HAL
open science

Identifier des signaux faibles avec Twitter : comment développer à partir de la Social Network Analysis des méthodologies spécifiques

Camille Baulant, Guillaume Sylvestre

To cite this version:

Camille Baulant, Guillaume Sylvestre. Identifier des signaux faibles avec Twitter : comment développer à partir de la Social Network Analysis des méthodologies spécifiques. Workshop " Intelligence Economique 2020 ", 2020 – 2025 : Quelle recherche française et francophone en intelligence économique ?, Jul 2020, Paris, France. hal-03266182

HAL Id: hal-03266182

https://univ-angers.hal.science/hal-03266182

Submitted on 21 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identifier des signaux faibles avec Twitter : comment développer à partir de la Social Network Analysis des méthodologies spécifiques

Workshop « Intelligence Economique 2020 », 2020 – 2025 : Quelle recherche française et francophone en intelligence économique ? Thème 4 : « Les Outils de l'IE »

Le Centre de Recherche en Gestion (CEREGE), axe intelligence Stratégique Internationale et *Revue Internationale d'Intelligence Économique* Maison de la Nouvelle Aquitaine, Rue des Pyramides, Paris, le 2 juillet 2020, de 9h à 18h

Camille Baulant (1) & Guillaume Sylvestre (2)
(1) Professeure Université d'Angers, GRANEM, SFR Confluence,
UFR Droit, Economie et Gestion, 13 Allée F Mitterrand, 49000 Angers,
camille.baulant@univ-angers.fr
(2) Directeur de l'innovation numérique à l'ADIT, 27 bis Quai Anatole France 75007 Paris, guillaume.sylvestre@laposte.net

21 juin 2021

Auteurs

Camille Baulant, fondatrice et directrice du Master Intelligence Economique et Stratégies Compétitives à l'International à l'Université d'Angers, professeure agrégée des Facultés en Sciences Economiques (promotion 1994). Rattachée au laboratoire Granem au sein de la SFR Confluence, ses champs de recherche combinent et l'économie de la connaissance (réseaux, clusters, leadership …) et l'économie internationale (compétitivité, Investissement Directs à l'Etranger, risque pays …). Elle participe et publie régulièrement à des colloques de niveau international. camille.baulant@univ-angers.fr

Guillaume Sylvestre a plus de dix ans d'expérience dans l'analyse des données issues du web en général, et des réseaux sociaux en particulier. Directeur de l'innovation numérique à l'ADIT, leadeur européen de l'intelligence économique, il développe des méthodologies innovantes d'analyse de l'information combinant intelligence humaine et outils de collecte et d'analyse des données. Son expertise couvre notamment l'impact stratégique de l'innovation sur les secteurs économiques et l'identification des signaux faibles via les réseaux sociaux. Il donne des cours à l'Ecole de Guerre Economique et l'IHDEN sur l'analyse des données et la datavisualisation, et intervient à l'Université d'Angers sur des projets de recherche liant Big Data, Réseaux sociaux et IA pour le Master Intelligence Economique et Stratégies Compétitives à l'International. guillaume.sylvestre@laposte.net, 06 22 03 21 82

Classification JEL : L14, L17, L12, M15

Résumé
L'article montre l'intérêt d'utiliser des méthodologies avancées pour analyser les données du réseau social Twitter propres à prévoir les nouvelles tendances des informations futures. Deux types d'information sont étudiés : la viralité des Fake News sur internet et la prévision de start-up montante. Dans les deux cas, nous utilisons la collecte automatique de données par Visisbrain et le logiciel Gephi en data visualisation pour étudier en profondeur les Social Graphs dans le but de caractériser par une analyse humaine le rôle des différents leaders sur Twitter. La première partie présente les outils de veille adaptés pour le réseau social Twitter. La seconde partie applique la théorie des leaders d'opinion à Twitter en combinant les outils de la Social Network et l'analyse des réseaux économiques. La troisième et quatrième partie proposent des applications concrètes de cette nouvelle méthodologie sur des cas précis : la viralité des Fake News au sujet du Pacte de Marrakech et la prévision de la the start-up Momenprunt jouant désormais un rôle clef sur le marché de l'immobilier.

Mots-clés : Big data, Intelligence Artificielle, signaux faibles, Twitter, start-up, fake news

Abstract
The paper shows the interest of using advanced methodologies for analyzing Twitter data in order to better forecast new trends about future information. We study two kinds of information: the Fake New Viral Process and the Up-grading Startup Forecasting. In both of cases, we use automatic data collect given by Visisbrain and Gephi Data Visualization Tools in order to deeply study Social Graphs by identifying the role of different leaders on Twitter network in using human analysis. The first part presents adapted monitoring tools for the Twitter social network. The second part applies the opinion leader's theory on Twitter analysis by combining Social Network Analysis tools and human's economic networks analysis. The third and four parts present concrete applications of this new methodology for analyzing the Marrakech Pact Fake News and the forecasting of the up-grading start-up, named Momenprunt, which today plays a key role in real estate market.

Key words: Big Data, Artificial Intelligence, weak signal, Twitter, Start-up, Fake News.

Introduction

Les traces numériques en ligne laissées par les internautes, qui permettent de nous renseigner sur l'activité sociale réelle (Rogers, 2015), sont extrêmement nombreuses et riches dans les réseaux sociaux Facebook, Twitter ou Linkedin qui disposent de milliards d'utilisateurs dans le monde, dont plusieurs dizaines de millions en France. Ces services stockant les données des messages postés par leurs utilisateurs et les interactions liées, ils se prêtent à des analyses quantitatives et automatisées sur des grands ensembles de données (Cha, Haddadi, Benevenuto, Gummadi, 2010). L'analyse de tels volumes de données rend obligatoire l'usage d'outils de traitement Big Data automatisés. L'interprétation des informations issues des réseaux sociaux est complexe. Le réseau social Twitter ne reflète pas l'opinion publique sur un sujet donné, et dispose de moins d'utilisateurs que Facebook en France (environ 10 millions contre 33 millions pour Facebook, la majorité des utilisateurs de Twitter étant par ailleurs inactifs). C'est un outil exploité avant tout par les politiques et les journalistes (Rieder et Smyrnaios, 2012), et qui permet une très forte exposition publique pour un coût dérisoire. Twitter peut donc être exploité utilement pour anticiper les « signaux faibles » correspondant à l'action d'un groupe / d'une communauté spécifique sur un sujet précis, par exemple l'émergence des enjeux liés aux migrations climatiques avant la conférence sur le

climat de Paris (Sylvestre, 2016). Twitter est également le terrain, comme Facebook, de campagnes de désinformation ou « fake news », qui rencontrent un succès certain (Keller, Schoch, Stier & Yang, 2020).

Les algorithmes des réseaux sociaux, en particulier Twitter, vont déterminer quelles publications obtiendront ou non de la visibilité, et seront donc potentiellement virales (Jammet, 2018). Ce fonctionnement favorise notamment les réactions émotionnelles à l'actualité (Alloing, Pierre, 2017), ou la controverse entre deux camps irréconciliables (Antonilos-Basso, Paddeu, Douay et Blanc, 2018). Les sujets mis en avant en temps réel par les réseaux sociaux correspondent souvent à des réactions à chaud exploitables uniquement en gestion de crise. S'il est nécessaire de suivre ces réactions très reprises ponctuellement, il est important pour l'analyse stratégique de pouvoir anticiper des évolutions. Un bon dispositif de veille sur Twitter doit être capable de repérer des sujets moins porteurs, moins d'actualité, mais qui peuvent préfigurer les enjeux de demain. Aujourd'hui, les outils de veille et d'analyse de l'information des réseaux sociaux sont adaptés à l'identification en temps réel des réactions émotionnelles / des controverses, mais pas à la détection des signaux faibles / de nouvelles tendances. Ces plateformes sont très efficaces pour suivre des termes / comptes précis, voir identifier des variations de volumétrie (Guilain, 2015). La détection de l'inconnu, de la nouveauté, ne fonctionnera pas ou seulement dans des cas bien délimités.

Dans cet article, nous nous intéressons à une autre approche : l'analyse et la compréhension des interactions sur Twitter pour identifier des signaux faibles spécifiques à certains risques ou opportunités. Plus que le contenu du message, c'est la dynamique qui en ressort, entre des individus qui échangent, s'interpellent, se soutiennent, se confrontent qui nous intéresse. Cette dynamique, si elle se décline de différentes façons, se retrouvera quelle que soit la langue ou les termes employés. Elle permettra de mettre en avant des caractéristiques spécifiques des comptes essentiels dans la viralité de l'information, et donc d'anticiper des signaux faibles correspondant à des problématiques concrètes. Nous entendons par « signaux faibles » toutes les informations détectables permettant, selon un contexte et des objectifs spécifiques, d'anticiper des risques ou des opportunités pour des organisations (Ansoff, 1984 ; Lesca, 1995). Identifier des signaux faibles est par essence très complexe, puisque leur interprétation sera variable selon le contexte. Plus l'information sera potentiellement novatrice ou disruptive pour une organisation, moins il y aura de chances qu'elle soit traitée. (Sylvestre, 2016). Ces difficultés ont amené certains auteurs à considérer inutile la recherche de signaux faibles, car impossible à valoriser, voir à identifier dans un contexte big data (Alloing, Moinet, 2016).

Dans ce travail, nous postulons qu'identifier des signaux faibles par des informations issues des réseaux sociaux est possible, dès lors que deux conditions sont remplies : définir précisément le type d'informations recherchées en lien avec un objectif précis, et mettre au point une méthodologie d'analyse spécifique à ce que l'on recherche, combinant expertise humaine et algorithmes big data. Pour éviter les biais liés à la collecte de données limitée offerte par des outils gratuits de scraping, nous avons fait appel à la plateforme Visibrain qui a accès à toutes les données Twitter. Visibrain intègre l'accès à toutes les données Twitter sans limite dans le temps, et à des dizaines d'opérateurs pour une collecte fine et exhaustive de l'information. Visbrain permet aussi l'exportation de données réexploitables pour l'analyse des Big Data par les outils de la Business Intelligence ou de la Social Network Analysis. Enfin, Visibrain offre différents modules d'analyse pour faciliter l'interprétation des échanges sur Twitter.

La première partie du travail présentera les outils de veille adaptés pour le réseau social Twitter. La seconde partie appliquera la théorie des leaders d'opinion au réseau Twitter, en combinant les outils de la social network et l'analyse des réseaux économiques. La troisième et quatrième partie présenteront des exemples concrets d'analyses qui montrent la nécessité d'adapter les méthodologies aux types de signaux faibles recherchés : l'identification de la circulation de Fake News avant que celles-ci deviennent virales, et la détection de start-up prometteuses mais peu visibles. Les deux exemples traités correspondront à deux types de communautés spécifiques exploitant Twitter pour développer leur réseau et leur influence : les militants de groupes extrémistes, et les dirigeants de start-up.

I. Twitter, un réseau social qui nécessite une approche de veille adaptée

Le réseau social Twitter permet un accès payant à l'intégralité des données publiées par ses utilisateurs. C'est une différence fondamentale avec Facebook, qui donne accès à très peu de contenu, tout comme Instagram, ou encore Linkedin qui permet des recherches sur le contenu publié mais uniquement de façon limitée. Les données des tweets ne peuvent être obtenues que via des fournisseurs de service partenaires de Twitter, qui devront respecter des conditions d'utilisation strictes sur la réutilisation des données et la protection des informations des utilisateurs. Outre les messages, URL, photos, vidéos et émoticônes partagées, l'accès à ces plateformes permet d'exploiter les métadonnées des tweets : localisation lors de la publication, localisation indiquée par le compte, nombre d'abonnés et d'abonnements, etc …

L'analyse de sentiments, mise en place depuis une dizaine d'années, n'a jamais été fiable malgré l'évolution des outils de data mining. La raison en est simple : la compréhension du langage humain et de son contexte n'est pas adaptée aux algorithmes d'analyse automatique de l'information, dès lors que l'interprétation peut varier d'une personne à une autre. Plusieurs travaux soulignent ainsi la difficulté d'analyser la polarité des messages, notamment d'identifier le sarcasme, l'ironie … mais plus simplement d'en comprendre le contexte. Notons que même pour des experts humains, il n'est pas évident de qualifier dans la même catégorie certaines informations (Boyadjian, Velcin, 2017).

De plus, ce type de technologie basé sur des corpus d'informations déjà qualifiés n'arrivera pas à identifier de nouveaux concepts / des signaux faibles. Par contre, il sera très efficace pour détecter et analyser des éléments connus : problématiques de service client, identification de produits, de personnalités … l'intelligence artificielle permettra d'effectuer une pré-analyse de ces données. Certaines plateformes de veille intègrent des modules d'intelligence artificielle (IA) pour identifier de nouvelles informations, en comparant en temps réel les informations partagées avec le bruit ambiant. Les limites sont là aussi nombreuses, notamment le problème de la langue, puisque les outils sont souvent paramétrés avec des exemples en anglais.

Ces modules exploitent deux types d'application de l'IA. La première concerne une analyse sémantique du texte des publications. Cette analyse sera effectuée via un thésaurus permettant de reconnaitre des mots-clés ou combinaisons de mots-clés spécifiques, ou avec des algorithmes complexes de machine learning, intégrant du natural language processing. La seconde application consiste à exploiter des statistiques des publications. Il s'agit ici de combiner les totaux de tweets, d'abonnés, de mentions, de publications de tweets … pour

établir les caractéristiques génériques d'un corpus ou de certains comptes Twitter. L'IA pourra ici être employée pour identifier des « anomalies » correspondant à de nouveaux évènements.

CAS D'USAGE DE L'IA	VALEUR AJOUTEE	LIMITES
Analyse sémantique de l'historique de la veille pour automatiser la surveillance	Identifier automatiquement des documents similaires	1. Imprécis si pas de possibilité d'adapter l'IA 2. Ne permet pas d'identifier de nouvelles informations
Analyse sémantique des « sentiments » des publications	Visualiser la part des éléments négatifs et positifs	Ne fonctionne que sur des typologies de données déjà prédéfinies
Analyse sémantique de l'information recherchée selon des catégories précises	Mettre en avant les partenariats, les fournisseurs, etc … autour d'une société	Très efficace sur les sujets d'entrainement de l'IA
Exploitation des statistiques des publications pour identifier des informations virales	Alerter sur l'apparition de signaux non anticipés	1. Beaucoup de bruit 2. Sensibilité à adapter selon le sujet
Exploitation des statistiques des publications pour qualifier des « influenceurs » sur un corpus	Pointer les utilisateurs les plus influents sur un sujet donné	1. Risque de biais via des robots / opérations d'influence 2. Critères non transparents

Camille Baulant et Guillaume Sylvestre, 2020

Ces approches ont en commun le fait qu'elles ne sont pas adaptées à l'identification de signaux faibles. Elles seront très efficaces pour automatiser la détection d'éléments connus, mais pas pour identifier de nouvelles informations. Il faudra les calibrer pour chaque thématique / sujet novateur, puisque les termes employés ou les statistiques de publication n'auront pas forcément la même signification selon le contexte. L'exploitation des statistiques des publications comporte également des biais importants, notamment le risque de qualifier des robots comme comptes influents, alors qu'ils n'ont pas réellement de portée en dehors de leur communauté artificielle. La recherche d'influenceurs est souvent limitée à des informations quantitatives, et donc facilement biaisées. Plusieurs études ont ainsi montré que le nombre d'abonnés / de like d'un compte sur les réseaux sociaux ne préjuge pas de son éventuelle capacité à influencer (Cha, Haddadi, Benevenuto, Gummadi, 2010 ; Beauvisage, Beuscart, Couronné, Mellet, 2011). Si les retweets ou mentions montrent une proximité et des interactions entre des acteurs (Boyd, Golder & Lotan, 2010), il est essentiel de vérifier le contexte : on peut aussi bien mentionner un compte Twitter par soutien / approbation que dans le cadre d'une controverse (Smith, Rainie, Shneiderman, Himelboim, 2014).

II. L'identification des leaders d'opinion sur les réseaux sociaux, une nouvelle méthodologie combinant outils de Social Network Analysis et étude des réseaux économiques humains

1. Exploiter les algorithmes de la Social Network Analysis pour qualifier les communautés Twitter

La très grande quantité de données issues des réseaux sociaux en général, et de Twittter en particulier, permet de disposer de modèles d'analyse ne se limitant pas à l'exploitation du texte ou l'utilisation des méta-données. Il est également possible de détecter des communautés et d'identifier les leaders au sein de ces communautés. Les interactions au sein des réseaux sociaux peuvent être représentées et analysées par la Social Network Analysis, c'est-à-dire l'étude des interactions entre les membres d'un réseau. D'abord appliquée aux réseaux humains dans le cadre de la théorie des graphes, la Social Network Analysis est aujourd'hui appliquée aux données Twitter (Grandjean, 2016). Le champ de recherche de la Social Network Analysis exploite un outil d'analyse des communautés par la datavisualisation : Gephi. Cet outil, libre et gratuit, développé à l'origine par des étudiants de l'université de Compiègne, est de plus en plus utilisé par les professionnels de l'analyse des réseaux sociaux, car il permet d'interpréter de façon très efficace et précise des grands corpus de données Twitter.

En proposant une cartographie des échanges Twitter à partir des tweets et mentions des comptes, Gephi permet de visualiser le réseau des acteurs (Grandjean, Jacomy et Girard, 2016). La position des comptes Twitter est déterminée par l'algorithme de spatialisation Force Atlas 2 (Jacomy, Venturini, Heymann, Bastian, 2014), qui met en évidence les communautés correspondant aux interactions. Cela permet de constater visuellement l'effet « bulle de filtre » des échanges sur les réseaux sociaux : il est très difficile pour des publications issues d'une communauté donnée d'être partagées par d'autres communautés. La cartographie des réseaux sociaux permet également, en visualisant les communautés et leurs interactions, d'identifier des typologies d'échanges permettant de caractériser ceux-ci (Smith, Rainie, Shneiderman, Himelboim, 2014).

Gephi exploite plusieurs calculs statistiques pour analyser les interactions entre les tweets étudiés :

-*Modularity Class* : identifier des communautés au sein des échanges sur Twitter (Blondel, 2008). C'est un calcul statistique qui décompose le réseau en sous-structures selon un score affecté à chaque entité.

-*Eigenvector Centrality* : prendre en compte l'influence des comptes Twitter au sein de leur propre communauté. Ce calcul croise le nombre de liens (mentions / retweets) d'un compte Twitter avec la qualité de ses liens dans le réseau (le fait d'être relié à des entités ayant elles-mêmes beaucoup de liens).

-*Betweenness Centrality* : identifier les acteurs au centre des échanges entre les différentes communautés d'un corpus de tweets.

La combinaison de ces calculs statistiques à partir de corpus Twitter permet de mettre en évidence des campagnes / stratégies d'influence, ponctuelles ou structurées sur le long terme, l'emploi de robots / comptes inactifs pour pousser une thématique et l'apparition de

sous-communautés liées à un sujet donné et donc de signaux faibles concernant la montée en puissance d'un nouveau sujet / d'une thématique spécifique.

Il est nécessaire de disposer d'une expertise avancée de Gephi pour exploiter au mieux ces algorithmes, notamment en utilisant les filtres pour combiner certains éléments. Il est ainsi possible de définir des méthodologies pour anticiper des tendances à travers l'évolution des échanges sur des corpus de données précis, qu'il s'agisse d'identifier de nouveaux acteurs militants (Sylvestre, 2016) ou de vérifier l'influence et la qualité du réseau d'une start-up (Sylvestre, 2020). Si l'on s'intéresse aux comptes Twitter capables de diffuser l'information auprès d'autres communautés, et donc de contribuer à la viralité d'un sujet, on recherchera en premier lieu ceux qui sont à la fois influents au sein de leur communauté et repris par d'autres communautés. Ces comptes Twitter sont importants pour expliquer la viralité d'une Fake News, pour développer le réseau d'une start-up ...

Identification des types d'influenceurs sur Twitter via Gephi

	Relayé par d'autres communautés	**Relayé uniquement par sa communauté**
Influence importante dans sa communauté	**Compte Twitter très influent sur le sujet étudié (Information virale, transpartisan …).**	Compte Twitter très repris par ses abonnés mais qui ne dépasse pas son réseau (leader politique …).
Influence faible dans sa communauté	Compte Twitter peu influent dans son réseau mais capable d'être relayé ponctuellement (expertise, information virale …	Utilisateur classique de Twitter.

Source : Camille Baulant et Guillaume Sylvestre, 2020

Nous allons voir que ces caractéristiques correspondent également à de réels leaders dans les réseaux économiques : leur activité et leur influence sur les réseaux sociaux sont la transposition de leur leadership.

2. Le leader go-between, passerelle de l'information entre communautés

Les échanges sur les réseaux sociaux en général, et Twitter en particulier, correspondent principalement à des interactions entre des individus. Ces interactions sont généralement dominées par des « influenceurs », véritables leaders d'opinion qui seront repris par leur communauté, voir d'autres communautés. C'est dans ce dernier cas qu'ils auront vraiment de l'influence, car ils deviendront alors des « diffuseurs » d'information et permettront à un sujet de devenir viral. Nous nous intéresserons ici à deux principaux types de leaders d'opinion que l'on retrouvera sur Twitter : les « leaders économiques » (Muller, Pénin, 2008 ; Baulant 2015) et les « leaders go-between » (Amisse, Baulant, Muller, Vargas, 2008 ; Baulant, 2017).

Les « leaders économiques » se distinguent des « leaders hiérarchiques » de la société fordiste car ils peuvent se définir par deux grandes capacités inimitables : « voir loin » et « s'assurer que tous les membres du réseau auront une retombée réelle de leurs actions ». Les leaders économiques travaillent essentiellement sur des concepts assez généraux et mobilisent un réseau assez dense. Une attention particulière est donnée aux « liens faibles », acteurs difficiles à identifier au préalable car souvent en retrait mais dont l'action a été mise en évidence dans les travaux de Granovetter (1973) et de Crozier (1977). L'objectif stratégique est de réussir à « amalgamer » plusieurs causes (par exemple environnementale, sociale, politique, économique, locale…) de façon à pouvoir avoir une résonnance multiple et à tous les niveaux (local, national et mondial). Ces « leaders économiques » correspondront sur les corpus étudiés à des acteurs publiques / des entités très connues et visibles, intervenant sur de multiples sujets et repris à minima par leur propre communauté. Faciles à identifier via leur nombre d'abonnés et de mentions Twitter, ils n'expliquent pas à eux seuls la viralité des échanges.

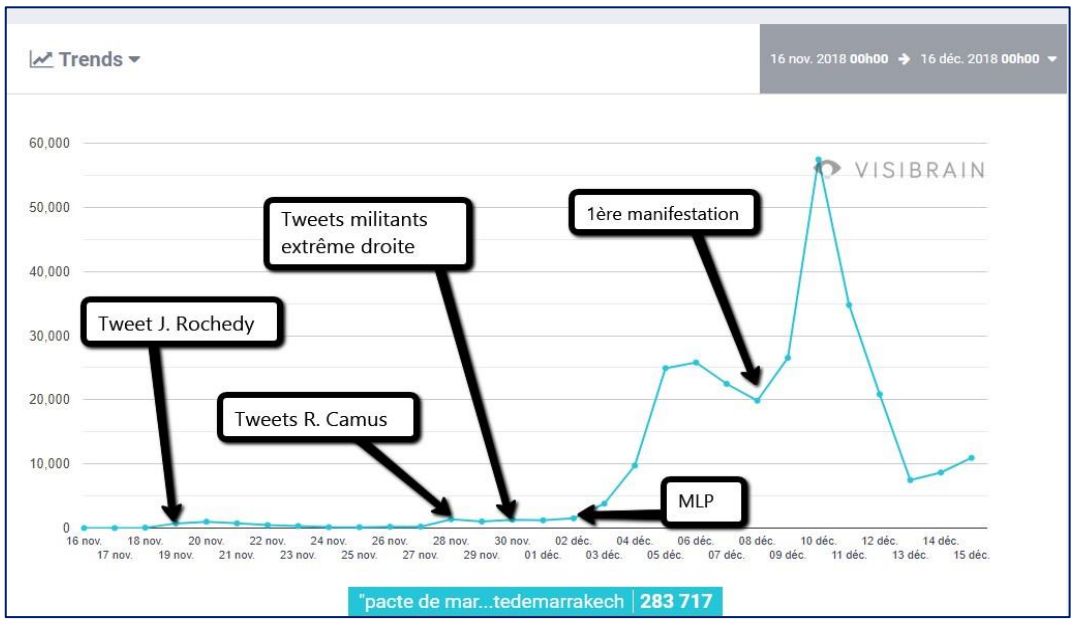
Le second type de leader travaille plus à court terme. Ils sont souvent plus difficiles à identifier car leurs relations sont essentiellement personnelles, humaines, non linéaires (Rallet, Torre, 2004 ; Kotter, 2007). Un mot « banal » pourra, grâce à une vidéo ou un contexte particulier, lancer avec succès une manifestation ponctuelle qui aura des conséquences pouvant dépasser largement l'évènement initial. En économie, on les appelle des « leaders go-between » (Amisse, Baulant, Muller, Vargas, 2008 ; Baulant, 2017) car ils ont des compétences interpersonnelles bien précises (Goleman, 2011) qui ont le pouvoir « d'unir les acteurs dans l'action », même si ces derniers peuvent diverger sur les stratégies de long terme. Ils travaillent, même quand ils agissent sur les réseaux sociaux, essentiellement pour le niveau « local » de façon à mobiliser tout un tissu d'acteurs (élus, citoyens, ONG, entreprises…) qui pourra avoir une influence non négligeable sur des niveaux supérieurs ou des intérêts financiers puissants. Ces leaders go-between sont cruciaux pour expliquer l'apparition de nouvelles thématiques, du fait de leur capacité à mobiliser différentes communautés via leurs connexions. Ils seront influents sur des sujets précis, et pas sur d'autres, à court et à moyen terme.

L'application des calculs statistiques et des filtres de Gephi permet ainsi d'identifier, après analyse, les leaders go-between. Leurs caractéristiques permettent ensuite, via des méthodologies adaptées, de qualifier différents types de signaux faibles.

III. Application à la diffusion de Fake News : Le Pacte de Marrakech

Le « Pacte mondial pour des migrations sûres ordonnées et régulières » (Pacte de Marrakech) a été signé lors d'une conférence intergouvernementale qui a eu lieu du 10 au 11 décembre 2018 à Marrakech. Il a pour objectif de clarifier les questions de la migration à l'échelle internationale. Il est le fruit de plusieurs mois de travail des Etats membres des Nations unis. Dans plusieurs pays, des politiques et partis opposés à l'immigration vont propager des rumeurs sur les conséquences de ce pacte : un abandon de souveraineté de la France à l'ONU sur les questions migratoires, l'ouverture totale des frontières, etc … en s'appuyant sur le faible écho médiatique du sujet. Cette stratégie de manipulation de l'information opérée par des militants d'extrême droite va permettre en quelques semaines à ce sujet de prendre une ampleur telle qu'une manifestation de gilets jaunes sera organisée le 8 décembre 2018, et que le site officiel de l'Elysée devra démentir la Fake News le 10 décembre 2018.

Sur Twitter, le hashtag *#Pactedemarrakech* est apparu le 18 novembre via le parti belge d'extrême droite Vlaams Belang, opposé à la signature du Pacte. En France, avant l'apparition du hashtag, c'est le militant d'extrême droite Julien Rochedy qui a le premier donné de la visibilité au sujet. Le Pacte de Marrakech restera ensuite relativement peu mentionné en France, jusqu'à la reprise du sujet par des élus nationaux à partir du 3 décembre, devenant ensuite viral jusqu'à une première manifestation organisée en Belgique le 8 décembre. En tout, près de 300 000 tweets en français vont mentionner le pacte de Marrakech en 3 semaines.

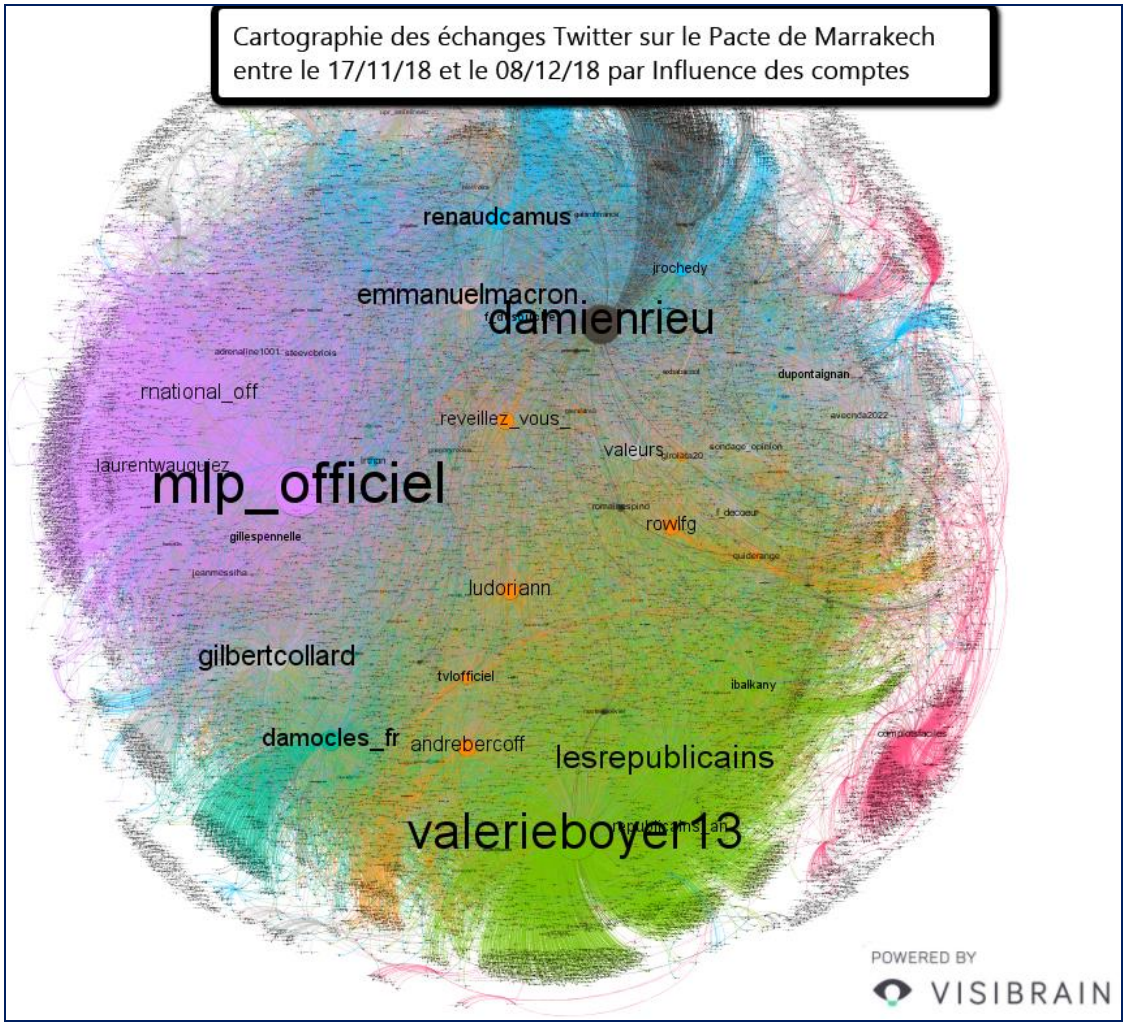


Nous allons constater en analysant les tweets avec Gephi que les vrais propagateurs de cette fake news ne sont pas les politiques qui ont repris l'information en étant mentionnés uniquement par leurs communautés, mais des figures et militants de l'extrême droite correspondant à des leaders go-between.

1. <u>Les leaders go-between essentiels pour la propagation de l'information</u>

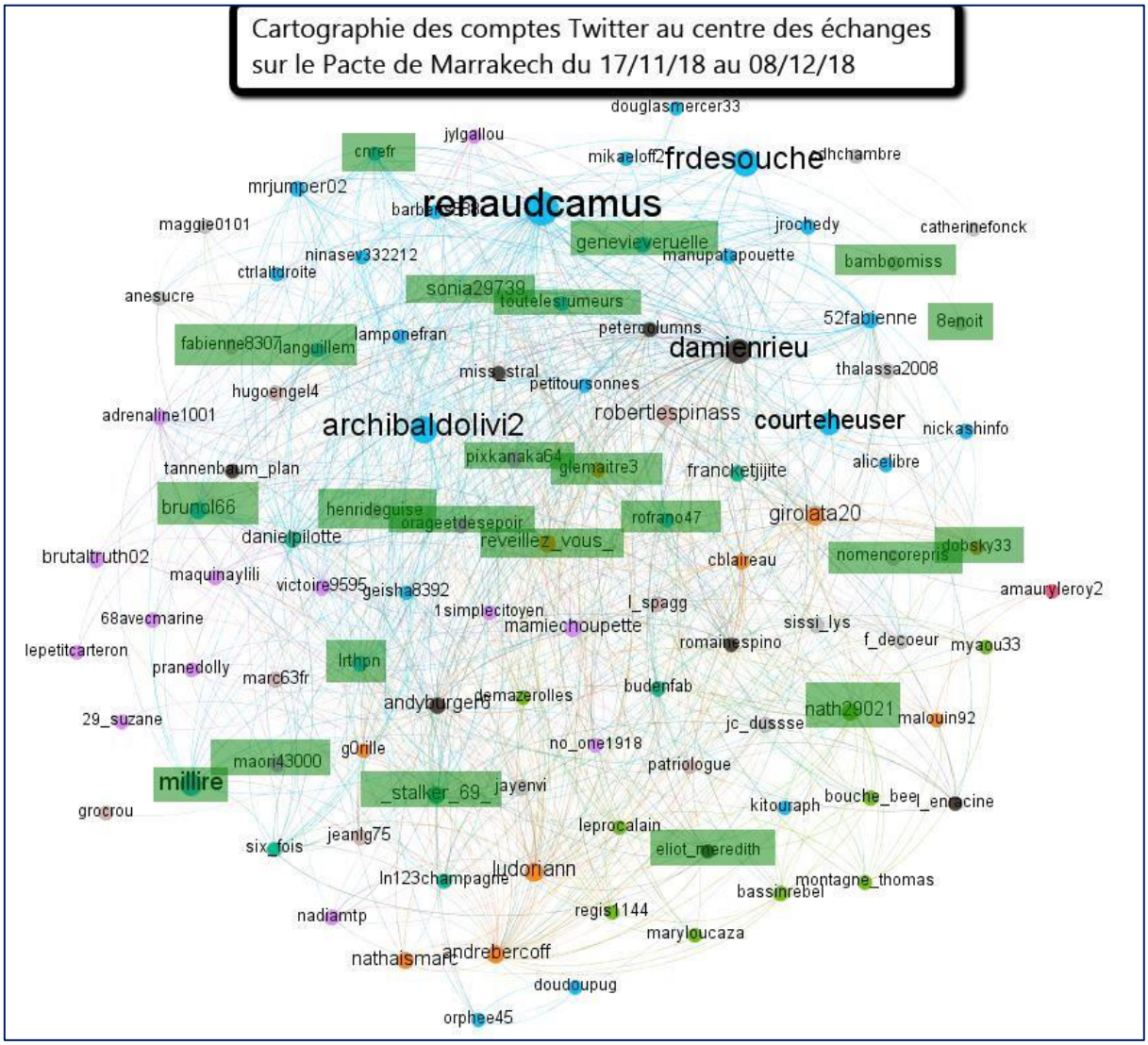
Du 17 novembre au 8 décembre 2018, 67 450 tweets mentionnant le Pacte de Marrakech ont été diffusés par 16 335 comptes Twitter actifs. On identifie avec le logiciel de Social Network Analysis Gephi 700 communautés, les 8 premières représentant plus de 70 % des comptes Twitter. Seules les communautés disposant de leaders influents sont très reprises. Les deux principales communautés, en violet et en vert, correspondent respectivement aux militants du Rassemblement National et à ceux des Républicains. Leur poids est similaire, autour de 16 % du total des comptes Twitter.

Cartographie via Gephi et Visibrain des comptes les plus mentionnés par leurs communautés sur le Pacte de Marrakech du 17 novembre au 8 décembre 2018 :



Pour autant, cette cartographie des comptes les plus mentionnés ne permet pas d'identifier les comptes qui ont été les plus actifs pour que le sujet devienne viral. En appliquant le calcul de betweeness centrality et un filtre, nous pouvons faire ressortir les comptes au centre des échanges. Nous avons surligné en verts ceux qui étaient effectivement actifs avant le 2 décembre :

Cartographie via Gephi et Visibrain des comptes au centre des échanges sur le Pacte de Marrakech du 17 novembre au 8 décembre 2018 :



On constate que les leaders go-between ne correspondent pas aux principales communautés politiques, mais à des groupes militants. Sur les trois premières semaines, du 17 novembre 2018 au 2 décembre 2018, 8226 tweets ont été partagés, par 3894 utilisateurs. C'est cette période qui a été déterminante pour que le Pacte de Marrakech devienne un sujet viral. Nous allons maintenant voir qu'il était possible d'anticiper le risque de propagation de cette Fake News dès la première semaine de sa diffusion, en combinant l'analyse via Gephi à un algorithme d'identification des opérations d'influence.

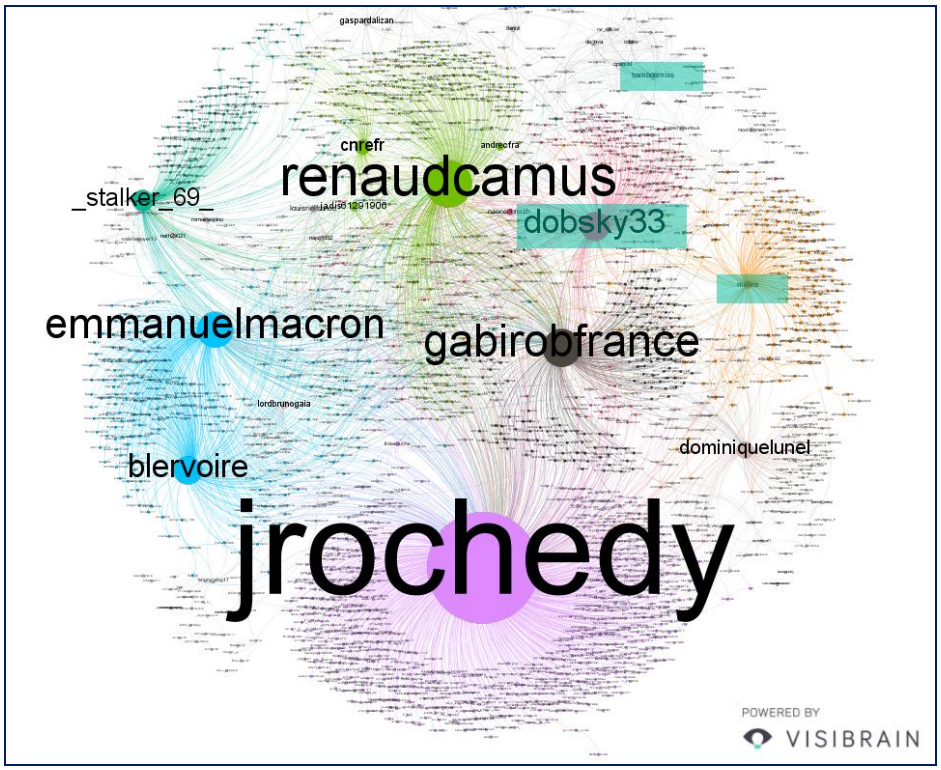
2. <u>Détection des stratégies de manipulation de l'information en combinant un algorithme et la détection des leaders go-between via Gephi</u>

Le Computationnal Propaganda Research Project de l'université d'Oxford a défini le « coefficient de manipulation Twitter » (Nimmo, 2019). En comparant via un calcul statistique la dynamique des échanges Twitter sur différentes thématiques, notamment des

sujets poussés par l'extrême droite en France, ce coefficient défini un seuil de « manipulation du trafic Twitter ». Au-delà d'un score de 12, on peut suspecter une manipulation.

Nous avons calculé ce coefficient sur les tweets mentionnant le Pacte de Marrakech du 17 novembre au 23 novembre 2018 inclus. Sur un total de 2908 tweets, il y a eu 1833 utilisateurs, pour 89 % de retweets. La variable centrale pour montrer la manipulation, à savoir le pourcentage de tweets provenant des 50 utilisateurs les plus actifs, est de 16,95 %. Ces éléments donnent un coefficient de manipulation Twitter particulièrement élevé : 27,44. Pour autant, cette manipulation n'est pas forcément efficace : si elle est générée par une seule communauté ou des robots, elle ne deviendra pas virale. Il faut donc effectuer une analyse avec Gephi pour vérifier si des leaders go-between sont présents pour propager la fake news à d'autres communautés. La cartographie via Gephi de ce corpus permet d'identifier 30 communautés, en cohérence avec un sujet émergent et invisible de la sphère publique. Les trois premières communautés en vert, bleu et violet regroupent 50 % des échanges. Nous avons surligné en vert les comptes qui correspondent à des leaders go-between.

Cartographie via Gephi et Visibrain des comptes les plus mentionnés par leurs communautés sur le Pacte de Marrakech du 17 au 23 novembre 2018 :



Renaud Camus, que nous qualifierons d'influenceur des réseaux d'extrême droite, a lui aussi touché d'autres communautés que la sienne. Il a également diffusé le sujet sur Facebook le 20 novembre, avant que les publications se multiplient fin novembre sur le réseau social. Dans le cas de la propagation d'une fake news, la coordination entre Facebook et Twitter est essentielle, et sera d'autant plus efficace si elle est assurée par des leaders go-between.

Nous avons ici combiné notre méthodologie d'identification des leaders go-between à un

algorithme exploitant les données brutes d'un corpus Twitter, le « coefficient de manipulation Twitter ». Il serait intéressant de l'intégrer à une plateforme de veille pour envoyer des alertes dès qu'un seuil est franchi. Un analyste pourra ensuite vérifier via notre méthodologie s'il y a un risque de viralité ou non sur la diffusion de la fake news repérée. Des recherches avancées sur les pages publiques de Facebook, intégrées à des plateformes comme Visibrain, permettront en complément de vérifier la propagation en dehors de Twitter. Le Pacte de Marrakech a commencé à être très commenté sur Facebook le 28 novembre, soit 4 jours avant la reprise du sujet par des leaders politiques. Cette méthodologie sera également efficace pour détecter des opérations d'astroturfing sur un sujet particulier. L'astroturfing est une technique utilisée à l'origine par des organisations militantes, mais aujourd'hui par tous types d'acteurs, des entreprises aux gouvernements, pour simuler un mouvement d'opinion. Ce mécanisme est très utilisé par l'extrême droite (Vanderbiest, 2018).

Le « coefficient de manipulation Twitter » sera mécaniquement élevé sur des sujets sociétaux / politiques, sans qu'il y ait forcément de manipulation : la controverse génère de l'engagement. En comparant ce coefficient entre deux périodes données, ou deux sous-thématiques, on pourra par contre caractériser une action d'influence. Gephi permettra avec une méthodologie adaptée de faire apparaitre les acteurs qui ont été les plus actifs sur un sujet donné, qui seront donc derrière l'opération d'astroturfing.

IV Application à la détection de start-up leaders : La Social Graph Due Diligence

Twitter a très vite été considéré dans les travaux de recherche comme un outil capable de prédire les tendances, qu'elles soient politiques ou économiques (Bollen, Mao, Zheng, 2010 ; Asur, Huberman, 2010). Nicolas Vanderbiest, qui s'est penché notamment sur les analyses Twitter de la société Filteris qui prédisait la victoire de François Fillon aux présidentielles de 2017, montre néanmoins que les données brutes sont facilement manipulables (Vanderbiest, 2018) et qu'elles doivent donc être interprétées avec précaution. Cela parait évident pour des groupes militants, mais c'est évidemment possible pour des investisseurs, notamment activistes, ou des personnalités de la tech, comme Elon Musk, souvent repris à l'ordre par la SEC pour ses tweets qui ont un impact direct sur le cours en bourse de ses sociétés (cours en bourse sur lequel toute sa rémunération est indexée dans le cas de la société Tesla). Pour autant, il est possible d'utiliser Twitter pour anticiper des tendances, dès lors que l'on prend en compte les biais inhérents au réseau social (Phillips, Dowling, Shaffer, Hodas, Volkova, 2017) :

Quelle que soit la thématique étudiée, l'essentiel des données Twitter correspond à du "bruit", c'est-à-dire des éléments peu ou pas intéressants par rapport à ce que l'on recherchera. Ce sera d'autant plus le cas sur des sujets innovants, où les effets d'annonces et les campagnes marketing sont démultipliés. Les algorithmes de Twitter, comme ceux des autres réseaux sociaux, amplifient la viralité de certains échanges dès lors qu'ils suscitent de nombreuses réactions. Les contenus à fort potentiel de réaction émotionnelle ou proposés par des leaders dans leur secteur obtiendront donc énormément de visibilité, favorisant les experts en communication qui sont influents dans leur communauté, mais rarement en dehors. Aucun modèle d'analyse "clé en main" ne fonctionne pour détecter les signaux faibles sur Twitter : chaque problématique et chaque thématique nécessiteront

d'affiner une méthodologie rigoureuse pour obtenir des résultats pertinents.

A partir de ces biais, et en identifiant les leaders go-between, nous avons développé une méthodologie de « Social Graph Due Diligence » pour détecter dans un secteur donné les start-up prometteuses. Cette méthodologie a déjà été appliquée aux secteurs des Fintech, de la Cybersécurité, et de la Blockchain (Sylvestre, 2020). Pour chaque cas d'étude, nous avons identifié une start-up qui était peu visible et influente, mais qui bénéficiait d'un réseau d'experts pour la soutenir, voire de l'influence de ses fondateurs. Plus que le volume de mentions, qui valorise en règle générale les grands groupes peu innovants mais dotés de moyens de communication importants, c'est la qualité des interactions que l'on va analyser. Nous allons revenir sur la Fintech du secteur immobilier identifiée en mai 2019 dans notre analyse des Fintech françaises (Sylvestre, 2020), pour vérifier son développement et donc la pertinence de notre méthodologie.

1. Monemprunt, une jeune pousse en pleine croissance dans les Fintech de l'immobilier

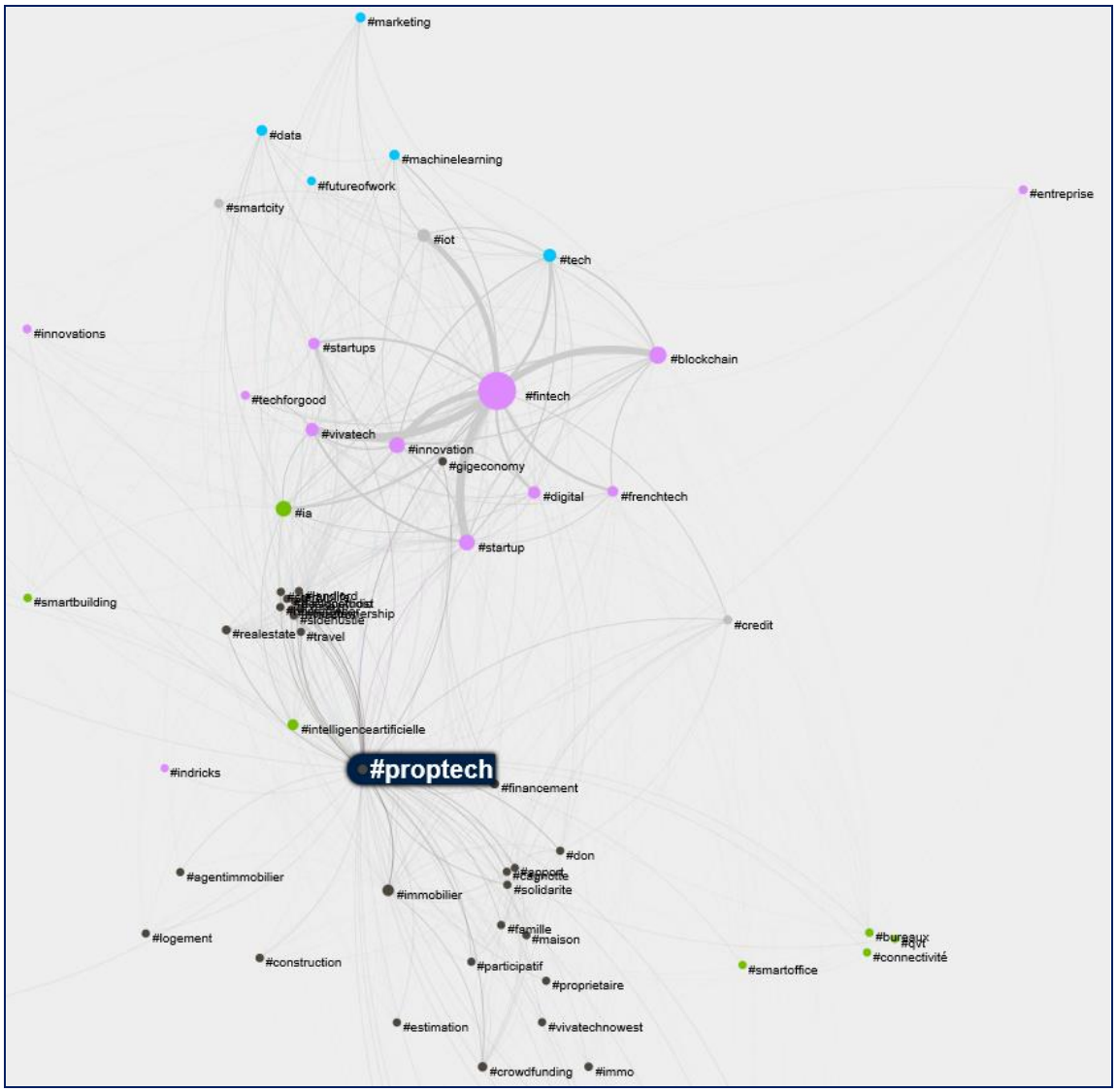
Monemprunt, Fintech lancée en 2016 par Emmanuel Frattini et Arnaud Guilleux, est une plateforme de courtage en ligne qui utilise le machine learning pour identifier les offres de banques les plus adaptées aux dossiers de ses clients. Son principe est d'allier l'innovation technologique à l'accompagnement humain via le courtage immobilier. La plateforme a effectué un tour de table en novembre 2018 de 1,5 millions d'euros avec BPIFrance, Go Capital et Sogescot. Start-up rennaise, elle est par ailleurs très bien ancrée dans son tissu local.

Sans être expert du secteur, nous avons sélectionné cette start-up en juin 2019 pour analyser plus en détail son réseau et ses interactions. Pour ce faire, nous avons identifié des Fintech françaises prometteuses lors de CES de janvier 2019, à partir des tweets en français sur les Fintech. Via notre méthodologie de « Social Graph Due Diligence », nous avons ensuite identifié les leaders go-between correspondant aux start-up également actives durant les échanges du salon Vivatech de mai 2019. Bien que peu mentionnée, Monemprunt est ressorti de nos analyses via notamment les compte Twitter de son fondateur Arnaud Guilleux, qui était intégré à une autre communauté (gris pour @monemprunt et ses autres dirigeants, vert pour @arnaud_guilleux).

Cartographie dynamique via Gephi, Sigma et Visibrain des comptes Twitter ayant interagi avec Arnaud Guilleux durant le salon Vivatech 2019 :

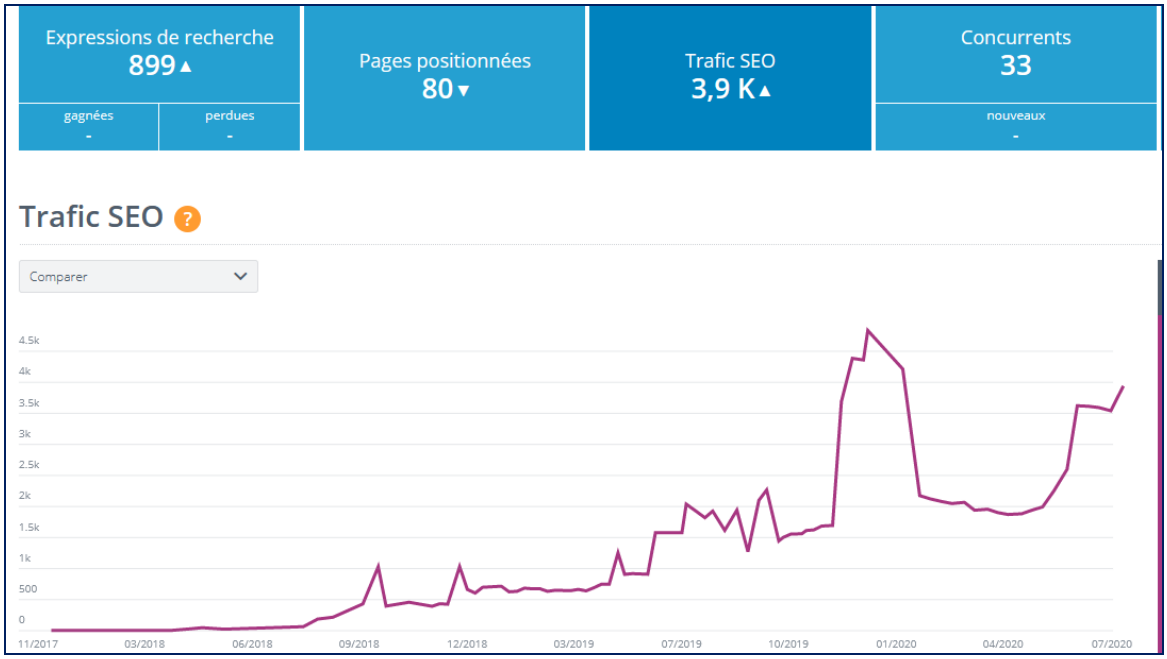
@monemprunt a donc bénéficié du réseau Twitter de ses dirigeants, experts de leur secteur et donc leaders go-between sur les Fintech dans l'immobilier. Trois des quatre directeurs sont actifs sur Linkedin et Twitter, et ils interagissent régulièrement avec leurs pairs. L'incubateur LeSwave a également fait bénéficier de son réseau la start-up. Plus que les termes liés à la Fintech, c'est le hashtag dédié #PropTech qui rassemblait en mai 2019 les mentions liées à ce secteur innovant :

Cartographie dynamique via Gephi, Sigma et Visibrain hashtags utilisés avec #Proptech dans les tweets échangés durant le salon Vivatech 2019 :



Depuis, la start-up rennaise exploite son réseau local pour gagner en visibilité, en intégrant la French Proptech en janvier 2020. Sa page Facebook est suivie par près de 3700 personnes, sa page Linkedin a 526 abonnés, son compte Twitter est suivi par 3200 personnes. La visibilité du site Internet monemprunt.com sur les moteurs de recherche a également fortement augmenté, malgré un creux pendant le confinement en cours de rattrapage.

Trafic du site monemprunt.com via le référencement sur les moteurs de recherche – source yooda.com



2. Un succès qui se confirme après le confinement, période complexe pour le secteur immobilier

Plus récemment, MonEmprunt a annoncé avoir effectué le 6 juillet 2020 une nouvelle levée de fonds auprès de ses actionnaires historiques, en intégrant également MBA Holding, pour un million d'euros. Malgré le Covid et l'incertitude sur le marché de l'immobilier, la start-up semble avoir les reins solides et dispose de plusieurs soutiens, à la fois locaux et sectoriels. Le 26 juillet, elle a également bénéficié d'une analyse positive de Michel Nizon, cofondateur de Kchehck, une start-up spécialisée dans l'analyse des start-up (qui avait notamment critiqué le manque de transparence de Yuka) et dont l'expertise est reconnue. Cette analyse a été saluée par Stéphane Bunouf, PDG du spécialiste en financement de l'innovation Melyad. Entretemps, Arnaud Guilleux a été convié sur BFM Business le 20 juillet pour intervenir sur le marché de l'immobilier post covid.

Le soutien de ces différents leaders go-between de la frenchtech, de la proptech et de la tech rennaise permet à MonEmprunt de renforcer son réseau alors qu'elle ne dispose pas encore d'une réelle stratégie de communication. Cela démontre une reconnaissance de son potentiel commercial et de l'expertise de ses dirigeants, dont la stratégie est validée par leurs pairs. A contrario, la start-up Ashler et Manson, fondée par Aymerick Penicaut, qui propose le même type de services avec une IA, et que nous avions identifié dans la même analyse comme ayant un réseau moins développé et structuré sur Twitter, ne semble pas avoir décollé. Pire, son trafic web a chuté sans reprendre depuis la crise Covid, et est loin derrière celui de MonEmprunt. Pourtant, Ashler et Manson avait bénéficié de mentions dans des articles et La Tribune et des Echos en 2019, contrairement à MonEmprunt. Cela indique également que les

médias classiques, même ceux ciblant les décideurs, ne sont plus prescripteurs. Aujourd'hui, les experts et les influenceurs business sont actifs sur Linkedin et Twitter. Il est essentiel pour les start-up d'intégrer leurs réseaux, d'être repris par les leaders go-between de leur secteur, locaux comme nationaux, et de développer leur propre démarche d'influence. Les investisseurs doivent à minima analyser les interactions sociales de leurs cibles pour vérifier si elles sont adoubées par leurs pairs via la Social Graph Diligence. S'ils souhaitent repérer de futures pépites, ils doivent mettre en œuvre une méthodologie similaire à celle de nos travaux, qui devra être adaptée à leurs objectifs et au secteur ciblé.

La même démarche sera également pertinente pour vérifier si des ETI ou des grands groupes qui investissent dans des secteurs innovants sont jugés crédibles par les experts de ce secteur. Leur couverture médiatique et leurs moyens importants en communication assurent une forte reprise de leur actualité, mais si celle-ci ne convainc par les vrais innovateurs, cela signifie que les effets d'annonce sont jugés peu crédibles. On notera qu'à la différence des Fake News, pour lesquelles il est essentiel de confirmer les signaux identifiés sur Twitter avec Facebook, c'est ici sur Linkedin qu'il faudra vérifier le réseau d'une start-up et de ses dirigeants. On pourra notamment identifier des échanges en lien avec des évènements professionnels, et des partenaires informels non visibles sur le site Internet ou Twitter.

Conclusion

Le travail effectué a permis de montrer l'intérêt des méthodologies avancées d'analyse des données Twitter pour anticiper des tendances, sur deux sujets très différents mais stratégiques : l'anticipation de la viralité des Fake News et l'identification de start-up prometteuses. Ces méthodologies, si elles s'appuient sur un outil de collecte automatisée des données, Visibrain dans cette étude, et un outil de datavisualisation permettant d'analyser le graphe social des acteurs sur Twitter, ne peuvent pas être déclinées automatiquement. C'est un des principaux enseignements de nos recherches. L'outil sert ainsi toujours de support à l'analyse humaine. Il est possible d'automatiser la détection de certains signaux, mais l'obtention d'informations stratégiques nécessitera une méthodologie avancée déclinée par un expert. Pour répondre à des enjeux stratégiques précis, les travaux de recherches s'appuyant sur les données issues de Twitter ou d'autres réseaux sociaux devront définir des méthodologies adaptées à la fois à la problématique posée ET au secteur concerné. La façon dont les parties prenantes, en particulier les leaders go-between, interviennent sur le réseau impactera in fine l'analyse. Si Twitter semble le seul réseau social pertinent pour la détection de signaux faibles, car il permet à la fois un accès exhaustif aux données et une exploitation fine avec des outils big data de ces données, il est pourtant essentiel d'enrichir la qualification de ces signaux faibles des informations collectées par d'autres sources appropriées. Au-delà de l'exploitation des traces numériques, c'est en interrogeant des sources humaines, des experts, que l'on aura au final les informations les plus pertinentes. Ceux-ci exploitant de plus en plus les réseaux sociaux, on pourra facilement interagir avec les acteurs identifiés par l'analyse d'un signal faible pour mieux qualifier ces derniers.

Néanmoins, la recherche pour automatiser la détection de signaux faibles présente des perspectives intéressantes. Le « coefficient de manipulation Twitter » pourrait être ainsi intégré à des plateformes de veille, en identifiant des seuils pertinents selon les thématiques

surveillées, pour alerter sur une manipulation possible de l'information. L'utilisation de la Social Network Analysis pour vérifier les échanges permet ensuite de caractériser l'alerte. L'IA est appliquée aux calculs statistiques de Gephi, et non au texte des publications Twitter ou aux données personnelles des comptes du corpus. Exploiter l'IA pour analyser la dynamique des échanges Twitter, plutôt que la sémantique, semble donc une piste sérieuse pour obtenir des résultats pertinents. Des travaux de recherche ont déjà essayé d'appliquer l'IA à l'analyse des interactions Twitter, dans le cadre de la détection de Fake News (Monti, Frasca, Eynard, Mannion, Bronstein (2019). Pour notre part, nous avons effectué un projet de recherche visant à appliquer l'IA à la détection des leaders go-between, en ciblant plusieurs types de polémiques sociétales (Baulant & Sylvestre, 2020). Plusieurs dizaines d'algorithmes de machine learning ont été testés pour maximiser les chances de succès de ce travail de recherche. Les résultats obtenus indiquent une possibilité prochaine d'automatiser la détection de certains signaux faibles liés aux échanges Twitter.

Bibliographie

Alloing C., Moinet N. (2016). « Les signaux faibles : du mythe à la mystification », 2016/3 (n° 76), p. 86-92.

Alloing C., Pierre J. (2017). « Le Web Affectif : Une économie numérique des émotions », INA.

Amisse S., Baulant C., Muller P., Vargas A. (2008). « The Concept of Cluster and its Two Logics ». EAEPE Conference Labour, Institution in a Global Knowledge Economy, Rome.

Ansoff Igor. (1984). « Implanting strategic management ». Prentice / Hall International.

Asur S., Huberman B.a. (2010). « Predicting the Future with Social Media », Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Washington, vol. 1, pp. 492-499.

Baulant C. & Sylvestre G. (2021) Analyse des Réseaux Sociaux pour améliorer l'Anticipation et l'Adaptation de la réponse d'ordre public, rapport scientifique Projet RSSA, Centre des Hautes Études du Ministère de l'Intérieur, mai 2021, 33 pages.

Baulant C. (2017). « How Happiness can lead to more Efficiency: A New Paradigm Adapted to the World Knowledge Economy », The *American Review of Political Economy, pp110-125.*

Baulant C. (2017). « Rethinking the links between Human Relationships and Economic Efficiency using the Local Micro institutions: the case of two emerging countries », *Journal of Economics Issues*, 51-3, pp 651-662.

Baulant C. (2015). « The Role of Networks for Helping Firms and Countries Invent New Competitive Strategies Well Adapted to the World Knowledge Economy », *Journal of Economics Issues*, 49-2, p 563- 573.

Beauvisage T., Beuscart J-S. Couronné T., Mellet K. (2011). « Le succès sur Internet

repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », *Tracés. Revue de Sciences Humaines*, pp 151-166.

Blondel M., Kazuhiro S., Kuniaki U. (2013). « Block coordinate descent algorithms for large- scale sparse multiclass classification. » *Machine Learning* 93.1 :31-52.

Blondel V., Guillaume J-L, Lambiotte R., Lefebvre E. (2008). « Fast unfolding of communities in large networks », J. *Stat. Mech*. P10008.

Bollen J., Mao H., Zhang X.-J. (2010) « Twitter moods predicts the stock market », *Journal of Computational Science*, 2 (1), pp. 1-8.

Boullier D., Lohard A. (2012). « Opinion mining et Sentiment analysis », *Collection Sciences Po- médialab*.

Boyadjian J., Velcin .J. (2017). « L'analyse quantitative des médias sociaux, une alternative aux enquêtes déclaratives ?. La mesure de la popularité des personnalités politiques sur Twitter », *Questions de communication*, 2017/1 (n° 31), p. 111-135.

Cha M., Haddadi H., Benevenuto F., Gummadi K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy., *Conference: Proceedings of the Fourth International Conference on Weblogs and Social Media*, ICWSM 2010, Washington, DC, USA, May 23- 26, 2010.

Crozier, Michel. (1977). L'Acteur et le Système. (en collaboration avec Erhard Friedberg), Paris, Le Seuil.

Goleman D. (2011). « What Makes a Leader? », *On leadership, HBR's 10 Must Reads*, Boston, MA, Harvard Business Review Press, 1996, second edition, 2011, p 1-22.

Grandjean M. (2016). A Social Network Analysis of Twitter: Mapping the digital humanities community, *Cogent Arts & Humanities* 3: 1171458

Grandjean M., Jacomy M., Girard M. (2017). Visual Network Analysis with Gephi, *Digital Humanities*, Krakow, Poland. pp.930-931

Granovetter, MS. (1973). « The strength of weak ties », *American Journal of Sociology*, 1973, 78(6): 1360-1380.

Guilain Y. (2015). Le Seeking Mantra au service de l'e-reputation, *I2D – Information, données & documents* 2015/2 (Volume 52), p 46-48.

Kotter, John P. (2007). « Leading Change: Why Transformation Effort Fails. » *Harvard Business Review*, January, 96-103.

Jacomy M, Venturini T, Heymann S, Bastian M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* 9(6): e98679. doi:10.1371/journal.pone.0098679.

Jammet, T. (2018). Vers une communication de marque dictée par les algorithmes ? Les relations publiques 2.0 face au Big Data. *Communication et organisation*, n°54, p.93-105.

Lesca, H., Schuler, M. (1995) Veille stratégique : comment ne pas être noyé sous les informations ? », *Actes du Colloque : VSST'95*, Toulouse, pp.45.

Keller F., Schoch D., Stier S., & J Yang J. (2020). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign, *Political Communication*, 37:2, 256-280.

Monti F., Frasca F., Eynard D., Mannion D., Bronstein MM. (2019). Fake News Detection on Social Media using Geometric Deep Learning, *Arxiv*, https://arxiv.org/abs/1902.06673

Muller P., Penin J. (2008). Why Do Firms Disclose Knowledge and How Does It Matter?. *Journal of Evolutionary Economics*, pp.85-108.

Nimmo B. (2019). « Measuring Traffic Manipulation on Twitter. » Working Paper 2019.1. Oxford, UK: *Project on Computational Propaganda*. comprop.oii.ox.ac.uk. 35 pp.

Phillips L., Dowling C., Shaffer K, Hodas N., Volkova S. (2017). *Using Social Media to Predict the Future: A Systematic Literature Review*, *Arxiv*, https://arxiv.org/abs/1706.06134.

Rallet A., Torre, A. (2004), Proximité et localisation, *Economie rurale* n°280, p 25- 41.

Rieder B., Smyrnaios N. (2012). Pluralisme et infomédiation sociale de l'actualité : le cas de Twitter, Réseaux 2012/6 (n° 176), 105-139.

Rogers R. (2015). Digital Methods for Web Research, Emerging Trends in the Social and Behavioral Sciences.

Smith M., Rainie L., Shneiderman B., Himelboim I. (2014). Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters, Pew Research Center.

Sylvestre G. (2016). Détecter des signaux faibles en cartographiant des tweets, *I2D – Information, données & documents 2016/3 (Volume 53)*, p 12-15.

Sylvestre G. (2020). La Social Graph Due Diligence sur Twitter, une méthodologie d'analyse stratégique qui combine algorithmes de sciences sociales, datavisualisation et expertise humaine, *Cahiers de la Documentation*, 2020/1.

Vanderbiest N. (2018) « Les institutions démocratiques : l'influence des réseaux sociaux durant une élection présidentielle », dans : Stéphane Taillat éd., *La Cyberdéfense. Politique de l'espace numérique*. Paris, Armand Colin, « U », p. 181-188.