Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur

DOCUMENT D'AUTOÉVALUATION

Équipe ComplexNetworks

Campagne d'évaluation 2023-2024 — Vague D

Table des matières

1 INFORMATIONS GÉNÉRALES SUR L'ÉQUIPE COMPLEXNETWORKS **3**
1.1 Les thématiques scientifiques et leurs enjeux . . . 3
Thématiques scientifiques . . . 3
Positionnement par rapport au contexte de recherche . . . 4
Avancées scientifiques majeures . . . 5
Animation scientifique . . . 5

2 INTRODUCTION DU PORTFOLIO **6**

3 AUTOÉVALUATION DU BILAN **7**
3.1 Autoévaluation de l'équipe . . . 7
Domaine 2. Attractivité . . . 7
Domaine 3. Production scientifique . . . 8
Domaine 4. Inscription des activités de recherche dans la société . . . 10

4 RÉFÉRENCES BIBLIOGRAPHIQUES EXTERNES **11**

5 RÉFÉRENCES BIBLIOGRAPHIQUES SIGNIFICATIVES DE COMPLEXNETWORKS **12**

A ANNEXE — MEMBRES PERMANENTS AU 31/12/2022 **13**

1 INFORMATIONS GÉNÉRALES SUR L'ÉQUIPE ComplexNetworks

Nom de l'équipe : Complex Networks (ComplexNetworks)

Responsable de l'équipe : Lionel Tabourier

	2017	2018	2019	2020	2021	2022
PR	0	0	0	0	0	0
MCF HDR	0	0	1	1	1	1
MCF	1	2	1	1	0	0
DR	2	2	2	2	2	2
CR HDR	0	0	0	0	0	0
CR	0	1	1	0	0	0
Total permanents	3	5	5	4	3	3
Émérites	0	0	0	0	0	0
Doctorants	8	9	5	3	2	2
Ingénieurs CDD ou hors tutelles	2	0	0	0	1	1
Post-doc, ATER, etc.	1	2	2	4	2	2
Stagiaires	4	4	1	5	0	5
Total non permanents	15	15	8	12	5	10
Total avec émérites	**18**	**20**	**13**	**16**	**8**	**13**
Equivalent temps plein recherche	2.5	4.0	4.0	3.0	2.5	2.5

TABLE 1 – Personnels ComplexNetworks sur la période 2017-2022 (au 1er juillet de chaque année)

1.1 Les thématiques scientifiques et leurs enjeux

Thématiques scientifiques

Description générale des activités de l'équipe

L'équipe Complex Networks travaille sur la représentation, la description et l'analyse de grands jeux de données d'interactions issus du monde réel, dans le but de gagner une meilleure compréhension de leurs caractéristiques et de leur fonctionnement. À cet effet, elle développe des méthodes et des algorithmes pour décrire ces données sous forme de graphes pour rendre compte de leur structure. Elle accorde aussi un intérêt particulier à leur dynamique, notamment via le formalisme des flots de liens (ou *link streams*). Les activités de l'équipe comprennent un volet applicatif, couvrant des cas d'études variés (détaillés ci-après).

Pôles structurant les activités de l'équipe

Pour la période 2017-2022 (et à venir), les activités de l'équipe peuvent être décrites autour de trois pôles principaux :

- un pôle fondamental, notamment structuré autour du formalisme des flots de liens pour décrire la dynamique d'interactions,
- un pôle algorithmique orienté vers une approche pratique de l'algorithmique, spécifique aux grands graphes et aux flots de liens issus de données réelles,
- enfin un pôle applicatif, qui a une forte dimension interdisciplinaire, concernant par exemple les réseaux sociaux en ligne ou les réseaux de transactions financières.

Notre approche consiste à faire des aller-retours entre les questions fondamentales et les problèmes pratiques, les questions fondamentales émergeant de l'étude d'applications et le développement de méthodes conduisant à des applications, ainsi qu'à la validation des méthodes proposées.

Les cas d'études qui ont fait ou font l'objet du travail de l'équipe relèvent de domaines très variés. Parmi les applications étudiées durant la période 2017-2022, notre activité s'est focalisée sur quelques thématiques, notamment dans le cadre de projets spécifiques :

- l'évaluation de la diversité dans les algorithmes du web (dans le cadre du projet Algodiv),
- les dynamiques d'opinions et des conflits culturels dans l'espace européen (projet Odycceus),
- la détection d'anomalies et de fraudes dans les transactions financières (Labcom FiT),

- et de manière plus ponctuelle : l'analyse du trafic dans des réseaux routiers et des flux d'individus, la propagation de virus dans les réseaux de transports d'animaux, les réseaux de décision juridique.

Positionnement par rapport au contexte de recherche

Positionnement vis-à-vis des axes du laboratoire

L'équipe occupe une position transversale par rapport aux axes du laboratoire ; son activité s'insère principalement dans les axes AID (*Artificial intelligence and data science*) et TMC (*Theory and mathematics of computing*), et de manière plus secondaire, via les cas d'étude applicatifs, dans les axes SSR (*Safety, security and reliability*) et ASN (*Architecture, Systems, and Network*). À noter que les contributions liées à l'axe ASN étaient marginales pour la période 2017-2022.

Positionnement à l'échelle nationale

Les activités de l'équipe Complex Networks relèvent du domaine des systèmes complexes, c'est-à-dire l'étude de systèmes constitués d'un grand nombre d'éléments en interaction, dont les propriétés émergentes sont difficiles à prédire depuis le comportement de chaque élément individuellement. Défini de manière si générale, le domaine de recherche est très vaste et à cheval entre plusieurs disciplines. De ce fait, les unités de recherche en France travaillant sur les systèmes complexes de manière transversale sont rares. On peut toutefois nommer l'Institut des Systèmes Complexes de Paris Île-de-France (ISC PIF), unité d'appui et de recherche CNRS dont la mission est de contribuer au développement d'une recherche inter-institutionnelle sur les systèmes complexes en Île-de-France [1], ou l'IXXI, son équivalent dans la région Rhône-Alpes [2]. Nous sommes en contact étroit avec ces structures.

En informatique, peu d'équipes travaillent sur le domaine à proprement parler, mais plutôt sur des sous-domaines de la discipline dont certaines activités recouvrent en partie celles des systèmes complexes, et plus précisement l'analyse des réseaux complexes. Les aspects algorithmiques et fondamentaux sont étudiés dans certains laboratoires spécialisés dans la théorie et l'algorithmique de graphes, par exemple l'IRIF (Université Paris Cité), le LABRI (Université Bordeaux II), ou le LIP (ENS Lyon). Sur les aspects plus appliqués, Complex Networks est proche thématiquement de groupes spécialisés dans la fouille de données et l'apprentissage sur graphes ; on peut citer le LTCI (Télécom Paris), SAMOVAR (Télécom Sud Paris), le LIPN (Université Paris-Nord), le CEDRIC (CNAM), le LIRIS (INSA Lyon), le LIB (Université de Bourgogne), le LIA (Université d'Avignon) ou le LITIS (Université de Normandie). L'équipe a aussi des liens sur certaines questions de recherche avec des laboratoires d'autres disciplines, par exemple ARAMIS (ICM, SU), le Centre de Physique Théorique (Université d'Aix-Marseille) ou Décision et Processus Bayésien (Institut Pasteur).

L'équipe a fait le choix depuis sa création en 2009 de travailler de manière transversale sur les réseaux complexes, tout en restant ancrée disciplinairement dans l'informatique, ce qui fait sa spéficité. Cela permet une bonne visibilité sur le sujet dans la discipline, et à ce titre nous sommes sollicités régulièrement par d'autres acteurs académiques et industriels pour des collaborations, détaillées plus loin dans ce rapport.

Positionnement à l'échelle internationale

À l'international, l'environnement est difficile à décrire en raison du caractère interdisciplinaire du domaine. En Europe, certaines institutions ont investi significativement dans les systèmes complexes, souvent en mettant un accent particulier sur un sous-domaine spécifique (réseaux sociaux, réseaux de communication, réseaux de mobilité,...). On peut noter parmi les groupes les plus importants en taille : Complex Systems and Networks Group (Queen Mary University), Social Dynamics (Cambridge University), Social Networks Group (ETZH), ISI Turin, IC Team (Université Catholique de Louvain), Department of Network and Data Science (Central European University), Complex Systems (Aalto University). D'autres centres très actifs du domaine se trouvent hors d'Europe et en particulier aux États-Unis : Center for Complex Network Research (Northeastern University), SNAP (Stanford University), CNetS (Indiana University) entre autres. Les moyens imposants de ces groupes rendent le domaine très concurrentiel, notamment pour les questions d'accès aux données et aux moyens de calcul. Le positionnement de notre équipe, qui consiste essentiellement à proposer des outils et méthodes pour décrire et comprendre, plutôt qu'à réaliser des implémentations très performantes à visée industrielle, permet d'éviter une confrontation défavorable sur les questions qui concentrent le plus d'activité.

1. https://iscpif.fr/presentation/
2. https://www.ixxi.fr/

Avancées scientifiques majeures

Sur la période 2017-2022, l'équipe a travaillé à développer le formalisme des flots de liens pour la description de grands jeux de données d'interactions dynamiques. Sur cette question, des avancées significatives ont été réalisées que ce soit pour la définition de concepts clefs [Latapy et al., 2018, Lamarche-Perrin, 2020] ou la réalisation d'algorithmes pour leur calcul [Viard et al., 2018, Crescenzi et al., 2020].

Par ailleurs, Complex Networks reste pionnière sur l'algorithmique de grands graphes issus de données réelles, domaine qui fait partie de son identité depuis la création de l'équipe. Dans cette direction, des travaux importants, repoussant les limites de l'état de l'art en termes d'efficacité sur certaines tâches spécifiques ont été réalisés : énumération de sous-graphes denses [Danisch et al., 2018, Sun et al., 2020], détection communautaire sur des plongements de graphes [Bhowmick et al., 2020].

Enfin, et bien qu'il s'agisse d'avancées difficiles à quantifier, l'équipe a contribué au développement d'un cadre théorique et d'outils pratiques pour l'étude de concepts complexes utiles à la description des comportements humains sur le web, tels que la diversité proposée et consommée en ligne par les utilisateurs [Ramaciotti Morales et al., 2021].

Animation scientifique

L'équipe organise régulièrement des séminaires (en moyenne tous les 2 mois) où nous invitons des membres extérieurs au laboratoire afin de faire connaître la communauté (principalement francophone) du domaine aux membres de l'équipe et plus largement au laboratoire.

Par ailleurs, les membres de Complex Networks contribuent ponctuellement aux manifestations liées à la communauté française des systèmes complexes (aux feuilles de route du Réseau National des Systèmes Complexes, ou à la récente *French Regional Conference on Complex Systems*). Ainsi, un des membres de l'équipe a fait partie de 2014 à 2021 du comité de pilotage de l'ISC PIF mentionné plus haut.

Enfin, les membres de l'équipe participent également à l'organisation de conférences dans notre domaine scientifique, ce point sera décrit plus en détails dans le Domaine 2 de la Section 3.

2 INTRODUCTION DU PORTFOLIO

- *Élément 1 (projet ou collaboration) :* Document scientifique de montage du LabCom FiT. LabCom financé par l'ANR entre Complex Networks et la société Bleckwen sur le sujet de la détection d'anomalies et de fraudes dans les transactions financières.
- *Élément 2 (publication) :* Article intitulé "Stream graphs and link streams for the modeling of interactions over time" par M. Latapy, T. Viard et C. Magnien, publié dans la revue Social Networks Analysis and Mining en 2018, définissant des concepts importants autour de la notion de flots de liens (*link stream*), centrale dans le projet scientifique de l'équipe.
- *Élément 3 (publication) :* Article intitulé "Listing k-cliques in sparse real-world graphs", par M. Danisch, O. Balalau et M. Sozio, publié dans les actes de la conférence WWW 2018, sur le listage efficace des cliques dans les grands graphes issus de données réels, représentatif des activités de l'équipe en algorithmique.

3 AUTOÉVALUATION DU BILAN

3.1 Autoévaluation de l'équipe

Domaine 2. Attractivité

Référence 1. L'unité est attractive par son rayonnement scientifique et s'insère dans l'espace européen de la recherche.

Le rayonnement scientifique de l'équipe s'est traduit par plusieurs éléments quantitatifs au cours de la période 2017-2022. Nous soulignons en particulier, en termes de reconnaissance par la communauté, que deux membres de l'équipe ont participé à l'article *"Avoiding traceroute anomalies with Paris traceroute"*, en collaboration avec des membres de l'équipe NPA, qui *a reçu le "Test of Time Award" de l'ACM Internet Measurement Conference* en 2022 ; ainsi que le fait qu'un des membres de l'équipe était *finaliste du Bell Labs Prize 2019* et qu'un autre *était sélectionné pour la finale du SAGE Prize for Innovation and Excellence*.

En termes d'organisation d'événements, outre les éléments précisés dans le document partagé de l'unité, l'équipe a participé à l'organisation d'un certain nombre d'évenements nationaux ou locaux comme l'organisation de la conférence AlgoTel 2018, en lien avec la communauté d'algorithmique de télécommunication ; ou dans le cadre du projet européen Odycceus, l'équipe a participé à l'organisation de conférences et d'une école d'été sur les thématiques du projet.

Enfin, les membres permanents de l'équipe participent régulièrement au Program Committee de conférences importantes des différents domaines auxquels nous sommes liés : ASONAM (2020-2022), ICWSM (2022), KDD (2021-2022), WWW/TheWebConf (2017 à 2022), VLDB (2018), WebSci (2020-2021), WSDM (2021-2022).

Référence 2. L'unité est attractive par la qualité de sa politique d'accompagnement des personnels.

Politique d'intégration des nouveaux permanents. Étant donnée la petite taille de l'équipe, il est possible de mener une politique d'intégration assez personnalisée. Ainsi, les anciens membres ont cherché à aider et encourager les nouveaux membres permanents à déposer des projets de recherche qui leur sont propres. C'était le cas par exemple du projet ANR JCJC LiMass.

Politique d'équipe. Nous cherchons à maintenir une bonne connaissance des activités des autres membres et jusqu'à présent l'équipe n'a pas connu de problème de membre (permanent ou non) non-publiant.

Sauf exception, chaque publication est portée par un(e) doctorant(e) ou post-doctorant(e) en premier auteur, il ou elle est assisté de coauteurs plus expérimentés. En général plusieurs permanents sont impliqués sur chaque publication (en raison de co-encadrements ou de collaborations).

Par ailleurs, l'équipe maintient une base d'informations générales (sous la forme d'un wiki) à destination des nouveaux arrivants, dans le but de faciliter la recherche d'informations sur les moyens propres à l'équipe, le fonctionnement des procédures administratives, la structure de l'environnement scientifique du domaine de recherche, etc.

Enfin l'équipe soutient les membres non-permanents dans leur recherche d'un futur emploi, en les orientant vers des sources d'information adaptées, en les faisant bénéficier du réseau de l'équipe, en les aidant dans la conception des dossiers de candidature ou pour la préparation aux oraux.

Résultats. Cette politique d'équipe a mené à l'obtention de 9 thèses et 1 HDR sur la période 2017-2022.

Immédiatement après leur passage, les 16 membres non-permanents passés par l'équipe durant cette période ont trouvé des activités dans les secteurs suivants : 6 en entreprise, 4 en CDD dans le domaine académique, 2 dans la fonction publique académique, 3 dans la fonction publique hors-académique et 1 en reprise d'études.

Mise à disposition de la communauté. Les articles sont systématiquement mis à disposition de la communauté via HAL (et arXiv). Par ailleurs, l'équipe publie autant que possible ses codes de travail, en général sous licence Creative Commons ; certains de ces programmes ont ensuite été réemployés dans des bibliothèques logicielles, telles que NetworkX (une des principales bibliothèques pour l'analyse de graphe en python [3]), ou Gephi (un logiciel courant de visualisation de graphe [4]). Enfin, l'équipe met également à disposition de grands jeux de données, ainsi

3. https://networkx.org/
4. https://gephi.org/

le jeu de données “Full Bitcoin Blockchain Data Made Easy” [5], contenant une mise en forme de la blockchain bitcoin en février 2021, a été partagé dernièrement.

Référence 3. L'unité est attractive par la reconnaissance de ses succès à des appels à projets compétitifs.

Politique de réponse aux projets. Le faible effectif de l'équipe nous contraint à faire des choix dans nos candidatures à des projets de long terme. Le temps de recherche disponible des permanents permet en moyenne d'avoir 2 à 3 projets actifs simultanément et plusieurs membres de l'équipe travaillent sur le projet. Le choix des projets auxquels l'équipe participe tient le plus souvent au type d'application envisagé.

Projets obtenus sur la période 2017-2022. Trois projets ont été obtenus :

- Le laboratoire commun (LabCom) FiT financé par l'ANR avec l'entreprise Bleckwen, dont un membre de l'équipe est responsable scientifique, sur la détection d'anomalies et de fraudes dans les transactions financières. Période 2019-2024.
- Le projet ANR JCJC LiMass, en tant que coordinateur, sur les algorithmes linéaires pour les grands graphes de données réelles. Période initialement prévue : 2019-2023, terminé de manière anticipée en octobre 2021 suite au décès du porteur, Maximilien Danisch, en mars 2021.
- Le projet européen Odycceus, en tant que participant, sur les dynamiques d'opinion et les conflits culturels dans l'espace européen. Période 2017-2021.
- Enfin, plusieurs financements plus ponctuels par les tutelles ou d'autres organismes : 1 projet DIM, 1 financement de thèse par la DGA, 3 par l'EDITE, 4 projets LIP6.

Référence 4. L'unité est attractive par la qualité de ses équipements et de ses compétences techniques.

L'équipe dispose de ressources de calcul propres : deux serveurs, dont un dédié préférentiellement au calcul et l'autre au stockage de données. Ces serveurs sont gérés par la DIN du laboratoire. En 2020, l'équipe a renouvelé son serveur de stockage pour remplacer deux serveurs de la génération précédente. Les membres de l'équipe sont également utilisateurs des clusters du laboratoire.

Domaine 3. Production scientifique

	2017	2018	2019	2020	2021	2022
Articles (revues)	1.60	2.00	1.75	1.66	2.00	0.40
Communications (conférences)	4.00	2.50	1.75	1.33	1.20	1.60

TABLE 2 – Publications par ETPR par an entre 2017 et 2022

Référence 1. La production scientifique de l'unité satisfait à des critères de qualité.

Durant la période 2017-2022, l'équipe a contribué à plusieurs publications dans des revues ou des conférences de renom de domaines de l'informatique connexes à l'analyse des réseaux complexes, notamment (les rangs des conférences sont ceux indiqués par CORE, les quartiles des revues sont ceux de Scimago pour l'année concernée) :

- en fouille de données : WWW 2017 et 2018 (A*) [Danisch et al., 2017, Danisch et al., 2018], WSDM 2020 (A*) [Bhowmick et al., 2020] Proceedings of the VLDB Endowment (Q1) [Sun et al., 2020], Machine Learning (Q1) [Tabourier et al., 2019]
- en algorithmique et informatique théorique : Algorithms (Q3) [Crescenzi et al., 2019, Crescenzi et al., 2020], Theoretical Computer Science (Q2) [Ramaciotti Morales et al., 2021, Lamarche-Perrin, 2020].

Travaux de collaborations de l'équipe. Ces publications de l'équipe sont issues de collaborations internationales et nationales, voici la liste des institutions ayant participé aux collaborations précédemment mentionnées :

- IIT Kharagpur (Inde) [Bhowmick et al., 2020]
- Université de Namur (Belgique) [Tabourier et al., 2019]

5. `bitcoin.complexnetworks.fr`

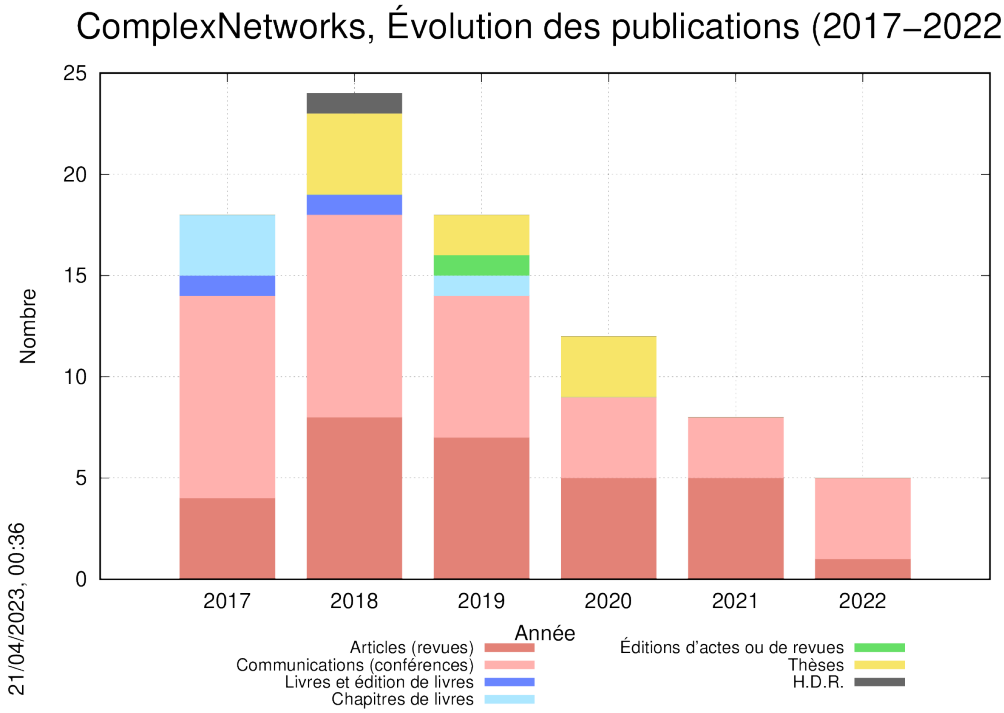

FIGURE 1 – Évolution des publications entre 2017 et 2022

- Sciences-Po Paris et l'ENS Paris Saclay [Ramaciotti Morales et al., 2021]
- LTCI (Télécom Paris) [Danisch et al., 2017, Danisch et al., 2018, Sun et al., 2020]
- IRIF (Université Paris Cité) et *Università degli Studi di Firenze* (Italie) [Crescenzi et al., 2019], [Crescenzi et al., 2020]

Dynamique de production. Les éléments quantitatifs montrent une baisse de production sur la période 2019-2022. Cette baisse du volume de publication est assez fortement corrélée aux effectifs de l'équipe qui ont diminué sur cette période (en nombre de permanents et de non-permanents), ainsi qu'au fait qu'une des membres de l'équipe a la responsabilité de directrice adjointe du laboratoire.

Référence 2. La production scientifique de l'unité est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels.

Comme précisé dans la section 3.1 (référence 2), les productions sont en général construites autour du travail d'un(e) non-permanent(e), encadré par un ou plusieurs permanents. Cela représente environ les 2/3 des publications de l'équipe sur la période 2017-2022. Par ailleurs, ce sont, autant que possible, les membres non-permanents qui vont présenter les travaux de l'équipe en conférence le cas échéant, éventuellement avec un ou plusieurs membres de l'équipe encadrante.

À noter que tous les permanents participent à la production de l'équipe de manière relativement homogène (si on prend en compte le temps alloué à la recherche de chacun).

Référence 3. La production scientifique de l'unité respecte les principes de l'intégrité scientifique, de l'éthique et de la science ouverte. Elle est conforme aux directives applicables dans ce domaine.

Outre la mise à disposition précédemment mentionnée des codes, en vue de permettre la reproduction des travaux réalisés dans l'équipe, nous cherchons également à encourager le travail de reproduction, notamment en début de thèse quand cela présente un intérêt pédagogique pour l'étudiant. C'est notamment le cas pour une publication dans la revue ReScience [Lécuyer et al., 2021], et dans d'autres cas ce travail a été intégré à des publications apportant aussi une contribution originale.

Par ailleurs, l'équipe cherche à s'orienter progressivement vers des modes d'éditions plus satisfaisants, notamment dans des journaux en Green ou Diamond Open Access. Cependant, le succès est mitigé dans ce domaine,

notamment à cause des contraintes de publications dans des conférences ou des journaux dans la thématique et de qualité.

Enfin l'équipe s'implique dans des initiatives d'*"ethical publishing"*, notamment via l'un de ses membres, membre de l'"advisory committee" du Free Journal Network (depuis 2020), membre du "managing board" (depuis 2020) et de l'"editorial board" (depuis 2021) du Peer Community in Network Science, et enfin membre du comité pour une recherche éthique de Sorbonne Université depuis 2022.

Domaine 4. Inscription des activités de recherche dans la société

Référence 1. L'unité se distingue par la qualité et la quantité de ses interactions avec le monde non-académique.

Sur la période, l'équipe a établi ou prolongé des collaborations avec le monde non-académique, via des collaborations industrielles à l'occasion de projets de long terme : le LabCom CLEAR avec Thalès (terminé en 2020), le LabCom FiT avec l'entreprise Bleckwen, le projet ANR Algodiv dont Orange était un partenaire. Les collaborations avec les acteurs de l'industrie se sont faites également à l'occasion de projets plus ponctuels : thèse CIFRE co-encadrée avec Thalès, collaboration avec la start-up Delight pour un encadrement de stage. L'équipe collabore également avec des acteurs institutionnels, tels que la DGA pour la réalisation de projets de thèse.

L'équipe souligne le soutien reçu de la part de la Direction de la Recherche et de la Valorisation de Sorbonne Université, qui a joué un rôle essentiel dans le montage du LabCom FiT auprès de l'ANR. Par ailleurs, l'équipe remercie également la DRV de son aide dans la négociation avec l'ANR sur la terminaison du projet JCJC LiMass suite au décès de son porteur.

Référence 2. L'unité développe des produits à destination du monde culturel, économique et social.

Implication dans la recherche sur des questions de société. Certains aspects de la recherche de l'équipe touchent directement à des questions sociétales, notamment au travers de plusieurs projets (Aldodiv, Odycceus, FiT). À ce titre, l'équipe participe à des débats de société tels que les conflits d'opinions dans les espaces numériques, la question de la description et de la mesure de notions comme la diversité sur les plateformes en ligne. Par ailleurs, le travail réalisé dans le cadre du LabCom FiT, sur la détection de fraudes dans les transactions financières au travers des modèles de flot de liens se veut aussi une contribution à un important problème de société.

Référence 3. L'unité partage ses connaissances avec le grand public et intervient dans des débats de société.

Actions de diffusion scientifique au grand public. Le caractère interdisciplinaire de l'analyse des systèmes complexes amène à se questionner sur le raisonnement scientifique en général et la philosophie des sciences. Plusieurs membres de l'équipe partagent donc un intérêt pour la philosophie des sciences. L'un d'eux co-organisait pendant la période 2017-2022 le séminaire de philosophie de l'Informatique Codes Sources. Par ailleurs, un des membres associés à l'équipe était membre du CA de la Société de Philosophie des Sciences.

4 RÉFÉRENCES BIBLIOGRAPHIQUES EXTERNES

5 RÉFÉRENCES BIBLIOGRAPHIQUES SIGNIFICATIVES DE ComplexNetworks

[Bhowmick et al., 2020] Bhowmick, A. K., Meneni, K., Danisch, M., Guillaume, J.-L., and Mitra, B. (2020). Louvainne : Hierarchical louvain method for high quality and scalable network embedding. In *Proceedings of the 13th international conference on web search and data mining*, pages 43–51.

[Crescenzi et al., 2019] Crescenzi, P., Magnien, C., and Marino, A. (2019). Approximating the temporal neighbourhood function of large temporal graphs. *Algorithms*, 12(10) :211.

[Crescenzi et al., 2020] Crescenzi, P., Magnien, C., and Marino, A. (2020). Finding top-k nodes for temporal closeness in large temporal graphs. *Algorithms*, 13(9) :211.

[Danisch et al., 2018] Danisch, M., Balalau, O., and Sozio, M. (2018). Listing k-cliques in sparse real-world graphs. In *Proceedings of the 2018 World Wide Web Conference*, pages 589–598.

[Danisch et al., 2017] Danisch, M., Chan, T.-H. H., and Sozio, M. (2017). Large scale density-friendly graph decomposition via convex programming. In *Proceedings of the 26th International Conference on World Wide Web*, pages 233–242.

[Lamarche-Perrin, 2020] Lamarche-Perrin, R. (2020). An information-theoretic framework for the lossy compression of link streams. *Theoretical Computer Science*, 806 :90–115.

[Latapy et al., 2018] Latapy, M., Viard, T., and Magnien, C. (2018). Stream graphs and link streams for the modeling of interactions over time. *Social Network Analysis and Mining*, 8 :1–29.

[Lécuyer et al., 2021] Lécuyer, F., Danisch, M., and Tabourier, L. (2021). [re] speedup graph processing by graph ordering. *The ReScience journal*, 7(1) :3.

[Ramaciotti Morales et al., 2021] Ramaciotti Morales, P., Lamarche-Perrin, R., Fournier-S'Niehotta, R., Poulain, R., Tabourier, L., and Tarissan, F. (2021). Measuring diversity in heterogeneous information networks. *Theoretical computer science*, 859 :80–115.

[Sun et al., 2020] Sun, B., Danisch, M., Chan, T. H., and Sozio, M. (2020). Kclist++ : A simple algorithm for finding k-clique densest subgraphs in large graphs. *Proceedings of the VLDB Endowment (PVLDB)*.

[Tabourier et al., 2019] Tabourier, L., Bernardes, D. F., Libert, A.-S., and Lambiotte, R. (2019). Rankmerging : a supervised learning-to-rank framework to predict links in large social networks. *Machine Learning*, 108 :1729–1756.

[Viard et al., 2018] Viard, T., Magnien, C., and Latapy, M. (2018). Enumerating maximal cliques in link streams with durations. *Information Processing Letters*, 133 :44–48.

A ANNEXE — MEMBRES PERMANENTS AU 31/12/2022

La table ci dessous liste les membres permanents de l'équipe ComplexNetworks.

NOM	Prénom	Corps	Employeur
LATAPY	Matthieu	DR	CNRS
MAGNIEN	Clémence	DR	CNRS
TABOURIER	Lionel	MCF (HDR)	Sorbonne Université

ÉLÉMENT DE PORTFOLIO 01

Projet ou collaboration

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Document scientifique de montage du LabCom FiT

Fichier de élément : Fichier pdf intitulé `LabCom_FiT_document_scientifique.pdf`

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Le LabCom FiT avec la société Bleckwen est un projet au cœur de l'activité de l'équipe pour la période 2019-2024. Tous les membres permanents et plusieurs membres non-permanents de l'équipe sont impliqués dans le projet. Par conséquent, ce projet joue un rôle important dans la trajectoire du cycle actuel et le prochain. Il concerne les trois pôles d'activité de l'équipe : fondamental par le développement de concepts sur les flots de liens utiles à la description des anomalies recherchées, algorithmique par la nécessité de developper des algorithmes passant à l'échelle sur des flux de données financières, applicatif par son sujet même, la détection de fraudes.

3 PRÉSENTATION DE CET ÉLÉMENT

L'élément proposé est le document scientifique du montage du LabCom FiT. Il en donne une vue générale en décrivant les objectifs, le contexte scientifique et institutionnel, sa feuille de route, ainsi que son mode de fonctionnement pratique. Nous reprenons ici très brièvement quelques éléments saillants développés dans le document.

L'objectif du LabCom FiT est de développer des méthodes de détection d'anomalies dans de grands jeux de données de transactions financières. C'est une question de recherche importante étant donné le coût pour la société de la fraude financière.

Les interactions financières ont une structure complexe et une forte dynamicité, suggérant que les approches à base de flots de liens sont pertinentes pour les décrire. L'équipe Complex Networks apporte son expertise sur les modèles de flots de liens pour proposer des caractéristiques pertinentes des données, ainsi que des algorithmes efficaces pour les calculer.

L'entreprise Bleckwen, Fintech née en 2016 et incubée dans la société ERCOM, utilise des méthodes d'apprentissage automatique pour lutter contre la criminalité financière. Elle fournit donc à Complex Networks d'une part des données d'entraînement et d'autre part une expertise métier pertinente sur le sujet. À terme, il s'agit pour Bleckwen d'exploiter les concepts et algorithmes développés au LIP6 pour les intégrer aux produits destinés à leurs clients (banques, organismes régulateurs, etc.)

La feuille de route du LabCom se décompose en différents chantiers, eux-mêmes répartis en deux axes principaux : le premier étant la détection de fraudes et le second la détection du blanchiment. Chacun des deux axes cherche à faire emploi de métriques de description de graphes et de flots de liens, jusqu'à présent sous-utilisées, dans une perspective de conception de métrique (*feature engineering*). Ces métriques seront employées par des algorithmes de détection d'anomalies dans des contextes difficiles, de part la masse et la sensibilité des données analysées. Les différents chantiers cherchent à décrire avec un niveau croissant de sophistication les différents aspects de ces phénomènes et la façon dont ils se traduisent dans les données.

Le LabCom est né de la volonté commune du LIP6 et de Bleckwen d'établir une collaboration scientifique de long terme. Il s'insère dans le projet de Sorbonne Université de contruire des relations fortes avec le monde industriel, en particulier dans le domaine de l'intelligence artificielle. Pour Bleckwen, il s'agit de bénéficier de l'expertise du monde académique afin de devenir une entreprise au premier rang de la lutte contre la fraude et le blanchiment.

Pour la réalisation de ce projet, il est prévu un investissement de 6.2 ETP, dont 3.4 du LIP6 et 2.8 de Bleckwen sur la durée prévue du LabCom (60 mois). Dans ce cadre, l'ANR finance 54 mois de post-doctorat. Il y a cependant pour les différents acteurs du projet une volonté de péreniser le LabCom au-delà du financement par l'ANR, notamment par les retours financiers liés à la commercialisation des produits de Bleckwen.

ÉLÉMENT DE PORTFOLIO 02

Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : *Stream graphs and link streams for the modeling of interactions over time*

URL de l'élément : `https://hal.science/hal-01665084`

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Cet article, écrit par Matthieu Latapy, Tiphaine Viard et Clémence Magnien, et publié dans la revue Social Networks Analysis and Mining en 2018, définit des concepts importants autour de la notion de flots de liens (*link stream*). Cette notion joue un rôle central à la fois dans les travaux passés de l'équipe et dans son projet pour le cycle à venir. Par ailleurs, l'article a reçu une certaine attention de la communauté du domaine depuis sa publication si l'on en juge par le nombre de citations reçues (environ 200 en avril 2023 d'après *Google Scholar*).

3 PRÉSENTATION DE CET ÉLÉMENT

L'article introduit d'une part les concepts fondamentaux du formalisme des flots de liens, qu'on peut voir, à l'intar des *time varying graphs* [1] ou de notions moins formalisées comme les *temporal networks* [2], comme une généralisation des graphes à des contextes dynamiques. Ce formalisme met en avant le caractère dual des graphes et des séries temporelles et cherche donc à définir des concepts tels que les cas limites permettent de retrouver les définitions classiques de ces deux domaines. Par exemple, à un instant donné, tous les liens d'un flot de lien définissent un graphe, et les diverses définitions pour un flot de liens doivent être équivalentes aux définitions correspondantes en théorie des graphes.

Les concepts définis sont donc des généralisations de concepts connus en théorie des graphes ou en analyse de séries temporelles. Parmi ceux-ci, on peut citer, de manière non-exhaustive, la densité, le voisinage, les cliques, les chemins temporels, les k-cores ou les notions de centralité.

4 RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5) :387–408, 2012.

[2] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3) :97–125, 2012.

ÉLÉMENT DE PORTFOLIO 03

Publication

1 DÉFINITION DE CET ÉLÉMENT

Titre de l'élément : Listing k-cliques in sparse real-world graphs

URL de l'élément : `https://www.complexnetworks.fr/wp-content/uploads/2019/04/main2.pdf`

2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Cet article, écrit par Maximilien Danisch, Oana Balalau et Mauro Sozio à l'occasion d'une collaboration avec l'équipe DIG de Télécom Paris, et publié dans les actes de la conférence (A*) WWW 2018, porte sur le listage efficace des k-cliques (sous-graphe complet de taille k) dans les grands graphes issus de données réelles. Ce travail est représentatif des activités de l'équipe dans le domaine de l'algorithmique, dans la mesure où il s'attache à un problème général d'algorithmique de graphes, mais l'examine dans le cas spécifique des grands graphes issus de données réelles, dont les particularités structurelles ont un impact sur la conception des algorithmes.

Par ailleurs, cette contribution reçoit un écho positif de la communauté de par les perspectives qu'elle ouvre. En effet, elle a reçu un nombre de citations relativement conséquent (environ 110 en avril 2023 d'après *Google Scholar*), et surtout a servi de base à d'autres articles qui cherchent à améliorer les performances sur la question considérée (e.g., [1]). Notre équipe utilise également les résultats de cet article pour certains de ses travaux actuels.

3 PRÉSENTATION DE CET ÉLÉMENT

L'article considère un problème bien connu de la littérature : le listage exhaustif des cliques de taille k d'un graphe. En faisant usage de l'ordre de dégénerescence des nœuds dans le graphe, il apporte à ce problème une contribution théorique, en calculant une borne supérieure à la complexité temporelle du listage des k-cliques. Surtout, il propose un algorithme et une implémentation parallèle très efficace pour résoudre le problème sur des cas pratiques de grands graphes issus de données réelles, qui ont la particularité d'être relativement peu denses, tout en présentant des zones localement denses et de grandes cliques. Un protocole expérimental très complet démontre que l'implémentation proposée bat significativement l'état de l'art du domaine.

4 RÉFÉRENCES BIBLIOGRAPHIQUES

[1] Ronghua Li, Sen Gao, Lu Qin, Guoren Wang, Weihua Yang, and Jeffrey Xu Yu. Ordering heuristics for k-clique listing. *Proc. VLDB Endow.*, 2020.