

- > Az-Eddine Bennani
- > Pascal Jollivet-Courtois
- > Yann Moulrier Boutang

Identifier les signaux faibles de tendances émergentes des Fintech via l'intelligence collective du Web social ?

Une Analyse Structurale de Réseau sur le big data de fils de Reddit



- > #Numéro 1
- > Capitalisme cognitif
- > Notes de recherche
- > CRI - Complexité, Réseaux et Innovation (Costech-UTC)
- > Capitalisme cognitif - > Mégadonnées - > Méthodologie de recherche

Références de citation

Bennani, Az-Eddine., Jollivet-Courtois, Pascal., Moulrier Boutang, Yann. "Identifier les signaux faibles de tendances émergentes des Fintech via l'intelligence collective du Web social ?. Une Analyse Structurale de Réseau sur le big data de fils de Reddit", 15 juin 2017, mäj 0000, *Cahiers COSTECH*<http://www.costech.utc.fr/CahiersCOSTECH/spip.php?article42>

Résumé

Qu'est-ce que le Traitement Automatique des Langues (TAL) peut bien avoir à faire avec le programme de recherche du Capitalisme Cognitif ? Si l'on en croit Christian Marazzi beaucoup. Marazzi considère que le capitalisme contemporain prend radicalement un « tournant linguistique », dans lequel « tous les domaines où l'activité humaine est matière à langage peuvent être réquisitionnés » (4^e de couverture). Ainsi, à notre sens, le TAL serait au capitalisme cognitif (à la manière de HAL de l'Odyssée de Kubrick) ce que le moteur à explosion a été au capitalisme industriel. La mobilisation dans cet article des sciences des réseaux (Analyse Structurale de Réseau) - de surcroît appliquées à des interactions sur des Forum de discussion - se comprend plus aisément. La encore, la forme réseau -qu'elle s'instancie comme réseaux d'entreprises, entreprise-réseau ou réseau social - serait la forme organisationnelle privilégiée du capitalisme cognitif, permettant notamment le travail coopératif réticulaire (Jollivet, 2001). L'intérêt pour la détection de signaux faibles et de tendances émergences via le big data des dialogues des multitudes sur le Web social se comprend alors mieux : il s'agit bien, dans une société pollen (Jollivet & Moulier Boutang, 2010 ; Moulier-Boutang, 2009) d'accéder à une forme d'intelligence collective (ou intellectualité diffuse) où le revenu de contribution prend tout son sens. Pourquoi l'objet des Fintech cependant ? Pour Moulier-Boutang, l'industrie financière est subsumée par les dynamiques de valorisation à l'œuvre du capitalisme cognitif. L'étude des mutations en cours dans la finance, à travers des innovations sociotechniques radicales de type blockchain, peut alors enrichir ... notre compréhension des mutations plus globale du capitalisme contemporain.

Auteurs



Pascal Jollivet-Courtois est Maître de Conférence en Sciences Économiques (GNU section n°5) ses recherches, en économie de l'Innovation et économie du travail, portent sur les mutations des dynamiques et processus d'innovation dans les économies et organisations apprenantes globalisées contemporaines (paradigme du numérique et du réseau, externalisation et socialisation de l'innovation, formes réseaux et communautés de l'innovation).



Az-Eddine Bennani est professeur associé (PAST / Habilité à Diriger des Recherches) à l'UTC, chercheur au Costech-UTC. Ses recherches portent sur les systèmes de management de l'information, le commerce et les phénomènes numériques, les technologies du soin aux personnes.



Yann Moulier-Boutang Économiste, professeur à l'Université de Technologie de Compiègne et directeur de la publication de la revue "Multitudes". Il est membre du comité d'orientation de Cosmopolitiques. Auteur de "La Révolte des banlieues ou Les habits nus de la République" (2005), "Le Capitalisme cognitif : la nouvelle grande transformation" (2007), "L'Abeille et l'Économiste", (2010).

Introduction problématique : l'intelligence collective plus puissante que les experts pour la détection de signaux faibles ?

Nombreux sont les experts qui se sont jusqu'à présent exprimés sur la « révolution » du phénomène Blockchain et sur ses impacts disruptifs via les nouveaux services proposés par les Fintech, depuis 2008-2009. Mais peu de travaux ont porté sur l'analyse de ce qu'en disent les gens « ordinaires », dans leurs conversations « communes ». Quels sont les expériences, les témoignages, les ressentis, les conversations des gens qui, sans être des « sachants » institués, des journalistes spécialisés, sont *liés* à ces innovations sociotechniques (Latour, 2005) en cours des Fintech ? Un traitement adapté de ces « sources » alternatives permettrait-il de détecter des *signaux faibles* et des *tendances* émergentes auxquels l'accès n'est pas possible via des techniques conventionnelles d'expertise ou de sondages (Taleb, 2006) ? Ce document de travail rend compte d'un travail en cours, à caractère exploratoire, sous une forme volontairement concise, destiné à être développé dans une étape ultérieure.

Nous présentons quelques résultats temporaires et partiels d'une recherche commencée en 2014 (Jollivet P., Ghitalla, & Moulier-Boutang, 2015) s'inscrivant dans la continuité d'un travail initié par Ghitalla (F. Ghitalla, Diemert, Maussang, & Pfaender, 2004), visant à capter et analyser les expressions spontanées de *profanes* sur le Web social¹ - via des techniques de web-scraping, d'ingénierie linguistique, et d'analyse structurale de réseaux sémantiques - afin de détecter des signaux faibles et des émergences de tendances, socio-économiques et/ou technologiques. Nous partons en effet du postulat que les citoyens-internautes disposent d'un savoir profane spécifique et que c'est par leurs interactions, sociales et sémantiques, leurs conversations (Laugier, 2011, 2013), notamment médiées par Internet, qu'est générée une forme d'intelligence collective, susceptible d'être bien différente de celle des experts habituellement consultés (Cardon, 2010 ; Moulier-Boutang, 2007).

Nous montrons qu'il est possible de développer un accès à cette intelligence collective – ici, concernant le phénomène Fintech/Blockchain, incluant les impacts organisationnels - à travers l'écoute et l'analyse instrumentée et monitorée de Forum et Fils de discussions (ici, sur Reddit) (Jollivet P., 2016). Ces Forums constituent en effet des lieux privilégiés de réalisation d'une cognition sociale par l'interaction, une innovation par l'apprentissage par l'interaction sociale

(Lundvall, 2010), dont un accès est possible via les traces numériques laissées par les contributeurs.

1. Positionnement théorique : le constructivisme pour accéder à une intelligence collective de réseau via l'exploration de données textuelles d'interactions sociales

La littérature mobilisée ici est assez multidisciplinaire, et nous aspirons en effet à contribuer à une recherche finalisée développant la *transdisciplinarité* (Nicolescu, 1996). Ainsi, un certain éclectisme peut se dégager des références provenant tant des sciences économiques et de gestion², des sciences de l'information et de la communication, que des sciences des réseaux et des statistiques. Pourtant, il s'agit bien d'explorer ici « le paradigme » du réseau (Castells, M., 2010) – fut-il social ou sémantique, économique ou entrepreneurial - susceptible de constituer un renouveau heuristique, conceptuel ou phénoménologique (Cowan, 2005; Easley & Kleinberg, 2010; Benkler, 2006). Pour ce faire, nous nous engageons dans un cheminement de recherche *constructiviste* qui tente de travailler en coévolution les concepts, les méthodes et les données (Charmaz, 2014a). Le présent article tend cependant à se focaliser³ plutôt sur les aspects méthodologiques de cette exploration du réseau - que nous pensons originaux - à travers une étude « expérimentale » sur les Fintech/Blockchain, incluant les impacts organisationnels.

La posture théorique à travers laquelle nous réalisons notre activité d'exploration de données est donc constructiviste tout particulièrement au sens de l'école de la *théorisation ancrée* (Charmaz, 2014b). Ainsi, tout en nous appuyant fortement sur des données, textuelles et de première main, collectées en masse⁴, nous ne nous inscrivons pas dans une perspective d'induction pure, contrairement à Anderson (Sterling, 2008). Symétriquement, bien que mobilisant des outils conceptuels, nous ne suivons pas un réductionnisme hypothético-déductif si pratiqué dans les sciences économiques (Uzunidis, 2012)..

2. Méthodes mixtes : du quali-quantitatif couplant Traitement Automatique des Langues, Sciences des Réseaux et Données Massives

Les méthodes mixtes ambitionnent de dépasser le clivage entre

méthodes qualitatives versus quantitatives (Johnson & Onwuegbuzie, 2004 ; Teddlie & Tashakkori, 2008). Sur le plan méthodologique, l'approche en termes de théorisation ancrée (op. cit.) s'articule bien avec les méthodes mixtes⁵.

Nous avons mis au point une méthode originale⁶, consistant en un processus de traitement de données massives (Big Data) issues du web social et d'analyses. Ce processus s'étend depuis l'amont de la captation de données textuelles, jusqu'à l'aval de la détection de signaux faibles et d'émergences, en passant par l'étape d'interprétation de ces données et des motifs lexicaux sur différentes cartes sémantico-sociales que nous avons préalablement générées.

Plus précisément, le processus de traitement et d'analyse se déploie selon les étapes suivantes.

a) Design de requête & captation (web-scraping) sur mesure

Cette étape consiste à *traduire* la thématique sur laquelle on souhaite procéder à l'écoute (ici l'enjeu des Fintech à travers la technologie blockchain) en des requêtes – plus ou moins complexes - compréhensibles par un moteur de recherche. Dans un premier temps (2015) nous formulons nos requêtes à travers un service Web⁷ permettant une formulation en logique booléenne. Nous mettons au point une unique requête, présentant une structure relativement complexe d'expressions imbriquées constituées d'opérateurs de type booléens (ET, OU et NEAR). Aujourd'hui, nous sommes à même de passer outre ce fournisseur spécialisé en mode en passant directement par des moteurs de recherches génériques (Google, Bing ..). Cependant, ces moteurs n'offrant pas autant de richesses dans l'emploi des requêtes booléennes complexes, nous devons concevoir plusieurs requêtes différentes, que nous agrégeons après l'étape de captation.

L'expérience nous a appris qu'il convenait d'associer, dans les requêtes, d'un côté certains mots clefs assez généraux (par ex., ici, blockchain, organisation ...) pour s'ouvrir à l'imprévu de signaux faibles, avec d'un autre côté des expressions clefs assez « pointues » pour être sûr d'accéder à certains sujets précis souhaités (par ex DAO).

L'étape de web-scraping de *données massives* (big data) nécessite la mobilisation de « robots » logiciels automatisant la captation et la

transformation de pages web (de type html) en base de données (principalement textuelles). Différents services Web fournissent cette prestation. De la qualité de la matrice de captation (développée sur mesure) va dépendre la qualité des données textuelles de verbatim récoltées, et l'ampleur du travail ultérieur de curation de données.

b) Traitement linguistique monitoré

Une originalité de notre processus de traitement porte sur la profondeur du traitement linguistique effectué sur les corpus textuels issus de la base de données de captation Web. Cette profondeur n'est pas courante quand on opère dans un univers de données massives. Nous avons recours à des techniques de Traitement Automatique des Langues (Computational Linguistics) réalisant notamment de l'analyse morpho-syntaxique. Cette analyse, monitorée et paramétrable, permet d'obtenir tout d'abord une lemmatisation automatique assez fine de termes. Mais elle autorise surtout une personnalisation du processus selon le sujet traité, à travers le rajout de termes et expressions spécifiques dans un dictionnaire particulier.

De surcroit, la méthode utilisée permet de rajouter à cette catégorisation linguistique des « étiquetage » (tags) - « manuels » ou automatisés - qualifiant les unités documentaires (ici, les post des contributeurs à un fil de discussions).

Nous utilisons ainsi comme variables (dont les tags sont des instanciations) qualifiant nos post :

- la date de la contribution (« posts »),
- le titre du fil de discussion (« thread) au sein duquel la contribution a été postée,
- le forum dans lesquels ledit fil figure,
- la requête qui a permis de capter cette contribution.

c) Traitement d'analyse structurale de réseau (« Structural Network Analysis »)

L'analyse structurale de réseau est de plus en plus utilisée pour traiter les entités de type réseaux sociaux (réseaux de personnes sur LinkedIn ou Facebook ...). Il est relativement original de procéder comme nous le faisons à l'analyse structurale de réseaux sur des entités « documentaires » (ou textuelles)(Pincemin, B., Heiden, S., 2008 ; Vergne

& Swain, 2016)⁸. En se focalisant sur telle ou telle variable de notre base documentaire de corpus, nous générons différentes cartes. Ces cartes sont tout d'abord qualifiables de « sémantique » (plus strictement des cartes de « lemmes »). Pourtant, la mobilisation des variables « Titre de Fil » et surtout « Forum », permettant de remonter à un niveau plus élevée de l'arborescence de la structuration des discussions sur la plateforme Reddit, fournit quelques indications sur des communautés (sociales) qui se structurent sur certains grands thèmes⁹. C'est ainsi que nous usons parfois de l'expression « cartes socio-sémantiques ».

Mais l'intérêt majeur pour la détection de signaux faibles de l'analyse structurale de réseau textuel (parfois appelé analyse de similitudes) est qu'elle ne met pas en exergue des « tendances centrales » (à l'instar des statistiques classiques)¹⁰ mais des *dissimilitudes* d'agrégats (« clusters ») et des *singularités* de motifs (« patterns ») (Granovetter, M. S., 1973 ; Watts, 2003, 2004).

d) Analyse et détection de signaux faibles « candidats »

Une première analyse, visuelle et descriptive, en termes de géographie de l'information et d'analyse structurale de réseau (distances entre lemmes et des motifs que forment ensemble certains lemmes, indicateurs de centralité structurale et d'intermédiarité structurale) permet d'identifier des clusters et des motifs « intrigants » ou « non-triviaux ». Ils constituent pour nous des potentiels *signaux* faibles de la carte (Franck Ghitalla, 2003 ; Ghitalla, G., 2013).

e) Hypothèse interprétative temporaire

La seconde étape d'analyse relève d'un processus interprétatif, et nécessite donc dès lors une connaissance « métier » du sujet traité¹¹. Il est donc souhaitable ici que soient mobilisés des acteurs à positionnement « professionnel » ou « praticiens », pour le moins pour leur soumettre les hypothèses interprétatives formulées¹².

f) Confrontation aux verbatim et infirmation/confirmation/évolution de l'interprétation du signal faible

Nous arrivons à l'étape clé de confrontation des hypothèses interprétatives temporaires (issues de l'analyse cartographique) au matériau « brut » constitué par les verbatim des « posts ». Un outil spécialisé d'exploration de corpus textuel assisté par ordinateur nous

permet de retrouver quels sont les différents « posts » correspondant au motif de lemme que l'on a repéré. L'analyse sémantique (humaine) de ces quelques « posts » permet alors de confirmer, infirmer ou de faire évoluer l'hypothèse temporaire interprétative de ce motif, et d'en valider au cas échéant le caractère de « signal faible » de tendance émergente.

3. Premiers résultats : « *Transculturalité* » / « *Blockchain* », signal faible ?

Quels sont les résultats principaux, temporaires, issus des traitements et analyses, sur notre sujet des Fintechs/Blockchain/Organisation lui-même. Les premiers résultats, exploratoires et demandant à être confirmés, suggèrent notamment un lien - topographique et socio-sémantique – entre les lemmes « *Blockchain* » et « *Transculturalité* ». Cette proximité socio-sémantique, plutôt imprévue, formant un motif lexical qui semble singulier, est-elle robuste ou constitue-t-elle un artefact (lié à la captation par ex.) ? Si c'est le cas, quelle hypothèse interprétative peut-on lui donner ? Enfin et surtout, cette interprétation de motif lexical se trouve-t-elle corroborée (ou non) par un retour à la base de données textuelles de captation et par la confrontation « sémantique » aux verbatim des discussions ? En l'étape actuelle du travail, ce couplage entre motif lexical visible sur la carte et verbatim associés semble suggérer un sens qui paraît bien susceptible de constituer un signal faible présentant un intérêt en termes d'émergence de tendance.

Ainsi, les technologies des Blockchain seraient bien considérées par les contributeurs-internautes de Reddit comme étant un fort vecteur d'innovation dans les organisations elles-mêmes, fussent-elle dans le domaine financier ou non. Mais cette innovation du Blockchain, considérée ainsi par les contributeurs de post comme nécessairement techno-organisationnelle, est également considérée comme étant intimement liée à des dimensions *culturelles* des organisations, et plus précisément à l'enjeu de la transculturalité (figure 1 : carte sémantique générée)

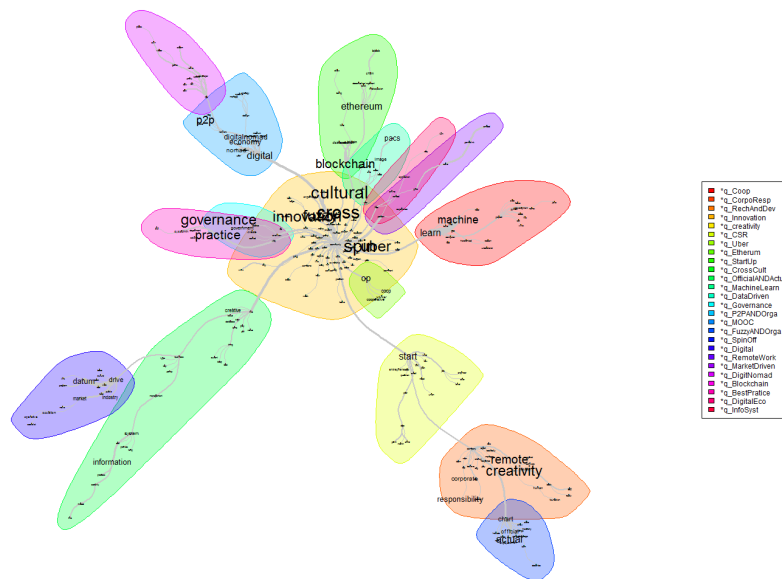


Figure 1 : carte sémantico-sociale générée (focalisation sur la variable Requête)
 (cliquer sur l'image pour l'agrandir)

*4. Discussion et interprétation : le Blockchain, l'émergence d'une « trans-confiance » ?

L'avancée des travaux de couplages ne nous permet pour l'instant que de faire des hypothèses interprétatives (que nous espérons valider pour la version finale de notre article) de ce « candidat signal-faible ». Le statut de la transculturalité dans l'innovation techno-organisationnelle « blockchain » tiendrait à la fois de la *condition* et du *résultat*, témoignant d'un processus de coévolution. La transculturalité serait une *condition* à la Blockchain car cette technologie, telle quelle s'exprime par exemple dans les Digital Autonomous Organisations (DAO), ne pourrait se réaliser qu'en sein d'une communauté d'acteurs réellement transculturelle. Cette nécessité d'interculturalité pourrait être interprétée comme condition d'appropriation et d'acceptabilité par les acteurs des organisations (Venkatesh et al., 2003,...).

La transculturalité pourrait ensuite constituer une *cause* de la Blockchain, car cette technologie serait également ressentie sur Reddit comme ayant pour *cause* la transculturalité croissante de la génération Y et surtout de la génération Z des *digital nomads*. Cette génération transculturelle, souhaitant trouver des alternatives aux dispositifs de confiance culturellement contingents des habituels « tiers de confiances »

nationaux, serait un des moteurs à l'origine du blockchain. Ainsi, à la manière du passage de pluri-culturalités à des trans-culturalités, nous assisterions au dépassement de pluri-confiances basées sur des plusieurs cultures restant nationales et à l'émergence de formes de trans-confiances, dont le Blockchain serait un élément constitutif.

Travail en cours, limites et perspectives : une meilleure focale *Fintech* grâce à une agrégation de deux corpus

Le travail restant à faire pour une version finalisée et développée de la recherche présentée dans cette proposition d'article tient surtout à la nécessité d'agréger deux vagues de captation de verbatim sur le Web social Reddit que nous avons effectuées : une vague de captation portant sur les relations entre Blockchain et mutation dans les Organisations d'une part (que nous avons exploitée ici) et une autre vague portant sur les relations entre secteur bancaire, apprentissage et Blockchain d'autre part. D'autres signaux faibles *confirmés* devront émerger. Les résultats partiels présentés ici laissent donc présager d'une moisson plutôt riche au cours des semaines prochaines.

Le positionnement de ces travaux de recherche nous paraît présenter certaines limites mais ouvrir également des perspectives de recherches appliquées stimulantes. Des limites portent sur le plan méthodologique – surtout à l'étape de captation sur le web (*webscraping*) – et sur le plan du cadrage de la problématique traitée (les Fintechs). Les techniques de *webscraping* que nous avons mobilisées nous paraissent générer en effet à l'heure actuelle un peu trop d'*artefact* de captation¹³. D'autre part, un cadrage plus focalisé sur la problématique des Fintechs (ici réalisé surtout à travers le sujet du Blockchain) nous apparaît, a posteriori, nécessaire, par une meilleure explicitation de ses contours dans notre étape de design de requêtes. Egalement, l'existence possible de biais cognitifs (Mitchell et al., 2002) - lors de l'expression des internautes en interaction et lors de l'interprétation des cartes et des verbatim par l'analyste – constitue également une limite à considérer.

La perspective principale qui nous semble ouverte par ces travaux (et que nous souhaitons explorer dans le futur proche) porte sur le prolongement dynamique de ces analyses de cartes socio-sémantiques web-sourcées, avec l'intégration des commentaires et interprétations de différentes parties intéressées dotées de compétences « métiers ». Nous

serions ravis de coopérer si des partenaires partageaient cet intérêt.

Bibliographie

- Benkler, Y. (2006). *The Wealth of Networks : How Social Production Transforms Markets and Freedom*. Yale University Press.
- Cardon, D. (2010). *La démocratie Internet : promesses et limites*. Seuil.
- Castells, M. (2010). *The Rise of the Network Society : The Information Age : Economy, Society, and Culture Volume I, 2^d Edition with a New Preface*. Consulté 31 mai 2017, à l'adresse <http://www.wiley.com/WileyCDA/WileyTitle/productCd-1405196866.html>
- Charmaz, K. (2014a). *Constructing Grounded Theory* (2 edition). London ; Thousand Oaks, Calif : SAGE Publications Ltd.
- Charmaz, K. (2014b). *Grounded Theory in Global Perspective : Reviews by International Researchers*. *Qualitative Inquiry*, 20(9), 1074-1084. <https://doi.org/10.1177/1077800414545235>
- Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets : Reasoning About a Highly Connected World*. Cambridge University Press.
- Economists are prone to fads, and the latest is machine learning. (s. d.). Consulté 21 mars 2017, à l'adresse <http://www.economist.com/news/finance-and-economics/21710800-big-data-have-led-latest-craze-economic-research-economists-are-prone>
- Ghitalla, F. (2003). *L'outré-lecture : manipuler, (s')appropriier, interpréter le Web*. Bibliothèque publique d'information.
- Ghitalla, F., Diemert, E., Maussang, C., & Pfaender, F. (2004). *Tarente : an experimental tool for extracting and exploring web aggregates*. In *2004 International Conference on Information and Communication Technologies : From Theory to Applications, 2004. Proceedings* (p. 627-628).
- Ghitalla, G. (2013, juillet 30). *Expédition cartographique dans l'univers des brevets*. Consulté 22 octobre 2016, à l'adresse <https://ateliercartographie.wordpress.com/2013/07/30/expedition-cartographique-dans-lunivers-des-brevets/>
- Granovetter, M. S. (1973). *The Strength of Weak Ties*.
- Johnson, R. B., & Onwuegbuzie, A. J.(2004). *Mixed Methods Research : A Research Paradigm Whose Time Has Come*. *Educational Researcher*, 33(7), 14-26. <https://doi.org/10.3102/0013189X033007014>
- Jollivet P. (2016). *Proceeding of the Third Annual International Conference on Business, Law & Economics (AICBLE)*. Athens Institute for Education & Research. Consulté à l'adresse <http://www.atiner.gr/abstracts/2016ABST-BLE.pdf>

- Jollivet P., Ghitalla, F., & Moulier-Boutang. (2015). Signaux faibles, émergences et prospective sur la mutation du travail avec écoute monitorée du Web. Uteam.
- Latour, B. (2005). *Reassembling the Social : An Introduction to Actor-Network-Theory*. OUP Oxford.
- Laugier, S. (2011). Le commun comme ordinaire et comme conversation, The Common as the Ordinary and as Conversation. *Multitudes*, (45), 104-112.
- Laugier, S. (2013). *Why we need ordinary language philosophy*. Chicago ; London : The University of Chicago Press.
- Lundvall, B.-Å. (2010). *National Systems of Innovation : Toward a Theory of Innovation and Interactive Learning*. Anthem Press.
- Mitchell, R. K., Busenitz, L., Lant, T., McDougall, P. P., Morse, E. A., & Smith, J. B. (2002). Toward a theory of entrepreneurial cognition : Rethinking the people side of entrepreneurship research. *Entrepreneurship theory and practice*, 27(2), 93–104.
- Moulier-Boutang, Y. (2007). *Politiques des multitudes : démocratie, intelligence collective & puissance de la vie à l'heure du capitalisme cognitif*. Amsterdam.
- Nicolescu, B. (1996). *La transdisciplinarité : manifeste*. Monaco : Editions du Rocher.
- Pincemin, B., Heiden, S. (2008). Qu'est ce la textométrie,. Consulté 16 janvier 2017, à l'adresse <http://textometrie.ens-lyon.fr/spip.php?article69>
- Sterling, A. B. S. B. (2008). The End of Theory : The Data Deluge Makes the Scientific Method Obsolete. Consulté 29 mai 2017, à l'adresse <https://www.wired.com/2008/06/the-end-of-theo/>
- Taleb, N. N. (2006). *The Black Swan : The Impact of the Highly Improbable* by Taleb, Nassim Nicholas (1 edition). Random House.
- Teddlie, C., & Tashakkori, A. (2008). *Foundations of Mixed Methods Research : Integrating Quantitative And Qualitative Approaches In The Social And Behavioral Sciences*. Los Angeles : SAGE Publications Inc.
- Uzunidis, D. (2012). De la méthode de recherche économique, On Method of economic research. *Marché et organisations*, (5), 101-106.
- Varian, H. R. (2014). Big Data : New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3-28. <https://doi.org/10.1257/jep.28.2.3>
- Vergne, J.-P., & Swain, G. (2016). *Categorical Anarchy in the U.K.? The British Media's Classification of Bitcoin and the Limits of Categorization* (SSRN Scholarly Paper No. ID 2800329). Rochester, NY : Social Science Research Network. Consulté à l'adresse <https://papers.ssrn.com/abstract=2800329>
- Wagner-Pacifi, R., Mohr, J. W., & Breiger, R. L. (2015). Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. *Big Data & Society*, 2(2), 205395171561381. <https://doi.org/10.1177/2053951715613810>

Watts, D. J.(2003). *Small Worlds : The Dynamics of Networks between Order and Randomness*. Princeton University Press.

Watts, D. J.(2004). *Six Degrees : The Science of a Connected Age* (Reprint edition). New York : W. W. Norton & Company.

1 Le Web Social peut être défini comme un espace de socialisation où les internautes interagissent en laissant des traces numériques constituant des données (qui, une fois interprétées, constituent de l'information et de la connaissance) et dont peut résulter de l'intelligence collective.

2 Au sein des sciences de gestion, nous ciblons tout particulièrement les recherches en management de systèmes d'informations

3 Cette focalisation présente est liée d'une part à des aspects prosaïques d'espace (la version initiale de ce document se devait être de 5 pages) et d'autre part à des aspects intellectuels. La méthodologie nous apparaît être un lieu et un temps de la recherche permettant de travailler de façon enrichissante sur les couplages ou relation entre travail théorique et travail empirique.

4 L'intégralité de la plateforme du Web social Reddit est crawlé, soit 190,227,552 posts à la date du 23, 2015. Source : <https://redditblog.com/2015/06/23/happy-10th-birthday-to-us-celebrating-the-best-of-10-years-of-reddit/>

5 On perçoit aisément, a contrario, que le raisonnement hypothético-déductif s'articule bien avec les méthodes quantitatives de type économétrique, via des tests paramétriques par exemple (Uzunidis, op. cit).

6 La mobilisation de ces méthodes dans des travaux incluant les sciences de gestions est ordinaire déjà par leur rareté. Autant les méthodes mixtes incluant l'analyse structurale de réseau et les données massives commencent en France à se développer via le renouveau des humanités numériques dans les disciplines de l'histoire, de la littérature, et plus récemment de l'information et de la communication, leurs présences en science de gestion y est limitée, et semble quasi inexistante, à notre connaissance, en science économiques (« Economists are prone to fads, and the latest is machine learning », s. d. ; Varian, 2014 ; Wagner-Pacifici, Mohr, & Breiger, 2015).

7 Le service Web Radarly de l'entreprise Linkfluence.

8 Parfois dénommé « text cluster analysis ».

9 Par exemple, la variable « Forum » permet de rattacher une contribution particulière d'un fil de discussion particulier à une des grandes thématique/communauté très structurée sur Reddit telles que « Politique » ou « Technologie » ou encore « Santé ».

10 Comme l'Analyse Factorielle de Correspondance (AFC) de Benzencri le permettait déjà.

11 Le statisticien-mathématicien français Benzencri soulignait également que l'utilisant de ses dispositifs d'exploration de données d'AFC et ACP exigeaient –contrairement aux statistiques et à l'économétrie – une connaissance fine du domaine étudié.

12 Une difficulté méthodologique émerge ici : comment peut on simultanément prétendre à assurer les « rôles » de Data Scientists, Data Analysts, experts métiers et .. chercheurs ? La question reste en grande partie ouverte, même si la polyvalence nécessaire au Data Analyst, ainsi que la mobilisation requise d'un collectif d'acteur, y répond en partie.

13 . Il s'agit par exemple d'éléments de pages web externes au fil de discussion lui-même codés en HTML, se retrouvant captés alors qu'elles ne devraient pas. L'identification et l'élimination de ces artefacts requièrent une attention humaine trop prenante et surtout source d'imprécisions.