

Sujet de thèse associé au projet APEX déposé dans le cadre de l'AAP Région
Bourgogne-Franche-Comté 2017

Titre de la thèse: Analyse physique des matrices de Google associées aux réseaux complexes réels

Encadrants de la thèse: Pierre Joubert (Professeur, UTINAM),
José Lages (Maître de conférences, UTINAM)

Thématique: systèmes intelligents et leur impact sociétal

Lieu: Institut UTINAM / Université de Franche-Comté

Date de début de thèse: 1er septembre 2017

Date de fin de thèse: 31 août 2020

Description du travail de thèse et intégration dans le projet général_____

Description du travail de thèse

Les systèmes complexes forment naturellement d'immenses réseaux comportant un nombre très important de nœuds interconnectés. Par exemple, dans les systèmes vivants, les protéines agissent entre elles via des réactions chimiques; dans le cerveau, l'influx nerveux est transmis aux neurones par les axones; dans les réseaux sociaux, les participants sont liés par relations; dans le commerce international, les pays exportent/importent entre eux des produits. . . De nos jours, l'archétype des réseaux complexes dirigés est le World Wide Web (WWW) contenant plus de 10^{12} pages reliées entre elles par des hyperliens. Ces réseaux, à la topologie extrêmement compliquée, possèdent des propriétés d'invariance d'échelle [1] et des propriétés dites de *ultrasmall world* [2]. L'étude des propriétés physiques de ces réseaux complexes dirigés nécessite l'utilisation d'outils statistiques spécifiques: en empruntant le paradigme du WWW, un *surfeur aléatoire* peut sauter d'un nœud A à un nœud B avec une certaine probabilité (chaîne de Markov). Les réseaux complexes dirigés peuvent alors être représentés par un opérateur stochastique – *la matrice de Google* – appartenant à la classe des opérateurs de Perron-Frobenius [3, 4]. L'analyse des réseaux complexes à l'aide de leurs matrices de Google permet de caractériser et de classer les quantités massives d'informations enfouies dans ces réseaux, et cela de manière extrêmement efficace.

L'objet de ce travail de thèse sera de produire des avancées fondamentales dans les directions de recherche suivantes:

Axe 1: compréhension des propriétés spectrales de la matrice de Google, notamment en étudiant la transition d'Anderson et la loi fractale de Weyl. Les propriétés universelles de la matrice de Google ainsi que la décomposition en valeur singulières de son spectre seront établies,

Axe 2: analyse des interactions entre cultures via l'analyse du réseau multilingue Wikipédia. La structure globale et les propriétés du réseau multilingue Wikipédia seront déterminées et l'intrication entre les différentes cultures sera étudiée. L'analyse des interactions entre pays, entre universités et entre personnages politiques sera obtenue avec la matrice *réduite* de Google,

Axe 3: analyse de la sensibilité du commerce international et du risque de contagion des crises en déterminant le spectre, les états propres et les communautés cachées de la matrice de Google du réseau multiproduit du commerce international (données OMC, ONU, OCDE). Une description des propriétés statistiques du réseau des transactions Bitcoin sera donnée. Les conditions de la stabilité des flux monétaires seront analysées sur l'exemple des transactions Bitcoin.

Ce travail de thèse est le maillon structurant du projet APEX puisque le doctorant ou la doctorante contribuera aux principales avancées scientifiques des axes de recherche **Axe 1**, **Axe 2** et **Axe 3** détaillés dans le descriptif du projet de recherche. Si le doctorant ou la doctorante possède des connaissances/notions de biologie des systèmes, il ou elle participera également à l'axe de recherche:

Axe 4: analyse des relations causales cachées entre protéines oncogène. Très récemment, des chercheurs de l'Institut UTINAM et de l'Institut Curie ont défini [5] les bases de l'application de la matrice (réduite) de Google dans le domaine des réseaux de données omiques et par conséquent ouvert la voie à de nombreux futurs résultats en biologie computationnelle.

En outre, ce travail de thèse interdisciplinaire bénéficiera de l'étroite collaboration déjà existante avec les laboratoires associés au projet ApliGoogle (<http://perso.utinam.cnrs.fr/~lages/projects.html>), lauréat du Défi MASTODONS du CNRS (<http://www.cnrs.fr/mi/spip.php?article819>) et co-porté par l'Institut UTINAM. Ces laboratoires sont le Laboratoire de Physique Théorique de Toulouse (UMR 5152 du CNRS), le groupe de Biologie Computationnelle des Systèmes du Cancer de l'Institut Curie, et l'Institut de Recherche en Informatique de Toulouse (UMR 5505 du CNRS).

Environnement du travail de thèse

L'équipe multidisciplinaire Physique théorique et Astrophysique (PhAs) de l'Institut UTINAM étudie les systèmes complexes à toutes les échelles, de l'atome à la Galaxie. Plus particulièrement, les physiciens théoriciens de l'équipe PhAs développent des outils mathématiques permettant de comprendre des phénomènes complexes tels que, sans être exhaustif, le phénomène de décohérence qui doit être contrôlé en information quantique, le chaos dans les systèmes dynamiques, notamment le chaos dans les systèmes exoplanétaires nouvellement découverts, ou le transport et l'extraction d'informations pertinentes enfouies dans les réseaux complexes réels (WWW, réseaux omiques, réseaux commerciaux, ...)

C'est dans ce très récent axe de recherche "réseaux complexes réels" développé et fortement soutenu à l'Institut UTINAM que s'insère ce travail de thèse. Cet axe, en plein essor, contribue au rayonnement international de l'Institut UTINAM. Le doctorant sera au cœur du développement de ce domaine, il bénéficiera d'une part du savoir-faire bisontin et d'autre part de solides collaborations avec des laboratoires nationaux (le Laboratoire de Physique Théorique de Toulouse, l'Institut de Recherche en Informatique de Toulouse, l'Institut Curie de Paris) ainsi qu'avec le laboratoire GIYA "Laboratorio Tandar" de Buenos Aires. Avec ces différents laboratoires, nous avons d'ores et déjà répondu à l'appel à projet franco-argentin de mobilité ECOS Sud - Argentine 2016 qui permettrait au doctorant sélectionné de passer plusieurs longs séjours auprès de nos collègues argentins spécialistes des réseaux complexes dirigés.

Profil du candidat

Le candidat devra être issu d'un Master 2 Physique et maîtriser les langages informatiques du type python, FORTRAN, C++, ... Des connaissances en théorie des réseaux complexes ainsi qu'en biologie des systèmes seraient appréciées.

Références générales

- [1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", Science, 286:509-512 (1999)
- [2] R. Cohen, S. Havlin, "Scale-free networks are ultrasmall", Phys. Rev. Lett. 90 (5): 058701 (2003)
- [3] A.M. Langville and C.D. Meyer, "Google's PageRank and beyond: the science of search engine rankings", Princeton University Press, Princeton (2006)
- [4] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine", Computer Networks and ISDN Systems v.30, p.107 (1998)
- [5] J.Lages, D.L.Shepelyansky, A.Zinovyev, "Inferring hidden causal relations between pathway members using reduced Google matrix of directed biological networks" (2016) soumis à eLife, preprint disponible à <http://perso.utinam.cnrs.fr/~lages/datasets/googlomics/>
- [6] L.Ermann, K.M.Frahm and D.L.Shepelyansky, "Google matrix analysis of directed networks", Rev. Mod. Phys. v.87, p.1261 (2015)
- [7] United Nations COMTRADE data base <http://comtrade.un.org>
- [8] OECD data base <https://data.oecd.org>
- [9] WTO data base <http://stat.wto.org>

- [10] Bitcoin web site <https://blockchain.info>
- [11] J.Bohannon, “The Bitcoin BUSTS”, Science v.351, p.1144 (2016)
- [12] K.M.Frahm and D.L.Shepelyansky, “ Reduced Google matrix”, arXiv:1602.02394 (2016)
- [13] J. Lages, A. Patt, D. L. Shepelyansky, “Wikipedia Ranking of World Universities”, The European Physical Journal B (2016) 89: 69, les données ainsi que le classement sont disponibles à <http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/>