables (homophilie) ou de la mémoire [Karsai] dans la création des arêtes conduit à leur intégration dans les modèles. Par ailleurs, plusieurs travaux relèvent le challenge majeur consistant à coupler les phénomènes de diffusion précédemment cités avec les réseaux temporels, en caractérisant l'interaction entre ces deux dynamiques [Aiello, Rocha, Masuda].

1. P.Holme, J.Saramäki, Temporal Networks, arXiv:1108.1780 (2011)
2. Q.Wang, E.Fleury, T.Aynaud and J.L. Guillaume, Communities in evolving networks: definitions, detection and analysis techniques, *Dynamics of Time Varying Networks* Springer (2012)
3. M.Karsai, N.Perra, A.Vespignani, The emergence and role of strong ties in time-varying communication networks , arXiv:1303.5966
4. L. Aiello, A. Barrat, C. Cattuto, R. Schifanella, G. Ruffo, Link creation and information spreading over social and communication ties in an interest-based online social network, *EPJ Data Science* (2012)
5. L.E.C. Rocha, V.D. Blondel, Bursts of vertex activation and epidemics in evolving networks, *PLoS Computational Biology* (2013)

6. N. Masuda, P.Holme, Predicting and controlling infectious disease epidemics using temporal networks, Prime Rep (2013)

► **Les multi-réseaux**

Depuis quelques années, de nouveaux développements traitent de l'analyse ou de la modélisation de réseaux multiples, sans les angles de la topologie [Berlingerio] ou de la dynamique [Nicosia] comme précédemment présenté par les réseaux uniques. On distingue les réseaux multiplexes ou multicouches ou multidimensionnels, avec des sommets identiques et différents types d'arêtes (par exemple, différents types d'interactions entre espèces [Kefi]). On parle aussi de réseaux superposés pour étudier la diffusion conjointe de plusieurs épidémies sur un même réseau social, la dynamique couplée du comportement social et de la diffusion d'une épidémie [Funk], ou des processus cognitifs parallèles dans les réseaux neuronaux. On parle par ailleurs de réseaux de réseaux pour caractériser des réseaux interdépendants, avec des arêtes de dépendances entre sommets des deux réseaux [Gao]. En particulier au niveau des infrastructures, la dépendance peut être critique et résulter de l'effet de la proximité dans l'espace [Schneider]. Si le panel des différents réseaux multiples est déjà large, le champ des mathématiques associées est encore à explorer.

1. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A. and Pedreschi, D., Multidimensional networks: foundations of structural analysis. World Wide Web Journal (2012).
2. Kéfi, S., E. Berlow, E. Wieters, [...] U. Brose. More than a meal... Integrating non-feeding interactions into food webs. Ecology Letters (2012).
3. V. Nicosia, G. Bianconi, V. Latora, and M. Barthelemy, Growing multiplex networks, Physical Review Letters (2013)
4. S.Funk, V.jansen, Interacting epidemics on overlay networks. Physical Review E (2010)
5. J. Gao, S.V. Buldyrev, H. E. Stanley, S. Havlin, Networks formed from interdependent networks , Nature Physics (2012)
6. C.M. Schneider, N. Yazdani, N.A.M. Araujo, S. Havlin, H.J Herrmann, Towards designing robust coupled networks, Scientific Reports (2013)

► **Construction des réseaux**

Au delà des protocoles et équipements pour collecter les données d'interactions, les mathématiques sont présentes pour la construction de réseaux à partir de la donnée brute, en particulier pour les données biologiques (puces à ADN par exemple [Chiquet]) où il s'agira de contenir les effets des irrégularités et bruits de mesures. On cherchera par exemple à prédire des contacts sociaux à partir de données à haut débit de GPS, RFID ou téléphone mobile [Scholtz]. De nombreux travaux portent enfin sur la prédiction d'arêtes manquantes ou erronées [Lu].

1. J. Chiquet, Y. Grandvalet, C.Ambroise, C., Inferring multiple graphical structures. Statistics and Computing (2011)
2. C. Scholz, M. Atzmueller, G. Stumme, A. Barrat, C. Cattuto New Insights and Methods for Predicting Face-To-Face Contacts, preprint (2013)
3. L. Lü, T.Zhou, Link prediction in complex networks: A survey , Physica A (2010)

Contacts en France

Bioinformatique : Daniel Kahn, Denis Thieffry

Ecologie : Stéphane Dray, Sonia Kefi

Mathématiques : Laurent Decreasefond, Catherine Matias

Informatique : Matthieu Latapy

Physique : Alain Barrat

SHS : Bertrand Jouve, Pablo Jensen