

LA SOCIAL GRAPH DUE DILIGENCE SUR TWITTER, UNE MÉTHODOLOGIE D'ANALYSE STRATÉGIQUE QUI COMBINE ALGORITHMES DE SCIENCES SOCIALES, DATAVISUALISATION ET EXPERTISE HUMAINE

Guillaume SYLVESTRE

Directeur de l'innovation numérique à l'Agence pour la Diffusion de l'Information Technologique (ADIT)

■ Les nouveaux secteurs technologiques innovants promettent monts et merveilles, mais l'atterrissage est parfois rude, comme on l'a encore vu récemment avec We Work. Outre la complexité des modèles économiques des startups, les bonnes idées ne déboucheront pas toujours sur des réussites commerciales. Nous proposons ici une méthodologie pour identifier les startups prometteuses via la cartographie de leurs interactions sociales. Il s'agit de vérifier la capacité d'une startup à influencer les acteurs de son secteur, à partir de l'analyse de la dynamique des échanges sur Twitter interprétée via la théorie des graphes. Les algorithmes de l'outil de datavisualisation *Gephi* permettent en effet d'analyser précisément les interactions d'un corpus Twitter pour identifier les acteurs réellement influents. La cartographie de *Gephi* permet ainsi de vérifier si une startup arrive à convaincre au-delà de son réseau propre d'experts et d'investisseurs, et de qualifier la pertinence de ses interactions.

■ De innovatieve nieuwe technologische sectoren beloven gouden bergen, maar de realiteit is soms hard, zoals we onlangs nog konden zien met We Work. Afgezien van de complexiteit van de economische modellen van de start-ups, leiden goede ideeën niet altijd tot een commercieel succes. Wij stellen hier een methodologie voor om de veelbelovende start-ups te identificeren door hun sociale interacties in kaart te brengen. Het komt erop aan om het vermogen van een start-up als influencer op de actoren van zijn sector te verifiëren, door de dynamiek van de uitwisselingen op Twitter te analyseren, die geïnterpreteerd wordt via de grafentheorie. De algoritmen van Gephi, het instrument om data te visualiseren, maken het inderdaad mogelijk om de interacties van een Twitter-corpus nauwkeurig te analyseren om te identificeren welke actoren effectief influencers zijn. Met de cartografie van Gephi is het zo mogelijk om te verifiëren of een start-up erin slaagt om mensen buiten zijn eigen netwerk van experts en beleggers te overtuigen, een om de relevantie van zijn interacties te kwalificeren.

Pour cette troisième analyse sur la détection des signaux faibles via Twitter¹, j'ai choisi de m'intéresser à la blockchain. Innovation popularisée via le bitcoin dont le cours suit les invectives du patron de la banque JPMorgan (Jamie Dimon avait effectué une série de sorties publiques contre le Bitcoin fin 2017, juste avant que la monnaie atteigne la valeur record de 20.000 \$²), la blockchain peine néanmoins à se démocratiser malgré toutes ses promesses ... car sans doute plus que dans d'autres secteurs innovants, il est pour le moins ardu de différencier les projets sérieux de ceux qui surfent sur cette mode pour lever des fonds. Je propose dans cet article une méthodologie pour évaluer la crédibilité des startups de ce secteur à partir de leurs interactions sur le réseau social Twitter

La blockchain, secteur prometteur pour l'innovation mais dont les applications peinent à se démocratiser

Une rapide définition pour commencer, via Blockchain France : "La blockchain est une technologie de stockage et de transmission d'informations transparentes, sécurisées, qui fonctionnent sans organe central de contrôle". Cette technologie repose sur la décentralisation du contrôle des actions effectuées, via un système peer to peer, qui permet donc de se passer d'une autorité de contrôle centralisée³. On peut donc, par extension, imaginer de l'utiliser pour toute action /

transaction nécessitant des opérations de contrôle, de vérification ... longues et coûteuses. Le secteur bancaire s'est retrouvé le premier concerné avec la création du bitcoin en 2008, première "crypto-monnaie" numérique s'appuyant sur la blockchain et fonctionnant sans autorité de régulation.

Si la technologie et ses promesses n'ont rien de récent, les applications destinées au grand public tardent à venir ... quand elles ne rencontrent pas le succès espéré. Axa avait été précurseur en lançant en 2017 *Fizzy*, une assurance avec paiement automatique pour couvrir les retards d'avion, issue d'un partenariat avec la startup lilloise Utocat. Le produit était un "smart contract", et s'appuyait sur la blockchain pour automatiser le paiement des souscripteurs dès qu'un retard d'avion de plus de deux heures était détecté via les données de Flightstats. Axa a mis fin au service en novembre 2019⁴, faute d'un nombre de souscripteurs suffisants, la police d'assurance couvrant par ailleurs un nombre de vols limités.

Il est donc difficile d'identifier aujourd'hui de réelles applications prometteuses car susciter l'intérêt des utilisateurs potentiels reste très compliqué. Si on s'intéresse aux mentions Twitter, il était d'ailleurs impossible pour Axa d'obtenir de la visibilité avec ce produit sans créer un compte dédié ... et in fine faute de créer des communautés en ligne d'utilisateurs sur

Twitter et/ou LinkedIn pour populariser la solution, celle-ci a rencontré un succès limité.

Les échanges sur Twitter, un big data complexe mais exploitable via l'analyse des graphes pour détecter des tendances émergentes

Je propose d'utiliser les conversations Twitter, analysées grâce à l'outil de veille *Visibrain*⁵, pour identifier les startups prometteuses. Dans un contexte où l'enjeu des startups est d'impressionner des investisseurs potentiels avec les bons "buzz words" et suffisamment de poudre aux yeux, Twitter est essentiel pour communiquer. Il faut bien évidemment se méfier des "metrics" usuels des réseaux sociaux : les retweets et les abonnements s'achètent, les comptes influents sont souvent liés à de gros acteurs très visibles (CAC 40 en France par exemple) mais au final peu innovants (les banques en étant un bon exemple sur les Fintech). On peut également faire appel à des agences de communication pour générer des interactions artificielles ... Il faut donc analyser plus en profondeur ces éléments pour leur donner du sens.

À noter que si LinkedIn va être de plus en plus utilisé pour la communication des entreprises, et notamment des startups, seul Twitter à ce jour permet d'obtenir l'accès à l'intégralité des échanges publics. On ne pourra donc faire d'analyse précise des tendances qu'avec Twitter, LinkedIn permettant ensuite de confirmer les enseignements des tweets.

La littérature scientifique sur l'utilisation des réseaux sociaux pour la prédiction⁶ indique l'importance de filtrer le bruit, de prendre en compte les biais propres aux réseaux sociaux, et d'adapter le modèle d'analyse à la spécificité des sujets traités :

- 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. Il faut arriver à supprimer ce bruit ;
- 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.

Je m'intéresse depuis plusieurs années à la détection des signaux faibles via la cartographie des échanges sur les réseaux sociaux par la théorie des graphes⁷. En effet, identifier par la seule analyse sémantique des nouveaux sujets est une promesse récurrente des outils de veille, mais ceux-ci ne permettent d'obtenir que des informations imprécises. Certains experts jugent d'ailleurs que la détection des signaux faibles est un mirage⁸ ... sans méthodologie spécifique et sélection précise des éléments à analyser, aucun outil ne permettra d'obtenir un signal pertinent.

Mon expérience d'analyse des données Twitter via le logiciel de datavisualisation *Gephi*⁹ et ses algorithmes issus de travaux en sciences sociales m'a permis de constater qu'il est possible de retrouver via la théorie des graphes des caractéristiques propres aux comptes Twitter dont le réseau / les sujets prennent de l'ampleur. Néanmoins, chaque thématique doit être analysée sous un prisme spécifique, ce qui rend très complexe l'automatisation de la détection d'informations pertinentes.

Gephi, outil d'analyse des graphes par des algorithmes de sciences sociales

Invisibles via des données quantitatives (comptes les plus mentionnés, hashtags les plus repris, voir thématiques émergentes / trending topics), les données liées à la détection de communautés et à leurs caractéristiques aident à détecter de nouveaux acteurs. Ces échanges ne seront pas visibles du grand public, mais auront su capter l'attention d'experts ou d'autres acteurs du secteur, améliorant ainsi le score des comptes émetteurs dans les différents algorithmes d'influence proposés par *Gephi*.

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)¹⁰. La position des comptes Twitter est déterminée par l'algorithme de spatialisation Force Atlas 2¹¹, qui met en évidence les communautés correspondant aux interactions.

Gephi permet en outre d'utiliser plusieurs algorithmes pour analyser les interactions entre les tweets étudiés :

- Modularity Class¹² : identifier des communautés au sein des échanges sur Twitter. C'est un calcul statistique qui décompose le réseau en sous-structures selon un score affecté à chaque entité. Chaque compte Twitter d'un corpus étudié sera ainsi affecté à une communauté selon les comptes avec lesquels il échange. Sur des corpus de plusieurs dizaines de milliers

de tweets, *Gephi* identifiera de 500 à 15000 communautés différentes, mais entre 50 et 75 % des échanges vont correspondre aux 10 premières communautés. On peut ainsi avoir un premier niveau d'analyse si ce score moyen évolue dans le corpus étudié (comme ce sera le cas pour cette étude sur la blockchain).

- **Eigenvector Centrality** : prendre en compte l'influence des comptes Twitter au sein de leur propre communauté. Cet algorithme 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). Cela permet de mettre en évidence les comptes Twitter très mentionnés (notamment par d'autres comptes très mentionnés) tout en effaçant les comptes très actifs mais peu repris par d'autres utilisateurs.
- **Betweenness Centrality** : identifier les acteurs au centre des échanges entre les différentes communautés d'un corpus de tweets. Cet algorithme permet de valoriser les acteurs au centre des échanges entre les communautés, sans être pour autant très mentionnés. Il est important pour détecter des comptes Twitter peu repris mais influents en dehors de leur réseau, mais il peut générer des faux positifs (un compte Twitter membre d'une petite communauté ayant deux ou trois mentions d'autres communautés aura un score très important). Il faut donc vérifier les résultats proposés par *Gephi* avec les données sources (issus de *Visibrain* en l'occurrence).

C'est la combinaison de ces algorithmes qui permet d'identifier des comptes Twitter pertinents dans la détection des signaux faibles. À noter que l'outil *Gephi*, s'appuyant sur la dynamique des échanges au sein d'un corpus Twitter, n'exploite pas les métadonnées personnelles des comptes Twitter : bio, abonnés, abonnements, localisation, date et heure des tweets ...

Un premier exemple correspondant à une analyse des tweets français sur diverses innovations technologiques en avril 2017 : la cartographie à gauche des comptes Twitter les plus influents dans leur réseau (Eigenvector Centrality) montre plusieurs acteurs, notamment des médias, qui bénéficient d'une forte audience. La cartographie à droite, qui utilise le score de centralité dans les échanges (Betweenness Centrality) et que l'on a filtrée pour plus de lisibilité, montre un résultat différent : certains des comptes qui paraissent très influents car ayant beaucoup de mentions, ne sont pas repris en dehors de leur communauté. D'autres, moins mentionnés, arriveront à toucher d'autres réseaux (surlignés sur la carto de droite). Les vrais influenceurs sont ceux qui ont à la fois un score important d'Eigenvector Centrality et de Betweenness Centrality (surlignés sur les cartos).

Pour aller plus loin et identifier des signaux faibles dans des secteurs technologiques, il est nécessaire de compléter les algorithmes de *Gephi* avec une expertise humaine pour sélectionner des startups pertinentes à étudier. Ce n'est en effet qu'à partir d'une connaissance préalable d'un sujet et des acteurs qui gravitent autour, que l'on pourra identifier précisément de potentiels signaux faibles.

Il faudra donc identifier ponctuellement les nouveaux entrants d'un secteur innovant pour ensuite analyser leur positionnement dans les échanges Twitter via *Gephi*, afin d'analyser si leur influence sur Twitter et donc leur crédibilité évolue ou non en fonction de leur actualité. Twitter est un outil de communication, où on retrouvera la participation à des événements divers et variés pour promouvoir une startup, établir des partenariats, lever des fonds ... et, selon le succès de ces événements, des interactions nouvelles naîtront, que l'on pourra analyser avec *Gephi* pour en évaluer la qualité.

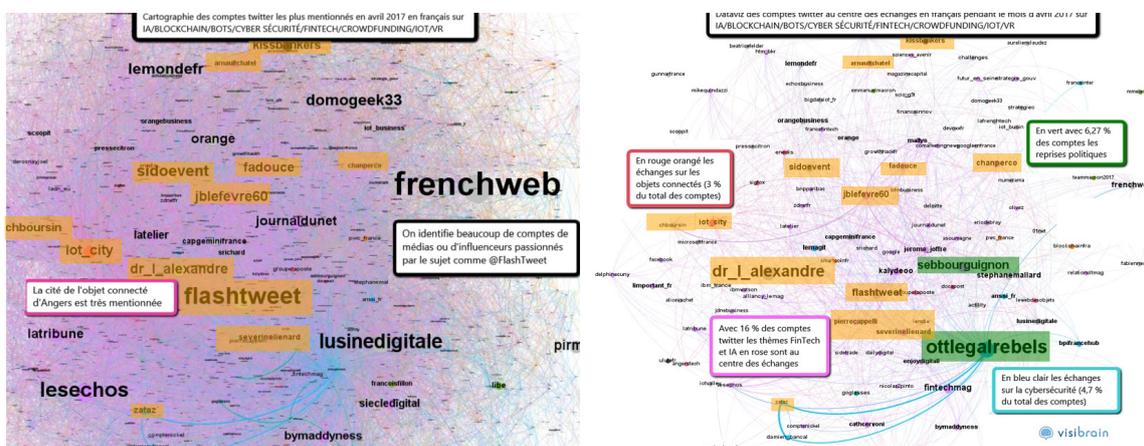


Fig. 1 : Identification de comptes twitter influents par la cartographie des échanges

Les tweets en français sur la blockchain ont diminué entre 2017 et 2019

Dans un premier temps, il est intéressant d'étudier l'évolution des volumes de tweets échangés pour comprendre la dynamique autour d'une thématique. La blockchain est de moins en moins mentionnée dans les tweets français : on passe de près de 240.000 mentions entre le 1er juin et le 31 décembre 2017, à moins de 200.000 du 1er mai au 1er décembre 2019, identifiées via le logiciel de veille sur Twitter *Visibrain*.

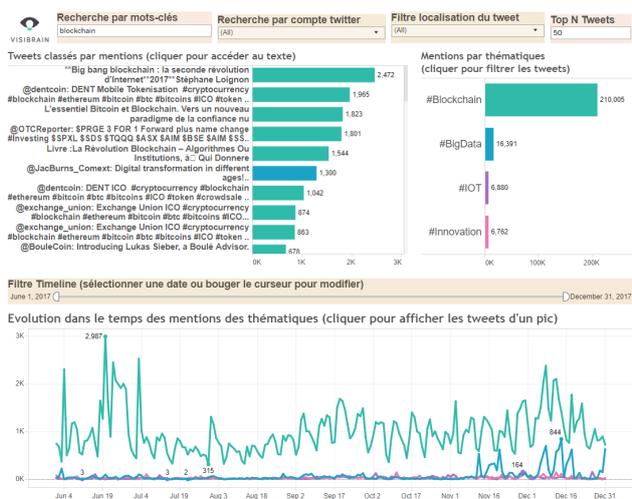


Fig. 2 : classement des tweets les plus mentionnés en français sur la blockchain de 01/06/2017 au 31/12/2017

Les mentions du bitcoin déclinent également, avec près de 60.000 mentions sur la période de 2017, et près de 40.000 mentions entre le 1er mai et le 1er décembre 2019.

Pour obtenir une analyse plus fine, je vais filtrer toutes les mentions liées au bitcoin et aux autres crypto-monnaies, qui risquent de rendre plus difficile l'identification de signaux faibles sur la blockchain. Cela ne veut pas dire qu'il n'y a pas d'innovations dans ce domaine mais les crypto-monnaies ne sont plus un signal faible aujourd'hui. Ce sont les autres applications potentielles de la blockchain qu'il serait intéressant de découvrir. Le réel potentiel de la blockchain n'est pas tant sans sa supposée infaillibilité que sa capacité à simplifier les transactions, voire à produire plus de valeur ajoutée pour celles-ci.

Dans le cadre de cette analyse, je prendrai comme périodes de référence les tweets en français des mois de mai 2019 et septembre-octobre-novembre 2019. Après avoir identifié des startups qui me semblent intéressantes, je vais analyser via les algorithmes de Gephi l'évolution de leur influence sur Twitter sur ces deux périodes afin d'identifier celles qui gagnent

en visibilité, alors que, comme on l'a vu, le volume total de tweets sur la blockchain décline.

Du CES (janvier 2019) à Vivatech en mai 2019 : peu de startups françaises sur la blockchain captent l'intérêt des influenceurs

N'étant pas expert en blockchain, je suis parti des entreprises françaises mentionnées pendant le CES 2019 comme point de départ de cette analyse. La société française la plus mentionnée sur la blockchain ne propose d'ailleurs pas elle-même de service innovant : *Blockchain Studio* est un outil qui aide à créer des applications exploitant la blockchain. Il sera pour autant intéressant de suivre l'évolution de son réseau Twitter.

Le journaliste du magazine Capital Grégory Raymond cite, dans un des seuls tweets en français sur la Blockchain mentionnés plus de 10 fois, d'autres sociétés plutôt axées sur les crypto monnaies, où la startup BCDiploma, qui propose de certifier les diplômes, apparaît.

Enfin, on notera un tweet de la société BTU Protocol, qui propose un service de réservation d'hôtels s'appuyant sur la blockchain, et Embleema, qui protège les données médicales via la blockchain.

Ce sont donc les startups suivantes : Embleema, BTUProtocol, BCDiploma et BlockchainStud que je vais analyser, afin d'identifier si leur influence sur les échanges concernant la blockchain a évolué ou non depuis janvier 2019.

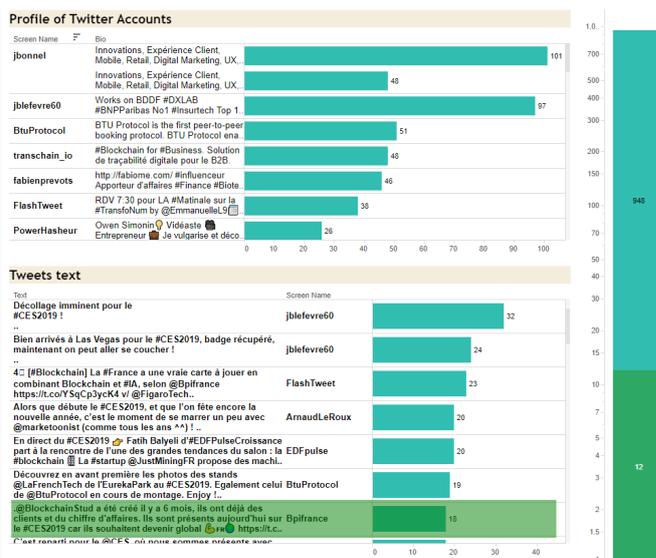


Fig. 3 : Identification des tweets et des comptes actifs sur la Blockchain en français durant le CES 2019



Fig. 4 : Tweet du journaliste Grégory Raymond durant le CES 2019

Pendant Vivatech en mai 2019, les startups blockchain restent peu influentes sur Twitter

Sur le mois de mai 2019, période où s'est tenu le salon Vivatech à Paris, le logiciel de veille Visibrain identifie 17.568 tweets en français sur la blockchain, en filtrant les mentions liées aux crypto-monnaies (29 689 mentions sans ce filtre). Il s'agit à 55 % de tweets originaux, publiés par 7432 utilisateurs.

L'analyse des communautés par Gephi montre un très fort morcellement des échanges, avec plus de 1.500 communautés détectées. Les 8 principales communautés, automatiquement coloriées par le logiciel, représentent entre 30 et 33 % des comptes Twitter. Sur des corpus de plusieurs dizaines de milliers de tweets, ces 8 communautés correspondent normalement à plus de la moitié des comptes, les échanges Twitter étant en général très polarisés. Cela signifie que la majorité des comptes actifs sur la blockchain ne sont pas repris en dehors de leur communauté, et sont donc très peu influents sur

les échanges. On peut le constater en comparant les scores d'Eigenvector Centrality (influence par les mentions) et Betweenness Centrality (capacité à toucher d'autres communautés).

Sans surprise, les startups qui nous intéressent sont très peu visibles. Seules BCDiploma et Embleema ont un début de communauté structurée, mais elles sont peu influentes sur les échanges. BTUProtocol et BlockchainStud quant à elles ont été très peu mentionnées.

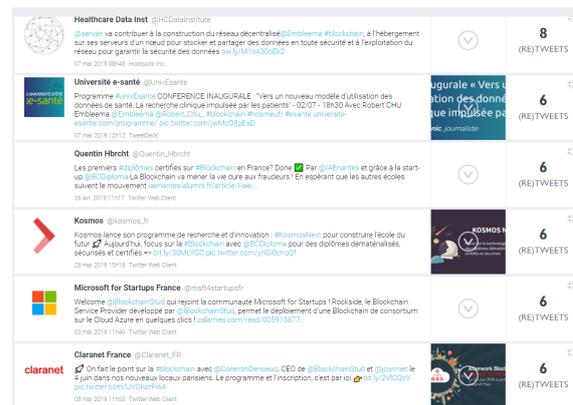


Fig. 6 : classement via le logiciel de veille Visibrain des tweets les plus mentionnés sur les comptes étudiés

Embleema a en effet noué un partenariat avec Servier, et a donné une conférence à l'université d'été de la e-santé le 7 mai 2019. Cet évènement a été twitté plusieurs fois par les organisateurs, générant quelques mentions. De la même façon, les partenariats de BCDiploma avec l'IAE Nantes et

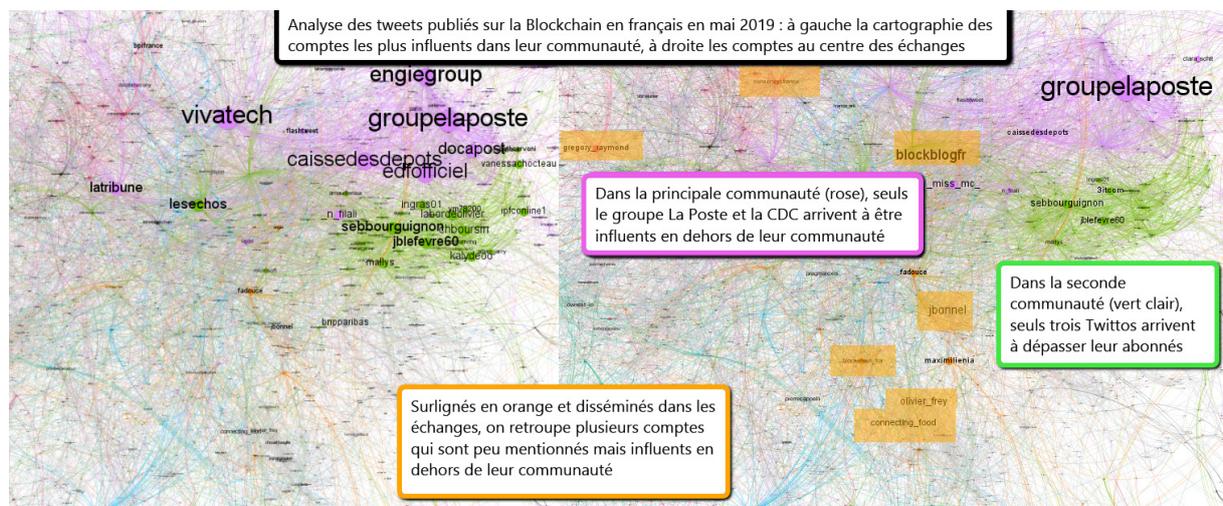


Figure 5 : identification des comptes et communautés françaises les plus influentes sur la Blockchain en mai 2019

Kosmos pour certifier les diplômes ont aussi généré quelques tweets. A contrario, bien que membre de la communauté " Microsoft for Startups France ", BlockchainStud a eu moins de visibilité, confirmant que le soutien par des noms très connus n'est pas une garantie de visibilité.

soit le double du mois de mai. Cela signifie que de nombreux twittos vont échanger ponctuellement sur la blockchain, sans pour autant s'impliquer sur le sujet. Ces nouveaux acteurs contribuent au morcellement des échanges, puisque Gephi détecte ici plus de 2.500 communautés, les huit premières totalisant toujours entre 30 et 35 % des échanges.



Fig. 7 : comparaison via Gephi des interactions Twitter entre les différents comptes cibles

L'analyse via Gephi ne montre pas de différence notable entre les réseaux de chaque startup, qu'on utilise l'algorithme valorisant les comptes les plus mentionnés (Eigenvector Centrality) ou ceux qui sont au centre des échanges (Betweenness Centrality). Aucune de ces startups ne se démarque vraiment, elles ne touchent que des réseaux liés aux évènements et partenariats qu'elles nouent.

J'ai surligné en orange les comptes ayant un score de Betweenness Centrality et d'Eigenvector Centrality importants, qui sont donc les plus influents sur le sujet. On notera ainsi, outre divers experts très actifs sur le sujet, l'importance du compte @BlockchainPartn, société de conseil sur la blockchain cofondée par Clément Jeanneau¹³.

Les réseaux de nos startups sont ici plus développés, en particulier pour BCDiploma et Embleema (en vert sur les cartos). La startup qui propose des solutions pour la certification des diplômes se retrouve ainsi avec un score élevé de Betweenness Centrality, ce qui signifie que @BCDiploma a été mentionnée en dehors de son réseau. @Embleema, de son côté, fait partie des comptes les plus influents, ayant été mentionnée par certains comptes très repris sur la Blockchain.

Deux startups commencent à avoir de l'influence sur les échanges Twitter et LinkedIn entre septembre et novembre 2019

Sur notre seconde période d'étude, 46 897 tweets ont été publiés avec le filtre (78.281 sans). On notera donc que Vivatech n'a pas réellement eu d'impact sur les mentions de la blockchain puisque la moyenne mensuelle de tweets publiés n'évolue pas. Il y a ici 53 % de tweets uniques, pour 13 835 utilisateurs,

On notera que si Embleema a un excellent score d'Eigenvector Centrality, c'est parce que la mention d'Eric Gaubert (directeur adjoint à l'innovation de Reinsurance Group of America) est très reprise par l'écosystème d'influenceurs Fintech français qui suit le sujet blockchain via notamment @jblefevre60 et @sebbourguignon. Pour autant, leurs retweets ne dépassent par leur réseau.

BCDiploma touche différentes communautés du monde de l'éducation grâce à Olivier Rollot (directeur exécutif HEADway) et l'Association Universitaire de Francophonie avec qui elle a noué un partenariat, ainsi que la CCI Hauts-de-Seine. Les communautés

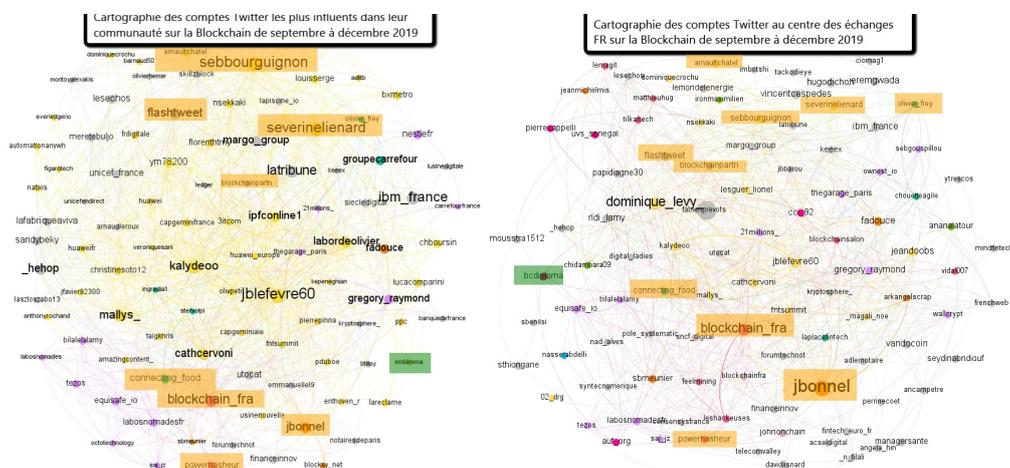


Fig. 8 : Identification des comptes Twitter français les plus influents sur la blockchain entre septembre et novembre 2019

	<p>Eric Gaubert at #HCE2019 @kalydeco</p> <p>La #startup de la semaine : @Embleema, la #blockchain entre patients et chercheurs https://t.co/PSTOLq3d via @la Tribune cc @seebourguignon @FGrillot @malys_ @plefevre60 @labondeviver #data #sante</p> <p>05 nov. 2019 18:52 Twitter for iPhone</p>		<p>14 (RE)TWEETS</p>
	<p>AUF @auf_org</p> <p>L'AUAF propose désormais à ses établissements membres, en partenariat avec @BCDiploma, un service #blockchain d'authentification numérique des #diplômes. a neuf auf.org/dineuf</p> <p>23 nov. 2019 10:09 Twitter Web App</p>		<p>12 (RE)TWEETS</p>
	<p>CCI Hauts-de-Seine @CCI_92</p> <p>#Maden92 @VilleDeGarches 1/2 finale Nord = Prix #innovation attribué à Luc Jarry-Lacombe pour @BCDiploma @villedeuel. Solution blockchain qui permet d'émettre et de partager des certificats numériques infalsifiables. Fin la fraude ! @Airbus @Entrepreneurs2 @Mede92 @cpmes2</p> <p>14 nov. 2019 20:41 Twitter Web App</p>		<p>9 (RE)TWEETS</p>
	<p>Luc Jarry-Lacombe, Cofounder & CEO @BCDiploma</p> <p>#blockchain. J'ai rejoint deux beaux dossiers dans la blockchain. La start-up @BCDiploma.com qui sécurise et certifie les diplômes et @vysalchain.fr qui, elle, garantit grâce à la blockchain la traçabilité des viandes. @ChantalBaudron1 forbes.fr/femmes-at-forb</p> <p>17 sept. 2019 15:55 Twitter Web App</p>		<p>7 (RE)TWEETS</p>
	<p>Olivier Rollot @O_Rollot</p> <p>Comment @BCDiploma BCDiploma garantit les diplômes sur la Blockchain blog headway-advisory.com/comment-bcdipl_pdc.twttr.com/vjGjggxPK</p> <p>04 nov. 2019 10:05 Twitter Web App</p>		<p>6 (RE)TWEETS</p>
	<p>AUF @auf_org</p> <p>En partenariat avec @BCDiploma, l'@auf_org propose à ses plus de 900 universités membres à travers le monde de déployer gratuitement l'application #BCDiploma http://twttr.com/O_Rollot/statuses/1191280303668490240</p> <p>04 nov. 2019 10:05 Twitter Web App</p>		<p>6 (RE)TWEETS</p>

Fig. 9 : Les startups @Embleema et @BCDiploma ont gagné en visibilité avec des mentions d'experts

respectives de ces acteurs ont aidé la startup à gagner en visibilité sur Twitter, d'autant plus qu'il s'agit d'acteurs locaux crédibles. A noter que dans le cadre d'un projet de recherche en cours du laboratoire Granem de l'université d'Angers, nous constatons que c'est la reprise d'une information sur Twitter par des leaders locaux qui en favorise la viralité, au détriment des "influenceurs" très suivis mais pas ou peu repris en dehors de leur communauté¹⁴.

Les deux startups sont par ailleurs très actives via LinkedIn, réseau social qu'elles privilégient pour partager leur actualité. Embleema dispose de plus d'abonnés que BCDiploma (1309 contre 578), et ses publications récentes suscitent de nombreuses réactions de professionnels de la santé et de la Tech. En complétant l'analyse Twitter des données issues des interactions LinkedIn, on peut donc considérer qu'Embleema est également capable de toucher plusieurs communautés qui jugent la startup crédible.

La Social Graph Due Diligence, nouvel outil d'aide à la prise de décision stratégique

Il est essentiel pour les entreprises et investisseurs européens de développer leurs propres méthodologies d'identification des sociétés innovantes performantes pour investir. Dans ce domaine, les GAFAM¹⁵ ont un énorme avantage : en se partageant la moitié

de l'hébergement cloud mondial, il est aisé pour eux d'identifier les plateformes innovantes qui ont du succès. N'oublions pas qu'Alexa est un outil d'analyse de la performance des sites web proposé par Amazon¹⁶ avant d'être un assistant vocal, et que Google Analytics reste l'un des outils préférés des webmasters ... Il faut donc développer d'autres méthodologies pour identifier les startups à succès faute d'accès à ces informations.

La Social Graph Due Diligence est une nouvelle méthodologie que je propose pour vérifier le potentiel des startups dans un secteur innovant. Twitter et LinkedIn seront en effet des outils essentiels pour valoriser leur activité, et les données d'interactions y sont très riches tout en étant liées à l'activité réelle des acteurs. Ce sont donc des indicateurs précieux pour vérifier la capacité d'une startup à se développer.

On pourra ainsi, en croisant les données de deux réseaux sociaux avec l'analyse de graphe via Gephi, vérifier si une startup arrive à générer des interactions naturelles avec des experts. A contrario, l'utilisation de faux comptes ou de robots sera immédiatement visible et devra être interprétée comme un signal négatif. La croissance de l'influence sur Twitter, accompagnée de l'évolution des interactions LinkedIn permettront de confirmer si le "signal faible" lié à une startup est bien pertinent.

Une vérification rapide et efficace, qui peut se révéler critique en cas de doute sur un investissement / une acquisition stratégique ... à compléter évidemment par une analyse critique des technologies mises en œuvre et de leur efficacité avérée, au-delà des belles promesses !

Guillaume Sylvestre
 Agence pour la Diffusion de
 l'Information Technologique (ADIT)
 62 rue de Miromesnil
 75008 Paris
 gsy@paris.adit.fr
 www.adit.fr

Décembre 2019

Notes

1. Deux exemples (Fintech et Cybersécurité) sont accessibles sur mon blog d'analyse des réseaux sociaux, Cartorezo, <<https://cartorezo.wordpress.com/>>
2. Agini, Samuel. Jamie Dimon vs bitcoin vs JPMorgan: a timeline of tantrums. *Fnlondon.com* [en ligne]. 15/02/2019. [Consulté le 24/12/2019] Disponible à l'adresse : <<https://www.fnlonon.com/articles/jamie-dimon-vs-bitcoin-vs-jpmorgan-a-timeline-of-tantrums-20190215>>
3. Pour plus d'information sur la Blockchain et ses enjeux : Brevet, Guillaume et Rioche, Julian. Les enjeux d'une économie positive : la blockchain. *Master Intelligence Economique et Stratégies Compétitives* [en ligne]. 19/02/2017. [Consulté le 24/12/2019] Disponible à l'adresse : <<https://master-iesc-angers.com/les-enjeux-dune-economie-positive-la-blockchain/>>
4. Raynal, Juliette. Clap de fin pour Fizzy, l'application phare d'Axa dans la Blockchain. *La Tribune* [en ligne]. 08/11/2019. [Consulté le 24/12/2019] Disponible à l'adresse : <<https://www.latribune.fr/entreprises-finance/banques-finance/assurance/clap-de-fin-pour-fizzy-l-application-phare-d-axa-dans-la-blockchain-832676.html>>
5. Seul logiciel de veille Twitter proposant des exports adaptés à la cartographie : <<https://www.visibrain.com/fr/>>
6. Phillips, Lawrence et Al. Using Social Media To Predict the Future : A Systematic Literature Review. *ArXiv* [en ligne]. 19/06/2017. [Consulté le 24/12/2019] Disponible à l'adresse : <<https://arxiv.org/abs/1706.06134v1>>
7. Sylvestre, Guillaume. Détecter des signaux faibles en cartographiant des tweets. *I2D - Information, données & documents*. ADBS, 2016/3 (Volume 53), p. 12-15
8. Alloing, Camille et Moinet, Nicolas. Les signaux faibles : du mythe à la mystification. Hermès, *La Revue*. Novembre 2016, n° 76, CNRS Editions, p. 86-92
9. Logiciel libre et gratuit : <<https://gephi.org/>>
10. Martin Grandjean, Mathieu Jacomy, Paul Girard, *Visual Network Analysis with Gephi*. [en ligne]. 11/05/2017. [Consulté le 24/02/2019] Disponible à l'adresse : <<https://hal.archives-ouvertes.fr/hal-01525569>>
11. Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, Mathieu Bastian (2014) *ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software*. [en ligne]. 10/06/2014. [Consulté le 24/12/2019] Disponible à l'adresse <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>>
12. Blondel, Vincent D, Guillaume, Jean-Loup, Lambiotte, Renaud, Lefèbre, Etienne. Fast unfolding of communities in large networks. *ArXiv* [en ligne], 25/07/2008. [Consulté le 24/12/2019] Disponible à l'adresse : <<https://arxiv.org/abs/0803.0476>>
13. Par ailleurs rédacteur d'analyses prospectives très intéressantes sur le site Signaux faibles, <<https://signauxfaibles.co/>>
14. Voir notamment sur l'importance des leaders locaux dans les réseaux d'entreprise : Camille Baulant, *How building Efficient Networks to help Firms to invent New Innovative and Competitive Strategies?* [en ligne]. 04/11/2016. [Consulté le 24/12/2019] Disponible à l'adresse : <http://okina.univ-angers.fr/publications/ua14095/1/c_baulant_march_2015.pdf>
15. Google (8% parts de marché CLOUD mondial en 2018, 75 % de parts de marché des smartphones avec Android en 2019, 32 % de parts de marché de la publicité en ligne en 2019), Apple (25 % part de parts de marché des smartphones avec iOS en 2019), Facebook (20 % parts de marché publicité en ligne en 2019), Amazon (32 % parts de marché CLOUD mondial en 2018, 4 % parts de marché publicité en ligne en 2019), Microsoft (18 % parts de marché CLOUD mondial en 2018) – source statista.com
16. <<https://www.alexa.com/>>