



*Visualisation de mèmes sur
Sina Weibo*

A DISSERTATION PRESENTED
BY
CLÉMENT RENAUD
TO
THE LTCI, SCIENCES ÉCONOMIQUES ET SOCIALES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTORAT
IN THE SUBJECT OF
SYSTÈMES D'INFORMATION

PARISTECH TELECOM
PARIS, FRANCE
JUIN 2014

© 2014 - Clément Renaud
UNDER CREATIVE COMMONS.

Visualisation de mèmes sur Sina Weibo

ABSTRACT

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Morbi commodo, ipsum sed pharetra gravida, orci magna rhoncus neque, id pulvinar odio lorem non turpis. Nullam sit amet enim. Suspendisse id velit vitae ligula volutpat condimentum. Aliquam erat volutpat. Sed quis velit. Nulla facilisi. Nulla libero. Vivamus pharetra posuere sapien. Nam consectetur. Sed aliquam, nunc eget euismod ullamcorper, lectus nunc ullamcorper orci, fermentum bibendum enim nibh eget ipsum. Donec porttitor ligula eu dolor. Maecenas vitae nulla consequat libero cursus venenatis. Nam magna enim, accumsan eu, blandit sed, blandit a, eros.

Quisque facilisis erat a dui. Nam malesuada ornare dolor. Cras gravida, diam sit amet rhoncus ornare, erat elit consectetur erat, id egestas pede nibh eget odio. Proin tincidunt, velit vel porta elementum, magna diam molestie sapien, non aliquet massa pede eu diam. Aliquam iaculis. Fusce et ipsum et nulla tristique facilisis. Donec eget sem sit amet ligula viverra gravida. Etiam vehicula urna vel turpis. Suspendisse sagittis ante a urna. Morbi a est quis orci consequat rutrum. Nullam egestas feugiat felis. Integer adipiscing semper ligula. Nunc molestie, nisl sit amet cursus convallis, sapien lectus pretium metus, vitae pretium enim wisi id lectus. Donec vestibulum. Etiam vel nibh. Nulla facilisi. Mauris pharetra. Donec augue. Fusce ultrices, neque id dignissim ultrices, tellus mauris dictum elit, vel lacinia enim metus eu nunc.

Sommaire

0.1	Introduction	1
1	Sina Weibo et le milieu numérique en Chine	3
1.1	L'Internet chinois : éléments de contexte	4
1.2	Médias sociaux en Chine : un paysage morcelé	9
1.3	Code, langage et milieu(x) numérique(s)	17
2	Les mèmes Internet, objets numériques culturels	31
2.1	Les mèmes : définitions et histoire	32
2.2	Mème, mémoire collective et culture	38
2.3	Textualité des mèmes et formes d'énonciations numériques	44
3	Méthodologie de recherche	57
3.1	Sciences sociales du réseau	57
3.2	Méthodes et outils d'analyse des réseaux sociaux	65
3.3	Design de la recherche : Méthodologie choisie et schémas d'analyse retenus	72
3.4	Méthodologie de traitement et de visualisation des mèmes	93
Table des Matières		102
Bibliographie		120
Liste des Figures		122
I Liste des mèmes		125

0.1 Introduction

1

Sina Weibo et le milieu numérique en Chine

Ce premier chapitre présenter le contexte et les enjeux de la présente recherche ainsi que les concepts majeurs qui y seront mobilisés. Nous débuterons ce travail en brossant un portrait de l'Internet en Chine et plus spécifiquement de *Sina Weibo*, service de réseau social et de microblog chinois qui fera l'objet de cette étude. En nous appuyant sur la littérature existante, nous porterons tout d'abord un regard historique sur l'Internet en Chine, en détaillant les enjeux idéologique, technologique et économique qui ont présidé à son développement. Ensuite, nous considérerons l'évolution des services de réseaux sociaux et plus particulièrement le cas de *Sina Weibo* qui sera abordé dans le détail au travers d'observations et d'exemples issus de la littérature. Nous discuterons notamment les différences et similitudes du service chinois avec ses homologues américains *Twitter* et *Facebook*. Après cette présentation factuelle du contexte de l'étude, la seconde partie de ce chapitre introduira l'appareil conceptuel qui sera mobilisé au cours de ce travail. En nous appuyant sur une sélection de travaux existants, nous considérerons l'Internet chinois successivement comme espace, lieu et territoire afin de comprendre comment s'articulent les usages de *Sina Weibo* dans le contexte chinois. Nous proposerons enfin le concept de milieu numérique pour décrire l'ensemble des protocoles et pratiques discursives qui préside à la création des objets numériques que nous décrirons sous la forme de modèles appelés *topogrammes*.

1.1 L'Internet chinois : éléments de contexte

La production de la recherche en sciences sociales concernant l'Internet et ses usages en Chine est largement structurée autour des intérêts politiques et économiques qui l'entourent. Les premières études académiques sont issues du champ de l'information-communication et plus spécifiquement du monde anglo-saxon (?). Leurs approches reposent sur le postulat qu'Internet serait un outil favorisant la démocratisation. Ainsi cette littérature peut se résumer en un échange d'arguments autour d'une alternative : l'entrée de la Chine dans la "société de l'information" va de pair avec la modernisation économique et la démocratisation ou à l'inverse avec un contrôle politique accru de l'information et global de la société. Ces questions de liberté d'expression et du rapport entre démocratie et médias occupent encore aujourd'hui une large place (MacKinnon, 2009; Douzet, 2007; Yang, 2008). Les autorités chinoises craignent en effet de voir se développer de nouveaux canaux de circulation et de nouvelles sources d'information échappant à leur autorité traditionnelle sur les médias - et contestant ainsi leur pouvoir. D'un autre côté, le média internet est également perçu par le gouvernement comme un agent indispensable du processus de modernisation politique et économique (dont l'ouverture de marchés commerciaux intérieurs) permettant de faire circuler l'information de façon optimale, en touchant ainsi l'ensemble des foyers. En somme, un Internet "sain" et maîtrisé soutiendrait le développement du système de valeurs gouvernemental, mais également de l'économie et de la culture. Ici, de nombreuses publications se sont également intéressées au rôle de l'économie en ligne dans la croissance économique de la Chine et aux liens étroits des entreprises web avec la censure (Dann et Haddow, 2008).

Les études propres aux usages d'Internet dans le contexte chinois sont toutefois moins nombreuses ou sont généralement le fruit des services marketing des entreprises locales ou internationales en quête de nouveaux marchés (?Bergstrom, 2012). Toutefois, on retrouve à la croisée de ces différentes discussions des études plus précises qui cherchent à mettre en relation les dimensions à la fois économique et d'usage du web chinois (Puel, 2009; Fernandez et Puel, 2010; ?). Dans le domaine précis de l'analyse des réseaux sociaux chinois, les publications internationales en sciences sociales restent encore peu nombreuses. Le champ de la recherche en informatique propose néanmoins plusieurs études, s'attelant notamment à décrire les dynamiques des échanges de contenus (Yu *et al.*, 2011) et les dispositifs de censure en place (?).

Afin de mieux se situer dans le large paysage de cette littérature, cette première partie rend compte des principaux éléments historiques, politiques et économiques d'infra et d'infostructure de l'internet chinois.

1.1.1 Petite histoire et évolution de l'Internet chinois

Avec plus de 560 millions d'utilisateurs connectés, l'Internet chinois est aujourd'hui le plus grand réseau national, dépassant depuis 2011 l'Amérique du Nord et l'Europe réunies

(?)¹. Alors que l'installation des infrastructures a débuté seulement en 1996, l'expansion en a été fulgurante (Fang et Yen, 2006). 51 millions de nouveaux internautes chinois ont fait leur apparition en 2012, soit une hausse de 10% par rapport à 2011 (?).

Comme ailleurs dans le monde, les débuts des réseaux informatiques en Chine se déroulent dans le contexte universitaire avec la création en 1994 du CERNET (*China Education and Research Network*) permettant de relier plusieurs grandes universités du pays. Le 17 mai 1994, la Chine effectue sa première connexion au réseau Internet en se reliant au *Stanford Linear Accelerator Center* (SLAC) de l'Université de Stanford aux États-Unis. L'année suivante, les infrastructures *backbones* nécessaire à l'installation de l'Internet à plus large échelle sont effectuées (ChinaNet, GBNet, CERNET) et la première licence d'exploitation commerciale est attribuée, effective dès 1996. Le 1er janvier 1997, le *Quotidien du Peuple* lance sa première version en ligne, devenant ainsi le premier site Internet d'information officielle du pouvoir central Chinois sur Internet. Dans le courant de l'année, le premier FAI privé chinois *China InfoHighway* voit le jour. En Novembre, le CNNIC (*China Internet Network Information Center*) est chargé de recueillir et publier des statistiques sur le développement de l'Internet en Chine (Dai, 2007).

Dès son commencement, la structuration et le contrôle du réseau Internet est un enjeu important pour le gouvernement de Pékin. Partie intégrante de leur programme de gouvernance, les dirigeants communistes ont appris depuis longtemps à considérer très sérieusement les nouveaux outils de communication, à la fois comme un risque et une opportunité. Le plus célèbre témoin de cette histoire est sans doute le cinéma soviétique mis en place par Lénine et Trotski dès leur arrivée au pouvoir en 1917 sous la forme des *agitki*, ces courts films de propagande qui révèleront par la suite de grands cinéastes comme Kuleshov, Vertov ou même Eisenstein (?). Au début des années 90, la Chine s'est engagée depuis plus de 10 ans dans de vastes réformes (*gaige kaifang*) pour se sortir de l'état de chaos où l'avait laissé la Révolution Culturelle. La politique d'ouverture du pays couplé à un fort protectorat économique ouvre l'ère du *Made in China* qui concerne l'ensemble des secteurs industriels, y compris ceux des médias et télécommunications dont le potentiel stratégique et économique est énorme. Les dirigeants de Pékin sont également bien conscients que la place forte que la Chine réclame dans le paysage mondial se dessinera notamment par une intégration accrue dans le réseau mondial des TIC. Ainsi, en Mars 2000 alors que la population des internautes chinois atteint 16,9 millions d'internautes, le premier ministre Jiang Zemin affirme : “*Internet technology is going to change the international situation, military combat, production, culture, and economic aspects of our daily life significantly.*” (?)

Alors qu'Al Gore et l'administration Clinton déploient les “*information superhighways*”

1. Et alors même que seuls 40% de ses habitants disposent à ce jour d'une connexion : « *La collection des données se fait grâce à des logiciels, questionnaires en ligne et sondages par téléphones (...) Pour le CNNIC, un internaute est un citoyen chinois âgé de plus de 6 ans qui utilise Internet au moins une heure par semaine ou a utilisé Internet durant les 6 derniers mois* » (d'après le site officiel de CNNIC, agence officielle du gouvernement).

aux États-Unis, le programme d’informatisation et de développement des TIC (*xinxihua*) devient une des clés du calendrier politique et économique chinois. Jiang Mianheng, le fils du premier ministre Jiang Zemin, est chargé du déploiement de ce vaste projet. Revenu en 1992 après plusieurs années passées dans les universités américaines puis dans la Silicon Valley chez HP, Jiang Manheng investit donc largement dans les infrastructures et soutient dès 1999 le lancement du haut-débit en Chine avec la compagnie *Netcom*, aujourd’hui encore deuxième opérateur du pays (Dai, 2007). Fortement influencé par les théories de Alvin Toffler (Tsui, 2007), le gouvernement de Jiang Zemin est bien décidé à ne pas rater la “*troisième vague*” de modernisation de l’industrie : l’informatisation. La mise en place d’une quinzaine de “*golden projects*” informe le mouvement de stratégie globale qui doit dynamiser transversalement tous les secteurs d’activité jusqu’au sein-même de l’administration chinoise. Alors que les “*golden customs*” se charge des données du commerce extérieur et que la “*golden sea*” devient un outil de communication entre cadres du Parti et administrations locales, les multiples “*golden projects*” sont soutenus par la volonté de mettre en place un *e-government*. Pékin veut faire de l’Internet chinois un espace de dialogue entre citoyens et organes du gouvernement avec notamment la mise en place de services administratifs en ligne et d’enquêtes transversales sur toute la Chine sur la qualité de vie dans les différentes villes. La gestion proactive de l’Internet national offre également au Parti un vecteur unique pour la diffusion de ses idées politiques. Plusieurs études montrent comment les idées et discours nationalistes sur l’Internet ont bénéficié d’un soutien constant du gouvernement chinois, avec notamment comme objectifs l’accès aux communautés chinoises émigrées à l’étranger (?). Finalement, le média Internet doit servir les intérêts du gouvernement et participer à l’unification territoriale, à l’instar des autres médias dits classiques comme la télévision.

1.1.2 Censure sur l’Internet chinois

Les modes d’adoption de la technologie Internet par le gouvernement de Pékin illustrent pourtant bien le dilemme constant entre ouverture et protectionnisme qui tire la classe politique chinoise depuis la fin de la Révolution Culturelle. L’ouverture au monde et la politique de réformes du “*nouveau départ*” s’accompagne pour la Chine de nombreux défis souvent considérés comme extérieurs, que résume avec clarté une phrase célèbre attribuée à Deng Xiaoping : “*Lorsque vous ouvrez une fenêtre pour avoir de l’air frais, vous devez vous attendre à ce que quelques mouches rentrent dans la pièce.*” Alors que les promesses d’expansion économique et politique de l’Internet sont bien au rendez-vous, les “mouches” entrées par les fenêtres des navigateurs web commencent à essaimer pour faire de plus en plus bruit.

Depuis 1998, le Ministère de la Sécurité Publique chinois travaille à la conception d’un projet intitulé *Golden Shield* qui constituerait un fichier global des citoyens chinois utilisable tant pour le contrôle de la démographie que le vol de véhicules ou la sécurité aux frontières (Lyons, 2009). Lors d’un Trade Show intitulé “*Security China 2000*” se déroulant

à Pékin en 2000, le projet est présenté publiquement comme une large base de données devant regrouper les informations administratives des citoyens et leurs activités en ligne dans le but de favoriser le travail de la police (Walton, 2001). Titanesque et complexe à réaliser, le projet est peu à peu modifié pour devenir un système de filtrage de contenus et de blocage de sites, basé sur le système du *firewall*. Fang Binxing, un professeur spécialisé dans la sécurité informatique à l'Université de Harbin est nommé chef-ingénieur du projet *Golden Shield*. Il recrute de nombreux ingénieurs et avec l'aide de l'université de Qinghua et de plusieurs entreprises occidentales (Nortel Networks², Cisco³) qui ensemble s'atèlent au développement technologique du projet qui devait devenir le système de contrôle de l'Internet chinois aujourd'hui en activité. Baptisé depuis de manière informelle “*The Great Firewall*” (GFW) par analogie avec la “*Grande Muraille de Chine*” (Great Wall), ce système « sociotechnique » hors du commun est aujourd'hui considéré comme une des plus grandes installations d'analyse et de traitement de données en activité. À chaque seconde, GFW traite et scanne des millions de chaînes de caractères issues des requêtes et pages vues par des centaines de millions d'internautes (Winter et Lindskog, 2012). Au-delà de la censure automatique, GFW emploierait aujourd'hui entre 30000 et 50000 personnes⁴ : ingénieurs, modérateurs, relecteurs, officiers de police, etc. Phénomène particulier, un groupe de rédacteurs est notamment chargé d'intervenir dans les discussions ou les forums en ligne pour faire valoir le point de vue officiel. L'adage veut que chacun des messages qu'ils postent soient rémunéré 50 centimes RMB, ce qui a amené les internautes chinois à baptiser ces représentants de l'ordre politique en ligne “*le Parti à 50 centimes*” (*wumao dang*).

Au-delà de l'activité manuelle de milliers d'employés, GFW opère également plusieurs type de blocages sur les contenus. Techniquement, la plupart du filtrage se déroule au niveau du fournisseur d'accès avec notamment des adresses particulières qui sont rendues inaccessibles : l'adresse *facebook.com* ou *youtube.com* renvoie une erreur “404 : Le site demandé n'existe pas”. Ainsi, de nombreux sites célèbres ne sont pas accessibles (*Twitter*, *Youtube*, *Facebook*, etc.) Les autres blocages effectifs s'effectuent selon les adresses IP ou les serveurs d'attribution de noms de domaine (DNS) menant parfois au blocage de serveurs entiers (Winter et Lindskog, 2012). Les URLs des pages de certains sites importants sont également filtrées. Sur Wikipedia notamment, la page “*Tiananmen Square protests of 1989*” est inaccessible depuis la Chine sans que le site Wikipedia soit pour autant

2. “*Nortel Networks signs contracts valued at over USD 120 Millions during Canadian Prime Minister Jean Chretien's visit to China*”, PR Newswire, <http://www.prnewswire.co.uk/news-releases/nortel-networks-signs-contracts-valued-at-over-us-dollars-120-million-during-canadian-prime-minister-visit-to-china-1500001.html>, consulté le 17 Février 2014 à 12 :18

3. “*Cisco Leak : ‘Great Firewall’ of China Was a Chance to Sell More Routers*”, Sarah Lai Stirland, 05.20.2008, <http://www.wired.com/threatlevel/2008/05/leaked-cisco-do/> consulté le 17 Février 2014 à 12 :39

4. “*What is internet censorship ?*”, Amnesty International Australia. 28 Mars 2008. Consulté le 17 Février 2014 à 15h12



Graphique 1.1.1 – *Jingjing et Chacha* sont deux figures créées par les autorités chinoises pour signifier la présence policière en ligne aux internautes. Le nom de ces veilleurs dessinés est une composition issue du mot “police” en chinois (*jingcha*). Source : <http://jswm.newssc.org/system/2008/10/29/011233423.shtml> consultée le 17 Février 2014, à 15 :32

intégralement bloqué. Également, des requêtes contenant des mots “interdits” sur les moteurs de recherche peuvent conduire à des micro-coupures ou un accès restreint au web pendant parfois plusieurs minutes⁵. La liste des sites et mots bloqués n'est pas publiée par le gouvernement et l'ajout sur ses listes s'effectue a priori sur la demande de différentes agences gouvernementales chinoises, sans notification publique. Essentiellement, il s'agit de sites à caractère pornographique (la pornographie est illégale en Chine), de sites liés aux groupes dissidents chinois (Falung Gong, Dalai Lama, Tibet Libre...), de sites du gouvernement taiwanais et d'autres sites revendiquant la liberté d'expression pour la Chine⁶.

Si le GFW est un outil de contrôle politique, il participe également largement au protectorat économique chinois. L'expansion rapide du vaste marché de l'Internet en Chine s'est faite sous un strict contrôle politique et si l'État a largement financé les infrastructures, GFW a été l'un des moteurs de la croissance des grandes sociétés du web chinois. L'absence du géant *YouTube* a notamment permis aux acteurs locaux de la vidéo en ligne de se développer rapidement, comme *Youku*, aujourd'hui sujet à une valorisation colossale sur les marchés d'affaires. La situation est similaire pour les réseaux sociaux. L'absence de concurrence étrangère due à l'interdiction de *Facebook* en 2008 puis *Twitter* en 2009 a permis aux acteurs chinois de se développer. Leur poids sur le marché intérieur les autorise aujourd'hui à rivaliser avec leurs concurrents américains au niveau mondial tant en nombre d'utilisateurs (?) qu'en revenus directs et indirects générés (?).

Ainsi, GFW a affecté l'économie du pays en profondeur et à ce titre notamment continue d'être une préoccupation première du pouvoir politique. L'évolution technologique de GFW suit de près l'évolution des moyens de contournement des blocages, qui sont nombreux. Ainsi, le célèbre logiciel *Tor* garantissant l'anonymat sur Internet est aujourd'hui

bloqué en Chine (Winter et Lindskog, 2012) ainsi que d'autres technologies communes d'anonymisation (comme le proxy notamment). Pourtant, il reste très facile de "faire le mur" (*fanqiang*) et d'accéder aux contenus en contournant les limites du GFW. Les solutions techniques à disposition sont multiples et souvent peu couteuses à mettre en place ou à utiliser. De nombreux services commerciaux proposent de se connecter depuis d'autres pays à l'aide d'un VPN (*Virtual Private Network*) pour un coût très faible. Le VPN permet de se connecter depuis une machine située dans un autre pays et de bénéficier ainsi de l'accès tel qu'il existe dans le pays où se situe la machine. Le flou juridique qui entoure l'existence de services commerciaux de contournement de GFW témoigne de l'intérêt du gouvernement chinois à protéger largement le marché intérieur en supprimant l'accès aux services majoritaires du web occidental, sans pour autant exercer une traque systématique de chaque personne voulant utiliser *Facebook* ou *Gmail*. De plus, l'absence totale de moyens de contournement interdirait l'accès à des sources précieuses d'informations professionnelles (notamment *Twitter*) et se ferai donc au prix d'une perte d'avantages concurrentiels décisifs pour les entreprises chinoises, que les autorités chinoises ne semblent pas désirer. Une étude de l'OpenITP parue en 2013 analyse l'usage des outils de contournement de la censure auprès d'un échantillon de 1175 utilisateurs en Chine. Au-delà des solutions technologiques variées, on peut noter que la première raison pour contourner le blocage de l'Internet et l'utilisation des services de *Google* (notamment de *Gmail*, dont le débit est très lent depuis la Chine), suivi de la volonté de se rendre sur les sites des réseaux sociaux américains comme *Facebook* et *Twitter*, puis de l'accès aux contenus d'actualité, de vidéo en ligne et de matériel à caractère pornographique. Les utilisateurs souhaitant accéder à des contenus à caractère politique ou utilisant l'Internet de manière anonyme pour communiquer de façon plus sécurisée représente moins de 10% de la population étudiée (OpenITP, 2013). Il est également important de noter que l'impressive majorité des internautes chinois n'utilisent pas de système de contournement des blocages de l'Internet.

1.2 Médias sociaux en Chine : un paysage morcelé

Alors qu'un blocage officiel s'applique donc sur les plus célèbres services Internet américains, de nombreux services se sont développés pour répondre aux besoins et intérêts des internautes chinois. Cette situation particulière influe sur les contenus diffusés et les usages possibles du réseau par ses membres. Au-delà des caractéristiques politiques et économiques du web chinois, d'autres phénomènes liés à ce développement technologique particulier viennent influencer la culture des entreprises web et l'être-ensemble des usagers en ligne. Dans la seconde partie de ce chapitre, nous allons présenter les différents services de réseaux sociaux en ligne (Social Network Services : SNS) les plus utilisés en Chine. Nous nous attarderons sur le cas de *Sina Weibo* qui sera utilisé comme support pour la suite de la présente étude.

Profitant de l'absence des grands noms du réseau social en ligne, de nombreux services ont vu le jour sur la Toile chinoise. L'importance du *guanxi* (Yu et King, 2008), élément profond de la culture traditionnelle poussant chaque Chinois à entretenir et exposer avec soin ses relations en société, peut également avoir contribué à créer un terrain idéal pour le développement rapide de ces sites (?). Plutôt morcelé, le paysage des SNS en Chine offre une variété de services et d'acteurs qui rassemble les internautes chinois selon leurs centres d'intérêts. *Douban* offre aux jeunes « branchés » de partager lectures, films et musique. *Kaixin001*, plus centré sur les jeux, propose un espace ludique pour les trentenaires au bureau. *Renren* (anciennement *Xiaonei*) est, quant à lui, un véritable clone de *Facebook* et se focalise sur le monde étudiant chinois (Renaud, 2011). Malgré ces nombreux concurrents, le service de messagerie instantanée *QQ* reste le leader incontesté du marché chinois. Aujourd'hui classé 8ème site le plus visité au monde⁵, *QQ* dénombre jusqu'à 100 millions d'utilisateurs connectés simultanément⁶. Considéré comme la quatrième plus grande firme du web mondiale, son créateur le géant *Tencent Holdings Limited* a investi depuis le début de l'Internet en Chine dans de nombreux domaines des TIC : jeux, publicité en ligne, e-commerce, etc. Plus qu'une simple messagerie de chat, les services de *QQ* sont multiples : la page de profil de chaque utilisateur (*QZone*) permet de maintenir un blog et d'écouter de la musique (*QQmusic*). Chaque utilisateur peut se créer un avatar en ligne pouvant recevoir de nombreux vêtements et accessoires vendus en ligne (*QQshow*) mais aussi participer à de nombreux jeux multi-joueurs pour tous les âges (*QQ Entertainment*). Le système de monnaie virtuelle (*QCoin*) mis en place pour les achats en ligne a généré dès son lancement en 2005 un nombre important de transactions⁷, poussant *Tencent* à obtenir une licence bancaire. L'utilisation de la messagerie *QQ*, réseau social avant l'heure, est devenu un véritable phénomène de société porté par la diversification de *Tencent* dans de multiples secteurs sous une marque unique. Au-delà des jeunes et des professionnels de l'Internet, le réseau *QQ* comptait en juillet 2011 plus de 812,3 millions de comptes actifs, faisant de lui le deuxième réseau social mondial après *Facebook*. Du magasin de photocopie de quartier au réseau de prostitution clandestin, *QQ* héberge les discussions quotidiennes et fait pour ainsi dire partie intégrante du paysage des villes modernes. Les chinois échangent plus volontiers leurs numéros de *QQ* que ceux de leurs téléphones portables et ce mode de communication est souvent préféré au mail dans les échanges au bureau. Convergeant rapidement avec la croissance fulgurante du e-commerce en Chine, les produits dérivés estampillés *QQ* sont devenu une véritable mode en Chine : voitures, téléphones, boissons, etc. Le groupe *Tencent* poursuit son évolution avec le lancement en 2008 de son service de microblog *Tencent Weibo*, qui a connu un véritable succès dès les premiers mois. Aujourd'hui, la firme de Shenzhen poursuit la conversion

5. D'après Alexa.com, consulté le 3 Février 2013.

6. Voir <http://im.qq.com/culture>, consulté le 14 Février 2013.

7. CentralBankalerton "virtualmoney", People's Daily, 12 Janvier 2007, http://english.people.com.cn/200701/12/eng20070112_340681.html consulté le 28 Mai 2014

de ces utilisateurs *QQ* vers sa plateforme mobile *WeChat* qui connaît actuellement une très forte croissance, au point de voir les autres services de microblog mis au banc par les utilisateurs. A l'origine application de messagerie écrite et vocale, *Tencent* continue de se diversifier en offrant désormais d'utiliser son compte *QQ* comme moyen de paiements pour de nombreux services du quotidien (taxis, nourritures, etc.)⁸.

1.2.1 Microblog en Chine et Sina Weibo

Les sites de *microblogging* (en chinois *weibo*) permettent aux utilisateurs de poster de courts messages composé de photos ou de texte de 140 caractères maximum, puis de les commenter et de les partager avec leurs lecteurs. A l'image de *Twitter*, chaque utilisateur peut souscrire aux fils d'info d'autres utilisateurs afin de recevoir leurs messages et mises à jour.

L'histoire du microblog en Chine débute en 2007 avec plusieurs services se présentant alors comme des clones de *Twitter*. Le service *Fanfou* connaît notamment un succès rapide alors que de nombreux journalistes l'utilisent pour enquêter et coordonner leurs actions lors de l'arrivée du SRAS ou le tremblement de terre de Wenchuan dans le Sichuan en 2008. Le service qui connaît un rapide succès est fermé sur ordre du gouvernement en juillet 2009, suite aux nombreux commentaires suscités par des émeutes s'étant déroulé à Urumqi dans la province du Xinjiang. A peine un mois après cette fermeture, la firme *SINA Corporation* saisit l'opportunité et s'installe en lançant son propre service de microblog intitulé *Sina Weibo*. *Sina Weibo* connaît une croissante soutenue avec plus de 10 millions de nouveaux inscrits par mois et devient en 2012 la plateforme de microblog la plus utilisée en Chine avec 250 millions d'utilisateurs (?). Les revenus de *Sina Weibo* ne cessent alors de croître (+19% en 2012⁹) alors que plus de 86 millions de messages sont postés chaque jour sur ce service¹⁰.

Figure historique de l'Internet chinois, *SINA Corporation* est célèbre pour son portail *sina.net* et son immense plateforme de blogs qui en font le fleuron des fournisseurs de contenus en ligne en Chine. Spécialisé dans “l'infotainment” (un mélange très tabloïd d'actualité et de news people), *SINA* est la première compagnie nationale chinoise à avoir été listée au *NASDAQ* dès Avril 2000. Avec son *Weibo*, la firme réussit un coup de force commercial et prouve une fois encore combien la censure gouvernementale est bénéfique à l'industrie du web chinois. Néanmoins, la réussite de *SINA* et de son service de microblog ne se fait pas sans connaître de nombreux ajustements parfois chaotiques. En effet, la stratégie agressive d'acquisition d'utilisateurs qui pousse la croissance *Sina*

8. “21 million taxi rides have been booked on WeChat in the past month”, Tech in Asia February 12, 2014 <http://www.techinasia.com/wechat-21-million-taxi-rides-booked> consulté le 17 Février 2014 à 15 :36

9. Voir http://corp.sina.com.cn/chn/Annual_Report_2011_Final.pdf, consulté le 16/02/2013.

10. D'après Sina Corp. Earning Calls - <http://phx.corporate-ir.net/phoenix.zhtml?c=121288&p=irol-EventDetails&EventId=4727394>, consulté le 16/02/2013

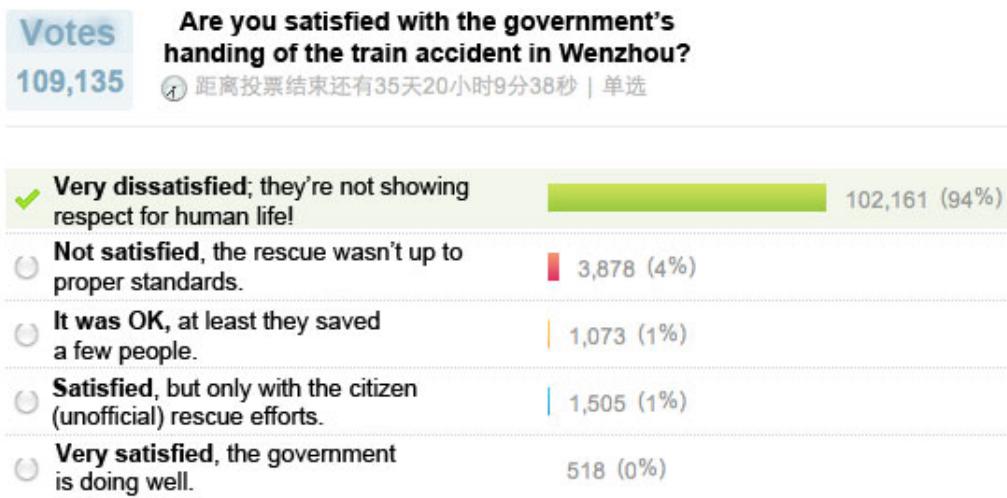


Graphique 1.2.1 – Capture d'écran de Sina Weibo, réalisé le 9 Avril 2013 à 08 :59

Weibo se fait avec la garantie pour les utilisateurs de pouvoir discuter et s'informer plus facilement en ligne. Dès le début de l'année 2010, les suppressions de comptes utilisateurs et de messages non désirés commencent à se répandre dans le service. Les discussions politiques sont régulièrement effacés et *Sina* se voit contraint de mettre en place un système de censure efficace sous la pression du gouvernement de Pékin. Néanmoins, afin de continuer à garantir la croissance du service, la firme de Pékin laisse une relative liberté aux utilisateurs en étant plutôt large sur la surveillance des discussions et les actions prises. Des personnalités publiques ou journalistes devenues “weibo-stars” mobilisent régulièrement l'opinion publique autour des sujets d'actualité attirant souvent des millions de lecteurs et de commentaires. Plusieurs scandales éclatent en ligne, mettant en cause des officiels et leur famille¹¹. Le 23 Juillet 2011, deux trains déraillent sur la ligne reliant Ningbo à Wenzhou qui venait d'être inaugurée en fanfare quelques jours auparavant, faisant près de 40 morts et 200 blessés. La colère gronde alors que le gouvernement tarde pendant plusieurs jours à prendre la parole sur ce sujet d'actualité épique. Sur la toile et *Sina Weibo* en particulier, les discussions vont bon train et les internautes indignés commentent le dernier drame du développement trop rapide de la Chine, où se mêle détournement d'argent public, corruption et sécurité des passagers.

Sina Weibo désactive alors la fonction de commentaires des messages. Dans les jours qui suivent, le gouvernement fait enfin une déclaration officielle sur les causes de l'accident de train puis se décide à agir en mettant en garde les internautes trop audacieux de représailles à venir. Les messages controversés sont supprimés, plusieurs comptes utilisateurs sont fermés, la police convoque les meneurs des discussions et d'autres mesures d'intimi-

11. Le fils d'un haut-cadre du Parti arrêté ivre par la police après avoir renversé 5 personnes et annonce : "Mon père s'appelle Li Gang" et se voit immédiatement libéré, provoquant un tollé chez les internautes http://www.chinadaily.com.cn/china/2011-03/02/content_12099500.htm



Graphique 1.2.2 – Un sondage publié sur *Sina Weibo* (traduction C. Custer, Tech in Asia, 1er Aout 2011), consulté le 24 Février 2014, à 22h12.

dation sont menés auprès des journalistes et des stars du weibo qui se seraient exprimés un peu trop directement à l'égard du Parti. Dans le même temps, le gouvernement a du mal à se saisir de ce nouvel outil. Alors que les administrations locales, universités et les médias d'État ont plutôt bien réussi le virage de leur stratégie de communication vers le microblog, les membres du gouvernement de Pékin ouvrent des comptes où ils sont parfois raillés, tournés en ridicule et harassés de questions. En Février 2012, les quatre plus grosses sociétés de microblog (dont *Sina Weibo*) annoncent que chaque utilisateur est maintenant contraint de modifier son profil pour mentionner son véritable nom, prénom ainsi que son numéro de carte d'identité. Cette velléité de vérification échoue et est abandonnée quelques semaines plus tard face à la mobilisation des utilisateurs et la difficulté de faire appliquer de telles mesures. Le gouvernement de Pékin édite pourtant une série de règles “*Several Regulations on Microblog Development and Administration Enacted by the Beijing Government*” dont la plus notable sera la possibilité de condamner tous ceux qui auront participé à la diffusion d'information considérées comme fausses, erronée ou mensongères. Face à la multiplication des actions gouvernementales et à l'apparition d'autres plateformes, la croissance du nombre d'utilisateurs de *Sina Weibo* est désormais stoppée pour aborder une phase de déclin estimé à près de 10% dans les deux premiers mois de 2014¹².

A la lecture de l'histoire de *Sina Weibo* on constate l'ambivalence des actions officielles du gouvernement chinois dans la réussite économique des entreprises d'Internet. Si la firme *SINA* a bénéficié de prime abord d'un avantage compétitif notoire par l'élimination de la concurrence, elle a par la suite souffert des conséquences du contrôle politique de l'Internet, avec notamment une perte de ses utilisateurs.

12. D'après le CNNIC cité dans l'article “*China's Twitter is bleeding users*”, 17 Janvier 2014, <http://blogs.marketwatch.com/thetell/2014/01/17/chinas-twitter-is-bleeding-users>, consulté le 17 Février 2014 à 18 :17



Graphique 1.2.3 – Cours de l'action *SINA* au Nasdaq entre 2009 et Février 2014 - Source : Google Finance, consulté le 17 Février 2014 à 15 :28

1.2.2 Sina Weibo, un usage plus ludique que Twitter

Dans son article nommé A Tale of two microblogs, Jon L. Sullivan (?) raconte comment l'évènement historique de la fermeture de *Twitter* en Chine a vu la communautés des microbloggers chinois se scinder en plusieurs groupes distincts :

- *Twitter* rassemble une communauté avide de libres discussions, souvent très politisée voire radicalement en opposition avec le gouvernement chinois.
- *Tencent Weibo* est utilisé par les utilisateurs de *QQ*, typiquement des personnes aux revenus plus faibles qui accèdent au web depuis leurs mobiles.
- *Sina Weibo* est le favori des travailleurs urbains, souvent plus jeunes ou éduqués, représentant davantage la classe moyenne montante.

Dans la littérature en sciences informatiques, plusieurs articles proposent des comparaisons entre *Twitter* et *Sina Weibo*. Une large analyse quantitative et comparative menée avec des jeux de données des deux services (?) nous apprend que le contenu de *Sina Weibo* est davantage corrélé avec des sentiments positifs (analysés automatiquement). Les utilisateurs de *Sina Weibo* parlent davantage de lieux et de personnes alors que les utilisateurs actifs sur *Twitter* s'intéressent plus aux organisations. Également, *Sina Weibo* connaît un pic d'activité le week-end alors que *Twitter* affiche généralement une baisse de régime dans les fins de semaine. Ces différentes indications suggèrent une tendance où *Sina Weibo* serait davantage utilisé pour des activités de loisir quand *Twitter* se destinerait à un usage plus professionnel. Une étude s'intéressant aux tendances sur *Sina Weibo* (?) indiquent que la majorité des comptes les plus influents de *Twitter* ont été vérifiés contrairement à *Weibo* où le taux est plus faible chez les grands utilisateurs. La vérification d'un compte se fait par l'authentification auprès du fournisseur de service afin d'attester la véracité de la personne utilisant le compte. C'est un enjeu important pour les figures publiques (marques, stars, homme politiques, etc.) Cet indicateur nous montre donc l'intérêt professionnel fort qui entoure *Twitter*, moins pressant dans le cas de *Sina Weibo* où moins de personnes ont ressenti la nécessité de faire officialiser leurs comptes. Sur les deux services de microblog, les utilisateurs inscrits possèdent un réseau de relations identifiables par leur souscription aux fils d'infos d'autres utilisateurs (*follow*). La relation peut être inexiste (*none*), mutuelle (*friend*) ou unidirectionnelle (*follow* : un utilisateur suit un autre mais n'est pas suivi par ce dernier). La comparaison d'échantillons des graphes sociaux issus

des deux services (?) montrent comment les relations sur *Sina Weibo* sont plus dissymétriques et moins réciproques, reflétant une hiérarchie plus forte entre les utilisateurs que *Twitter*.

Un autre facteur important de différentiation entre les deux services est la nature de la diffusion des contenus postés sur *Sina Weibo*. Contrairement à *Twitter* où le texte domine, la majorité des posts de Weibo contiennent des images ou des vidéos (Zhao et al., 2012). Les posts possédant des contenus multimédia (images, vidéos...) sont plus susceptibles d'être diffusés largement et restent en moyenne actifs pour une durée plus longue (Zhao et al., 2012). Également, les contenus sur *Sina Weibo* possèdent une proportion moins élevée de retweets et de commentaires que sur *Twitter* (Zhao et al., 2012; Gao et al., 2012). L'activité de la population de *Twitter* est plus intense, moins tournée vers la diffusion de masse et plus réactive aux influx de nouveaux contenus.

Nous voyons donc que le paysage de *Sina Weibo* se constitue autour de stars et célébrités qui concentrent l'attention avec des contenus à la diffusion très large. Moins tournés vers l'actualité et la conversation que son homologue *Twitter*, *Sina Weibo* agit comme véhicule de contenus à grande audience, souvent publiés par des personnalités publiques célèbres. Des études quantitatives montrent bien que les contenus les plus échangés et discutés concernent les loisirs et divertissements, la mode, la santé, etc. (?). Les messages à caractère humoristique (texte, images et vidéos) occupent également une place prépondérante dans les échanges des utilisateurs, contrairement à son homologue américain *Twitter* dominé plutôt par les sujets d'actualité (?). *Sina* poursuit ainsi son rôle historique de leader dans le domaine de l'infotainment.

Pourtant la population de jeunes urbains qui soutient la croissance de son service de microblog reflète aussi les transformations en cours dans la société chinoise. Les journalistes et spécialistes de l'information sont les premiers à se saisir de ce nouveau média. Dans un discours à Stanford en 2013, le PDG de *Sina* Charles Chao explique :

“Le plus grand changement apporté par le microblog en Chine concerne d'abord l'industrie des médias elle-même. Aujourd'hui, plus de 30% des actualités ont d'abord été reportées sur *Sina Weibo* avant d'atteindre les médias traditionnels. Le rôle des médias traditionnels a été déplacé vers un traitement des informations en profondeur (in-depth reporting).”¹³ (traduction de l'auteur)

L'omniprésence des supports mobiles (smartphones, tablettes)¹⁴ permet en effet des modes de traitement de l'information jusqu'ici inconnus qui bousculent les hiérarchies très contrôlées des salles de rédaction chinoises. Alors que la population urbaine croît

13. Charles Chao, PDG de Sina pendant la Stanford Graduate School of Business China 2.0 tenue le 3 Octobre 2013. Disponible en vidéo <http://www.youtube.com/watch?v=tlliivJKHk8>, consultée le 19 Février 2014 à 11 :23

14. Selon l'Universal Telecommunication Union, « la Chine dépasse 1 milliard d'abonnements mobile, avec 400 millions d'utilisateurs d'Internet mobile dépasse ainsi les États-Unis comme leader du marché des smartphones », <http://mobithinking.com/blog/china-top-mobile-market> consulté le 24 Février 2012.

rapidement, le smartphone est “*the first big urban purchase*” (Wallis, 2013) pour ceux qui arrivent en ville et représente un outil indispensable de participation à la société. En 2008, la Chine était le seul pays en Asie où les moins de 30 ans possédait plus d’amis en ligne que hors ligne (?). Ainsi, les réseaux sociaux jouent un rôle primordial dans la socialisation urbaine et viennent changer les modes d’expression. Il est à noter que les spécificités de l’écriture chinoise rendent possible l’écriture d’un court texte en 140 caractères, alors qu’une telle longueur autorise seulement une courte phrase dans une écriture utilisant un alphabet latin. Alors que le lectorat chinois a perdu toute confiance dans la plupart des médias traditionnels suite à l’absence répétée de courage et à la rétention d’information cruciale dans les dossiers importants qui animent le pays, le microblog s’installe comme une nouvelle source de confiance pour des millions de citoyens voulant comprendre et prendre part aux changements cruciaux de la Chine moderne.

Un rapport de l’Institut de Journalisme Reuters à l’Université Oxford paru en 2013 montre comment les usages du microblog ont amené des transformations dans le quotidien des journalistes chinois. Le journalisme d’investigation a notamment connu un essor important avec le renouvellement des sources grâce à une large diffusion en ligne des sujets. *Sina Weibo* n’a pas amélioré nécessairement la qualité de leurs investigations, mais a par contre permis une plus grande dissémination. La mobilisation des utilisateurs pour la protection des journalistes a également joué un rôle important ainsi que le renforcement de procédés de vérification existant depuis longtemps sur les forums du web chinois. Très populaire dans les années 2000, Le “*moteur de recherche de viande humaine*” (*renrou sou-suo*) est une forme de traque d’individu en ligne réalisée par un large nombre d’internautes à partir d’un nom ou d’une photo. Il s’agit souvent de retrouver quelqu’un désigné comme “coupable” (d’adultère, de corruption, etc) en réunissant un maximum d’informations à travers la Toile afin d’identifier ou de localiser la personne. Devant les dérapages rapides de ce type de procédés, les questions d’éthique du journalisme en ligne sont au cœur des discussions pour les journalistes et les titres de presse. En effet, l’usage des médias sociaux permis à certains journalistes de faire pression sur les pouvoirs publics, amenant parfois à la censure de leurs travaux, mais a également permis une très large auto-promotion pour de nombreux journalistes devenus des stars de *Weibo*.

Le contrôle des contenus sur *Weibo* est donc une réalité quotidienne et a été depuis son lancement la source de plusieurs études. Les formes les plus courantes sont : la suppression de posts, la suppression de comptes utilisateurs et le blocage de mots-clés. Le blocage de mots-clés s’effectue dans le moteur de recherche interne du site (“pas de résultats” quand vous chercher un mot bloqué) et plus récemment par l’impossibilité de poster un message contenant des mots ou des adresses web bloqués (Ng, 2013). La pratique de la suppression de comptes s’est intensifiée en 2013¹⁵ avec notamment la suppression de

15. “Over 100,000 Sina Weibo Accounts Shut Down or Penalized for Govt Rules Violations” par Gabriela Vatu, 14 November 2013 <http://news.softpedia.com/news/Over-100-000-Sina-Weibo-Accounts-Shut-Down-or-Penalized-for-Govt-Rules-Violations-400289.shtml> consulté le 17 Février à 16 :42

millions de “zombies” présents sur le site. Les “zombies” sont des comptes utilisateurs créés par des robots qui se chargent de reposter automatiquement des contenus, souvent afin d’augmenter le trafic sur le site. Les exigences des annonceurs publicitaires de plus en plus présents sur le site ont obligé *Sina Weibo* à faire la chasse aux robots sur son site, faisant ainsi diminuer le nombre de comptes actifs de manière significative. La firme Sina est garante auprès du Ministère de la Sécurité Publique chinois des contenus qu’elle diffuse et effectue à ce titre une surveillance constante pour supprimer les messages “non-conformes”. Lors de nos recherches, nous avons constaté que l’interface de *Sina Weibo* garde la trace des commentaires supprimés par le système d’administration. Dans les messages supprimés se trouvent à la fois des posts d’utilisateurs “zombies” et les posts jugés incorrects par les administrateurs. Au total, la suppression des messages s’effectue avec un taux estimé à environ 16%, allant jusqu’à plus de 50% dans certaines provinces comme Ningxia ou le Tibet contre seulement 12% à Beijing ([Bamman et al., 2012](#)).



Graphique 1.2.4 – La trace des commentaires supprimés par Sina est encore visible - Page 677 à 708, les commentaires ont été supprimées, soit approximativement 4% messages supprimés (589 messages sur 13452, à raison de 18 à 20 messages par page). Capture d’écran effectuée le 29 Janvier 2013 à 12 :32 :42, <http://www.weibo.com/1701401324/zeoBquVKi>, consulté le 16/02/2013.

1.3 Code, langage et milieu(x) numérique(s)

L'espace d'expression offert par l'Internet chinois et ses services de réseaux sociaux possède donc de multiples particularités et amène une variété de pratiques, de questions et de réflexions qui nécessitent d'être appréhendées avec des outils de problématisation et d'analyse spécifiques. Dans la troisième et dernière partie de ce chapitre introductif, nous allons donc nous pencher sur les concepts existants dans la littérature scientifique qui peuvent nous éclairer sur l'existence *in situ* des objets numériques. Afin de mettre en perspective le Web chinois et l'histoire de ces objets, nous introduirons notamment le concept de *milieu numérique*, héritier de l'idée complexe de milieu que nous explorerons

ci-après.

1.3.1 Lieu, espace, territoire et technologies

Géographie, management et diffusion de l'innovation, histoire des technologies, *cultural studies* ou études "nationales", les travaux qui s'intéressent aux relations entre technologies, espace, lieux et territoires sont nombreuses et offrent un paysage riche où se croisent de nombreuses disciplines scientifiques. L'influence des réseaux de transports sur l'expérience humaine et le développement des villes a notamment été largement étudiée (Offner, 1993; Doulet, 2001). L'Internet lui-même a également fait l'objet de nombreuses études monographique (par pays) ou comparative, une étude à l'échelle mondiale présentant en effet des problèmes de données et évidemment d'échelle (Dupuy, 2004). En Chine où l'urbanisation produit actuellement une vaste mobilité notamment des ruraux qui arrivent en ville, l'utilisation des réseaux sociaux médiatise bien souvent les choix de lieux et les rencontres des nouveaux arrivants. De nombreux groupes de discussions réunissent par exemple les nouveaux acheteurs d'immobilier qui échangent leurs stratégies d'achat et de défenses de leurs droits et de leurs biens, organisant aussi des rencontres (Li *et al.*, 2013). L'usage des réseaux sociaux revêt pour les nouveaux arrivants une importance capitale, notamment dans la recherche de groupes similaires et l'échange d'expériences. La disponibilité des données et le rôle croissant de la cartographie dans l'usage d'Internet viennent produire un « *Geoweb* » constitué de données et métadonnées spatiales (Crampton, 2009). Les vastes quantités de données produites par les multiples humains et machines sont souvent réunies et comprises dans leur attachement commun à un lieu physique (Torrents, 2010) offrant un point d'entrée unique (Open Data territorial, geo-localisation, GIS, POI¹⁶, etc.) par le marqueur spatial du *geotag* (?).

La carte notamment a joué sur le web un rôle important d'abord en tant qu'illustration, puis plus récemment d'interface avec le réel. Loin d'être figé par le territoire, la carte le décrit sous un ou des angles particuliers (Brunet, 1987; Jacob, 1992). Le développement de standards comme le système GPS (Haklay *et al.*, 2008) et de services et outils de cartographie ont contribué à une appropriation de la pratique cartographique par un nombre croissant de personnes (?). De nouvelles formes de données produites par les utilisateurs parfois appelées « *volunteered geographic information* » (Elwood, 2008) utilisent les services en ligne comme Google Maps pour dessiner un « miroir du monde offline » (Graham et Zook, 2011). Le courant dit de la néogéographie utilise abondamment le GIS et les outils en ligne (Google Maps, Flickr, etc.) pour comprendre les pratiques de ces nouvelles formes de « *géographies volontaires* » (?). Cette présence accrue dans le réseau dessine l'enjeu non seulement de cartographier le monde, mais également de cartographier le réseau lui-même, ouvrant ainsi de nouvelles voies pour découvrir la construction sociale des espaces par des pratiques individuelles et de groupe. Dans cette étude, nous avons

16. GIS : Geographic Information System ; POI : Point-Of-Interest

donc choisi d’interroger les objets numériques afin de comprendre comment se structurent la parole et la conversation dans le contexte unique de l’Internet chinois. Afin d’articuler les multiples dimensions d’analyse qui viennent enrichir notre réflexion, il nous faut donc brosser un portrait en large de l’Internet chinois, en le considérant tour à tour comme un espace *structurant* pour les actions des internautes qui le pratiquent, comme un territoire *sujet* aux relations de pouvoir actualisées par les groupes et individus et enfin comme un lieu *habité* par ceux qui y construisent chaque jour des significations communes. Nous proposons donc ici une revue sélective des quelques travaux à même de nous apporter des éclairages pertinents dans la vaste littérature s’intéressant aux dimensions géographiques des TIC.

Code / space : l’espace transductif des TIC

Dans leurs recherches autour de la géographie des technologies numériques, Dodge Kitchin ont travaillé à développer le concept de code comme un élément fondateur des espaces modernes dans lesquels nous évoluons. Reflétant l’importance croissante accordée aux TIC dans l’environnement urbain, ils citent un travail sur la production automatique des espaces : “*De plus en plus, les espaces de la vie quotidienne nous parviennent chargés de logiciels (software)*” ([Thrift et French, 2002](#)) En effet, si le projet urbain a été guidé pendant le demi-siècle dernier par l’apparition de la technologie automobile dans l’espace des rues, les TIC paraissent prendre le relais avec l’idée d’une ville intelligente et connectée, connue sous le nom de « *smart cities* » ([Ascher, 2009](#)). Dodge & Kitchin ont donc fait du *code* une des pierres d’angles de l’appréhension de l’espace dans leur travail en le définissant comme suit :

« an instruction or rule that has a single outcome determined by a binary logic (yes/no). The combination of these individuals logic rules produces code (program)” ([Kitchin et Dodge, 2011](#)).

La part croissante des TIC dans nos espaces quotidiens amènent les auteurs à envisager l’espace dans son interaction avec le code, symbolisant la suite d’instructions machiniques et électroniques qui permettent à un espace de remplir sa fonction. Dans un article intitulé *Flying through code/space : the real virtuality of air travel*, Dodge & Kitchin analyse la structure des espaces aéroportuaires. De l’achat des tickets jusqu’au vol des avions en passant par la gestion des bagages, le bon fonctionnement d’un aéroport est entièrement conditionné par le bon fonctionnement des longues successions d’instructions du code. Ici, Dodge et Kitchin propose le concept de code/space pour décrire ce type d’espace spécifique où lorsque le code échoue (*failure*) alors le code/space tout entier échoue ([Dodge et Kitchin, 2004](#)). L’exemple de l’aéroport est parlant : si le système de check-in des bagages ou les machines responsables du contrôle de sécurité des passagers ne fonctionnent pas, alors l’espace aéroportuaire ne peut exister en tant qu’aéroport. L’analyse de la spatialité ne se situe alors plus dans un domaine sémantique ou narratif, mais plutôt dans

les processus et opérations qui s'y déroulent et le code et les technologies y jouent un rôle primordial :

« Code is employed as the solution to a problem, a particular kind of transduction is occurring.» (Kitchin et Dodge, 2011).

L'espace n'est pas un donné mais s'explique plutôt comme : “*une forme d'ontogénèse (en perpétuel devenir-au-monde), l'espace est une pratique ; un faire ; un évènement (...) qui ne pré-existe pas à son faire (doing)*”. L'espace est considéré non pas comme une production, mais comme une transduction. Reprenant le travail de Simondon sur l'individuation par la technologie, Dodge Kitchin présente l'espace comme une pratique qui comprend les actes, actions, occurrences, mémoires, perceptions, etc. d'un groupe d'individus s'y trouvant. La fonction de l'espace structure et est structurée par les individus et le code y est considéré comme une entité agissante. Dans le *code/space*, la relation dyadique entre code et espace est bijective : l'un ne peut aller sans l'autre. En terme simondonnien, la *transduction* ne peut être assurée sans code. Si l'exemple de l'aéroport illustre bien cette nécessité du code dans le devenir-espace, Dodge Kitchin ont également identifié d'autres catégories où cette relation est plus ténue : les *coded spaces*, qui peuvent poursuivre leurs fonctions même lorsque le code échoue ; les *background coded spaces* où les processus de transduction induit par l'espace ne s'appuie pas nécessairement sur le code, mais propose néanmoins des possibilités de l'activer (machine éteintes ou inactives, etc.)

L'analyse fonctionnelle des rapports entre espace et technologie de Dodge Kitchin offre à voir comment les TIC peuvent être un facteur *transductif* pour les individus se mouvant dans les espaces de leurs vies quotidiennes. Si nous appuyons pleinement ce constat, il nous semble que le parti-pris des auteurs de considérer le “code” comme une abstraction incluant uniquement les instructions ou “logiques machiniques” ferme la porte à l'immense densité des activités symboliques qui se jouent dans l'usage des technologies. Comment notamment considérer les “contenus” du web dans cette grille de lecture ? Comment résister dans une perspective historique les logiques de médiation de l'espace par les technologies de l'écriture ? Il nous semble en effet que la faillite fonctionnelle (*failure*) des code/spaces précède l'arrivée des technologies et s'opère déjà à un niveau symbolique - la fonction de l'espace du Palais du Louvre après la chute des rois de France se voit radicalement modifiée. La transduction opérée lors de la pratique d'un espace s'effectue donc dans un jeu d'appropriation symbolique qui passe notamment mais pas seulement par les technologies. Ici les technologies du langage et de l'information jouent notamment un rôle crucial dans l'affirmation du récit symbolique (*narrative*) qui construit l'espace. L'activité du code dans la structuration des code/space de Dodge Kitchin existe donc sous une forme non seulement fonctionnelle mais également sémantique, voire phatique ou même esthétique comme l'a décrit Jakobson dans ces analyses des fonctions du langage (Jakobson et Halle, 1956). Au-delà de sa dimension machinique, le code possède les caractéristiques d'une *poiesis* dépassant l'idée simple de fonctionnalité pour exister dans

la complexité d'une écriture comme traduction du langage humain et machine.

Codes, discours et territoires des technologies

Le code serait davantage à comprendre comme un mode d'expression humain à travers la technologie, actualisant l'*episteme* décrit par Foucault dans *Les Mots et Les Choses* comme l'élément fondamental de la pensée d'une époque et sa considération pour le monde (Foucault, 1996). Définie comme l'état des connaissances scientifiques et littéraires, l'*episteme* existe comme la somme des savoirs d'une époque présupposée traduite en un regard sur le monde. Le *code* exprimé dans de nombreux langages est une écriture où s'expriment les savoirs d'aujourd'hui. Le code source d'une page Internet, d'un programme informatique ou d'un driver hardware ne s'écrit pas seulement en langage "machine" mais mélange langages informatique et humain. A la fois production savante, outil scientifique, vecteur d'expression et interfaces des savoirs, le code constitue l'expérience narrative du monde par les TIC. Possédant de nombreux mots, aspects et syntaxes issus de multiples langues humaines, les replis de l'écriture informatique laisse transparaître de tout bord leur origine humaine. Les standards d'encodage des bases de données comme l'Unicode deviennent presque le nouvel alphabet de notre écriture (Guichard, 2014). Une minorité de personnes savantes, "lettres" de l'informatique déchiffrent et écrivent le code alors que l'immense majorité se voit frappée d'illettrisme devant l'arrêt soudain de l'ordinateur ou la perte d'un fichier. Ainsi, la définition du "code" de Dodge Kitchin doit être étendue pour recouvrir plus largement les pratiques symboliques liés aux activités de l'écriture du code dans ces espaces.

Le code ainsi redéfinit nous ramène alors à une lecture foucaldienne du discours dans sa relation intime avec le territoire (Foucault, 2004). Dans ces nombreux travaux sur la généalogie, Michel Foucault cherche à comprendre comment les relations de pouvoir créées par les discours portés sur les objets préside à la production de territoires, d'interdits comme autant de sujets de ces discours. Nous définissons la *discursivité* comme le processus de construction de ces discours, et ce faisant le processus de définition des territoires de l'Internet issus de la pratique du code. Historiquement, l'Internet a été très tôt sujet à l'appropriation par le discours de nombreux groupes actifs dans leur volonté de territorialisation. La métaphore géographique et spatiale a structuré le vocabulaire de l'Internet dès sa création (Graham, 1998) : *site*, cyberspace, etc. L'*Electronic Frontier Foundation* se charge de protéger l'u-topie qu'est Internet avec JP Barlow qui rédige la fameuse *Déclaration d'Indépendance du Cyberespace* (?). A l'opposé du spectre, les autoroutes de l'information in-forment le paysage comme autant de géogrammes massifs (Berque, 1999). L'appropriation des protocoles du réseau, la lutte pour la défense de standards ouverts sont autant d'expression qui s'ancrent dans les pratiques du discours, dont les mots free et open cristallisent l'histoire de luttes de revendication des territoires de l'Internet (?). L'autre grande métaphore constitutive de l'Internet est textuelle avec ses pages, langages et hypertextes (?). La formule choc « *Code is law* » (?) résume l'idée que

les processus textuels du code mettent en jeu un ensemble de rôles, protocoles et mises en scène qui agissent comme autant d'autorités à travers le discours. La jurisprudence fait loi comme écrit sur les murs des bureaux de *Facebook* à Palo-Alto : « *Code wins arguments* »¹⁷. La territorialisation de l'Internet se fait donc au travers d'un ensemble de pratiques discursives, métagrammaire des discours en ligne où les luttes symboliques se jouent dans les pratiques discursives du code. La confrontation symbolique au sein des territoires numériques se poursuit dans le discours, réifié dans les pratiques du code. Les structures fonctionnelles et symboliques deviennent les stratégies de production de territoires qui se définissent dans l'usage des langages.

Sur l'Internet chinois, les pratiques de censure de l'écriture en sont le reflet le plus frappant. Blocage de mots-clés, détournements de langage, suppression et modification de texte sont l'expression de cet affrontement de discursivités parfois antagonistes. La Grande Muraille gouvernementale scanne les masses de texte pour reconnaître et stopper des mots tels que « *Printemps arabe* » ou « *événements de Tian-An Men* » (?). Néanmoins, l'état actuel des techniques de *data mining* ne permet pas encore de déceler les phénomènes langagiers comme les jeux de mots ou l'ironie. Bien souvent, les internautes chinois choisissent l'humour pour permettre à leurs idées de se frayer un espace. Revêtant leurs masques de chat, les internautes chinois sont devenus spécialistes dans la publication de jeux de mots, chansonnettes et petites vidéos d'animaux, comme autant de couperets cinglants pour railler les officiels trop pompeux de Pékin. Dans la guerre de l'information que se livrent sans cesse censeurs et internautes, les véritables héros sont bien souvent de simples photos truquées de crabes et de lamas. Ces blagues numériques, d'apparence bien souvent inoffensive, font chaque jour le tour de la Toile chinoise, portant en elles toute la subversion d'internautes aspirant à plus de liberté. Début 2010, alors que pleuvaient les longs discours pieux du Parti sur l'harmonie de la nouvelle société (en chinois *hexie*), on voit apparaître en ligne des essaims de crabes de rivière (se prononçant également *hexie*) couverts de chaînes en or criant : “*Vive l'harmonie*” au volant de leur limousine. Devenus aujourd'hui une image vivante de la corruption des hauts dignitaires du Parti, on croise régulièrement dans les commentaires d'un article officiel un petit crabe de rivière, comme un petit rappel posté par un lecteur.

On voit bien comment le code décrit ici un territoire sujet à l'autorité politique sous la forme d'un système de *data mining* cherchant à mettre en forme le discours. La fonction du *code/space* sémantique que forme ici l'espace de l'Internet est réécrit par un jeu de langage et la circulation d'objets digitaux permet de reterritorialiser cet espace en apparence régit par un code strict de censure.

17. Dans la Lettres aux Investisseurs écrite par M. Zuckerberg pour l'IPO de *Facebook* http://www.sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm#toc287954_10



和谐 (hexie) : Harmonie 河蟹 (hexie) : Crabe de

(b) Discours du PCC sur l'harmonie

(c) Mèmes / humour des internautes

(c)

Graphique 1.3.1 – Mot de la semaine : Crabe de Rivière - China Digital Times du 21 Mars 2012 - Licence Creative Commons <http://chinadigitaltimes.net/2012/03/word-of-the-week-river-crab>, consulté le 15 Février 2013.

Les lieux des technologies comme

Dans son célèbre livre sur les Arts de Faire, De Certeau (?) considère la ville comme un texte dont chaque piéton énonce et révèle (*performe*) des sens nouveaux par son activité de marcheur. Actualisant l'espace urbain par sa marche, l'habitant de la ville s'approprie des lieux qui restent néanmoins partagés avec d'autres. Au détour des rues, le sens commun des lieux urbains se construit avec les multiples énonciations de ceux qui les habitent et les font vivre. Cette magnifique image de la poésie du texte urbain met en lumière la dualité que nous avons abordé précédemment avec la *transduction* de Simondon : l'espace ne peut fonctionner sans les pratiques de ceux qui l'habitent. Plus encore, l'être-ensemble et le devenir-soi procèdent de la construction de lieux communs, *poiesis* des espaces habités. Les TIC font aujourd'hui souvent partie intégrante des lieux que nous habitons. Graham (Graham, 1998) dans son travail sur l'étude des lieux et de leur rapport à la technologie identifie trois types majeurs d'approches dans la littérature :

1. L'approche « *substitutive* » ou « *transductive* » qui voit dans l'arrivée des TIC la disparition de la valeur des lieux, dans un idéal de proximité utopique (?) ou un discours dystopique sur leur proche disparition (Virilio, 1998; ?). Ces considérations sont les formes traditionnelles du débat accompagnant l'innovation dont sont férus la communication industrielle et la critique des médias (?), restant souvent purement prospectif et faisant peu de cas des usages.
2. Plus modérée, l'approche qualifiée par Graham de « *co-évolutionniste* » s'interroge sur la façon dont les interactions dynamiques des espaces virtuels *space of flows* et réels *space of places* (?) produisent de nouveaux lieux . Ces études économiques

et sociales prennent la forme d'une médiologie de l'espace adaptée aux études stratégiques pour l'urbanisme et l'implantation des télécommunications. Son interprétation par les aménageurs lui donne parfois une dimension déterministe peu utile à la compréhension des phénomènes liés à l'appropriation et l'historicité des technologies (Offner, 1993).

3. Une dernière approche plus récente se cristallise autour de l'idée de relations et de réseaux. Afin d'éviter l'écueil des causalités directes et de la notion fataliste d'"impact", les lieux comme « *des moments articulés dans un réseau de sens et de relation sociales* » (Massey, 1993), des assemblages entre objets matériels "actants" (Latour, 1996), individus et groupes sociaux. Cette approche s'intéresse davantage au lieu comme une géométrie sociale en mouvement, liée au temps et à la situation (?) et refuse l'idée d'une existence "virtuelle" qui serait commune à différents objets (Bingham, 1996).

En privilégiant une approche dynamique et relationnelle des lieux comme constructions sociales de sens (Kyle et Chick, 2007), la technologie perd son rôle déterministe de productrice d'espaces et d'usages pour devenir une actualisation d'un espace-temps géographique et historique par des groupes d'individus. Comme le note Cresswell dans son travail sur les lieux : "places are practiced. People do things in place." (Cresswell, 2004). Il propose trois aspects pour décrire l'actualisation des lieux par leurs pratiques : location (un point dans l'espace, "the 'where' of place"), *locale* (les aspects visibles et tangibles du lieu, "the way a place looks") et *sense* ("the feelings and emotions a place evokes"). Néanmoins, cette définition ne permet d'appréhender l'existence des lieux en ligne (le site web d'un lieu est-t-il une *locale* ou une partie du *sense*?). Graham et Zook proposent le concept de DigiPlace afin de décrire plus spécifiquement l'existence des lieux en ligne : « *DigiPlace - that is, the use of information ranked and mapped in cyberspace to navigate and understand physical places (...) In other words, DigiPlace represents the simultaneous interaction with software (information) and 'hard-where' (place) by an individual.* » (Zook et Graham, 2007). S'inspirant des travaux de Harley sur le pouvoir du cartographe, leur travail sur le rôle de Google Maps dans la présence des lieux sur Internet met à jour l'interaction et l'hybridation entre existence spatiale et existence en ligne constituant des DigiPlace.

Les phénomènes de transduction à l'œuvre dans les pratiques spatiales de l'Internet sont donc le reflet des états du réseau à un moment donné, les lieux eux-mêmes formant ainsi un réseau basé sur leurs relations et similarités. Brunet dans son *Vocabulaire de la Géographie* propose notamment l'idée de synapses : « *espaces ou lieux par lesquels on passe, par où l'on communique, les isthmes, les détroits, les estuaires, les carrefours, les ports et les ponts, etc. ;* » (?). Envisagés par leur fonction "synaptique", les lieux deviennent alors un construit social au rôle clair, évitant l'écueil d'une qualification en-soi. Aéroports, hypermarchés, aires d'autoroutes ou zones industrielles ont été qualifiés de *non-lieux* par Marc Augé, produits d'une hyper-modernité « *qui ne peut se définir ni comme identitaire,*

ni comme relationnel, ni comme historique. » (?). Réfuté plus tard par l'auteur lui-même, cette idée de lieux simplement *produits*, et non construits fait l'impasse sur la tension d'usage, le devenir lieu que modèlent ses utilisateurs assidus ou épisodiques, ceux qui y travaillent voire même y habitent. Le lieu n'est pas nécessairement “patrimonial” comme produit d'une histoire, mais peut « *être ou ne pas être un non-lieu selon le statut de l'individu envisagé.* » (Debarbieux, 1993). Les hypermarchés et leurs galeries marchandes sont des haut-lieux de socialisation et de rencontres pour les adolescents qui vivent en banlieue (Matthews *et al.*, 2000) mais paraissent froids et inhumains aux habitants des rues du centre-ville.

L'étude menée par Puel, Pons et Xiaoting autour des pratiques sociales environnantes les café Starbucks de Beijing (Chine) montre également comment la stratégie marketing de la firme s'appuie précisément sur cette “absence patrimoniale” pour faire sa place au travers du vaste territoire chinois. La non-existence d'un “*bon café avec Internet*” dans les villes chinoises offre la possibilité aux clients, usagers de ces cafés récemment apparus de s'y rendre pour utiliser l'Internet ou retrouver leurs amis (Puel *et al.*, 2007). Ainsi, l'absence d'histoire n'interdit pas de la constitution de pratiques communes à forte valeur symbolique, qui sont à comprendre ici dans des réseaux d'appartenances mondiaux (jeune, dynamique, urbain, etc.). Dans son livre *The Great, Good Place*, Ray Oldenburg propose le concept de tiers-lieu pour décrire ces lieux qui, séparés de l'environnement de travail ou de la maison, permettent de se socialiser. Jouant un rôle majeur dans la construction de communautés et la constitution d'une société civile, ces tiers-lieu doivent répondre à certains critères comme : un coût d'accès nul ou modeste, une très bonne accessibilité, la présence régulières des « *habitues* », un lieu accueillant et confortable et enfin un lieu où l'on peut rencontrer facilement de nouvelles personnes (Oldenburg, 1999). L'idée de tiers-lieu virtuels a également été mentionnée pour désigner les chatrooms ou les plateformes de réseaux sociaux en ligne qui rempliraient des fonctions sociales similaires (Soukup, 2006).

Nous voyons donc qu'en considèrent les lieux de l'Internet, nous mettons à jour un ensemble de pratiques qui ne s'intéresse plus nécessairement à l'ordre du discours (et aux pratiques de censure notamment) mais plutôt aux phénomènes d'individuation, notamment au travers des multiples actes d'énonciation qui forment les pratiques et usages du web. Dans cette étude nous cherchons donc à comprendre de manière plus profonde comment les internautes habitent leur Internet. Il ne s'agit pas d'analyser le discours en terme de relations de pouvoir mais plutôt d'essayer de distinguer comment les circulations des objets numériques sont structurantes pour les pratiques des internautes, comme autant de lieux habités quotidiennement.

1.3.2 Le milieu : richesse et désuétude

Afin de problématiser les relations entre protocoles du discours et pratiques locales d'appropriation, nous avons choisir d'introduire le concept de *milieu numérique*. Nous définirons d'abord brièvement ce concept avant de présenter un regard historique sur

l'idée de milieu dans les sciences. Nous discuterons ensuite de l'acception particulière que nous avons choisi de défendre ici et nous verrons comment cette notion sera utile pour la suite de notre étude sur les réseaux sociaux en Chine.

L'idée de milieu numérique est a priori définie dans les termes suivants : "The multiple networks, which are connected by protocols and standards, constitute what I call a digital milieu." (Hui, 2012)

L'usage de multiples interfaces et le dédale des réseaux TIC constitue "*un nouveau milieu perceptif*" (?). Notre milieu physique est aujourd'hui (déc)ouvert par l'existence de notre milieu digital qui nous aiguille : rencontres, restaurants, voyages, etc. sont bien souvent médiatisés par l'Internet en premier lieu. Ainsi, nous évoluons dans un milieu numérique qui agit comme support des processus de transduction et de connaissance du monde. Le code prend ici pleinement part à la construction de ce milieu digital, à la fois déterminant pour la production des actes de discours et ouvert à l'appropriation des pratiques et usages du quotidien.

Historiquement, le concept de milieu se détache du centre pour résister et mettre en perspective la relation des êtres à leurs environnements sous des jours parfois contradictoires. Débattue puis écartée mais toujours très usitée, la notion de milieu introduit autant le déterminisme d'un combat pour la survie et l'adaptation, que la liberté créatrice du sujet dans un univers ouvert à sa volonté.

Pour illustrer au mieux l'extraordinaire fécondité philosophique de la notion de milieu, nous allons tout d'abord essayer de comprendre la trajectoire de ce mot durant les siècles derniers (Canguilhem, 1965). Sans pour autant remonter à son aube étymologique, nous nous apercevons que le mot milieu décrit un trajet singulier dans le monde des sciences. Employé dès le 16e siècle par Descartes dans son Traité de la lumière, il représente pour Newton une mesure de distance dans l'éther, cette non-matière qui structure la gravitation et fait se mouvoir les objets. Défini plus tard par d'Alembert dans son encyclopédie comme un : "*espace naturel dans lequel un corps est placé, qu'il se meuve ou non*", le mot connaîtra durant tout le XIXème un large développement sémantique en s'étendant de la physique à la biologie. Les naturalistes français de l'époque affirme que le milieu n'est pas seulement *environnant* mais influe sur les êtres vivants. Ainsi Lamarck dès 1809 écrira dans sa *Philosophie Zoologique* : "*le milieu a une grande puissance pour modifier les organes*". Par la suite dans l'histoire du XXème siècle, nous pouvons identifier deux moments marquants pour ce concept et structurants pour l'histoire des sciences dans son ensemble (Taylan, 2010). Le premier se joue sous la plume d'Auguste Comte qui, inspiré de la biologie naissante, articule le vital au social dans la sociologie naissante et décrit le milieu comme « *l'ensemble des circonstances extérieures (...) nécessaires à l'existence de chaque organisme déterminé* » (Comte, 1838). Sans sombrer pour autant dans un déterminisme total, Comte introduit une première dialectique des rapports avec le milieu comme conditions de possibilité de la vie : « *Tout être vivant (...) modifie sans cesse son milieu.* », écrit-il alors. Le concept qui connaît un succès croissant en France est

peu de temps après réapproprié par les penseurs d'Outre-Rhin qui lui donnent alors un sens différent. S'opposant au francisé *Der Milieu*, le géographe Ratzel introduit dans son *Anthropogéographie* (1899) le mot *Umwelt* qui se démarque rapidement par sa dimension fortement déterministe. Dans un même mouvement, le physiologue et biologiste Jakob von Uexküll étudie dans son laboratoire la tique et se rend compte que le milieu de la tique se définit non pas par tout l'environnement qui l'entoure mais seulement par ce qui lui est utile et approprié. Le milieu devenu *Umwelt* s'oppose alors à l'environnement indifférencié et devient l'ensemble des éléments qui sont porteurs de significations (*Merkmalsträger*) pour un être. Uexküll propose ainsi une « *biologie subjective* » qui étudierait les relations de chaque espèce avec son milieu. Dans l'Allemagne du siècle débutant, Uexküll cherche à diffuser largement sa théorie qui s'adresse pas tant aux animaux qu'à l'humain dont le milieu sera la Patrie (*Heimat*) (Feuerhahn, 2009). Reflétant les débats guerriers entre la *Kultur* allemande et la *Civilisation* française (Elias, 1975), se cristallise dans l'idée de Milieu une tension politique sur les relations entre nature et vivant qui devra déchirer l'Europe pendant longtemps encore. Foucault dans son cours au Collège de France du 11 Janvier 1978 parle de l'influence de l'idée de milieu sur la conception du territoire pour les urbanistes du XVIIIème siècle. Quand sous Louis XIV, les villes étaient construites d'après un espace conçu comme vide (voir *Richelieu* en Indre-et-Loire), la question de l'urbaniste du XVIIIème est de comprendre la ville dans son évolution future. L'enjeu est devenu l'*adaptation* du milieu existant (à la fois urbain et naturel), la transformation du *donné* compris comme un élément qu'on peut venir modifier. L'idée sous-jacente de milieu ouvre la possibilité de l'appropriation de la nature. En terme foucaldien, cette nouvelle bio-politique se fonde sur la territorialisation du milieu comme nouveau centre des enjeux de pouvoir. L'ère industrielle réalise ce projet d'une adaptation à la fois *au* et *du* milieu vu comme tension nécessaire de l'évolution, passage obligé vers la civilisation. Alors que l'humain s'est vu déplacé dans son rôle central par l'astronomie galiléenne puis l'évolution darwinienne, l'idée de milieu pose comme enjeu majeur du vivant la maîtrise de l'environnement - la lutte pour ne pas être maîtrisé. Poursuivi par la psychanalyse de Freud qui introduit l'Autre au sein du sujet, l'entreprise de décentrement de l'humain vers son milieu se joue dès l'abord dans les termes de la vie ou de la mort. Plus tard, Lacan identifiera la transition de la petite enfance à l'enfance par le « stade du miroir » comme moment où l'enfant différencie enfin le Milieu (*Umwelt*) du Soi (*InnenWelt*) (Lacan, 2001). Ce "passage au milieu" est donc d'une importance capitale puisque s'y joue la constitution de l'être.

L'approche du milieu comme élément unificateur des sciences est également un sujet toujours en discussion. L'écologie notamment a largement recentré le milieu sur le rôle déterminant de l'Homme par l'introduction du concept d'*environnement* (?). Les conséquences de ce passage de l'idée de nature à celle de milieu restent profondes, notamment dans le droit civil où nous sommes passés d'un rapport du « droit imposé » de la nature au « droit négocié » du milieu (Papaux, 2008). Méta-réflexion, la discussion sur le "milieu

académique" donne lieu à d'intéressants échanges (Stengers *et al.*, 2009) qui interrogeant notamment la définition trop abrupte des disciplines scientifiques et leur herméticité. Parfois nommé *mésologie*, l'étude du milieu se donne pour mission de réconcilier des pratiques diverses de la biologie à la sociologie en imaginant une étude par le milieu (Stengers, 2003). Les œuvres de Deleuze Guattari, dont la mésologie se revendique, parlait déjà de la pratique d'une philosophie du milieu : "*Partir au milieu, par le milieu, entrer, sortir, non pas commencer ni nir, [...] renverser l'ontologie, destituer le fondement, annuler n et commencement. [...] C'est que le milieu n'est pas du tout une moyenne, c'est au contraire l'endroit où les choses prennent de la vitesse.*" (?) La réflexion sur la technique et les technologies s'est notamment saisie à bras le corps de cette notion, avec notamment l'idée de *milieu technique* par Friedman et Leroi-Grouhan (Stiegler, 1998). Tout geste (du plus banal au plus rare) s'effectuerait dans un milieu technique qui le rend possible. Gilbert Simondon dans son livre Du mode d'existence des objets techniques continue cette réflexion avec ce qu'il appelle le milieu associé :

« médiateur de la relation entre les éléments techniques fabriqués et les éléments naturels au sein desquels fonctionne l'être technique. (...) C'est ce milieu associé qui est la condition d'existence de l'objet technique inventé. » (Simondon, 1989).

Simondon problématise le milieu associé comme vecteur de l'*individuation*, où se produit la rencontre entre objets et individus pour chacun les actualiser. Poursuivant ce travail, Stiegler comprend les technologies de l'information comme un milieu essentiellement social, comprenant à la fois ce qui est autour de l'individu (environnement) et entre les individus (medium) (?). Sans sa lecture critique des industries culturelles, Stiegler formule l'idée qu'un milieu est *associatif* s'il permet l'individuation. A l'inverse, certains milieux seraient *dissociatifs* car ils ne permettraient pas le devenir individu, à l'image des mass media qui divisent producteurs et consommateurs de symboles. La dynamique industrielle des deux derniers siècles a entraîné une massification des phénomènes culturels, créant un milieu technique considéré comme largement dissociatif car dé-réalisant les individualités (Simondon, 1989). Néanmoins, le renouveau technologique porté par l'apparition des technologies numériques ouvre aujourd'hui une page nouvelle pour l'individuation en offrant un milieu extrêmement associatif, fondé pour ainsi dire sur le lien. Voyant un nouvelle âge des Lumières (Stiegler, 2012), Stiegler conçoit l'Internet comme un milieu qui ne serait pas structurellement dissociatif et pourraient donc recréer de nouvelles formes plus horizontales d'économie symbolique où existent davantage de symboles partagés.

1.3.3 Cyberespace et milieu numérique

Associateur ou dissociateur, les protocoles qui régissent l'accès au milieu sont donc les enjeux politiques du milieu numérique, conditionnant l'existence des objets numériques.

L'étude des relations entre espace et dispositifs socio-techniques ne se comprend donc pas seulement en termes d'infrastructures, mais plus finement dans l'observation et la description d'une géographie du réseau. Les débuts de la géographie au XIXème siècle définissent en effet cette discipline comme la science des milieux. Vidal de la Blanche cherche alors à expliquer comment les actions humaines sont déterminées par des faits "naturels" pré-existants. Alors que la sociologie fait école, les géographes aux aussi commencent à considérer les œuvres humaines comme partie intégrante du ou des milieux qui les produisent (?). Cette nouvelle géographie humaine cherche à inclure à mettre en relation : "*l'étude des relations verticales qui se développent au sein de chaque milieu, et celle des relations horizontales qui mettent en relation les milieux*" (Claval, 1990). Les réflexions géographiques sont nourris par de vastes controverses sur de nouveaux paradigmes : l'espace, le territoire, le paysage, les lieux qui estompent peu à peu l'idée déterministe de milieu pour développer un appareil conceptuel plus complexe. Le développement méthodologique avec notamment la géomatique et les outils d'analyses issu de la statistique permettent de proposer des lectures variés de faits géographiques divers. Brunet définit les *chorèmes* comme des "structures élémentaires d'organisation de l'espace" (Brunet, 1980), Berque parle de géogrames définis comme "*motif éco-techno-symbolique (...) au sein de la relation qu'est l'écoumène*" (Berque, 1999). Inspiré de la philosophie japonaise moderne, l'écoumène de Berque se rapproche de l'idée de milieu et est décrit comme une vaste matrice relationnelle des choses et des êtres - une dimension écologique générale, une "*trajectivité*" (Watsuji et al., 2011). Au sein de cet écoumène, le géogramme offre un modèle des faits géographiques où se rejoignent les aspects techniques, symboliques et sociaux (les relations humaines). Alors que la notion de milieu disparaît peu à peu pour être remplacée par celle d'environnement(?), l'espace des géographes se voit bientôt augmenté d'une nouvelle réalité à prendre en compte : le cyberspace. Espaces, sites, routes, les nombreuses métaphores géographiques viennent remettre en question des pans entier de la discipline et renouveler les approches. Le cyberspace, "*hallucination consensuelle*" décrite par Gibson (?) déclare bientôt son indépendance (Barlow, 2001) et dessine ainsi une géographie virtuelle (Batty, 1997) qui s'interroge sur les dimensions de ces nouveaux espaces décrits par la circulation d'information. Les structures spatiales et économiques préexistantes semblent être renforcées par les stratégies territoriales et les équipements des acteurs. L'importance des Etats-Unis (Zook, 2001; ?) dans la localisation des flux Internet (capital, data centers, noms de domaines...) montre bien comment les évolutions technologiques participent à la fragmentation territoriale à l'échelle mondiale. Néanmoins, les routes de l'Internet s'écartent aussi bien souvent des autoroutes de l'information pour venir construire des sens beaucoup plus locaux par les nombreux mécanismes des activités en ligne. "*cyberspace is 'made real' through the language of place*", comme l'écrivent justement Dodge & Kitchin (Dodge et Kitchin, 2007). S'opposant aux objets naturels et techniques, les objets numériques sont à comprendre dans leurs relations matérielles et temporelles avec les infrastructures de leur production et archivage dans les mémoires des

données du web. Ainsi, le milieu numérique dans lequel chacun évolue se présente sous la forme d'objets numériques actualisés qui fondent le milieu associé. À l'instar des géogrammes du paysage de Berque ou des chorèmes de l'espace de Brunet, nous pouvons imaginer ici des “*topogrammes*” qui permettront de considérer les faits et objets digitaux. Le milieu numérique, tenant à la fois du cyber-espace (le lieu physique où se situent les machines) et constitués des objets digitaux qui lui sont spécifiquement associés. Le topogramme en tant que modèle permet de décrire et considérer sous un jour commun des objets digitaux dissemblables.

Afin d'observer et de décrire le milieu numérique en Chine et de tester l'hypothèse méthodologique proposé par la notion de topogrammes, nous avons choisi d'étudier de façon empirique les dynamiques à l'œuvre dans et autour de certains objets numériques particuliers sur les réseaux sociaux : les mèmes Internet.

2

Les mèmes Internet, objets numériques culturels

Afin de mener à bien cette étude, nous avons choisi de nous intéresser à un objet numérique particulier : les *mèmes Internet*. Courts messages se propageant rapidement sur la Toile, les mèmes proposent une illustration pertinente des discursivités multiples qui prennent place dans les échanges en ligne. Plus que de simples blagues de potache, nous verrons ici comment ces constructions collectives éphémères laissent parfois des traces symboliques qui structurent le milieu numérique qui les produit. Nous présenterons tout d'abord la notion controversée de *mème*, définie d'abord comme une unité minimale de diffusion des cultures. Nous verrons ensuite comment ce concept teinté d'un évolutionnisme peu convaincant flotte depuis sa création en marge de la littérature scientifique. En introduisant ensuite les *mèmes Internet*, nous observerons le fort regain de popularité dont a bénéficié ce mot ces dix dernières années. Nous envisagerons ensuite le rôle des mèmes Internet - et à plus forte raison des technologies numériques - en résitant leur étude dans la perspective historique plus large des questionnements sur la formation de mémoires collectives. En interrogeant notamment la notion de "performativité", nous verrons comment les pratiques *d'énonciation* qui entourent les mèmes Internet forment une pratique rhétorique et une activité symbolique importante dans l'actualisation et l'échange de signes. Enfin, nous détaillerons les processus discursifs entourant les mèmes en proposant une catégorisation selon l'intention des énoncés, comme prélude méthodologique à notre démonstration.

2.1 Les mèmes : définitions et histoire

Le dictionnaire d’Oxford¹ donne deux définitions du mot *mème* :

Meme (n.)

1. Un élément d’une culture ou système de comportement passé d’un individu à un autre par imitation ou par d’autres moyens non-génétiques.
2. Une image, vidéo, morceau de texte, etc., la plupart du temps de nature humoristique, qui est copié(e) et propagé(e) rapidement par les utilisateurs d’Internet, souvent après avoir été modifié(e).

Cette définition nous renseigne sur l’usage de ce terme, en le définissant à la fois comme un élément culturel transmissible et comme une forme particulière de contenus diffusé sur Internet. En nous appuyant sur son évolution dans la littérature, nous allons tout d’abord essayer de voir comment les deux versants de ce concept se sont historiquement articulés et à plus forte raison comment cette articulation peut nous servir pour comprendre les phénomènes à l’œuvre dans les réseaux sociaux en Chine.

2.1.1 La mémétique : une éthologie culturelle teinté d’évolutionnisme

Le premier usage du concept de *mème* est souvent attribué au biologiste Richard Dawkins dans son livre *Le Gène égoïste* (1976). Dawkins s’inspire des théories évolutionnistes de l’éthologie moderne pour proposer le *mème* comme un élément moléculaire de la culture qui permettrait sa transmission, semblable au gène des individus biologiques. Considéré comme une « *unité d’information culturelle qui peut être copiée, située dans le cerveau* » (Blackmore, 2001), le *mème* serait la fondation de pratiques culturelles qui évolueraient selon des variations de la sélection naturelle. Représenté comme une « *unité distincte de la pensée* » (?), le concept se fonde sur l’analogie entre les processus de transmission culturelle et génétique : “*Cultural transmission is analogous to genetic transmission in that, although basically conservative, it can give rise to a form of evolution*” (p.72). Cette approche éthologique de la culture considère donc le *mème* à la fois comme un élément transmissible et un facteur de transmission doté de la capacité de se reproduire lui-même. Élément actif, le *mème* serait donc un « *gène égoïste* » agissant de manière isolée et distincte, spécificité d’une “culture”. Plus encore, son but unique serait sa propre pérennisation par sa propagation de cerveau en cerveau (Blackmore, 1997). Il agit donc comme un agent culturel possédant une forme de volonté propre pour se propager. Souvent représenté grâce à l’image du virus, il donne l’idée d’une propagation de la culture sous forme de contamination avec en arrière-plan une lutte pour la survie et la fécondité des idées.

1. D’après *British & Words English* publié par Oxford University Press en 2014, <http://www.oxforddictionaries.com/definition/english/meme>, consulté le 24 Février 2014 à 21 :50

Le même est un “réplicateur culturel”, une extension à part entière du vivant au-delà du biologique (Bloom, 2002).

Le québécois Fernand Dumont définit la culture comme cette « *maison où l'on habite ensemble* » (?). Nœud dans une topologie sociale, le même pourrait donc également se présenter comme un point d’entrée, une porte entrouverte vers ce lieu où d’autres sont déjà passés et se trouvent encore. Le même devenu particule culturelle définit un seuil, ce lieu de passage si particulier qui « *fonde les espaces* » (Bonnin, 2000) et invite ou interdit d’entrer. Comme on enlève ces chaussures au dojo et qu’on sonne à la porte, les mêmes sont peut-être à envisager comme des rites de franchissement de seuils culturels, pratiques de liaison du vivre-ensemble politique d’Arendt (2001). Pour le gène comme pour le même, il ne s’agit pas de considérer la fonction mécanique d’un “réplicateur” mais d’observer l’altération qui se déroule lors de son actualisation pour en comprendre les limites et le rôle. La présence *in potentia* d’une unité culturelle identique ne constitue pas nécessairement une réalité in-formante pour des groupes sociaux ou des individus (Lissack, 2004). Les sciences de la communication ont largement étudié depuis 50 ans les modalités de transmission des informations. Les études sur la réception notamment ont bien montré qu'il ne suffisait pas qu'un message soit émis pour être décodé et compris (?). Déjà avec Shannon et Weaver (Jakobson, 1960), l'environnement exprimé par le concept de *bruit* vient altérer largement les phénomènes de transmission tout au long de leurs diffusions (?). La mémétique, faute d'étude de cas conséquentes et d'applications théoriques réelles (Jouxtel, 2014) a subi de nombreux revers conceptuels en s'appuyant notamment sur l'image peu crédible d'une transmission par réplication quasi-mécanique. Ignorant la dimension poétique des actes de transmission, cette vision mécaniste issue d'une rationalisation excessive reflète pourtant les écueils non-dits des approches scientifiques modernes. Thierry Bardini dans son livre *Junkware* (2011) effectue une recherche extensive sur les discussions et considérations qui entourent la partie non-codante de l'ADN appelée “*junk ADN*”. Analysant les discussions dans les publications scientifiques, il montre comment plus de 80% des éléments structurant l'ADN ont été très tôt étiquetés comme “*bruit*” puis “*junk*”, car il était impossible d'identifier leur participation active au codage de protéines. La métaphore de ce “*junk non-codant*” si envahissant et la relative facilité avec laquelle nous nous permettons de l'ignorer démontre la nécessité d'une approche renouvelée des phénomènes complexes du vivant, et notamment de ceux de la transmission culturelle. Déjà clairement identifiées dans les études en communication, les fonctions non-langagières notamment sont indispensables au bon déroulement d'un acte de langage. Ainsi si la suppression du bruit est souvent un préalable méthodologique pour l'étude scientifique, elle peut souvent fausser l'approche expérimentale et les conclusions théoriques en refusant d'admettre sa partialité. L'étude des mêmes est encore largement en quête de reconnaissance scientifique et si la construction théorique permettant d'isoler des éléments culturels pour l'étude paraît intéressante, elle manque d'une réelle prise sur l'observation et l'analyse par l'étude de cas notamment. La fermeture dès 2005 du *Journal of Memetics*, parution de référence

de la discipline naissante est annoncé dès 2002 par un article de B. Edmonds ([Jouxtel, 2014](#)). Intitulé *Three Challenges for the Survival of Memetics*, l'article exhorte les chercheurs intéressés à produire ce que Edmonds juge comme le minimum indispensable pour gagner la reconnaissance des milieux scientifiques : “*a conclusive case-study; a theory for when memetic models are appropriate; and a simulation of the emergence of a memetic process.*” ([Edmonds, 2002](#)). En effet, il paraît impossible d’assoir scientifiquement la légitimité du concept en se fondant uniquement sur une analogie de phénomènes. Si le même a donc raté sa cible dans le domaine scientifique, son acceptation plus récente sous la forme de contenus Internet a néanmoins redonné au concept une nouvelle vie dans la culture populaire. Dans le même temps, ce sens renouvelé a permis de définir précisément un domaine d’application idéal et a permis l’émergence de nouvelles études dont nous allons maintenant discuter.

2.1.2 Mèmes Internet : définition, littérature et exemples

Contrairement au concept éthérique de mème présenté dans la partie précédente, la définition des *mèmes Internet* est de prime abord plus pragmatique. Il s’agit de courts messages faits de texte, image, vidéo ou de son gagnant rapidement une forte popularité sur Internet en étant partagés, commentés, réappropriés puis transformés lors de leur diffusion. L’utilisation du terme *mème Internet* pour décrire la diffusion de messages ne recouvre pas nécessairement la dimension évolutive et culturelle du concept initial de Dawkins, mais garde l’idée générale d’une circulation « virale » d’idées parmi des groupes d’individus². Le concept de mème a très fortement gagné en popularité avec cette nouvelle acceptation. En 2012, il a notamment été sélectionné parmi les 10 mots les plus marquants de l’année par le prestigieux dictionnaire américain Merriam-Webster. Ce choix a été motivé par la très forte popularité sur Internet des images parodiques du politicien Mitt Romney après une bourde lors d’une intervention télévisée aux Etats-Unis³. Ainsi, le mot *mème* dans une acceptation que nous prendrons ici soin de nommer *mème Internet* est aujourd’hui entré dans le vocabulaire commun du Web. De nombreux sites spécialisés ([knowyourmeme.org](#), [quickmeme.org](#), [memefest.org](#), etc) ont vu le jour avec comme mission d’archiver et de collecter ces pièces de la culture web. Un des plus anciens mèmes Internet est certainement l’usage des émoticones ou smileys, ces petites figures qui servent à exprimer des émotions dans le contexte d’oralité écrite d’Internet. Apparu dans les premiers jours du réseau Internet, les émoticones répondent à un besoin d’expression non-verbale dans la communication en ligne. Très simples à utiliser ou à modifier, les *smileys* connaissent une popularité rapide et se diversifient partout autour

2. “*The meaning is not that far away from the original. It’s anything that goes viral.*”, Dawkins interviewé par le magazine [Wired](#) <http://www.wired.co.uk/news/archive/2013-06/20/richard-dawkins-memes>, consulté le 12/08/2013 à 7h53 GMT+8

3. “*Words of the year 2012*”, Merriam-Webster <http://www.merriam-webster.com/info/2012words.htm> consulté le 25 Février 2014 à 19 :01 GMT+1

de la toile.

19-Sep-82 11 :44 Scott E Fahlman :-)
From : Scott E Fahlman <Fahlman at Cmu-20c>

I propose that the following character sequence for joke markers :

:-)

Read it sideways. Actually, it is probably more economical to mark things that are NOT jokes, given current trends. For this, use

:-)

Graphique 2.1.1 – 19 Septembre 1982 : la première mention du smiley par Scott Fahlman⁴, que l'on retrouve quelques jours plus tard sur les mailing lists les plus utilisées de l'époque : Arpanet⁵ et Usenet⁶

L'usage des émoticones s'est aujourd'hui largement répandu, notamment chez les adolescents et jeunes adultes (D Derks, Bos, & Von Grumbkow, 2007). Une récente étude a même montré que les zones du cerveau stimulées par la vue d'émoticones étaient similaires à celles stimulées lors de la vue d'un visage, indiquant un ancrage symbolique profond de l'usage de ces signes (Churches *et al.*, 2014). Des émoticônes singuliers se sont également développés dans différentes langues pour exprimer des sentiments particuliers, propres au langage et à ses modes d'expression. Le caractère idéographique des langues chinoise et japonaise se prête particulièrement à ces jeux de dessin langagier. En chinois, le plus célèbre exemple est sans doute le caractère [56E7?] (jiong4) qui représente en langue ancienne une fenêtre d'où provient la lumière et signifie « lumineux ». Sa ressemblance avec une figure humaine (un émoticône) a fait renaitre ce caractère désuet qui signifie désormais qu'un utilisateur est agacé, embarrassé ou même choqué. La conjonction "[56E7?]rz" a même été inventé : [56E7?] représentant la tête et rz le corps agenouillé d'une personne ; elle signifie l'échec et le désespoir. Cette particularité des langues asiatiques donnent à l'émoticône un rôle central dans la communication en ligne qui se traduit dans le design des interfaces. Les réseaux sociaux chinois proposent tous par défaut de multiples jeux d'émoticônes disponibles pour l'utilisateur qui communiquent ainsi très rapidement en images.

L'exemple de l'émoticone illustre donc la manière dont un élément visuel et langagier

6. D'après <http://www.cs.cmu.edu/~sef/Orig-Smiley.htm> archive consultée le 10 Août 2013 à 09 :15 GMT+8

6. Usenet mailing list <https://groups.google.com/forum/#!msg/net.works/dzzAYYz9fYM/m6DZBY5ZpzUJ> archive consultée le 10 Août 2013 à 09 :25 GMT+8

6. Usenet mailing list <https://groups.google.com/forum/?hl=en#!topic/net.news/tywzP0ON0tw> archive consultée le 10 Août 2013 à 09 :20 GMT+8

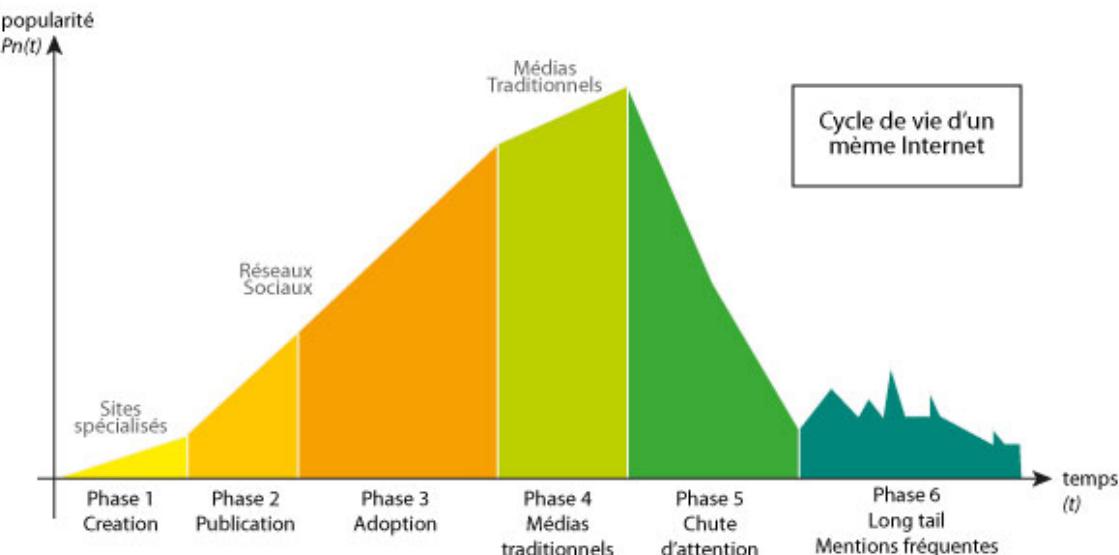


Graphique 2.1.2 – Un jeu d’émoticônes sur weibo.com – Consulté le 10 Aout 2013 à 10 :00 GMT+8

vient constituer les pratiques en ligne, en relation proche avec la culture et le lieu qui l’a fait naître.

Il est difficile de définir un mème a priori par la nature de son contenu. Néanmoins, la structure de la diffusion d’un message peut nous permettre de le décrire comme un mème. Reproduisant en grande partie le cycle de vie classique d’une information comme une rumeur ou une *news*, les mèmes possèdent des modes de diffusion en ligne assez déterminés et prévisibles. La plupart du temps, ils sont mis en circulation sur un petit nombre de sites spécialisés, avant d’être repris dans une première phase par un assez petit nombre d’utilisateurs qui se charge de les publier sur les réseaux sociaux (Bauckhage, 2011). Les réseaux sociaux agissent alors comme une chambre d’écho, qui détermine si le “proto-mème” encore en devenir deviendra mème ou restera simple message isolé. Durant cette phase souvent nommée *adoption*, le mème entre en concurrence avec d’autres informations sur les réseaux sociaux où les utilisateurs sont sans cesse sollicités par d’autres informations (Davenport & Beck, 2001). Si l’attention générée par le mème auprès des utilisateurs atteint un pic suffisamment important, il faut environ 2h30 pour que les mèmes rejoignent les pages des médias plus traditionnels en commençant par les blogs, puis les sites d’information (Leskovec, Backstrom, & Kleinberg, 2009) . Ensuite, l’attention envers le mème décroît fortement et rapidement. La présence épisodique de citations maintient l’existence du mème apparente dans des groupes définis (Buchel, 2012).

L’évolution du volume de la diffusion permet donc de définir un mème Internet. Néanmoins, il est impossible de donner une estimation du volume minimum pour devenir “mème” tant ce chiffre dépend de la population étudiée : il existe des mèmes à très forte diffusion comme le smiley ; d’autres mèmes se diffusent seulement au sein de groupes d’in-



Graphique 2.1.3 – Cycle de vie d'un mème Internet -Clément Renaud - 2013

dividus restreints sans s'étendre en-dehors. Ainsi, certains mèmes peuvent avoir connu une diffusion très importante dans un groupe, mais rester absolument inconnu du reste de l'Internet. Une étude de 2012 comparant la diffusion de nombreux mèmes sur Twitter montre que les utilisateurs tendent à choisir les mèmes selon la structure de leur réseau social et le moment d'exposition, produisant ainsi une grande hétérogénéité des mèmes dans le réseau (Weng, Flammini, Vespignani, & Menczer, 2012). Ainsi, il est hasardeux d'essayer de décrire le concept de mème par son contenu tant les sujets et les discussions varient.

Quelques éléments d'ordre grammaticaux peuvent néanmoins être observés dans la forme que prennent les contenus, appelé parfois “véhicule” du mème. Les mèmes qui sont diffusés les plus largement sont composés d'images et de vidéo. L'économie d'attention très limitée de l'Internet et les modes de lecture sur écran dans un contexte d'abondance d'information font que l'on privilégie souvent les médias visuels sur le texte (Goldhaber, 2006). Un autre élément important est la facilité avec laquelle un message peut être approprié par un utilisateur qui veut le modifier ou tout simplement le diffuser dans le réseau. L'existence des mèmes est en effet largement conditionnée par la possibilité d'une diffusion à moindre coût et effort pour l'utilisateur final, la plupart du temps non-rémunéré. Ici on voit émerger une structure visuelle caractéristique du mème : une image accompagnée d'une légende écrite en caractères blancs détournés de noir, ou de caractères blancs sur fond noir. L'utilisation de haut contraste de couleur dans les typographies permet de faire apparaître très efficacement des légendes juxtaposées à l'image.

L'usage d'images légendées est une des formes les plus communes pour les mèmes Internet, en particulier ceux de nature comique ou absurde. La mise en place de sites permettant de générer rapidement ce type d'images légendées (memegenerator.com, mememachine.com, etc.) renforcent l'unité formelle des mèmes sur l'Internet, ou plutôt sur



Graphique 2.1.4 – Exemples de mèmes Internet, d’après <http://knowyourmeme.org>, consulté le 12/08/2013 à 10 :01 GMT+8

l’Internet anglophone et francophone notamment. En effet, on constate que cette forme typique du mème ne se retrouve pas sur l’Internet chinois qui utilise plus volontiers des montages d’images ou des jeux de mots, avec davantage de diversité dans les formes que peuvent prendre les différents mèmes.

2.2 Mème, mémoire collective et culture

2.2.1 La mémoire comme trace

Si le concept de mème est souvent considéré comme très récent, on peut néanmoins le resituer facilement dans le vaste paysage des travaux sur la mémoire collective qui ont existé depuis le XIXème siècle (Laurent, 1999). Max Stirner dans son livre *The Ego and Its Own* (1844) énonce déjà l’idée que les individus sont sujets à la circulation de concepts issus de souvenirs communs ou illusoires, comme notamment le nationalisme et la religion. Le logicien Bertrand Russell reprend par la suite dans son livre *The Analysis of Mind* (1921) les travaux sur la mémoire et l’évolution sociale du physiologiste allemand Richard Semon (1904). Utilisant le concept central de *mneme* (du grec mneme, mémoire), Semon travaille sur l’idée de “*traces mnésiques*” laissées par les diverses expériences au niveau cellulaire comme au niveau de l’organisme tout entier. La psychanalyse a également cherché à saisir cette distance impalpable entre expérience et organisme en interrogeant les marques laissées par les souvenirs. Pour Freud comme pour Semon, “l’appareil psychique” de la mémoire se constitue sous la forme de “traces”, qu’il se refuse néanmoins à localiser dans des zones spécifiques du cerveau. Lacan après lui suggérera que la perception et la mémoire des expériences se structurent dans le langage lui-même, seul outil de connaissance du monde. Depuis les dix dernières années, plusieurs découvertes dans le

domaine de la neurologie viennent corroborer cette idée que la mémoire existe sous forme de traces. Les travaux autour de la *plasticité neuronale* montrent notamment l'existence de la mémoire sous la forme de connections, relations ténues ancrées dans notre réseau neuronal global (Magistretti & Ansermet, 2012).

Etapes constituantes de la mémoire	t=1	t=2	
Freud	expérience	perception	Traces mn
Lacan	expérience(signifié)	perception(signifié)	Signifiant (traces
Neurosciences	expérience	Perception	Traces synaptiq

fig. Convergence entre la trace psychique et synaptique (Magistretti & Ansermet, 2012)

Ces disciplines s'intéressent majoritairement à l'étude de la constitution d'une mémoire individuelle, ne nous livrant que peu de clés pour comprendre les éléments qui font qu'une mémoire devient commune. Le paléontologue Leroi Gourhan propose dans son livre *L'homme et la Matière* de considérer que les humains possèdent trois formes de mémoire : une mémoire individuelle sensible, stockée dans les organes du corps ; une mémoire héritée génétiquement, stockée dans l'ADN ; et une troisième forme de mémoire, transmise de générations en générations : la *technologie*. Dans sa lecture de Leroi-Gourhan, Stiegler (1998b) explique comment *l'objet technologique* porte en lui les traces des expérimentations, réussites et échecs passés, mémoire cumulative de temps et de sociétés passées, à la fois héritée et commune, transmissible par son usage.

Forme de mémoire	Contenu	Stockage
génétique	Particularités héritées des ancêtres	DNA
épigénétique	Mémoire sensible de l'expérience personnelle	Organes, nerfs, cerveau
technologique	Pratiques de la vie quotidienne en société (usages)	Objets technologiques

Fig. Les trois formes de mémoire d'après Leroi-Gourhan et Stiegler

2.2.2 Diffusion de mèmes et structuration d'une mémoire collective

La relation entre mémoire humaine et mémoire technologique est au centre de notre étude. La technologie a depuis toujours été considérée comme une mémoire extérieure. Dans *Phèdre*, Platon raconte l'histoire du roi égyptien Thamous recevant en cadeau du dieu Thot l'écriture, le remède (*pharmakon*) qui devait “*soulager la science et la mémoire*” (Platon, 274e). Le roi Thamous, effrayé par cette nouvelle technologie de la mémoire qu'est l'écriture se voit saisi de l'angoisse d'une perte de cette mémoire. Avec la fin de cette oralité, la disparition de la méthode active des antiques théâtres de la mémoire au profit d'une mémoire technologique inerte et extérieure à soi pourrait-t-elle sceller l'avènement d'une nouvelle bêtise ? Aujourd'hui, l'importance grandissante des bases de données et de connaissances soulèvent encore une fois les mêmes questions, toujours irrésolues. Nicolas Carr constate notamment que “*l'Internet nous rend stupide*” et que son usage répété entraîne une baisse drastique de nos facultés de concentration (Carr, 2010). A l'ère du Big Data et de l'expansion sans fin de notre mémoire numérique, la constitution de nos bases

de données interroge notre construction d'une mémoire collective. Les mèmes Internet, d'abord gravés dans les disques durs des serveurs, viennent être actualisés par ceux qui les partagent, les commentent, jouent avec et se les approprient. L'exemple de l'Internet chinois nous montre la volatilité de cette mémoire numérique, artefact historiographique d'une culture soumise au bon-vouloir des administrateurs du réseau. Le Manifeste de l'Archiviste publié par Yuk Hui (2014) s'ouvre sur l'angoissante interrogation deleuzienne :

Un nouvel archiviste est nommé dans la ville. Mais est-il à proprement parler nommé ? N'est-ce pas sur ses propres instructions qu'il agit ? (?) .

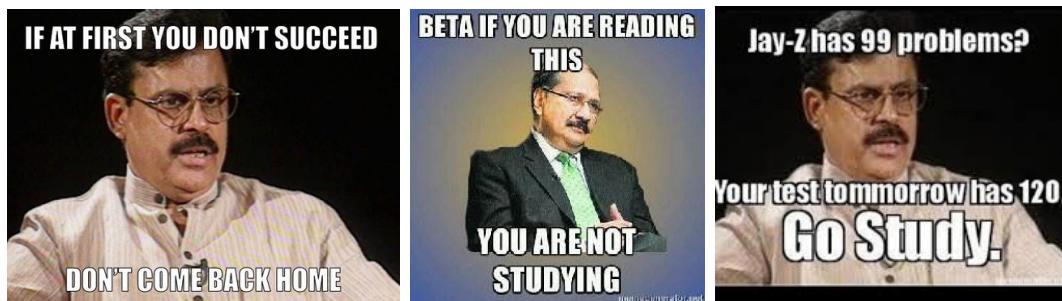
L'existence et l'usage quotidien des bases de données questionnent chaque jour l'assujettissement des symboles de notre mémoire à l'objet technologique, à la fois bâquille, prothèse et maquillage postiche de notre détestable devenir bête. L'étude des mèmes Internet nous offre ici une fenêtre pour porter notre regard sur ces objets dont le corps scellée dans les profondeurs glacées des *data centers* nous parvient en dansant, d'abord sur nos écrans puis dans un coin de notre tête. Avec l'usage répété des technologies et de l'écriture numérique, les frontières entre milieu numérique et mémoire collective s'estompent pour laisser entrevoir un enchevêtrement de silicium, d'idées et de chair, constitutifs de notre savoir moderne.

Poursuivant l'idée d'une archéologie du présent introduite par Foucault, il s'agit donc de documenter les processus par lesquels ces obscurs habitants des bases de données viennent laisser leurs traces pour constituer des bribes de nos mémoires collectives. Maurice Halbwachs dans son vaste travail aborde les façons dont l'histoire structure l'être-ensemble des groupes humains. En disant que "*l'histoire de notre vie fait partie de l'histoire en général*"(Halbwachs, 1947), il identifie une mémoire autobiographique (personnelle) et une mémoire historique (sociale). Les inquiétudes et considérations autour de la "vie privée" sur Internet illustrent les liens intimes entre ces deux mémoires aux frontières devenant aujourd'hui chaque jour plus poreuses. L'acte singulier et autobiographique devient sous l'effet du réseau un fait social, disponible à tout moment dans "l'historique" qui se déroule sous le curseur. L'oubli devient alors un commerce très prisé permettant de garantir la limite entre la mémoire autobiographique de la fin de soirée de samedi dernier et la mémoire socialement acceptable du CV du chercheur d'emploi. À l'inverse, les photos du dernier voyage en Papouasie ou la pose avec une star de la télé témoignent fièrement d'un lien mémoriel entre autobiographie et histoire commune. Les *mèmes* se propagent ainsi d'individus en groupes pour former peu à peu des éléments de mémoire commune. Objets, chansons, histoires, légendes, icônes... leur diffusion autour de la toile se fait par différents procédés tenant autant de la copie que de l'appropriation.

Le site de questions/réponses *Quora.com* offre un regard intéressant sur cette question puisque nous y trouvons une question : « *What are some quintessential Indian memes ?* ». Les utilisateurs répondent donc en ajoutant des exemples de même qui semblent appartenir dans leurs esprits à la "quintessence des mèmes indiens".



Graphique 2.2.1 – Exemples de réponses à la question "What are some quintessential Indian memes?" D'après <http://www.quora.com/India/What-are-some-quintessential-Indian-memes>, Consulté le 12/08/2013 à 0 :41 GMT+8



Graphique 2.2.2 – Exemples de mèmes du père indien autoritaire

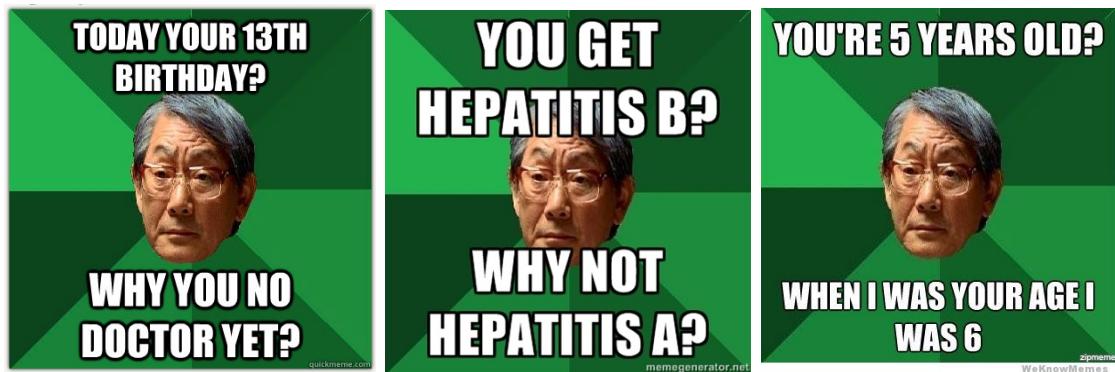
Avec plusieurs centaines d'images postées par des utilisateurs majoritairement indiens⁷, nous pouvons constater plusieurs choses :

- **Langue** : à part trois réponses, la totalité des réponses sont en anglais. Cela peut s'expliquer par le fait que l'anglais est une langue de communication majoritaire en Inde, et également par le fait que le site Quora.com n'accepte habituellement que des réponses en anglais.
- **Forme** : à l'exception de quatre réponses, les mèmes revêtent tous la forme « classique » : photo retouchée et légendé par un texte en anglais aux lettres blanches sur fond ou détour noir.
- **Humour** : la plupart des réponses sont de nature comique.
- **Réurrence** : certaines images sont très récurrentes et si les légendes diffèrent, le sens reste le même.
- **Thématiques diverses** : de nombreux messages traitent de la vie de famille (parents, mariage), de loisirs (cricket), de la vie quotidienne (achats, école , etc.) et un peu de politique.

L'image la plus citée représente la figure d'un père autoritaire :

Étonnement, il existe un mème plutôt similaire avec un père asiatique cette fois :

La définition des mèmes faites par Blackmore ne correspond pas nécessairement à ce que nous observons ici. L'unité formelle de l'ensemble de ces mèmes (image avec des caractères blancs cerclés de noirs) propose une définition bien plus restreinte. Néanmoins, nous pouvons comprendre qu'il s'agit bien de la manifestation d'une culture particulière.



Graphique 2.2.3 – Exemples du même "High Expectations Asian Father" D'après "What are the funniest High Expectations Asian Father meme images ?" sur Quora.com <http://www.quora.com/Memes/What-are-the-funniest-High-Expectations-Asian-Father-meme-images>, Consulté le 12/08/2013 à 10 :56

L'image d'une vache assise devant un sigle "Fuck The Police" serait en effet un absolu non-sens sans la référence à l'Inde où les vaches sont sacrées et jouissent de droits particuliers que même la police ne peut entraver. Ainsi, il existe bel et bien des pré-requis pour comprendre ou actualiser un même. Les plus évidents sont :

- **L'accès et l'usage de la bonne technologie** : Il est nécessaire pour un utilisateur de posséder et de savoir utiliser la technologie par laquelle le même est diffusé.
- **La langue** : Le message possède une légende donc l'utilisateur doit pouvoir la lire et la comprendre.
- **Les implicites** : L'utilisateur doit posséder le socle de références communes et d'implicites qui sont impératifs pour pouvoir comprendre le message.

Nous observons ici de très larges et vagues groupes (« *indiens* », « *asiatiques* »...) sans pouvoir vraiment comprendre dans le détail ce qui peut réellement constituer des éléments communs. Les travaux sur la formation des "communautés" en ligne ont montré comment la circulation des objets digitaux peut posséder une fonction de catharsis pour des groupes plus réduits (Steyer *et al.*, 2006). Néanmoins, on voit bien qu'ici l'appartenance pré-existe puisque le même nécessite de nombreux pré-requis pour le comprendre. Le langage et son expressivité par l'humour sont notamment des contraintes incompressibles pour l'actualisation de ce même par un individu. Néanmoins nous pouvons voir dans ce cas particulier que le même participe à l'affirmation de l'existence du groupe, avec la figure redondante de caractéristiques communes du père, affirmant par la dérision une forme de paternité commune aux membres de ce groupe. D'autres mèmes ne nécessitent pas tant de références, ce qui contribue largement à leurs diffusions. La vidéo du clip musical *Gangnam Style* du chanteur Psy a notamment atteint des records inégalés en termes de diffusion⁷ grâce à une très vaste campagne télévisuelle et sur internet. La diminution des implicites et la standardisation de l'écriture utilisée a sans doute contribué à la diffusion,

7. Première vidéo à avoir officiellement dépassée le milliard de vues sur Youtube. 1,733,769,243 vues, consulté le 13/08/2013 à 09 :35 GMT+8 <http://www.youtube.com/watch?v=9bZkp7q19f0>

avec un langage du corps quasi universel, le pas de danse. Formellement, il s'agit d'un vidéo clip très classique dont la structure et le montage sont largement familiers au public. Les attributs des personnages du clip sont également de grands classiques du vidéo clip commercial : voitures, belles filles et bijoux en or. La présence d'un quartier spécifique de Séoul en Corée du Sud agit ici comme un attribut du contenu mais aucune des références ne nécessite de préalable linguistique particulier. De plus, les mystères de la musique et de la scénographie agissent bien évidemment au delà de toute analyse formelle pour faire de ce hit un des mèmes Internet les plus connus dans le monde. Les méméticiens disposent classiquement de deux procédés pour analyser la diffusion des mèmes :

— **La contamination**

Le même se déplace à la manière d'un virus, en contaminant les sujets les plus susceptibles de l'être lors d'une phase d'exposition. L'exemple le plus classique pour ce modèle est la diffusion des croyances religieuses qui agirait par contagion (Dennett, 2006)

— **La réPLICATION**

Les activités culturelles humaines procèdent de l'imitation, notamment au travers de phases cruciales d'apprentissage. Ainsi, les mèmes existent et se diffusent dans toutes activités nécessitant une imitation : “*If we define memes as transmitted by imitation then whatever is passed on by this copying process is a meme.*” (Blackmore, 2006).

Le modèle épidémique de diffusion vient appuyer la vision éthologique du mème. Adaptée de la virologie, les « sujets à risque » seraient plus à même d'être « contaminés » par une « exposition » suffisamment longue à tel ou tel mème (Wang & Wood, 2011). Blackmore définit néanmoins trois phases indispensables pour reconnaître un mème comme tel :

Memes fulfill the role of replicator because they exhibit all three of the necessary conditions ; that is, *heredity* (the form and details of the behavior are copied), *variation* (they are copied with errors, embellishments or other variations), and *selection* (only some behaviors are successfully copied). (ibid.)

Les trois aspects sont indissociables et forment selon Blackmore un “*véritable processus évolutioniste*” (ibid.). La définition quasi tautologique du mème (« *Whatever is passed on*») montre bien comment le mème en tant que concept est considéré chez Blackmore non pas comme un phénomène mais plutôt comme un objet en-soi. La définition du mème en tant qu'objet autonome se heurte dès l'abord au risque de devenir un pur artefact de l'observation, n'existant que dans l'esprit de l'observant. Alors que Dawkins met en garde non sans humour que son livre *The Selfish Gene* “*devrait être lu presque comme s'il était de la science-fiction*” (?), la faiblesse des modèles de diffusion des mèmes vu comme un réplicateur ou comme un virus ne recouvre que partiellement la réalité observée empiriquement pour les mèmes Internet. Ainsi, l'appartenance des individus à tel ou tel groupe pré-existe au mème. Dans son appropriation se joue davantage l'expression

d'un sentiment d'appartenance qu'une contamination qui nierait les termes de sa volonté individuelle pour y substituer l'individu comme sujet du même.

2.3 Textualité des mèmes et formes d'énonciations numériques

2.3.1 Le même comme figure rhétorique de l'écriture intertextuelle

Nous tâcherons donc ici de définir le même comme une série d'actes *d'énonciation* qui contribuent à l'existence et la reconnaissance mutuelles d'individus comme groupe, à l'opposé d'un élément sémantique pure qui serait constitutif d'une hypothétique culture commune. Les partages, commentaires, réappropriations puis transformations des *mèmes Internet* nous serviront de supports pour comprendre les *actes d'énonciation*. Ces courts messages faits de texte, image ou vidéo sont en quelque sorte les voix, répétitions, annonces et redites d'une foule d'individus et de groupes qui habitent la Toile. Au-delà de l'idée d'un "objet" numérique qui réifierait les actions en une substance figée, nous nommons *énonciation* le moment d'existence observable où se manifeste un même. Comme pour les émoticones, les pratiques actuelles de l'écriture en ligne des mèmes Internet sont à envisager comme des formes renouvelées d'oralité. Email, chat ou réseaux sociaux, ces discussions font partie d'une "*oralité seconde*" (?) constituante des technologies de l'écrit à l'ère du numérique. Contrairement à l'oralité première des illettrés, cette oralité seconde est structurée par l'usage des technologies et notamment les structures formelles de l'écriture pour le média :

Telephone, radio, television and the various kind of sound tape, electronic technology has brought us into the age of 'secondary orality'. This new orality has striking resemblance to the old in its participatory mystique, its fostering of a communal sense, its concentration on the present moment, and even its use of formulas. But it is essentially a more deliberate and self-conscious orality, based permanently on the use of writing and print, which are essential for the manufacture and operation of the equipment and for its use as well. (?)

L'hypothèse de Ong est ici que l'oralité des écritures numériques est une forme manufacturée de l'expression orale, à laquelle préside la production (industrielle) des technologies. Les services de réseaux sociaux en ligne confirment l'hypothèse première de Ong puisque *Sina Weibo* ou *Twitter* contraint l'écriture à une longueur maximum de 140 caractères. Néanmoins, l'oralité du même Internet et plus généralement des écritures numériques ne procède pas seulement de cette écriture conditionnée mais également de la mise en relation des textes : l'intertextualité. Là où pour Ong l'industrie médiatique vient contraindre l'écriture pour la réifier en produit industriel simulant l'oral, l'intertextualité

vient subvertir ces limites imposées en renvoyant le lecteur à la page suivante.

En effet, le langage des nouveaux médias ne se contente pas de proposer des nouvelles opérations d'interactions et de navigations mais s'inscrit également dans de nouveaux modes de lecture et de narration (?). Une des grandes difficultés pour l'analyse textuelle et narrative dans le cadre des nouveaux médias est de comprendre où commence et où se termine la narration. La structure éminemment relationnelle du discours narratif en ligne et son intertextualité en font un objet mal défini, que ses auteurs n'ont pas signé d'un point final. L'étude des discursivités hypertextuelles s'apparenterait donc davantage à l'étude des formes dans les contes et légendes qu'aux études herméneutiques classiques sur des textes finis (Clément, 1995). Comme le note Lévi-Strauss dans son travail sur Propp, même si les contes possèdent souvent une structure similaire, leur étude en tant qu'élément culturel n'est révélatrice que dans un contexte précis et incarné. Il est inutile de vouloir extraire une supposée intention culturelle du texte car on se doit de le comprendre lors de son énonciation, sur la place d'un marché au Maroc avec les conteurs de Ben Jelloun ou au pied du lit d'un jeune Européen avec les histoires compilées par les frères Grimm. Héritant à la fois des pratiques culturelles et folkloriques anciennes tout en se renouvelant sous les nouvelles contraintes de la technologie (Barber, 2008), le même est lui aussi à comprendre dans son contexte d'énonciation. De récents travaux travaillent à comparer les modes de diffusion des mèmes avec ceux des traditions folkloriques (?). En considérant les mèmes Internet comme un "*folklore numérique*", on comprend mieux la nature presque auto-référentielle de la relation entre le même et la culture qui le voit naître. La transmission d'éléments particuliers dans la discursivité du même en fait une figure rhétorique d'énonciation de sa propre origine. Habituellement définie comme « *une forme typique de relation non linguistique entre des éléments discursifs.* »⁸, la figure rhétorique se produit dans l'intertextualité des redites et commentaires qui réaffirme dans le même l'existence d'un groupe qui le constitue.

Le large pouvoir fédérateur de certains mèmes fascine. Diffusés très largement, on se plaît à imaginer comment ils peuvent réunir en leur sein des groupes et individus distants, auparavant inconnus et étrangers. L'importance croissante d'Internet dans l'émergence de groupes d'intérêts et d'activités voit les études sur la formation de communautés en ligne fleurir. Néanmoins, il nous semble important de se questionner sur la nature des relations créée lors des discussions partagées en ligne. Est-ce bien là le fait d'une réelle rencontre comme le croient les plus enthousiastes ? Ou au contraire est-ce le produit d'une machine médiatique et décérémoniale qui produit du lien sans engendrer de rencontres comme le pensent les plus pessimistes ? Cette vaste question est sans doute un des enjeux centraux des questionnements autour de l'Internet, notamment pour le management et les sciences de gestion. Si l'énonciation est une pratique structurante pour un individu ("je suis"), elle peut l'être également pour un groupe ("nous sommes"). Dans la perspective empirique où nous nous situons, il nous faut tout d'abord interroger les pratiques du

8. *Les figures de rhétorique*, Laurent Jenny, Université de Genève, 2003

langage pour mieux comprendre comment les discursivités d'un même peuvent agir sur les groupes. Le même en tant considéré comme un acte d'énonciation se caractérise non seulement par sa manifestation langagière, mais également par l'intention qu'il contient. Dans le cadre des réseaux sociaux, nous ne disposons que de très peu d'éléments sur l'intention car les données disponibles sur les réseaux sociaux sont par définition le résultat d'actions passées (écriture, clics, etc.). Nous devons donc développer un modèle à la fois conceptuel et pratique pour nous permettre d'étudier ces phénomènes d'énonciation dans le cas particulier des mèmes Internet. Wittgenstein dans ses *Recherches Philosophiques* défend une analyse pragmatique du langage en écrivant : “*Don't ask for the meaning, ask for the use.*” (?). Ainsi, il s'agit de comprendre les jeux langagiers non pas comme un champ linguistique mais comme un ensemble d'actes qui font sens en contexte et possèdent une intention.

“Expressions have meanings even when they are not being used, but it is only in using expressions that a person means something.” ([Bach, 1994](#))

Austin dans ses lectures sur William James met au centre du langage sa dimension pragmatique et propose de comprendre comment on « *fait des choses avec les mots* » ([Austin, 1975](#)). Ses leçons présentent une manière nouvelle de catégoriser les différents actes d'énonciation et montrent la grande diversité des éléments non-linguistique présents dans ces actes. Austin introduit le concept de *performativité* pour nommer le processus qui permet de construire par le langage une réalité extérieure au langage. Les mots ne nomment pas seulement les choses mais peuvent également les faire changer, voire les fabriquer. Austin s'intéresse donc aux énoncés selon leurs *sens*, leurs intentions (la *force*) ou leurs *effets*. Le concept de *performativité* a depuis continué son chemin tant en linguistique qu'en sciences sociales, et plus récemment dans des champs aussi divers que le management ou l'étude du discours scientifique et des pratiques légales ([Denis, 2006](#)). Les économistes également ont beaucoup discuté de la performativité des discours économiques sur l'économie réelle (?). Les *cultural studies*, et plus précisément les *gender studies* ont aussi fait un large usage de ce concept pour exprimer l'influence de la matérialité des mots sur les comportements humains ([Butler, 1993](#)).

« Pour qu'ils deviennent de « véritables » performatifs, les faits, les théories ou les formules doivent circuler dans des chaînes de traduction qui consolident l'assemblage des entités qui le composent et leur permet d'acquérir le statut de « *matters of fact* » (...). C'est lorsqu'ils arrivent à durer, c'est-à-dire à s'inscrire dans le monde (par l'intermédiaire d'objets, de textes, de dispositifs techniques complexes) que leur performativité s'accomplit. » ([Denis, 2006](#))

Les actes d'énonciation répétés modèlent donc le corps des personnes et de la société, y *laissant* durablement leurs marques. Les énoncés actualisent les discours des groupes sociaux sur eux-mêmes ([Butler, 1993](#)) et ce caractère performatif est constitutif de l'énonciation. Les énoncés collectifs que sont les mèmes Internet possèdent également cette di-

mension performative qui actualise le discours de certains groupes en réalité tangibles. La circulation et la structuration de ces mèmes vient structurer le milieu numérique et peut ainsi influer sur la définition de groupes sociaux et des rapports qu'ils entretiennent. Une image ou un mot d'un même ne peut néanmoins devenir commune sans le préalable permettant de dénouer l'implicite de l'énoncé, faisant de l'énonciation un accroissement de proximité dans l'expérience dans le moment. Blackmore évoque en termes évolutionnistes le "potentiel transformatif" du même, avec ce qu'elle nomme la *variation* puis la *sélection*. Ici ce sont les actes d'*énonciation* qui forment une praxis du même. Pour nous, le même n'est pas un méta-symbole en évolution mais un réseau de praxis culturelles constitué de multiples d'actes d'énonciation. Ainsi, le même ne peut être simplement « copié » mais a besoin d'être acté pour exister. Dans la définition du même de Blackmore comme dans le cas des réseaux sociaux, nous observons que l'énonciation d'un même procède d'une variation parfois nulle, parfois minimale, parfois importante de sa forme d'origine. Cette déformation due à l'énonciation est le propre de la fonction d'apprentissage, notamment langagier. Le caractère performatif du même devient visible dans l'usage approprié de la variation qui en est fait.

Le succès de l'intention de l'énoncé est visible dans son imitation, avec comme garantie l'erreur ou la variation. Platon dans *La République* puis par la suite Aristote dans sa *Poétique*, s'interroge sur la notion de *mimesis* définie comme les formes d'imitation qui permettent soit de reproduire, soit de styliser la nature. Pour Aristote, le but singulier de la *mimesis* est de mettre à jour la dimension empathique cachée de la nature, de la styliser pour y révéler le continuum de l'expérience propre à tous les êtres. Plus l'artiste s'approche de la nature en se dirigeant vers une imitation « véritable », plus il s'éloigne de la réalité de la nature. L'importance d'une approche rhétorique comme par exemple la stylisation devient le véritable moyen d'accès à la signification profonde des choses. Freud réutilisera le concept de *mimesis* pour décrire l'énonciation du sens d'un évènement traumatisque passé dans la vie d'un individu au travers d'activités créatives (art, parole, rêves etc.). Dans la continuité d'Aristote, Freud comprend le rêve comme une *mimesis* du passé et du réel, révélant au sujet un objet symbolique enfoui. Peut-être est-il possible de considérer le même Internet comme une *mimesis* des groupes sociaux et médiatiques qui le produisent. Les processus de symbolisation et de stylisation jouent en effet un rôle déterminant dans sa diffusion. Acte d'énonciation, le même agit alors comme une *mimesis* des activités et états d'âme de groupes d'individus qui l'énoncent. S'il est possible de le revivre plusieurs fois, il prendra peut-être à chaque fois un sens différents. Comme le rêve freudien qui est une manifestation biologique de la mémoire inconsciente se produisant durant le sommeil, le même se manifeste sous des formes visibles symboliques incarnées, *mimesis* particulière du groupe des individus qui l'énoncent.

2.3.2 Typologie des mèmes Internet

En nous appuyant sur la littérature concernant les figures rhétoriques, nous pourrions chercher à définir plus précisément une typologie des mèmes fondée sur les formes du discours. En grec ancien, le topoï défini à la fois un lieu ou un endroit mais également un ensemble de formes rhétoriques utilisant des motifs particuliers lors de l'argumentaire afin de persuader lors de joutes oratoires. En littérature comme en mathématiques, le terme de *topos* définit également un ensemble de catégories ouvertes mais connues, “*un protocole de description des univers possibles*” (Badiou, 2006). Le même peut donc être compris comme un *lieu commun*, idée “reçue” utilisant des situations ou des images communes et stéréotypées pour opérer une transformation sémantique en jouant sur la répétition d’éléments (les sèmes du discours). En rhétorique, l’usage du *topos* a pour objectif de contribuer à la persuasion de l’auditeur par la mobilisation subtile d’éléments de culture commune. Tout l’art du rhéteur consiste à trouver un moyen subtile d’actualiser un lieu commun en une situation unique propre au contexte pour convaincre. Dans le discours, il prend bien souvent la forme de l’anecdote que la rumeur se charge de diffuser, sa diffusion étant d’autant plus efficace qu’il possède un caractère amusant ou railleur (Flaubert, 1997) Concevoir le même comme une forme de pratique rhétorique semblable au lieu commun nous permet de résituer le phénomène des mèmes Internet dans la continuité historique des pratiques de l’écriture et de l’énonciation rhétorique, ainsi que de ses modèles d’analyse socio-textuelle et littéraire (Plantin, 1993). Ainsi, nous pouvons aborder la lecture des mèmes Internet sous le jour de leur existence aussi bien formelle (textuelle) que rhétorique (comme actes d’énonciations et de persuasion). Les mèmes tout comme les lieux communs se donnent à voir d’abord sous la forme de paradoxes, qui deviendront eux-mêmes des lieux communs. L’originalité d’un lieu commun en devenir se définit dans une tension constante entre imitation et nouveauté, subversion et actualisation de formes canoniques : “*dialogue de l’horizon d’attente et de l’écart esthétique, c’est-à-dire le jeu du classique et du moderne, la tension entre le même et l’autre qui existe, dans tout texte et dans toute lecture, entre le plaisir et la jouissance, pour reprendre les mots de Barthes.* (Compagnon, 1997)

L’approche des mèmes renouvelée par la nature intertextuelle du langage multimédia montre des caractéristiques d’après une centaine de mèmes parmi les plus largement diffusés d’Internet :

1. **Humour** : Le même doit posséder une dimension comique et accrocheuse
2. **Intertextualité** : Le même met en jeu un ou des renvois à d’autres éléments culturels ou textuels, souvent implicites.
3. **Juxtaposition atypique** : Les éléments visuels ou sémantiques mis en jeu dans le même ne possèdent pas de corrélations apparentes et c’est la mise en relation de plusieurs objets improbables qui en fait un objet intéressant.

L’étude en question nous offre un début de critères mais porte seulement sur un type précis

Objectifs du mème	Exemples célèbres et références
Absurdiste, humour	LOLCats, Tumblr (Bauckhage, 2011)
Actualité, satire, commentaire social	CaoNiMa (xiao Mina, 2012), Cute Cat Theory (A)
Publicité, marketing viral	Gangnam Style (Bolsover et al., 2013), Memetic Market
Marketing politique, soutien, pétition	Obama Ohio Campaign (Walker, 2012), Pétitions en ligne (A)
Fan clubs, adoration	Fan-fiction , Machinima, etc.
Hoax, spam	Email spam, « Nigerian scam »

Graphique 2.3.1 – Les différents type de mèmes Internet observables - tableau réalisé d'après la littérature indiquée

de mèmes à caractère plutôt comique, ignorant les discussions plus sérieuses, d'ordre politique notamment. La forme particulière de juxtaposition atypique observée par les auteurs serait en rhétorique une “*métaphore in praesentia*” décrite comme une « *figure de rapprochement analogique entre deux représentations co-présentes* » (?). Nous nous trouvons donc en présence d'une forme particulière de mème seulement. Si l'humour sous toutes ces formes (blague, sarcasme, ironie, etc.) est un élément très répandu qui favorise la circulation des contenus en ligne, il nous semble néanmoins un peu réducteur de se limiter à cette définition. De nombreux mèmes Internet existent non pas grâce à l'humour mais grâce au *pathos* qu'ils dégagent. Le mème *Kony2012*, un des plus diffusés de l'histoire d'Internet, présentait le militaire Ugandais Joseph Kony dans une vidéo faite d'images de guerre et d'enfants en pleurs⁹. Ainsi, on peut dire que les mèmes Internet utilisent les formes classiques de la rhétorique adapté au langage médiatique moderne teinté d'humour ou de sensationnalisme. En se saisissant du monde social et politique, les mèmes Internet produisent d'intéressants discours sur les faits qu'ils relatent - et sur la technologie qui les produit. Alors que les révolutions du Printemps Arabe avaient notamment soulevé l'espoir d'un monde meilleur par l'usage des réseaux sociaux pour instaurer la démocratie (?), la communication instantanée via ces mêmes réseaux sociaux devenait quelques semaines plus tard une des causes majeures de la flambée de violence durant une série d'émeutes à Londres (Casilli & Tubaro, 2011). Les différentes intentions des discursivités et actes d'énonciation à l'œuvre dans un mème peuvent donc nous aider à dresser une typologie des mèmes. Une typologie des mèmes ne peut être considérée comme exclusive et les glissements sémantiques et symboliques qui peuvent s'opérer nous obligent à considérer une catégorisation non-exclusive. En nous appuyant sur différents exemples et sur la littérature, nous dressons une première ébauche de typologie des mèmes Internet qui sera ensuite approfondie par l'étude empirique des données le cœur de notre étude. Pour les exemples, nous essaierons de donner les références et implicites indispensables permettant de comprendre le mème dans son intertextualité.

Cette typologie nous permet de poser un premier regard sur des catégories non-exclusives constituant le paysage des mèmes Internet, inspirés d'exemples et de la lit-

9. *KONY2012 : See How Invisible Networks Helped a Campaign Capture the World's Attention, Social Flow*, <http://alturl.com/zniry> consulté le 28 Février 2014 GMT+1

tériture existante. Nous pouvons d'ores et déjà noter que l'ensemble de ces différentes pratiques pré-existent à l'Internet et sont souvent la reconduction sous une forme numérique de discursivités pré-existantes (penser à la publicité, la propagande politique ou même les rumeurs). La forme et l'échelle de la diffusion sont néanmoins des différences importantes ainsi bien sûr que les situations d'énonciation (dans une file d'attente ou devant son ordinateur).

Voici donc cette typologie plus en détails et illustrée d'exemples.

Absurdiste, humour

De nombreux mèmes Internet parmi ceux que nous avons présentés dans la première partie de ce chapitre viennent se classer dans cette catégorie. Connaissant parfois des succès planétaire, un des exemples les plus caractéristiques est celui des “LOLcats”, ces vidéos ou images de chats illustrés de citations et autres accessoires comiques. Omniprésent sur la Toile, l'immense répertoire de photos et de vidéos comiques de chats dans différentes situations (jouant du piano, sautant dans des boîtes en cartons, etc.) constitue un des plus visionnés de l'Internet (?).

Actualité, satire, commentaire social

En regroupant ces trois catégories, nous souhaitons considérer la pratique ancienne des discussions politiques ou polémiques autour de faits divers ou d'actualités que l'Internet vient aujourd'hui renouveler dans la forme. Nous avons vu précédemment avec les “crabes de rivières” (voir 1.2.2.) comment la satire politique en Chine se manifestait bien souvent sous la forme de mèmes Internet. Par les voies de l'Internet, de nombreux faits divers sont devenus en Chine des affaires d'État, influençant le jeu politique et offrant un nouveau canal de diffusion pour les pratiques subversives. Connue pour son œuvre artistique, l'artiste chinois Ai Weiwei est une des figures célèbres de l'Internet chinois. Utilisant abondamment blog et microblog, son travail artistique depuis plusieurs années a pris la tournure d'un bras de fer médiatique avec le gouvernement sur les grandes questions sociétales de la Chine d'aujourd'hui. Son père, poète célèbre du régime communiste sous Mao fut déchu durant la révolution Culturelle, transmettant à son fils à la fois un goût pour les arts et une violente amertume pour le pouvoir en place à Pékin. Critiquant ouvertement la corruption des officiels et de la police, Ai Weiwei a vu régulièrement ses comptes Internet de blog et microblog fermés, notamment par le service Sina Weibo. Sur Twitter néanmoins, il jouit d'une grande popularité avec près de 200.000 “*followers*” pour son compte officiel @aiww. Au fil de son blog relatant ses nombreux déboires avec le gouvernement, on croise bien souvent une figure mythologique née de la toile chinoise, le *caonima*.

Luttant contre les crabes de rivière, le *caonima* symbolise la lutte contre la censure pour un Internet libre. Apparu pour la première fois dans une vidéo virale¹⁰,

10. *Song of the Grass-Mud Horse (Cao Ni Ma)*, Youtube <https://www.youtube.com/watch?v=wKx1aenJK08>, consulté le 28 Février 22 :08



(a) Une image de la vidéo originale représentant le *caonima*. **(b)** Un caractère chinois composé spécialement par les internautes pour signifier cet animal mythique

Graphique 2.3.2 – Caonima : un animal mythique du web chinois, d'après le *Grass-Mud Horse Lexicon Classics* publié par *China Digital Times* (2013)

le *caonima* est un animal de la famille des camélidés (l'alpaga) courant fièrement dans les prés sur une petite musique de dessin animé aux paroles entièrement réécrites pour l'occasion. Littéralement “cheval d’herbe et de boue”, le mot “*caonima*” [FF08 ?][8349 ?][6CE5 ?][9A6C ?][FF09 ?] cache en fait un double sens puisque son homophone ([64CD ?][4F60 ?][5988 ?]) est une grossière interjection à l’intention des génitrices des censeurs de Pékin. Devenu aujourd’hui une véritable icône anti-censure, il n’est pas rare de le croiser sur un tee-shirt ou accroché à un sac dans de nombreux endroits improbables en Chine. Comme le note très justement An Xiao Mina : “les mèmes sont les graffitis du web censuré” (xiao Mina, 2012). Malgré le zèle des pouvoirs publics à les effacer le plus rapidement possible, ils témoignent de l’existence de réalités occultées qui se manifestent souvent de manière dérisoire, grotesque et improbable, mais sont énoncées malgré tout.

Publicité et marketing viral

La popularité des services de réseaux sociaux a entraîné de nombreuses marques à se concentrer sur ce nouveau média pour mener leurs campagnes de publicité et de promotion. Les utilisateurs chinois sont très fortement engagés dans la création de nouveaux contenus avec 76% des utilisateurs chinois créant davantage de contenus qu’ils n’en lisent, contre seulement 20% en France (?). Ainsi, de nombreuses marques cherchent à concevoir des interactions en ligne utilisant les mèmes comme véhicule de messages commerciaux afin de mettre à contribution les internautes dans la diffusion de leur marque. Une des stratégies marketing les plus répandues consiste à créer un *hashtag* amusant afin d’inciter les internautes à s’en saisir, le partager et créer de nouveaux contenus. La marque de préservatifs *Durex* a no-

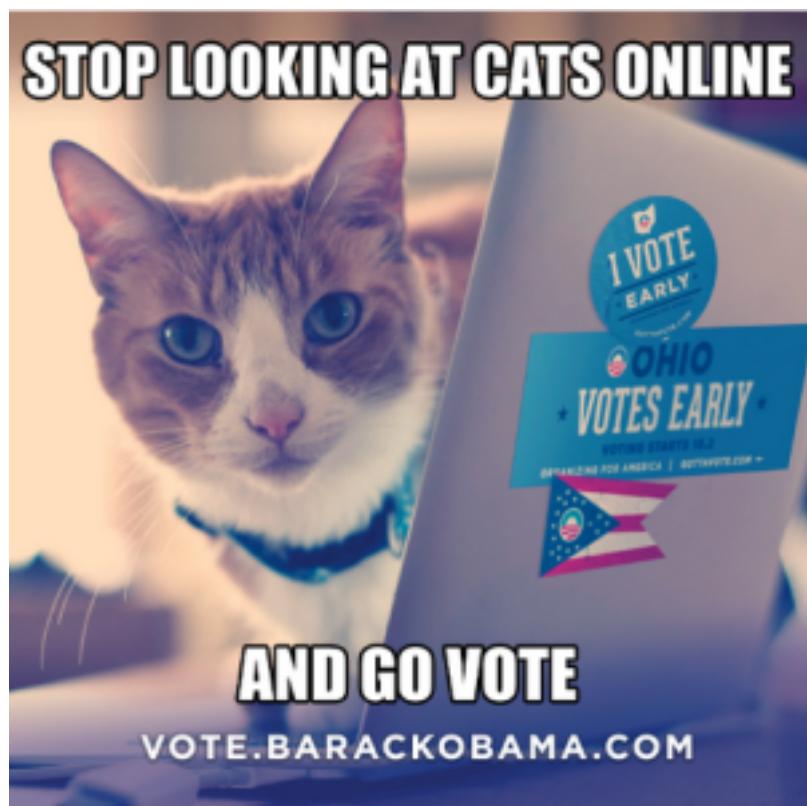
tamment connu un large succès avec le hashtag *#BienEtreNocturne* (en Chinois *#*[591C ?]*[798F ?]**[5229 ?]*) sur Sina Weibo. Drôle et un peu osé, cette campagne a généré un fort engagement des internautes, contribuant largement à la propagation du profil utilisateur et du nom de la marque (?). En récupérant les données des utilisateurs impliqués dans la diffusion, la marque peut également procéder à un travail statistique d'analyse pour mieux connaître les utilisateurs intéressés par ses produits. Ce type d'information s'avère précieuse dans un marché chinois en mouvement où il existe un fort besoin pour des études de marché précises auprès de segments de populations actifs en ligne (adolescents, jeunes, classe moyenne émergente...) (Bergstrom, 2012). Malgré le paysage morcelé des services de réseaux sociaux chinois, les utilisateurs chinois semblent davantage enclins à suivre et interagir avec les marques que leurs homologues américains notamment¹¹.

Marketing politique, soutien, pétitions

L'industrie n'est pas seule à s'être emparée des réseaux sociaux et les grandes entités politiques ont également saisi à bras le corps ce nouveau média. Les plus brillants exemples sont sans doute les campagnes menées par Barack Obama pour la présidence des Etats-Unis. L'importance capitale des réseaux sociaux a placé la stratégie de diffusion virale en haut de la pile des préoccupations pour l'organisation de ces deux campagnes électorales (Miller, 2008) avec notamment l'utilisation de nombreux mèmes pour fédérer les votants. Lors de l'étape de campagne auprès des électeurs de l'Ohio, un état décisif de la course présidentielle, l'équipe d'Obama a notamment fait usage d'un des fameux LOCats pur mobiliser son électorat.

En Chine également, les réseaux sociaux sont très utilisés pour la diffusion d'idées lors de campagnes politiques. Les groupes nationalistes soutenus par le gouvernement se saisissent régulièrement de l'actualité pour rebondir et rassembler les foules (?). Là encore, les mèmes jouent un rôle important dans l'appropriation des discussions au travers de la ré-énonciation de sujets controversés en des termes différents. Les tensions grandissantes entre les habitants du territoire de Hong Kong et ceux venus de la Chine intérieure ont été notamment le sujet d'une discussion intéressante par mèmes interposés. Les visites à Hong Kong sont très régulées pour les citoyens venus de Chine intérieure mais le flux massif de touristes n'a néanmoins cessé de croître depuis plusieurs années avec l'augmentation des quotas et l'accès aux congés dans les grandes villes de la RPC. De nombreux chinois se rendent donc à Hong Kong pour voyager mais également profiter de la détaxe des produits de consommation et des services publics de bien meilleure qualité. La qualité des hôpitaux de la ville et l'application du droit du sol dans la loi hongkongaise amènent de nombreuses jeunes mères venues de Chine à traverser la frontière pour

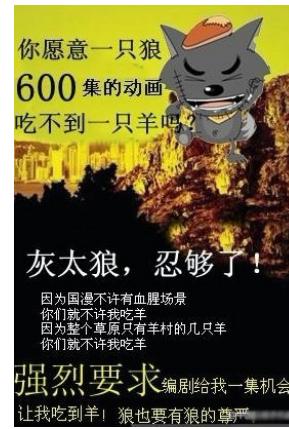
11. D'après Ken Hong, Sina Weibo General Marketing Manager <http://adage.com/article/global-news/questions-sina-weibo-s-ken-hong-china/239508/>, consulté le 19 Février 2014 à 12 :45



Graphique 2.3.3 – LOLCat utilisé durant la campagne d’Obama en Ohio - d’après *Obama Campaign Deploys Cat Meme to Get Out the Vote in Ohio* sur Politicker, consulté le 28 février 2008 à 11h41 GMT+1



(a) Le tract original



(b) Exemples de tracts réalisés par les internautes en réponse à la campagne



**Graphique 2.3.4 – Détournement de tracts hongkongais anti-chinois,d'après *The Civic Beat*,
<http://reader.thecivicbeat.com/2012/03/locusts-and-pandas-and-bears-??-o-mai/> consulté le 1er Mars 2014 à 22h58**

venir accoucher à HK. Cette pratique très controversée donne à l'enfant le passeport hongkongais et se monnaie à prix d'or, rendant l'accès aux hopitaux de plus en plus cher et attisant la colère des habitants de HK. En septembre 2012, une pétition lancée à Hongkong a cherché à recueillir des votes pour interdire l'accès aux hôpitaux aux chinois venus de RPC. Un tract publié pour cette campagne a largement circulé sur le web chinois ; on pouvait y lire : "Habitants de Hong Kong, nous avons assez souffert ! (...) Ne nous laissons pas envahir par la vermine venue de Chine intérieure."

Les internautes chinois choqués mais armés d'un humour toujours très cinglant ont donc entamé une contre-campagne en proposant des versions modifiées et réécrites du dit tract. Réactions épidermiques, propos nationalistes sommaires, mais aussi critique du tourisme de masse, dénonciation de la corruption et de la mauvaise qualité des soins hospitaliers en Chine et nombreuses blagues absurdes, les adaptations et réponses singeant le tract d'origine ont montré un panel d'arguments et de réactions qui a permis de crever l'abcès et d'ouvrir un débat national sur ce problème épique révélateur des griefs actuels des habitants de HK envers ceux de la RPC.

Fan clubs, adoration

Comme tout mass media, les réseaux sociaux possèdent une énorme quantité de contenus consacrés aux stars, à leurs vies, leurs coups durs et leurs derniers films et chansons. Développant des stratégies d'envergure sur les médias chinois, de nombreuses stars internationales ont fait leur arrivée sur Sina Weibo comme notamment Brad Pitt ou Kobe Bryant. Néanmoins, leur influence reste incomparable à celles des stars originaires de Chine, de Taiwan et de Hong Kong ou bien encore de Corée



Graphique 2.3.5 – Un message disant “Les peuples de la Chine et du Japon sont des amis” posté par la star Sola Aoi

du Sud, principal producteur de pop culture en Asie (Martel, 2010). Une des stars les plus plébiscitée dans les réseaux sociaux est pourtant une étrangère puisqu'il s'agit de la japonaise Sola Aoi, ex-actrice de film pornographique devenu une des 10 personnes les plus suivies sur Sina Weibo avec près de 15 millions de followers¹². Personnalité médiatique publique, la star a depuis plusieurs années utilisé les réseaux sociaux pour nouer contact avec le public chinois qui semble bien la connaître, malgré l'interdiction de la pornographie en Chine. Aujourd'hui retraitée du monde du X à 30 ans, Sola Aoi utilise sa popularité pour défendre une paix durable entre la Chine et le Japon. Alors que les tensions politiques exacerbées entre la Chine et le Japon ont mené à des rixes et des maltraitances envers les ressortissants dans les deux pays, elle a notamment contribué à créer un appel à la non-violence sous forme de mème qui a été extrêmement diffusé.

Ainsi, les fan clubs et les stars elles-mêmes jouent un rôle important dans la création et la diffusion des mèmes, occupant une large place dans le paysage médiatique et notamment celui des réseaux sociaux.

Hoax, spam

Un des plus célèbres exemples dans ce domaine sont les emails dits de “fraude 4-1-9” cherchant à extorquer de l'argent au destinataire. Le numéro 419 correspond

12. D'après Sina Weibo <http://www.weibo.com/1739928273/yBJYNt701>, consulté le 1 Mars 2017 à 23h11

I am Stella Amah 19 years of age the only daughter of late Mr Boni Amah whom was killed by the rebels that attacked our country cote d'Ivoire west Africa and took over our town (BOUAKE). I ran to Abidjan the economical capital of cote d'ivoire from were I am contacting you. Before the death of my father he told me that he has a sum of US\$9,000,000(Nine million united states dollars) kept in a private security company here in cote d'ivoire in my name as the next of kin...

Graphique 2.3.6 – Extrait d'un "Nigerian Scam", d'après <http://www.hoax-slayer.com/stella-amah-scam.shtml>, consulté le 2 Mars 2014 à 18h50

au numéro de l'article interdisant la pratique de l'escroquerie en ligne dans le code pénal nigérian. En effet, la forme la plus connue de cette arnaque est celle du "prince nigérian" demandant un numéro de compte bancaire pour y transférer rapidement des fonds :

Adaptés sous de multiples formes, ce type de messages existent également sous les réseaux sociaux sous la forme de *bots*, comptes tenus par des robots postant des messages promotionnels.

Au regard des différents exemples données ici, nous voyons qu'il est difficile de classer les mèmes selon des catégories précises et qu'en bien des endroits ces catégories se recoupent : les stars font de la politique alors que les politiques font dans le comique. Ainsi, il ne s'agit pas de dresser des catégories rigides mais de disposer d'une classification flexible pour pouvoir analyser les éléments discursifs présents dans les mèmes Internet. La suite de cette étude interrogera plus précisément la constitution de ces formes de discursivités dans leurs multiples empiriquement observables par l'analyse de données issues de *Sina Weibo* : structure du réseau social, contenu des mèmes, forme multimédia, éléments de contexte du message, etc. Nous allons donc maintenant présenter les méthodes d'extraction, de traitement et d'analyse de données en détaillant la méthodologie que nous avons retenue pour cette étude.

3

Méthodologie de recherche

Dans ce chapitre nous présentons la démarche expérimentale que nous avons choisie d'utiliser pour mener à bien cette recherche, ainsi que les résultats que nous avons obtenus. Nous avons choisi de nous saisir de différents outils informatiques, algorithmiques et de visualisation afin d'observer les mèmes circulant sur Sina Weibo. Nous débuterons ce chapitre en discutant des implications d'une méthodologie fondée sur l'analyse et la visualisation de données pour une recherche en sciences sociales. Dans un second temps, nous brosserons un panorama des méthodes utilisées pour explorer les données issues des réseaux sociaux. Enfin, nous présenterons notre démarche ainsi que les résultats de cette étude.

3.1 Sciences sociales du réseau

Pour l'empiriste, l'acte essentiel de la recherche est l'observation. Structure schématique faite de points et de lignes, le réseau n'offre que peu de prises pour une approche empirique. Pourtant, ce concept protéiforme traverse aujourd'hui les disciplines pour se retrouver au centre des débats scientifiques. Couplé à l'autre grand insaisissable de l'étude qu'est *l'information*, le modèle du réseau fait la promesse de nouvelles perspectives en offrant un cadre conceptuel commun pour les sciences et de nouveaux horizons méthodologiques grâce aux outils informatiques. L'interrogation autour des réseaux se trouve donc plus que jamais au cœur du devenir des sciences.

3.1.1 Le réseau comme enjeu pour l'étude

Durant la seconde guerre mondiale et jusqu'à la fin des années 1950, les esprits les plus brillants du siècle (Gödel, Von Neumann, Einstein..) se côtoient à l'*Institut des Etudes Avancées* de Princeton (IAS) aux USA et leurs travaux posent les bases logiques et technologiques d'où émerge la science informatique actuelle. Dans le domaine de la philosophie mathématique tout d'abord, les interrogations amorcées par Russel dans ses *Principia Mathematica* puis leur critique par Gödel dessinent les contours d'une "machine réursive", ancêtre mathématique de la machine de Turing et de l'ordinateur de Von Neumann. Considéré comme le père de l'ordinateur, Von Neuman s'exile d'Allemagne pour rejoindre l'IAS nouvellement fondé dès 1933 où il poursuit des recherches dans le domaine de la physique nucléaire. En 1943, Von Neumann intègre le *Projet Manhattan* dirigé par Oppenheimer où il est chargé de superviser l'immense processus de calcul nécessaire à la construction de la bombe qui s'abattra sur Hiroshima le 6 Août 1945. C'est durant cette période que Von Neumann développe l'architecture encore utilisée aujourd'hui comme base fondamentale du design électronique. Dans ses lectures à l'Université de Yale en 1957 parues sous le titre célèbre de *L'Ordinateur et le Cerveau* (?), Von Neumann identifie les différentes unités d'un ordinateur : unité d'arithmétique logique, unité de contrôle, mémoire et entrées/sorties. Les travaux teintés d'ombre de ce prestigieux mathématicien font écho à ceux de l'anglais Turing qui travaille également sur la question de l'intelligence des machines depuis le début de la guerre. Leur correspondance témoigne du respect mutuel qu'entretiennent alors les deux savants, ainsi que leurs nombreux questionnements sur la possibilité d'une machine intelligente (Istrail et Marcus, 2013). Au cœur de leurs discussions se trouve la quête d'un modèle universel capable d'expliquer le fonctionnement de l'activité cognitive.

Norbert Wiener, un autre prodige des mathématiques prend également part à cette recherche. Sa théorie qu'il nomme *cybernétique* place la notion d'information au cœur de la réflexion sur le fonctionnement des systèmes : "*Information is information, not matter or energy*" (?), p. 155). Utilisant les concepts de bruit, de messages et de *feedbacks*, il jongle entre mathématiques appliquées et sciences cognitives pour comprendre les phénomènes de transmission de signaux complexes. Dans une lettre du 29 Novembre 1946, Von Neumann écrit à Wiener que l'étude du cerveau "*the most complicated object under the sun, literally*" nécessite de poursuivre une réflexion transverse à de multiples disciplines scientifiques (Masani, 1990). Depuis le début de cette même année 1946, les "conférences cybernétiques" aussi connues sous le nom de "conférences Macy" ont débuté à New York dans le but de définir "*une science général de l'esprit humain*"¹. Regroupant physiciens, cogniticiens, biologistes, anthropologues et linguistes, ces conférences sont l'objet de nombreuses publications dont *The Human Use of Human Beings* (?) qui propose d'étudier la société en considérant les communications entre hommes et machines. Le travail mené lors

1. Foundation for Cybernetics, The Macy Conferences <http://www.asc-cybernetics.org/foundations/history/MacySummary.htm> consulté le 10 Mars 2014 à 16h16 GMT+1

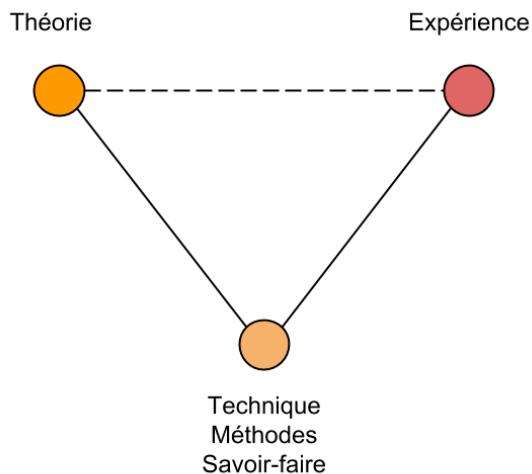
de ces conférences est souvent considéré comme la pierre fondatrice du champ des études sur la communication (??). Quelques années plus tard, McLuhan et son idée controversée de *village global* (?) amènent un large auditoire autour des œuvres de Wiener et l'étude des modes de communication. Progressivement se constitue un champ épistémologique pour l'étude de l'information dont la définition reste encore aujourd'hui un enjeu important (?). En France, les Sciences de l'Information et de la Communication se sont principalement structurées autour de la critique des médias (??) dans une tradition européenne déjà bien installée (?). Aux Etats-Unis, les *cultural studies* (?) s'interrogent davantage sur l'économie politique des symboles et trouvent leur lettres de noblesse dans les *media studies*, aujourd'hui devenue une discipline universitaire reconnue en Angleterre et aux USA. Toutefois, l'émergence d'un “*paradigme communicationnel*” (?) nécessite un dialogue pas toujours évident entre les différentes disciplines constituées autour de l'étude des pratiques de l'information et de la communication : sociologie des médias, études des systèmes d'information, informatique, sciences de l'information et de la communication, etc.

Alors que le programme fixé par la cybernétique a pour ambition d'explorer les structures, possibilités et contraintes des systèmes communicants, l'image fermée et cyclique du système devient rapidement trop étroquée pour une réflexion sur les relations en pleine expansion. Deleuze et Guattari proposent dans l'introduction de leur livre *Milles Plateaux* (?) de s'éloigner de l'arborescence classique du système auto-suffisant pour envisager les liens entre sujets et objets sous la forme ouverte et combinatoire du *rhizome*. Le rhizome se définit par son caractère non-fini et fait de l'étude des causalités un phénomène contextuel aux spécificités pas forcément reproductibles. Imperméable à la catégorisation et aux classifications ordonnés, l'objet d'étude in-formé devient *open-ended*, moment donné à voir et état unique d'un ensemble plus vaste. La vérité ou la validité n'est donc plus à découvrir dans l'analyse, mais plutôt dans l'articulation des contingences entre objets et environnements dont se saisit le champ émergent des études sur la complexité (?). Dans le même temps, les progrès de l'intelligence artificielle et de la robotique viennent également questionner notre définition du vivant face à ces nouveaux êtres (Hofstadter, 1999). Le modèle du *réseau* vient à son tour in-former la pensée scientifique comme un nouvel *episteme* foucaldien offrant une grille de lecture des faits sociaux renouvelée (?Latour, 1996). Aujourd'hui, la structuration d'un champ de recherche cohérent et multi-disciplinaire autour de l'étude des réseaux reste un enjeu important pour le monde de la recherche (Brandes *et al.*, 2013). Cette considération pour la complexité des phénomènes permet en effet d'imaginer une approche scientifique qui atténuerait les liens entre des disciplines dont le dialogue parfois difficile est néanmoins nécessaire, afin de mener vers une réconciliation des sciences humaines, des sciences naturelles et des sciences du vivant où cohabiteraient “*l'étude des organismes, des organes et des organisations*” (?)

3.1.2. Code, données et les nouveaux outils de l'écriture scientifique

Les objets ouverts et mal définis du réseau posent néanmoins la question d'une continuité avec le projet aristotélicien d'une compréhension du monde par le savoir analytique. La considération d'un objet comme moment du réseau rend caduque l'idée de sa définition *a priori*, en dehors des relations qu'il entretient avec son environnement. A l'inverse, une définition des objets entre origine, fonction et destination ne devrait pas échapper non plus la réflexion ontologique autour de son *être*. Le philosophe américain William James propose une méthode qu'il nomme *empirisme radical* où l'existence des objets doit être considérée par le prisme de l'expérience (James, 1912). L'expérience, même scientifique, existe dans un contexte, un lieu, un temps, un moment, une succession d'étapes, et c'est cet ensemble qui nous permet de percevoir les phénomènes étudiés. Si l'acte essentiel de l'empiriste et du philosophe est l'observation, alors elle doit être radicale dans son honnêteté et accepter l'expérience dans son unicité dont de nombreux facteurs sont non-reproductibles. Cette question de l'expérience et de sa reproductibilité est depuis toujours au cœur du débat sur les sciences et se pose comme un des ressorts fondamentaux du savoir scientifique. Les sciences dites "dures" comme la physique, la biologie, l'informatique ou même la géologie valident leurs hypothèses grâce à l'usage extensif d'appareillages multiples devenus une clé de la reproductibilité des expériences. Du simple microscope à l'accélérateur du CERN, la méthodologie d'observation est pour ces disciplines souvent indissociable de la technologie. Le travail de la recherche et de la preuve s'articule ici autour des trois pôles de la théorie, de l'expérience et de la méthode conditionnée à la technique.

En sciences humaines néanmoins, la tension entre la théorie (les concepts), l'expérience (le terrain) et la méthodologie s'articule plus rarement autour d'outillages élaborés. Les réflexions sur la méthodologie en sciences sociales portent plus largement sur des argumentations conceptuelles où le langage lui-même est considéré comme l'outil premier ou encore de nombreuses réflexions éthiques sur l'impact des pratiques d'observation dans le cadre de l'anthropologie notamment. En France, l'intérêt constant porté à l'utilisation des différentes formes de technologie en sciences humaines s'est souvent constitué autour de la critique de méthodologies quantitatives très répandues outre atlantique - la statistique en sociologie, l'étude comportementaliste en psychologie ou le film ethnographique (?). Comme le note justement Latour, la quête de légitimité des Sciences Sociales s'est souvent traduit par une capacité à se procurer et traiter des données : "*Sociology has been obsessed by the goal of becoming a quantitative science.*" (?). Pourtant, il serait illusoire de corrélérer la quantité de données à une quelconque objectivité de la démarche scientifique, tant les démarches et outils nécessaires pour la collecte sont en soi autant de biais importants. La question de la réussite des études utilisant les Big Data semblent davantage corrélée à la capacité d'une approche méthodologique forgée dans un échange interdisciplinaire et des pratiques renouvelées de l'écriture. La complexité des questions abordées lors du traitement des données, non seulement technologique mais aussi dans les interrogations sur la



Graphique 3.1.1 – Triptyque des pratiques scientifiques

légitimité de leur existence, leur utilisation et leur provenance (Boyd et Crawford, 2011) rendent nécessaire la discussion entre de multiples connaissances. Aujourd’hui, l’utilisation de l’ordinateur conditionne l’écriture scientifique de l’étude de terrain, la prise de notes, la rédaction ou la publication et se présente ainsi comme un impératif méthodologique pour la réflexion et le travail en sciences humaines et sociales (Wiewiorka, 2013). L’appareillage projeté au centre de la pratique quotidienne du chercheur vient modifier le travail de réflexion sur les phénomènes étudiés et s’accompagne de multiples contraintes. La prise en main de ce nouvel “*instrument intellectuel*” (Guichard, 2014) passe par une lente alphabétisation aux langages des machines. L’angoisse latente du “bug” n’est pas apaisée par les rares techniciens présents dans les centres de recherche en sciences humaines, souvent déjà dépassés par la diversité des demandes technologiques. L’écriture, savoir-faire indispensable de la recherche, possède une nouvelle matérialité dans les disques durs produisant une dépendance accrue aux réseaux d’ordinateurs. Cette empreinte “digitale” laissée par les calculateurs sur les pratiques scientifiques n’est ni anodine ni révolutionnaire et s’inscrit dans la longue tradition des cultures de l’écrit qui déjà bien avant les moines copistes “combinent les gestes de la main et les opérations de la pensée.” (Jacob, 2011)

Tim Berners-Lee, considéré comme l’inventeur du World Wide Web et du Web Sémantique définit la constitution d’un réseau mondial des savoirs comme un projet d’“*ingénierie philosophique*” (Halpin et Monnin, 2014). Mi-philosophe mi-ingénieur, le chercheur contribuant à ce réseau de savoir doit donc être en capacité de connaître les protocoles et les langages pour y accéder et s’y mouvoir aisément. La méthode scientifique s’écrit notamment avec de nouvelles formulations (fonctions, algorithmes, code...). Les données générées

par les usages d'un nombre croissant de machines communicantes et productrices d'information, souvent désignées par le concept « valise » de Big Data (Lohr, 2012) offrent des possibilités nouvelles pour l'étude en sciences sociales. L'analyse de ces vastes jeux de données s'accompagne également de nouveaux impératifs et questionnements sur l'observation des phénomènes humains qu'ils représentent. À la fois barrière et opportunité, une difficulté majeure réside dans le discernement nécessaire entre praxis des outils informatiques, fascination pour ces outils et réflexions pertinentes sur la qualité des méthodes employées. Le traitement quantitatif par ordinateur permet d'extraire de nombreuses connaissances utiles de jeux de données parfois très importants mais n'assure pas pour autant la qualité des résultats. Une bonne compréhension de la provenance et des méthodes de collection des données est nécessaire afin d'identifier des algorithmes de traitements intéressants, adaptés et efficaces parmi ceux disponibles (Rajaraman et Ullman, 2011). Les méthodologies d'exploration et de recherche utilisant le “Big Data” comme source nécessite la mise en œuvre d'une ingénierie complexe soutenu par une connaissance des technologies nécessaires à l'analyse de données. L'algorithme, la statistique, l'informatique mais également la cartographie et le design graphique doivent se conjuguer pour permettre de produire des résultats à la fois intéressants et fiables. Ce travail d'hypothèses et de vérifications pour l'analyse de données doit réunir de nombreuses compétences. La définition de la problématique la plus adaptée nécessite une connaissance aigüe du terrain et des outils et algorithmes qui seront articulés au sein d'un système ingénierique parfois très complexe. Le design et la lecture d'algorithme pour le “*data mining*” sont donc les clés pour le travail du chercheur confronté aux données. Néanmoins, ces algorithmes ne devraient être que la traduction de questions formulées grâce à une connaissance aigüe des problématiques du contexte et des objets étudiés - notamment pour identifier les données manquantes. La “science des données” promet donc d'apporter un véritable renouveau des méthodes et des résultats scientifiques, au prix d'un travail soutenu pour faire face à ces changements d'habitudes et de langages. Les applications statistiques du “big data” permettent aujourd'hui une fiabilité accrue des prédictions par l'augmentation du volume des corpus traités (Breiman, 2001). Le domaine de l'intelligence artificielle (AI) a grandement bénéficié de l'accroissement de la capacité de traitement des données notamment pour la prédiction grâce au techniques dites de *machine learning*. Néanmoins Peter Norvig, directeur de recherche chez Google, reconnaît lui-même : “*We could draw this curve : as we gain more data, how much better does our system get? And the answer is, it's still improving—but we are getting to the point where we get less benefit than we did in the past.*” (Somers, 2013). Comme le note Douglas Hofstadter, un des pairs de l'Intelligence Artificielle à propos du super-ordinateur qui venait de battre Kasparov aux échecs : “*Okay, Deep Blue plays very good chess—so what? Does that tell you something about how we play chess? No. Does it tell you about how Kasparov envisions, understands a chessboard?*” (Somers, 2013).

En effet, ces pratiques méthodologiques doivent réussir à s'inscrire dans la continuité

de l'historicité et des exigences des disciplines. La portabilité des méthodes et la disponibilité des données sont encore des questions centrales et non-résolues. En sciences sociales notamment, les services de réseaux sociaux en ligne offrent de très large corpus dont l'utilisation est régie par les exigences commerciales des sociétés privées qui les détiennent. L'analyse des données issues de service de réseaux sociaux en ligne est pourtant un champ d'études en rapide expansion (Nettleton, 2013). Pourtant, le débat sur la validité des éclairages apportés par l'analyse des données issues des services de réseaux reste encore largement ouvert et nous entendons dans ce travail de recherche y contribuer.

3.1.3 Visualisation et espace perceptif pour l'information

Face à de larges volumes de données, un des grands enjeux est d'en restituer une forme intelligible afin de d'identifier des tendances ou des motifs particuliers. La visualisation permet de produire une lecture particulière de parties intéressantes et intelligibles d'un jeu de données (Cairo, 2013). Définie comme « *a process that transforms data, information, and knowledge into a form that relies on the human visual system to perceive its embedded information.* » (Graffetti *et al.*, 2010), la visualisation introduit la question du design visuel au cœur de la problématique d'analyse (Wesolowsky et Tufte, 1992). La visualisation correspond à une série d'actions qui résulte dans la production de marqueurs visuels (points, ligne, aires, surface, volume) avec comme étapes la définition de leurs propriétés rétinienennes (couleur, taille, texture, etc.) et leur positionnement dans l'espace visuel (?) à définir les bases d'une grammaire de la visualisation en commençant par en identifier les formes syntaxiques :

1. les objets graphiques montrés (ex. point, flèche, pictogramme, etc.),
2. l'espace graphique donnant sens à l'organisation des objets (ex. systèmes de coordonnées géographiques, timeline, etc.),
3. les propriétés graphiques des objets (couleurs, tailles, etc.),
4. l'organisation des objets en différentes catégories (ex. cadre, liens, légendes, etc.).

Ces choix forment une tâche importante dont l'enjeu n'est pas seulement visuel, mais se joue également dans le champ de la représentation où les objets sont donnés à voir et par là-même donnés à comprendre. Les travaux sur l'apparition de la perspective dans la Renaissance italienne ont montré comment l'espace de la représentation visuelle fait écho aux changements sociaux profonds de l'époque (Raynaud, 2005). Au-delà d'une simple technique picturale, les œuvres des peintres du Quattrocento témoignent de changements profonds dans la perception : l'espace perceptif se structure désormais autour du sujet et de son "point de vue" qui construit l'ensemble de la représentation (Damisch, 1999). Empreinte de rationalité, la perspective construit comme au théâtre un espace de représentation centré autour du spectateur. En 1639, le mathématicien Desargues modélise les notions intuitives de perspective et d'horizon grâce à la géométrie projective qui permet d'étudier les propriétés inchangées des figures lors de leur projection. Cette géométrie d'un

genre nouveau se structure autour du *plan projectif*, élément topologique qui “rassemble en une seule surface l'imagination de tous les points de vue possible” (?). En construisant un plan géométrique fondé sur le point de vue, de nouveaux êtres géométriques aux propriétés étranges voient le jour, dont l'existence logique force notre représentation classique : la bande de Möbius ne possède qu'une seule face et il est impossible de distinguer l'intérieur de l'extérieur d'une bouteille de Klein. Dans ces surfaces dites *unilatères*, le local est traversé en tout point par un tout global. Nous ne pouvons pas traverser puisque nous sommes toujours sur la même face. Ni bord, ni extérieur, ni intérieur, le plan projectif apporte des éléments de réponses conceptuelles aux limites de la représentation dans l'espace.

Dans le contexte de systèmes et de réseaux complexes, la visualisation de données structure l'espace perceptif afin de construire une scène à n dimensions dont l'enjeu est la recherche d'un “point de vue” pour l'étude. Alors que la perspective prend pour parti de matérialiser le sujet au centre de la représentation par des points de fuite, la visualisation de données cherche elle à utiliser les objets comme dimensions du champ de la représentation, en structurant souvent l'espace autour de quantités. Néanmoins, la place du spectateur / utilisateur dans la visualisation reste un des enjeux majeurs encore à explorer. En élaborant sa “méthode graphique”, J.E. Marey ([Marey, 1885](#)) utilise la photographie pour créer un nouvel espace de représentation du mouvement et observer des phénomènes jusqu'ici invisibles. Si la temporalité de l'écrit ou de la voix est avant linéaire, l'espace visuel permet de manier le réel pour le décomposer en actes logiques. Charcot cherche dans les images de l'hystérie des témoignages de la folie et procède à la mise en scène de ses patients dans ce nouvel espace de représentation ouvert par la photographie (?).

La visualisation scientifique se prévaut donc d'une existence avant tout pratique dont le premier objectif serait de « *to effectively convey information* » ([Kelleher et Wagener, 2011](#)). Son caractère syncrétique et sa capacité à résumer une large masse d'information rapidement en font un des plus importants éléments de la publication en science notamment pour sa diffusion et la facilitation d'accès à une connaissance ([Ware, 2004](#)). Dans sa sémiologie graphique, Bertin ([Bertin, 1977](#)) distingue deux usages majeurs des graphiques de visualisation : 1) un moyen de communiquer des informations (dans le cas où l'information a déjà été comprise) 2) un moyen visuel de résoudre des problèmes logiques (quand le graphique est utilisé comme support de lecture et de manipulation d'informations). Ces deux caractéristiques peuvent coexister dans certaines pièces mais la transition entre les deux nécessite souvent un travail important de restructuration visuelle. Dans le traitement et la visualisation des données *l'interface* joue un rôle primordial. On peut désormais agir sur la 2ème catégorie de Bertin pour mieux explorer le sens (?). Manovich montre comment l'interface définie comme “*the ways to represent ('format') and control the signal.*” ([Manovich, 2013](#)). Ce formatage nouveau de l'information induit des changements dans la pratique de la lecture qui, toujours selon Manovich, s'apparenterait davantage

à de la reconnaissance de *pattern*, symbolisé par l'usage de l'icône et du menu en design d'interface. Ainsi si l'interface constraint la lecture, la prise en compte des formes narratives (les *patterns* de Manovich) prend une grande importance quand il s'agit de concevoir une visualisation d'information. L'usage des signes graphiques doit se faire avec une connaissance des usages de l'interface, afin de recréer la coopération textuelle des rôles de lecteur et de designer/auteur nécessaire pour la production un sens (Eco, 1985). La *citizen science* ou encore *night science* a fait de l'interface un paradigme en utilisant la visualisation pour amener un grand nombre de collaborateurs à explorer et analyser de vastes jeux de données en effectuant des tâches simples (Silvertown, 2009). Le projet *Eyewire* permet à des internautes de contribuer à la classification d'images du cerveau humain en vue de la réalisation d'un modèle 3D (Seung, 2012). Utilisant des scans de tranches de 1mm réalisés par l'institut Max Planck, la modélisation 3D d'un cerveau complet promet une belle contribution pour la découverte des fonctionnements cognitifs. Néanmoins, la tâche est colossale et nécessiterait plusieurs années pour une équipe classique de scientifiques. *Eyewire* propose donc une interface web où un simple jeu de coloriage permet d'identifier les neurones et de contribuer ainsi au dessin du modèle 3D.

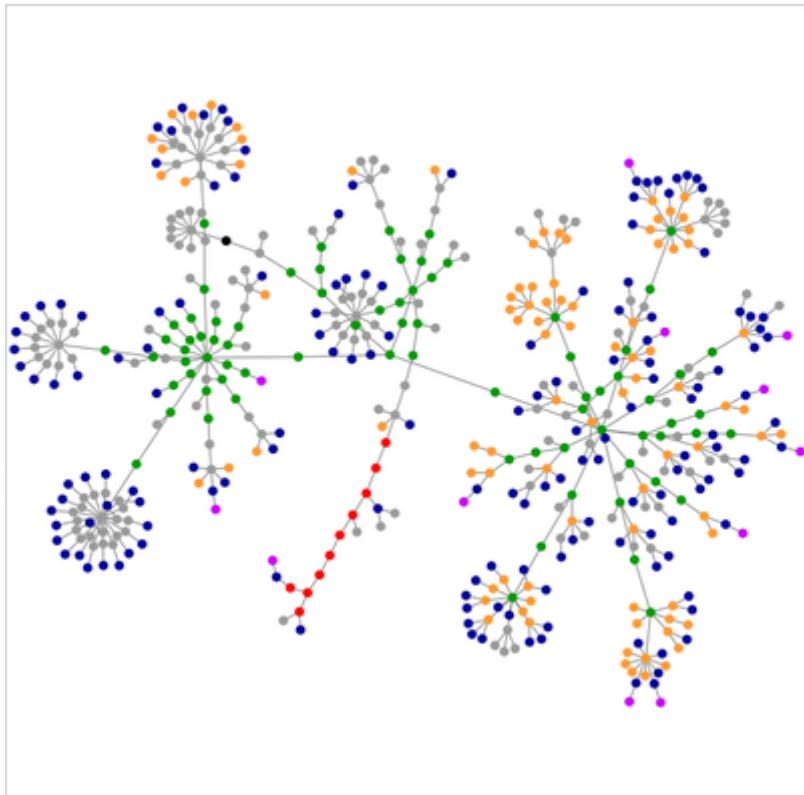
Cartes, code ou graphiques, les nouveaux outils numériques participent donc à structurer de nouvelles pratiques de l'espace grâce à la construction de nouvelles représentations.

3.2 Méthodes et outils d'analyse des réseaux sociaux

Nous allons maintenant regarder comment l'analyse des données des services de réseaux sociaux en ligne (SNA) peut permettre d'interroger les pratiques des technologies numériques pour mieux en comprendre les tenants.

3.2.1 Anatomie d'un réseau social

La représentation des relations sociales sous forme de graphe trouve son origine dans les travaux des psychologues allemands de la *gestalt* durant les années 1920-1930 (Scott, 1988). S'inspirant des études sur le cerveau, le psychologue J. L. Moreno s'applique notamment à comprendre les principes organisationnels holistiques des groupes humains et fonde la sociométrie avec comme objectif la qualification et la quantification des relations sociales (Moreno et Jennings, 1938). Moreno cherche à identifier et isoler des leaders de groupes sociaux définis en étudiant l'asymétrie ou la réciprocité de leurs choix et fréquentations amicales. Cherchant des moyens de représenter les tendances à l'auto-organisation qu'il observe, il cartographie les relations directes et indirectes entre personnes sous forme de *sociogrammes*. Les anthropologues s'emparent rapidement de ce type d'outils pour comprendre les formes tribales (Lundberg, 1936) et l'émergence progressive de la topologie comme domaine important des mathématiques vient définir de nouveaux types de relations entre objets disparates, avec notamment la théorie des graphes qui donne à

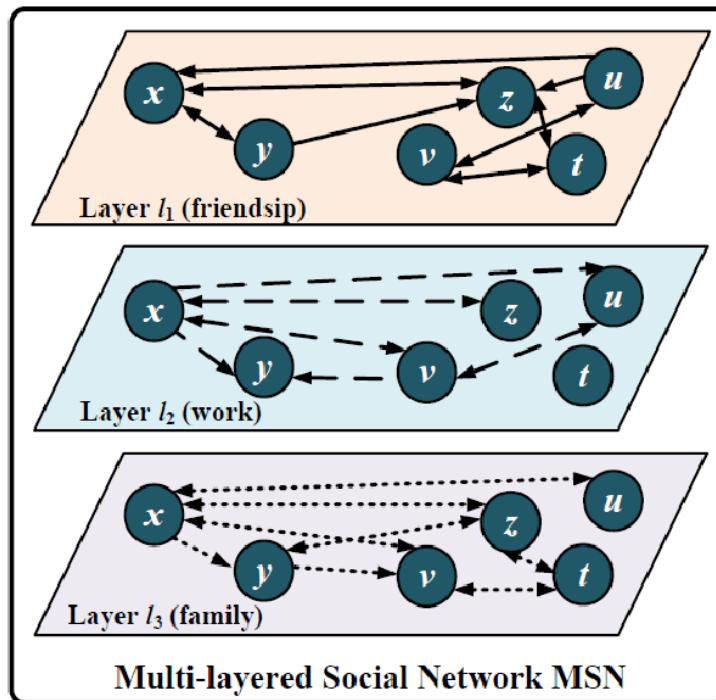


Graphique 3.2.1 – Représentation d'un réseau sous forme de graphes.

l'étude des réseaux ses modèles logiques ([Harary, 1977](#)). Milgram (?) voit les relations humaines comme autant de petits mondes (*small-worlds*) connectés entre eux. Granovetter s'intéresse à l'importance des relations ténues et lointaines (*weak ties*) dans l'acquisition d'informations importantes (?). L'influence de la théorie des graphes amène notamment les sociologues à expérimenter de nouveaux modèles venus de la physique ou de la biologie, en proposant de nouvelles pratiques comme celle de la simulation sociale ([Epstein et Axtell, 1996](#)).

La matérialité de l'image du *graphe* structure la représentation du réseau social. Dans la littérature concernant les réseaux, les notions de graphe et de réseau sont interdépendantes et la théorie des graphes sert notamment de systèmes de notation pour la mise en équation des réseaux ([Nettleton, 2013](#)). Si cette structure point-ligne semble toutefois revêtir une limite de taille pour la description de phénomènes humains (comment en effet réduire les relations humaines à une simple ligne ?), elle semble néanmoins aujourd'hui encore difficile à dépasser.

Un réseau considéré comme graphe, noté G , se compose d'un ensemble de nodes ou vertices (les points) et de liens ou edges (les traits). On représente ainsi un graphe sous la notation $G(V,E)$ où V est l'ensemble des nodes du réseau et E l'ensemble des liens décrivant leurs relations. E décrit les relations entre les nodes qui peuvent être directionnelles (paires de vertices ordonnées) dans le cas d'un graphe *orienté* ou accompagné de valeurs particulières dans un graphe dit *pondéré*. Les relations ainsi exprimées portent sur un



Graphique 3.2.2 – Un réseau social multi-couches, d'après (Brodka&al., 2013)

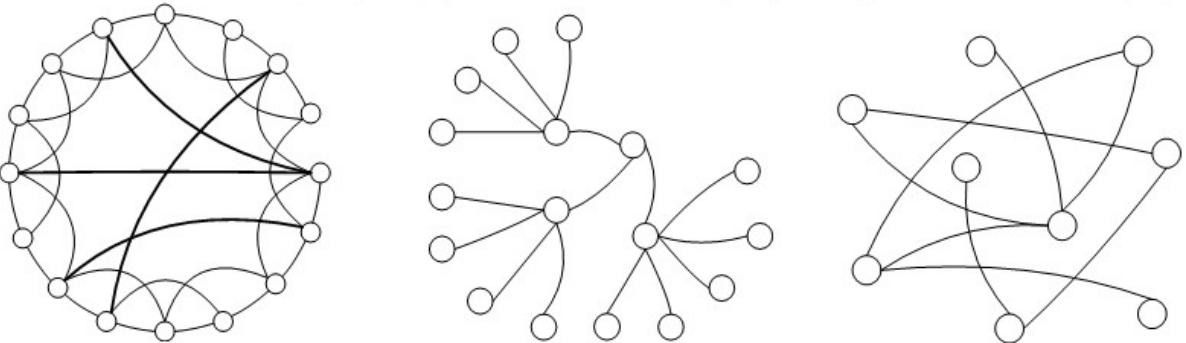
aspect unique, quantifiable et isolable. La prise en compte de facteurs multiples, comme notamment l'espace physique, le temps, mais également les multiples réseaux de relations qui peuvent exister entre deux acteurs nous amènent à considérer un graphe disposant de multiples couches (*multi-layered*) pour décrire l'ensemble des groupes de relations. Imaginons un graphe de personnes $G(V, E_n)$ où V sont les vertices représentant les personnes et E_n un nombre n d'ensemble de liens décrivant chaque type de relations spécifiques. Le graphe ci-dessous montre un exemple d'un graphe multi-couche où E_1 est l'ensemble des relations amicales, E_2 les relations de travail et E_3 les relations familiales.

Nous pouvons ainsi décrire différents jeux de relations entre un jeu d'acteurs finis, permettant notamment de mieux comprendre les relations entre ses différentes dimensions. Si cette approche résout momentanément la question des n dimensions à considérer (?) elle augmente également la complexité du graphe et la possibilité d'erreurs de lecture ou de typage des relations.

Afin de mieux comprendre l'organisation d'un réseau, nous disposons de plusieurs mesures pour décrire les relations et le rôle des différents acteurs : i

- *Degree* : (*degré* ou *valence*) mesure le nombre de connections d'un nœud dans un réseau. Cette valeur indique souvent une possibilité, le potentiel d'un node donné à interagir avec d'autres.
- *Closeness* : (proximité) mesure la facilité d'un node à se connecter à un autre. Dans un réseau en ligne, on calcule la proximité en estimant la distance la plus courte entre un node et un autre.
- *Betweenness* (*centralité*) mesure le degré d'importance d'un node dans le réseau en

(a) Small-World Network (SWN) (b) Scale-Free Network (SFN) (c) Random Network (RN)



Types communs de réseaux

tenant en compte le nombre de nodes dépendant de lui pour établir une connection entre eux. La centralité représente la capacité à bloquer ou laisser filtrer l'accès à certaines parties du réseau. Dans une entreprise par exemple, la secrétaire du CEO a par exemple une très haute centralité.

En observant ces différentes mesures, nous pouvons définir différentes structures types pour chaque réseau. La distribution des degrés dans le graphe permet notamment de comprendre les modèles qui régissent les connexions entre les nodes. L'exemple le plus simple est le *random network*, réseau où les acteurs sont connectés de manière entièrement aléatoire. Un réseau dont le degré de distribution correspond à une loi de puissance est appelé *scale-free network*. Un réseau dont seulement quelques nodes possèdent une centralité élevée et dont la structure d'ensemble est faite de groupes ou *clusters* interconnectés et appelé un *small-world network*.

Les services de réseaux sociaux en ligne sont structurés en *small-worlds*. Dans une étude analysant de larges corpus issus de différents services de réseaux sociaux en ligne, Kumar & al. ([Kumar et al., 2006](#)) ont montré que ces communautés possèdent toutes une structure et une évolution similaire. Les inscrits de chaque service se répartissent autour de trois grands groupes : *singletons* (membres isolés, inactifs), *giant component* (la majorité des utilisateurs actifs) et *middle region* (des communautés isolées qui interagissent entre elles mais pas avec le reste du réseau). D'après les auteurs, il existe très peu de chances que deux communautés isolées même très similaires se rencontrent dans ce type de réseau, car l'entropie de la structure *small-world* se renforce avec le temps en se fragmentant davantage. Une des raisons principales est que ces communautés isolées suivent souvent un modèle en *étoile*, c'est à dire qu'elles sont construites autour d'un individu central charismatique. Éventuellement, il leur arrive de rejoindre la masse du *giant component* mais elles en deviennent difficilement des acteurs majeurs et restent en périphérie. Dans un réseau social de type *small-world* comme les OSNS, les acteurs les plus influents sont ceux qui sont capables de 1) renforcer les liens dans un cluster (*closure*) et 2) développer les connections faibles entre des clusters (*brokerage*) ([Burt, Centola, & Kahl, 2008](#)). Ce "capital social" est inscrit dans la structure du réseau social lui-même ([Lin, 1999](#)) et n'a pas de relations avec le degré du node (son nombre de connections) ([Cha et al., 2010](#)).

L'individu le plus puissant d'un réseau est donc celui possédant le plus grand nombre de connections *potentielles*, proches et facilement accessibles et qui bénéficie d'une place privilégiée pour bloquer ou autoriser l'accès aux autres parties du réseau. L'analyse organisationnelle a notamment montré l'importance capitale des secrétaires pour le maintien d'une bonne circulation des informations dans le réseau de l'entreprise ou la nécessité d'un nombre très faible de connections directes pour les acteurs importants de réseaux terroristes ou mafieux (Russel, 2011).

3.2.2 Diffusion des discours dans les réseaux sociaux

L'analyse de réseau social permet de comprendre la structure d'un moment du réseau. En effet, les connections entre différents acteurs sont sans cesse en mouvement et l'information en circulant vient modifier l'activité du réseau et bien souvent sa structure même. Notre étude porte sur la diffusion de mèmes au sein du réseau social chinois Sina Weibo et s'inscrit ainsi dans la continuité des travaux s'intéressant aux rapports entre diffusion et technologies. Néanmoins, le champ de la diffusion de l'innovation, objet traditionnel de la géographie, s'est largement constitué autour des problématiques technologiques et organisationnelles liées à la situation spatiale du lieu et ses équipements (Crevoisier, 2004). Les études urbaines ont notamment montré comment la réalité économique et géographique présidait à la constitution des savoirs nécessaires à la diffusion de l'innovation technologique (Howells, 2002). Cette vision doit être modérée par une considération plus importante des modalités d'appropriation des technologies comme pratiques propres aux territoires (Fernandez et Puel, 2010). Plus largement, les études sur la diffusion sont dominées par l'analogie du virus comme modèle de propagation des messages, comportements et idées. Depuis la fin du XIXe siècle (Le Bon, 1895), ce modèle est prépondérant dans les recherches autour de la diffusion d'information en ligne (Goel *et al.*, 2012) Les membres de groupes dans un réseau seraient *exposés* à un message ou une idée avant d'être *infectés*, devenant alors porteur puis agent de sa diffusion. Ainsi, en considérant la position d'un individu au sein du réseau de diffusion, il serait possible de définir un "*degré d'infection*" (?) et d'anticiper la diffusion selon une « *probabilité immune* » qui déciderait de la « *qualité infectieuse* » de l'objet diffusé ou de la « *possibilité de rétablissement* » du sujet infecté (Wang et Wood, 2011). Cette analogie du viral propose une vision mécaniste qui fait peu de cas des facteurs contextuels ou psychologiques et ignorent ainsi les processus de décisions individuels en jeu (Jackson et Yariv, 2010). Pourtant, nous savons notamment que le choix des mots ou la situation des personnes diffusant un message sont des facteurs décisifs et structurant des processus d'adoption des messages en ligne (Conover *et al.*, 2013). En s'intéressant à la diffusion géographique de mots dans les réseaux sociaux, Eisenstein & al. (Eisenstein *et al.*, 2012) observe que leur diffusion se limite à un domaine géographiquement bien défini, dépendant de facteurs culturels et démographiques. Par exemple, les villes ayant d'importantes communautés afro-américaines ont davantage de chances d'adopter un même mot que d'autres parfois plus proches géo-

graphiquement. Cette question est au centre des études de marketing qui s'interrogent notamment sur l'adoption de nouvelles marques ou de slogans. Afin de déterminer statistiquement les possibilités d'adoption de produits et prévoir la pénétration dans un marché précis, la modélisation mathématique des effets de réseaux dans la diffusion est souvent utilisé (Bass *et al.*, 1994). Ici, l'analyse des données des réseaux sociaux est un grand enjeu pour la prospective économique et l'application de ces modèles mathématiques aux données utilisateurs permettant de considérer des segments précis de marché. Le marketing politique fait lui aussi grand cas de l'analyse de réseaux sociaux pour comprendre et orienter les discussions. La campagne de réélection d'Obama aux USA en 2012 a fait un usage extensif de l'analyse de données des réseaux sociaux pour identifier, déterminer et cibler des groupes sociaux particuliers grâce au travail d'une vaste équipe d'ingénieurs et de "data scientists"².

Une des grandes interrogations dans ce domaine est évidemment les rôles joués par les différents acteurs du processus de diffusion – et la manière de les identifier. Les études concernant les leaders d'opinion, traditionnelles en sciences de la communication (Katz et Lazarsfeld, 1955) trouvent une continuité directe dans l'étude des réseaux sociaux en ligne avec le domaine florissant des recherches sur l'identification des *influenceurs* (Bakshy *et al.*, 2011; Leavitt *et al.*, 2009). L'analyse quantitative permet notamment de mieux comprendre l'influence réelle des acteurs dans le réseau grâce à l'étude de leurs comportements. Le concept d'influence sur les réseaux sociaux revêt en réalité des formes très variables et procède notamment d'une légitimité construite autour de sujets précis par des personnes spécialisées devenues référentes (Cha *et al.*, 2010). D'autres "influenceurs" possèdent une grande capacité d'amplification pouvant par exemple initier le développement d'une "*masse critique*" autour d'une information, définie classiquement comme le seuil d'adoption à partir duquel la diffusion devient pérenne parmi une foule d'acteurs (Oliver et Marwell, 2001). L'image quantitative d'une "masse" uniforme et actionnable dans le réseau est néanmoins remise en question par l'étude de données dont l'analyse montre que les relations entre différents acteurs du réseau sont à considérer qualitativement en termes de relations de pouvoirs (Steyer *et al.*, 2006). Les dynamiques d'échanges ne répondent en effet pas tant à une relation pré-existente dans le réseau qu'à un ensemble de situations où les acteurs adoptent des comportements et des réactions particulières. La diffusion peut ainsi se comprendre comme une pratique du *bouche-à-oreille* éminemment contextuelle, où certains acteurs sont plus ou moins influents dans telle ou telle situation ou sur tel et tel sujet. Le temps joue néanmoins un rôle déterminant puisqu'en considérant des séries de résultats où des acteurs se côtoient durant plusieurs années, il est souvent

2. "Harper Reed, the chief technology officer for the Obama re-election campaign, who heads a team described as "100 data scientists, developers, engineers, analysts, and old-school hackers [that] have been transforming the way politicians acquire data—and what they do with it.", from The Blaze <http://www.theblaze.com/stories/2012/10/03/very-creepy-details-of-obama-campaigns-voter-data-mining-effort/> consulté le 12 Mars 2014 à 14 :50 GMT+1

Graphique 3.2.3 – Tableau récapitulatif de méthodes d’analyse de données de réseaux sociaux en ligne.

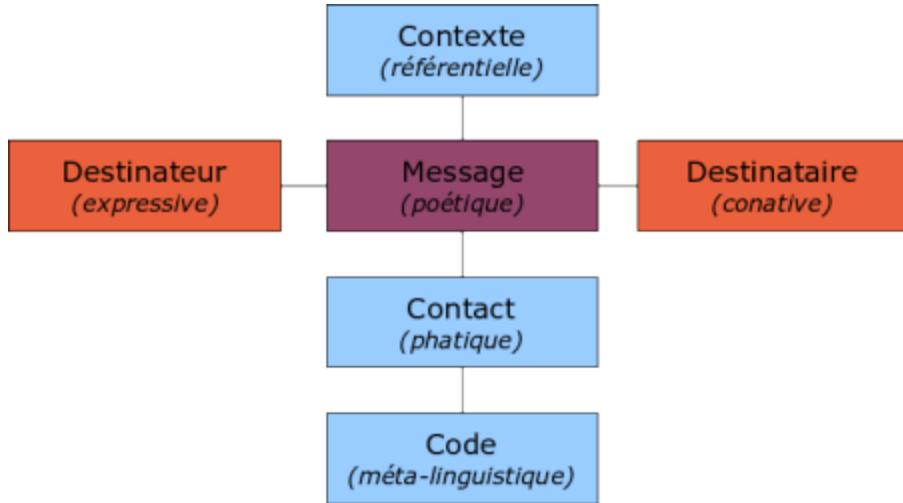
		Courant d’analyse	
1		Graphes sociaux de groupe définis (cartographie de réseau fini)	
2		Découverte de groupes par critères (communautés)	Cette méthode per...
3		Analyse sémantique des conversations (analyse de contenu)	En utilisan...
4		Analyse de la diffusion (évolution des relations d’après une conversation)	D’après ...
5		Analyse comportementale et agents de diffusion (classifications et mesures)	Ici on étu...
6		Analyse contextuelle et géographique	
7		Simulation sociale	Ce type d’analyse ...

difficile d’identifier quel acteur influence l’autre ([Aral *et al.*, 2009](#)).

3.2.3 3.2.3. Méthodes d’analyse de données des réseaux sociaux en ligne

Après moins d’un siècle d’existence, l’analyse de réseaux sociaux sous forme de graphes a donc connu une rapide évolution et une diversification dans de nombreux domaines de recherche. Les services de réseaux sociaux en ligne offrent notamment la possibilité d’obtenir des données sur les comportements de groupes sociaux en très vaste quantité. Véritable vivier d’études, ce champ de recherche en pleine expansion trouve ses origines dans des disciplines diverses qui poursuivent souvent des objectifs et des méthodes très différentes. Le tableau infra présente quelques-unes des méthodes d’analyse de données définies selon leurs domaines d’application. Ces méthodes coexistent souvent lors d’études utilisant le *data mining*; nous présentons ici des exemples qui permettront ensuite de mieux situer nos perspectives de recherche dans ce paysage de pratiques.

Notre recherche choisit donc de s’appuyer sur une étude de cas spécifique de Sina Weibo reprenant les éléments méthodologiques de l’étude du graphe social de la diffusion entourant les conversations (Fig. 1-4) et l’analyse sémantique des sujets dominants et sous-jacents aux conversations (Fig. 1-3). Également, nous souhaitons amener une plus forte contextualisation des usages (Fig. 1-6) en prenant en compte notamment les relations entre les dimensions sémantique et conversationnelle, mais aussi géographique de l’existence des conversations. L’approche de l’étude des données des SNS par les géographes notamment s’est jusqu’ici largement focalisée sur l’analyse des *geotag* et des cartes en ligne comme principale approche méthodologique ([Graham et Zook, 2011; Poorthuis et Zook, 2013](#)), réduisant considérablement les possibilités d’étude des données de réseaux sociaux en ligne ([Crampton *et al.*, 2013](#)). Nous souhaitons ici mettre en perspective ces différentes dimensions afin d’enrichir le modèle d’étude. En effet l’analyse de la diffusion utilise principalement les graphes des réseaux de diffusion mettant en scène les utilisateurs et leurs interactions en ligne. Loin d’être intérressant, ce type de schéma est néanmoins très réducteur car il se fonde sur un modèle communicationnel très primaire “émetteur-récepteur”



[Modèle de Jakobson] Modèle de Jakobson, intéressante que nous pouvons chercher à adapter dans le contexte des échanges en ligne.

dont on connaît les limites. Les travaux de Jacobson ont notamment permis d'étoffer ce modèle en considérant les différents aspects fonctionnels des actes de communication.

Ainsi, en nous basant sur les modèles des théories de la communication, nous pouvons peut-être améliorer les modèles méthodologiques d'analyse. Nous proposons ici la notion de *topogramme* comme modèle pour comprendre les motifs de diffusion des mèmes, considérés comme *topos* ou *lieux communs*. Le topogramme, en tant que représentation graphique des différentes dimensions et dynamiques lisibles dans les données nous permet donc d'approcher un travail d'observation précise de la diffusion des mèmes, voire par la suite de classification des actes de communication en ligne. Afin de prendre en compte, les différents aspects de la communication, il nous faut donc effectuer une analyse à plusieurs niveaux (multi-layers), regroupant un ensemble de réseaux à la fois sémantique, conversationnel et géographique.

3.3 Design de la recherche : Méthodologie choisie et schémes d'analyse retenus

3.3.1 Constitution et collection d'un corpus de données

Afin de poursuivre notre étude, nous allons donc procéder à l'analyse de données issues de services de réseaux sociaux en ligne. La plupart des services de réseaux sociaux en ligne offrent un large accès car il s'agit souvent de la fondation de leurs modèles d'affaires basés sur la valorisation et la revente de ces données pour le marketing ciblé (Ko *et al.*, 2010). Pour entrer en contact avec la base de données, les SNS mettent à disposition une API (*Application Programming Interface*) qui permet à un programme ou une autre application web de se connecter au service pour demander et obtenir des données. L'API est donc la première source d'obtention de données depuis les SNS. Néanmoins,

les données des réseaux sociaux sont soumises à d'importants enjeux et contraintes tant commerciales, éthiques que politiques dans le cas de la Chine notamment. Les conditions d'utilisations techniques et légales (*Terms of Uses*) de ces données sont également soumises à des changements fréquents, étroitement liés à l'évolution commerciale et technologique de compagnies souvent très jeunes. Voici une liste des limitations et écueils pouvant être rencontrés lors de l'extraction et de l'analyse de données des SNS : ii

- *Compatibilité* : Une solution technologique devient facilement caduque lors de l'évolution d'une API (ex. Twitter APi v1.0 n'existe plus, ainsi le code doit être réécrit pour la version 1.1).
- *Disponibilité des données* : Chaque API répond à des formats et critères précis et possède ses propres limitations. Pour accéder à l'API du moteur de recherche de Sina Weibo, il faut s'identifier auprès de la compagnie grâce à une carte d'identité et des paiements par requête sont exigés.
- *Limitations d'usage* : Afin de limiter le trafic et conserver le contrôle sur les données distribuées, les SNS mettent en place des limitations d'accès à leurs serveurs, notamment : limitation du nombre de requêtes par heure, limitation du nombre de requêtes par machine (basée sur l'adresse IP), limitation du nombre d'utilisateurs connectés. Ainsi, Twitter limite à 150 requêtes API par heure pour un compte non identifié, pouvant augmenter jusqu'à 500 après authentification. Les données datant de plus de 7 jours sont payantes, reflétant la valeur d'un accès en "temps-réel" aux données.
- *Légalité* : Les données sont soumises aux conditions de propriété décrites légalement par la firme qui les publient (Clifton *et al.*, 2006). Ces conditions sont susceptibles de changer. Ainsi, Twitter a exigé en 2012 le retrait a posteriori de nombreux jeux de données publiés par des chercheurs depuis plusieurs années parfois (McCreadie *et al.*, 2012). Actuellement, Twitter précise notamment dans ses *Terms of Use* : “*You may not resyndicate or share Twitter content, including datasets of Tweet text and follow relationships*”³.
- *Ethique* : Vous pouvez extraire depuis une API des profils d'utilisateurs contenant les informations qu'ils ont auparavant publié en ligne (Felt & Evans, 2008). En exposant les données personnelles des utilisateurs, le chercheur est responsable de l'usage qu'il fait des données (Rieder, 2005)

Afin d'obtenir des données et de contourner les limitations de l'API, la pratique dite du “*scraping*” permet d'obtenir des données à l'aide d'un robot qui lit et sauvegarde des parties ou l'intégralité de pages web. Les moteurs de recherche utilisent notamment cette technique pour l'indexation des pages. Ce type de pratiques est également soumis à des limitations par les services web (blocage de l'IP source) et se situe à la limite de la légalité, voire est explicitement interdite dans le cas de certains SNS (Petschulat, 2010).

3. Terms of use de Twitter, <https://dev.twitter.com/terms/api-terms>, consulté le 12 Mars 2013 à 17h08

Un programme appelé *spider* ou *crawler* est chargé d'effectuer des requêtes régulières à l'API afin d'obtenir et collecter les informations obtenues dans une base de données. Plusieurs approches existent dans les techniques échantillonnage de réseau social. La première, fondée sur des mots-clés extraits les posts contenant des mots ou des hashtags particuliers. La seconde utilise l'échantillonnage de graphe, collectant au fil des liens les conversations ou profils des utilisateurs. Classique des études statistiques, cet échantillonnage dit *boule de neige* “élargit l'échantillon en partant d'un node original pour s'éloigner vers ses voisins” (Rothenberg, 1995). Ici, deux grandes catégories s'opposent : les techniques traversales où les nodes sont classifiés après avoir été visités et les “walk” aléatoires où l'extension du graphe visité se fait de manière aléatoire (Gjoka et Kurant, 2011).

Au cours du travail préparatoire de cette recherche, nous avons tout d'abord expérimenté plusieurs algorithmes et outils de collection de données afin d'en comparer les résultats. Une première approche d'extraction par utilisateurs a été infructueuse car la sélection du groupe source (*seeds*) ne permettait pas d'obtenir des résultats cohérents⁴. Par la suite, une autre approche de collecte de données via le développement d'un plug-in pour le navigateur Google Chrome⁵ nous a permis de tester et d'apercevoir les limites de la collection de données par mots-clés. Cette étape nous a également montré l'intérêt que peut présenter une approche collaborative de la collection de données ou de seeds par un système de “curation” collaboratifs afin d'obtenir des éléments précis de contenus et de réduire la masse de données inutiles qui pollue souvent les jeux de données (Ding et al., 2013). Après de multiples tests et comparaisons d'outils et de librairies, plusieurs difficultés majeures limitaient l'obtention d'une quantité de données suffisantes, notamment la nécessité de ressources assez importantes (en terme de développement et de disponibilité des machines), un temps d'acquisition parfois très long et l'exigence d'une veille constante sur les SNS pour identifier un même au bon moment (les tweets de Sina Weibo devenant indisponibles via l'API au-delà de 7 jours). La première limite se situe bien sûr dans la capacité d'une personne seule à mener à bien cette large tâche. Nous avons donc choisi de considérer les jeux de données collectés sur Sina Weibo disponibles sur l'Internet. Une fois écartés les nombreux jeux tronqués, modifiés ou incomplets, nous avons pu obtenir plusieurs jeux de données provenant de recherches préalables dans le domaine particulier de l'échantillonnage (Ding et al., 2013) ou ayant servi de bases à des études précédentes (Gao et al., 2012; ?). Finalement, nous avons identifié le jeu de données constitué lors du projet *Weiboscope* de l'Université de Hong Kong comme répondant à nos besoins en termes de dimensions (temps, nombre d'utilisateurs observés), taille (nombre de tweets) et contenus (géo-localisation, présence des tweets censurés).

Notre travail d'analyse s'appuie donc sur ce jeu de données collecté sur le service de microblog Sina Weibo par le *Journalism and Media Studies Centre* (JMSC) de l'Université

4. Code disponible : <https://github.com/sharismlab/Pyweibo>, consulté le 14 Mars à 5h32

5. Code disponible : <https://github.com/sharismlab/battlefield>, consulté le 14 Mars à 5h12

mid		
retweeted_status_mid		Pseudo message ID of the origin
Uid		
retweeted_uid		Pseudo user ID of the original
Source		The app
Image		
text		body of the message. Any address handl
geo		GIS information. Please refer to
created_at		
deleted_last_seen		The last seen time before
permission_denied	'permission denied'	status is marked when the message was found missing in

de Hong Kong lors de son projet *Weiboscope*. Téléchargeable ouvertement, la publication de ce jeu de données a pour objectif de “enables academic use of the data for better understanding of the social media in China and making the Chinese media system more transparent.”⁶ Il s’agit d’un échantillonnage aléatoire de messages (*random sampling*) effectué quotidiennement durant toute l’année 2012 sur un panel d’environ 350 000 utilisateurs ayant au moins 1000 followers (Fu et al., 2013). La totalité du jeu de données comprend 226,841,122 messages répartis sur 52 semaines, dont des messages ayant été supprimés par les utilisateurs eux-mêmes ou par les administrateurs de Sina Weibo eux-mêmes - parfois sur ordre du gouvernement chinois (*ibid*, 2013).

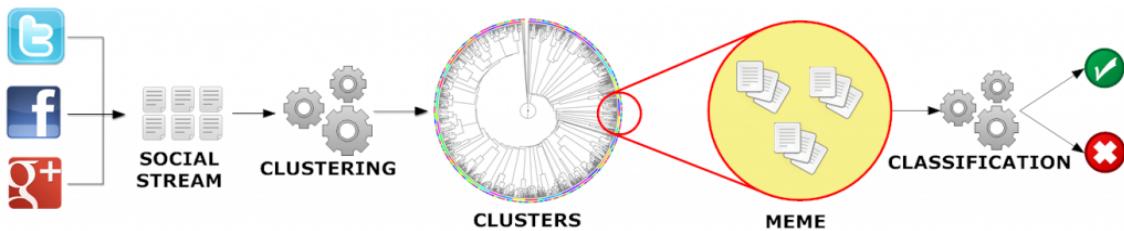
Ce jeu de données a été mis à disposition sous une forme anonymisée où les identifiants des messages et des utilisateurs ont été remplacés par des pseudo-identifiants. La collection des données a été effectuée sur une série d’utilisateurs (génération aléatoire d’identifiants dont l’existence est ensuite validée) pour donner “une image représentative des usages et utilisateurs de Sina Weibo (...) dont les études auparavant limitée à des analyses non-aléatoire (...) se cantonnaient aux utilisateurs les plus populaires” (Fu et al., 2013). Ainsi, plutôt que de considérer uniquement les “stars” de Sina Weibo, cet échantillon s’attache à refléter également les pratiques des utilisateurs “lambda”. Ce jeu de données a déjà été partiellement étudié dans le but de comprendre la démographie des utilisateurs de Sina Weibo, leurs activités et les comportements pouvant permettre de prédire les réactions notamment de censure (Fu et al., 2013). La démographie des utilisateurs se composent de 55% d’hommes habitant principalement dans les grandes villes de Chine (Pékin, Canton, Shanghai). Une des découvertes importantes est le très faible taux de création originale de messages malgré une activité importante des utilisateurs, indiquant que l’essentiel de l’activité sur Sina Weibo se constitue de re-posts et de commentaires. Le jeu de données est accompagné des informations succinctes de profil des utilisateurs dont le lieu, rempli par eux, sans néanmoins fournir leurs noms d’utilisateur vérifiables. Notre travail de recherche s’articule autour d’une nouvelle lecture de ce jeu de données unique.

6. Le jeu de données Weiboscope est disponible à l’adresse : <http://147.8.142.179/datazip/>, consulté le 14 Mars à 17h21

3.3.2 Identification des mèmes en tant que “clusters”

Une fois l’acquisition des données effectuées, il s’agit désormais de savoir les analyser correctement pour y déceler les mèmes que nous souhaitons considérer. Les travaux dans le domaine de la détection et l’identification de mèmes dans les données de réseaux sociaux restent encore peu nombreux. Une des études pionnières est l’outil *Meme Tracker* (devenu *NIFTY*) conçu en 2009 par le *SNA Project* de l’Université de Stanford (Leskovec *et al.*, 2009). Cet outil permet une étude sous forme de graphes de la diffusion de phrases dans un vaste corpus de texte mais n’est pas adapté à la langue chinoise. La discussion sur la modélisation mathématiques des mèmes (Ahmad et Teredesai, 2006; Nye, 2011) émane souvent de recherches en informatique cherchant à prendre en considération différents facteurs de diffusion lors de l’analyse machine de données (Zubiaga *et al.*, 2011; ?), considérant le même comme un vecteur de modification du réseau lui-même (Ienco *et al.*, 2010). Plus marginal, des études s’intéressent aux dynamiques géographiques des mèmes (Kamath *et al.*, 2013). Néanmoins, aucune de ces différentes approches ne permet d’apporter une réponse technologique ou algorithmique satisfaisante pour l’identification de mèmes Internet dans un ensemble de données issu des réseaux sociaux. Afin de déceler les mèmes dans un vaste ensemble textuel, nous devons y déceler des motifs de diffusion particuliers. La dénomination *machine learning* regroupe un ensemble d’algorithmes qui permettent d’explorer des jeux de données pour en extraire des représentations et y identifier des propriétés (*features*) particulières. Basé sur les sciences statistiques, ces algorithmes font usage de mesures de similarité pour classifier les éléments d’un corpus. Les catégories utilisées pour la classification peuvent être définies au préalable par l’utilisateur - c’est le “*supervised learning*” ou inférées du jeu de données lui-même dans le cas du “*unsupervised learning*” pour la détection de *clusters* (Breiman, 2001). La multitude d’algorithmes de *machine learning* disponibles pour la détection de clusters au sein d’un corpus textuel ou d’un réseau social (Nettleton, 2013; ?) rend l’identification d’une solution difficile. De plus, les algorithmes utilisés traditionnellement pour l’analyse de documents textuels (*topic modeling, LSA*) se heurtent au caractère très disparate des corpus issus des réseaux sociaux (*text sparsity issue*) faisant diminuer drastiquement leur efficacité (Hong et Davison, 2010).

Ferrara & al. (Ferrara *et al.*, 2013) propose dans un papier intitulé “*Meme clustering in social media*” un algorithme utilisant la classification automatique non-supervisée pour détecter des mèmes dans un corpus de données de réseaux sociaux. Ce travail récent propose de tenir compte non seulement des textes et hashtags, mais aussi des liens et des modèles de diffusion pour identifier des groupes de messages intéressants et procéder au *clustering* des mèmes. L’algorithme s’articule autour du concept de “protomèmes”, représentant les éléments fondamentaux d’un même en cours de création (?). Dans le contexte des médias sociaux, le protomème est défini par les entités (liens, hashtags...), mots-clés et séquences de conversation qui constitue le même en devenir. En identifiant puis comparant les différents protomèmes présents dans chaque tweet, il est possible d’y



Graphique 3.3.1 – Algorithme de reconnaissance de mème (clustering) (Ferrara et al., 2013)

déetecter des similarités et de deviner les mèmes en formation.

Cet algorithme suppose donc d'extraire dans un premier temps les éléments remarquables du corpus de tweets afin d'établir des représentations de ces protomèmes contenant les éléments à comparer : *phrases* (texte brut), *mentions* (@, RT), *hashtags* et *urls*. Pour constituer ces jeux de protomèmes, nous utilisons le pattern *map-reduce* qui permet de chercher et lister des éléments dans un vaste jeu de données (*map*) avant de les regrouper dans une liste (*reduce*). Une fois ces protomèmes constitués nous procédons à leurs comparaisons selon plusieurs critères :

- *Similarité de texte* : comparant le contenu texuel de chaque protomème
- *Similarité d'utilisateurs* : comparant les utilisateurs contenus dans chaque protomème
- *Similarités de tweet* : recherchant les tweets identiques dans différents protomèmes
- *Similarité de diffusion* : considérant les références aux utilisateurs contenus dans chaque protomème

Ici nous utilisons la *sémantique vectorielle* (Support Vector Machine) afin de comparer les éléments des protomèmes grâce à une représentation algébrique sous formes de vecteurs. Pratique ancienne de l'algèbre linéaire appliquée à la science informatique (Salton et al., 1975), il s'agit de permettre la conversion d'objets textuels (mots, identifiants, images...) sous une forme aisément comparable. Pour convertir le texte sous forme vectorielle, l'algorithme classique Tf-idf (*Term Frequency - Inverse Document Frequency*) est utilisé (Soucy et Mineau, 2005). Les autres mesures de similarité sont la *mesure cosine* (ou *similarité cosinus*) des protomèmes convertis sous forme de vecteurs binaires. Une fois ces différentes valeurs de similarité calculées, nous utilisons les scalaires définis dans le papier de référence pour assigner des poids à chacun des vecteurs et les combiner en une seule valeur (Ferrara et al., 2013). Cette matrice de valeurs de similarité nous permet alors de définir les protomèmes les plus similaires et d'identifier ainsi des *clusters* dans les données correspondant aux mèmes.

Si cette approche offre des résultats probants sur de petits volumes (quelques centaines de tweets), la très grande demande en puissance de calcul et ressources mémoires nécessaires rendent le traitement d'un jeu données plus vaste irréalisable. Les opérations de comparaison et le calcul de similarités sur de vastes volumes de données font croître très rapidement la complexité des algorithmes et ainsi la quantité de calculs à effectuer. Le calcul du coût d'un algorithme se fait au travers des notions dites de domination, avec

notamment le “grand O” exprimé $O(f(n))$ qui fait correspondre à la complexité d’un algorithme une fonction f de la quantité d’information manipulée n . Ainsi pour un algorithme courant de complexité $O(n^2)$, les ressources de calcul (*computation*) et de mémoire (RAM ou stockage) nécessaires augmentent de manière exponentielle à chaque élément ajouté au corpus n . L’algorithme de “meme clustering” décrit par Ferrara atteint donc un coût exorbitant devant un large volume de données comme celui du jeu Weiboscope. La limite physique de calcul est rapidement atteinte rendant impossible le franchissement du palier expérimental et la vérification des hypothèses de travail à une échelle suffisante.

3.3.3 L’identification des mèmes par hashtags

Les travaux en sciences humaines sur les mèmes (Bauckhage, 2011; Coscia, 2013; Knobel et Lankshear, 2007) et les sites spécialisés (Buchel, 2012; ?) présente bien souvent un même sous la forme d’un titre (mot ou groupe de mots) accompagné d’une collection d’images ou/et de vidéos, ainsi qu’un texte explicatif (mise en contexte) et des indications sur les dates de parution et son évolution dans le temps. Ici, un ou des auteurs se saisissent des matériaux bruts du web et les rassemblent pour reconstituer une vision particulière du même, aussi représentative que possible. Cette approche nécessite peu de déploiement technologique (un copier/coller suffit souvent), mais exige néanmoins une forte connaissance et un suivi régulier du terrain Internet pour y dénicher ces représentations du mèmes. Cette démarche que nous nommerons *ethnographique* s’apparente à de la curation de contenu (Buckingham *et al.*, 2006). Les *hashtags* (en français “mots-dièses”) sont utilisés dans l’écriture sur les réseaux sociaux et se présentant dans Sina Weibo sous la forme d’un mot entouré de deux dièses - ex. *#mot-clé#*. Marqueur particulier, le hashtag permet à un interlocuteur de procéder à une dénotation ou connotation du message original (Romero *et al.*, 2011) ou d’affirmer son caractère évènementiel (lors d’un évènement sportif, d’une conférence, etc). Facile à identifier dans la masse des données en ligne, il permet de désigner un ensemble de contenus sous un même signe. Ainsi, il est un vecteur important permettant de collecter simplement une large masse d’informations autour d’un même. La constitution d’un corpus autour d’un “hashtag” présente néanmoins plusieurs limites. Premièrement, le même est par définition un objet en mutation. Ainsi, il est souvent difficile de l’identifier une bonne fois pour toute par un ou plusieurs de mot-clés. De plus, le même existe bien souvent sous la forme d’images ou de vidéos qui ne sont pas nécessairement légendées ou référencées et donc peu accessible à une recherche “plein texte”. Ainsi, une approche pour la recherche de mèmes ne peut être entièrement textuelle et doit s’intéresser aux autres forme de contenus web (notamment les liens et les hashtags). De plus, l’ajout de hashtags dans les messages est un acte volontaire non systématique. Ainsi, si l’identification de certains mèmes peut se faire grâce à la recherche de hashtags, l’ensemble des messages contenant des hashtags ne recouvrent pas systématiquement un même. Comme nous le verrons, les hashtags sont bien souvent de simples artefacts de campagne marketing en ligne.

Afin de procéder à l'analyse des mèmes, nous avons donc indexé l'ensemble des contenus du corpus Weiboscope contenant des hashtags sur toute l'année 2012 (30 millions sur un total de 200 millions tweets environ). Nous avons donc dans un premier temps extrait l'ensemble des messages contenant un ou des hashtags de l'ensemble des données avant de les classifier pour obtenir des jeux de données cohérents par hashtags. Une première contrainte ici est propre au traitement informatique de la langue chinoise qui constitue le langage majoritaire de notre corpus. L'absence d'espace entre les différents caractères de la phrase en chinois oblige à prendre en compte la sémantique de la phrase pour procéder à sa segmentation. Cette question usuelle pour les usagers du NLP (*Natural Language Processing*) en chinois est un sujet de recherche encore très discuté (Qiu *et al.*, 2013). L'identification de la meilleure option parmi la multitude des librairies et logiciels disponibles pour la segmentation puis le NLP en chinois a été un aspect préliminaire importants de notre travail d'analyse de données. Nous avons ainsi réalisé de nombreux benchmark afin de comparer les différentes technologies existantes. Un travail mené avec l'équipe du site Internet d'actualité scientifique *Guokr* à Pékin nous a amené à choisir en premier lieu un algorithme de segmentation développé sur la base de textes en chinois ancien⁷. Après plusieurs tentatives, nous avons préféré le segmenteur open-source Jieba⁸ plus généreux en termes de fonctionnalités et permettant de meilleures performances. Une fois la phrase chinoise segmentée, la seconde étape consiste à supprimer tous les mots les plus communs (*stopwords*), ainsi que la ponctuation et d'autres caractères trop courants afin de préserver uniquement les mots riches de sens dans les phrases (les mots-clés). L'extraction des hashtags est effectuée grâce à une expression régulière qui scanne le texte pour identifier et préserver uniquement les caractères situés entre les deux signes dièse (#). Sur un total d'environ 30 millions de tweets contenant des hashtags, nous avons choisi de retenir seulement les hashtags possédant plus de 1000 messages. Nous avons ensuite décidé d'ignorer les 1000 hashtags les plus utilisés car ils ne présentaient pas d'intérêt pour notre étude, étant la plupart du temps des noms de marque ou des mots-clés trop généraux (par exemple : "bonne nuit", "nouvelles de sports", "photos de nourriture", etc.)

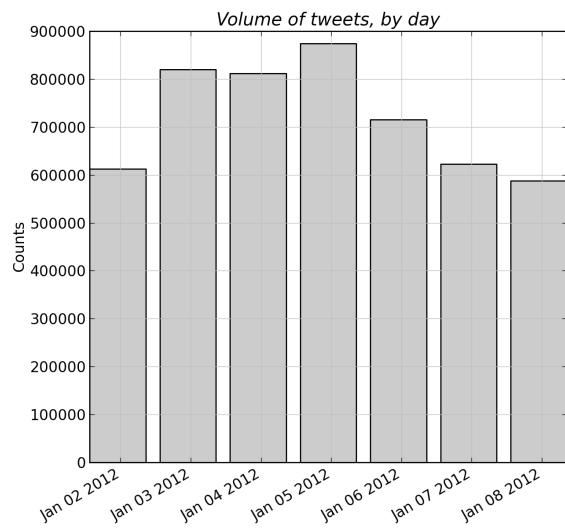
Notre étude vise à étudier les dynamiques conversationnelles et nous devons donc déterminer les plus adéquats parmi des hashtags de nature souvent très différentes. Pour ce faire, nous avons sélectionné pour chaque jeu de données (chaque hashtag) deux mesures significatives : premièrement, le volume de messages ; deuxièmement, la quantité d'échanges et d'interactions effectives entre les utilisateurs (commentaires, retweets, etc.). Ces deux mesures nous permettent de nous assurer que 1) nous possédons une quantité suffisante de messages pour mener à bien l'étude et que 2) la discussion a bien eu lieu et qu'il ne s'agit pas de messages redondants ou non reliés entre eux. Nous avons choisi d'ignorer les échanges dominés à plus de 80% par le même utilisateur pour éviter

7. gkSeg : "Yet another Chinese word segmentation package based on character-based tagging heuristics and CRF algorithm" <https://github.com/guokr/gkseg>, consulté le 14 Mars à 18 :28

8. <https://github.com/fxsjy/jieba>, consulté le 14 Mars à 18 :30

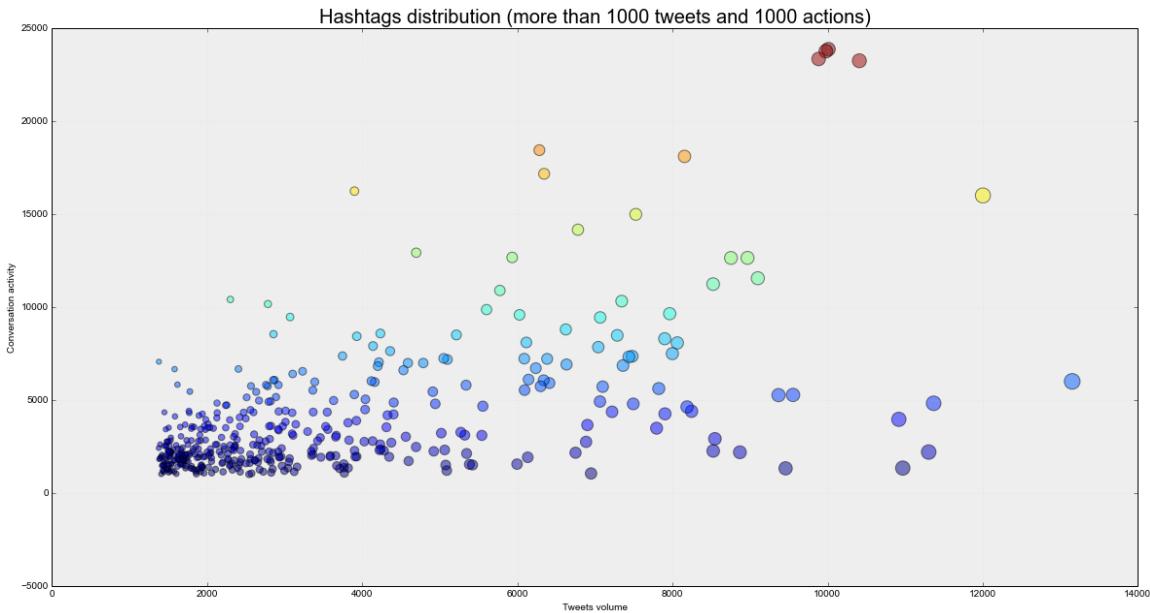
hashtags	users	actions	tweets
吴奇隆 201	13243	22349	
一起到老 182	0	364	
春运 92	13	256	
轻松一刻 92	11	240	
人品值分析 90	490	321	
朝阳区 88	49	165	
理性态小度 87	0	329	
美图GIF 87	101	404	
我正在听 86	6	330	
微盘签到 84	304	305	
2012来了 83	206	309	
中级达人 83	0	159	
分享 82	87	563	
星座 82	5	195	
极速互联随我行 81	317	701	
微信 80	11	288	
微博达人 80	9	364	
2012 79	37	270	
那些年 79	106	217	
晚安心语 78	27	150	
周杰伦 77	130	286	
多多cosplay 15	414	463	

(a) Hashtags les plus utilisés pendant la 1ère semaine de 2012



(b) Volume de tweets pendant la semaine

Graphique 3.3.2 – Les résultats concernant la 1ère semaine de l'année 2012 donnent un aperçu du volume analysé : 5,044,331 tweets, 398 392 utilisateurs uniques cités (dans un total de 2 115 544 mentions), 264 651 urls uniques (pour un total de 426 914) et 44 382 hashtags uniques (pour un total de 244285).



Graphique 3.3.3 – Distribution des 429 hashtags sélectionnés

la pollution de l'étude par l'activité de robots. Le graphe ci-dessous permet d'observer la distribution de 429 hashtags possédant tous plus de 1000 tweets et 1000 échanges : l'axe vertical représente la quantité d'actions (échanges) et l'axe horizontal le volume des conversations. Tout en bas du graphe se trouvent donc les hashtags ayant été les moins discutés, avec en haut ceux aux conversations les plus intenses. La taille des points illustre le volume de messages et la couleur la quantité de conversations.

En procédant à l'étiquetage des hashtags les plus actifs durant l'année 2012 sur Sina Weibo (figure 2 ci-dessus), nous constatons que la plupart sont associés à des activités commerciales, de loisirs ou de divertissement. Ici nous observons que les usages majoritaires du réseau social Sina Weibo correspondent pour la plupart à ceux d'autres mass-médias plus traditionnels de par le monde. Le commerce en ligne occupe notamment une place proéminente. La marque de téléphonie mobile chinoise *Xiaomi* est abondamment citée, reflétant son importance croissante dans le marché chinois et surtout sa stratégie commerciale qui cible abondamment les réseaux sociaux avec de nombreux hashtags très discutés (notamment “*Fans de Xiaomi*” [5C0F ?][7C73 ?][7C89 ?][4E1D ?]). Egalement, de nombreuses campagnes promotionnelles d’ouverture ou d’anniversaire de magasins ont réussi à se hisser dans le jeu de tête des hashtags les plus discutés. Radio-crochets ou chanteurs reconnus, les stars de la télévision et de la chanson sont aussi présents dans le peloton de têtes des discussions sur Sina Weibo. Le célèbre chanteur Han Geng notamment compte près d'une dizaine de hashtags le concernant parmi les 500 les plus discutés (“*Han Geng fait une pub pour Nokia*”, “*Han Geng va en Italie*”, “*Han Geng fait une pub Pepsi*”, “*Han Geng refuse une interview*”, etc.) Ici encore, le réseau social agit comme le prolongement des mass média traditionnels, élément-clé des nouvelles stratégies de publicités en ligne, parfois particulièrement agressives comme dans le cas de Han Geng. Les

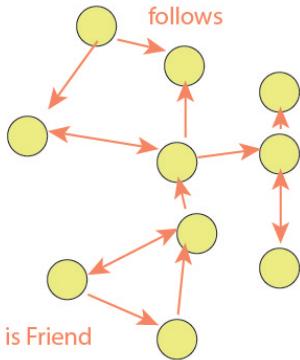
contenus de la télévision sont largement relayés et discutés, notamment les séries télévisuelles. Le cinéma est aussi représenté. Le film comique chinois *Lost in Thailand* sorti en Décembre 2012 dépeint les aventures d'un chinois en vacances en Thaïlande. Premier grand succès commercial du box-office chinois, sa popularité se reflète dans l'importance au sein des discussions en ligne. Les tendances des ventes du livre sont reflétées par de nombreux best-seller sur "l'amélioration de soi" ou la "réussite économique". Ce type de hashtags ne se limite pas au support web mais s'origine directement dans d'autres médias plus traditionnels. Le gouvernement lui-même utilise Sina Weibo pour faire passer ses messages avec un hashtag "*information officielle*" utilisé notamment pour des démentis publics ou droit de réponse par l'entreprise Sina, propriétaire du service. Également outil de conversation, les discussions sur les réseaux sociaux parlent de la vie de tous les jours. La situation routière et les bouchons dans chaque ville sont un des grands sujets de discussions. Ce sont dans ces échanges quotidiens que se cristallisent plus particulièrement les enjeux politiques et médiatiques des réseaux sociaux. Nouveau café du commerce, les commentaires sur les faits divers et l'actualité mettent souvent à jour les dysfonctionnements de systèmes politiques, urbains ou légaux. Il est intéressant néanmoins de noter que parmi les hashtags les plus discutés, les phénomènes de suppression de contenus par les administrateurs (censure) restent très marginal. Le *China Digital Times* de UC Berkeley maintient une liste des mots interdits sur Sina Weibo depuis plusieurs années (Ng, 2013). En comparant cette liste de mots censurés à celle des hashtags, nous avons pu voir qu'aucun des 3000 hashtags les plus utilisés en 2012 n'a été soumis à une interdiction même temporaire sur Sina Weibo. Les hashtags les plus sujets à la censure ne sont pas en lien avec des domaines politiques ou des sujets sensibles, mais plutôt avec des contenus à caractère pornographique (la pornographie est interdite en Chine). Reflétant les usages majoritaires (commerce, loisirs, etc.), les hashtags véhiculent des contenus souvent moins controversés et les "mots censurés" sont plus à même d'apparaître dans des discussions informelles.

3.3.4 Visualisation du graphe conversationnel d'utilisateurs

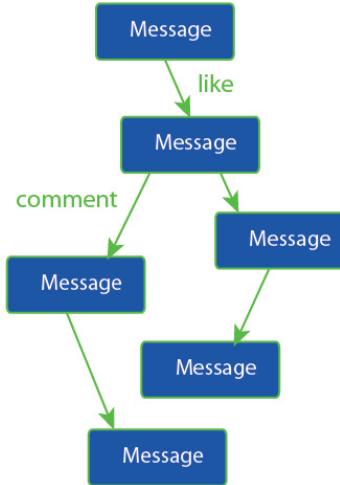
Après avoir identifié différents *mèmes parmi* les hashtags les plus discutés de l'année 2012, nous allons maintenant procéder à l'analyse du déroulement des échanges d'après chacun des différents corpus de données. Plusieurs types de graphes peuvent être extraits :

- **Graphe social statique** : représentant les relations pré-existantes dans la structure du réseau étudié (untel est ami avec untel, untel suit untel, etc.)
- **Graphe conversationnel** : représentant toutes les interactions qui entourent et structurent la diffusion de messages
- **Graphe de diffusion** : représentant les interactions qui se sont produites entre les acteurs durant la diffusion du message. Ce dernier type de graphe est un recouplement des deux autres.

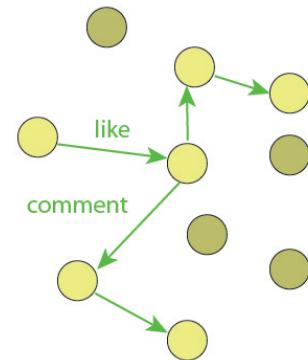
[Warning : Image ignored]



1. Static Social Graph



2. Conversation Graph

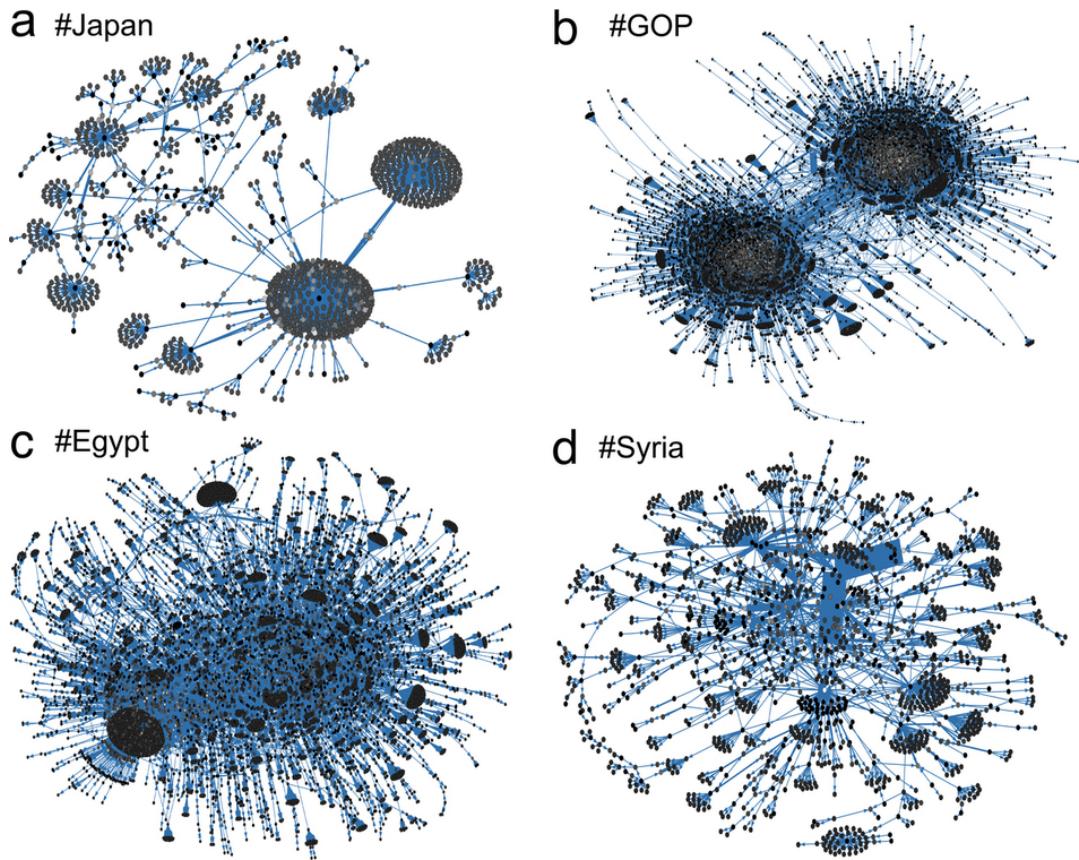


3. Spread Graph

Graphique 3.3.4 – Les 3 types de graphe classiquement extraits des données de réseaux sociaux

Dans notre étude, le graphe social statique ne présente pas particulièrement d'intérêt puisqu'il correspond à un ensemble de relations peu affecté par les discussions. De plus, nous ne disposons dans le jeu de données Weiboscope que de son état final qui ne témoigne pas de l'évolution des relations. Nous voulons obtenir ici les graphes de diffusion sous la forme de conversations structurées (graphe directionnelle des réponses et commentaires) de l'ensemble des messages. Dans un article paru dans *Nature* (Weng *et al.*, 2012), les chercheurs du *Centre de Recherche sur les Systèmes Complexes* de l'Université d'Indiana identifient les caractéristiques des mèmes connaissant le plus de succès (la plus large diffusion). Un travail de visualisation de mèmes identifiés par des hashtags sur Twitter leur permet notamment de mettre à jour des motifs particuliers dans la structure des conversations.

Afin de mettre à jour le graphe conversationnel entourant les hashtags sélectionnés sur Sina Weibo, nous avons choisi d'extraire la séquence d'interactions des messages (mentions, retweets) composant chaque même. Cette structure de graphe nous permet de représenter la diffusion de chaque même sous forme d'un graphe contenant un node par utilisateur et un ensemble de relations correspondant aux échanges visibles dans les textes des messages. Dans un premier temps, le logiciel *Graphviz* nous a permis d'obtenir une représentation basique du graphe conversationnel afin d'avoir un aperçu sur la nature des conversations par l'observation des motifs qui la compose. En effet, si les deux mesures identifiées à l'étape précédente (volume de messages et volume d'échanges) nous permettent d'effectuer un premier tri parmi les hashtags, cette première visualisation nous permet de considérer la nature des échanges et l'implication des utilisateurs d'après la structure des motifs conversationnels. Chaque utilisateur est symbolisé par un point et chaque message par un trait reliant deux utilisateurs. Un motif très compact reflète

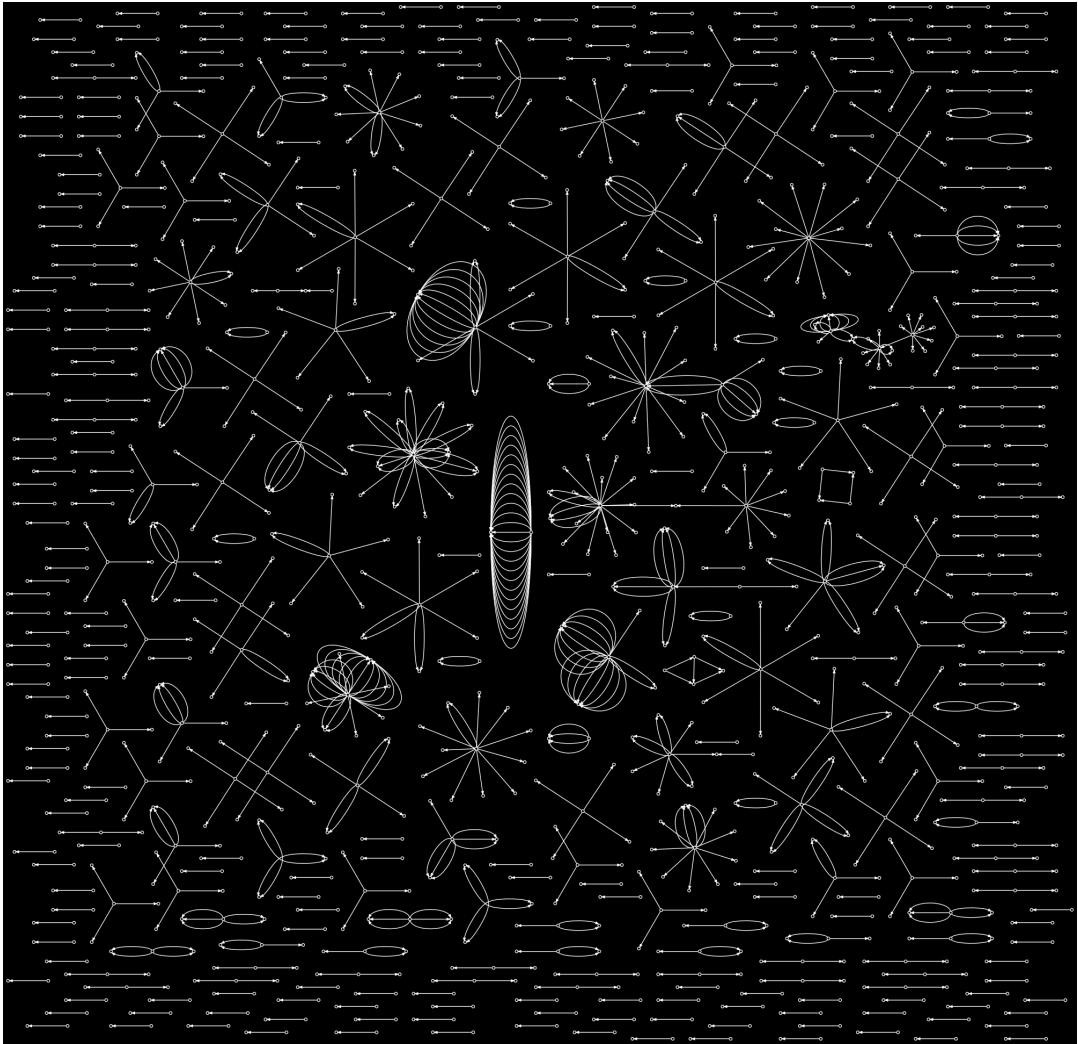


Nodes represent Twitter users, and directed edges represent retweeted posts that carry the meme. The brightness of a node indicates the activity (number of retweets) of a user, and the weight of an edge reflects the number of retweets between two users.

- (a) The *#Japan* meme shows how news about the March 2011 earthquake propagated.
- (b) The *#GOP* tag stands for the US Republican Party and as many political memes, displays a strong polarization between people with opposing views.

Memes related to the Arab Spring and in particular the 2011 uprisings in (c) *#Egypt* and (d) *#Syria* display characteristic hub users and strong connections, respectively.

Graphique 3.3.5 – Graphe de diffusion de hashtags sur Twitter d'après (Weng et al., 2012)

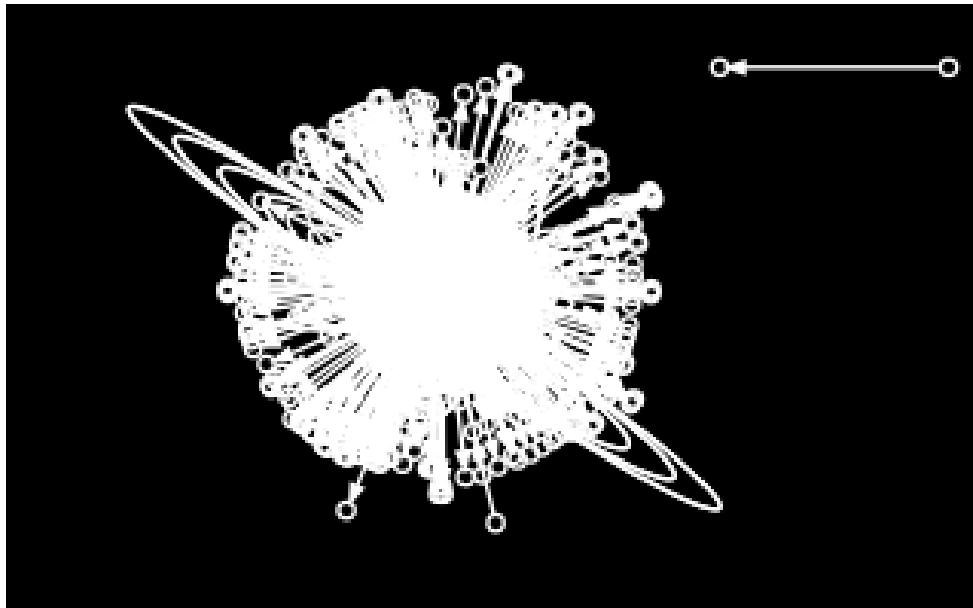


Graphique 3.3.6 – Fig. Visualisation simple du hashtag “WeicoPlus”, un trait représente un échange entre deux utilisateurs

une conversation animée entre des utilisateurs peu nombreux échangeant beaucoup. A l’inverse, un motif disparate reflète des échanges plus brefs et morcelés.

WeicoPlus est une application mobile permettant d’utiliser Sina Weibo. Le hashtag #WeicoPlus# est ajouté automatiquement quand les utilisateurs postent des photos depuis ce service. Ainsi, on remarque que le graphe conversationnel entourant WeicoPlus se compose essentiellement de messages simples, mais ne donne pas lieu à une conversation structurée - à l’exception de quelques rapides échanges entre un nombre réduit de personnes.

A l’inverse, le hashtag “Veuve d’enfant unique”[FF08 ?]#[5931 ?][72EC ?][6BCD ?][4EB2 ?]#[FF09 ?]c le débat en une forme très dense qui reflète une surenchère de commentaires et d’actions autour du hashtag, propre d’une conversation animée.



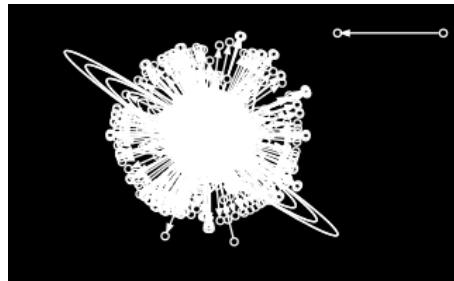
Graphique 3.3.7 – Visualisation simple des conversations autour du hashtag Veuve d'enfant unique

3.3.5 Premiers éléments d'analyse : visualisation de mèmes

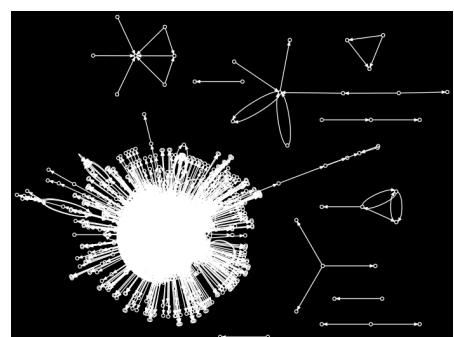
Les différents modèles de conversation que nous obtenons dans cette première étape se présentent sous une forme schématique et peu détaillée. Afin de comprendre plus en détails les dynamiques conversationnelles qui les entourent, nous avons choisi de sélectionner trois exemples parlants de hashtags dont les graphes conversationnels présentent des particularités des modèles de diffusion dissemblables et organisés. Pour chacun d'eux, nous allons procéder à une analyse plus détaillée des graphes conversationnels afin de considérer les différences entre ces différents modèles.

Nous voyons que les trois représentations des graphes ci-dessus donnent à voir des structures plus ou moins morcelées, avec un ensemble de points et de lignes très compactes qui représentent la majeure partie de la conversation. Afin de visualiser plus précisément les groupes et les dynamiques qui constituent les discussions autour de chaque hashtag, nous allons utiliser le logiciel Gephi (?) afin d'examiner de plus près la composition de ces graphes. Pour ce faire, Gephi va nous permettre d' “étaler” le graphe en repositionnant les nodes et en les coloriant pour en identifier les composantes et les tendances.

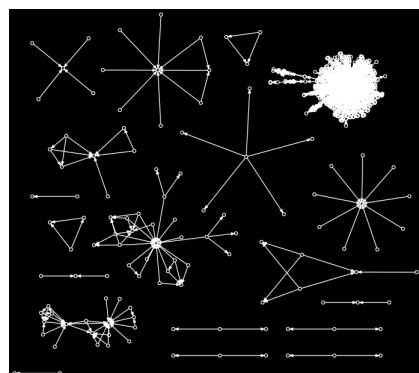
Chaque utilisateur est représenté sous la forme d'un point. La taille des points correspond à l'importance de l'utilisateur dans le réseau total des échanges, caractérisé par son degré de centralité intermédiaire (*betweenness centrality*), une mesure topologique correspondant au nombre de plus courts chemins du graphe passant par cet utilisateur. La couleur est utilisée pour représenter la *modularité* du réseau, c'est à dire le nombre de communautés engagées dans la conversation définies comme les cliques d'utilisateurs constituant plus de 1% du réseau total d'échange (Blondel *et al.*, 2008). La position des nodes est calculée grâce à l'algorithme *Force Atlas 2* (?) utilisant une modélisation phy-



(a) Pluie torrentielle à Tianjing



(b) Veuve à l'enfant unique



(c) Abolition des lois sur la prostitution

Graphique 3.3.8 – Visualisation des réseaux de conversation



Graphique 3.3.9 – Downpour bypasses Beijing, batters neighbor, in Qinghua News le 2012-07-26 13 :29 :59, http://news.xinhuanet.com/english/china/2012-07/26/c_131740415.htm consulté le 27 Juin 2014.

sique où les nodes peu connectés entre eux se repoussent et ceux très connectés s'attirent. Ainsi, la proximité de deux nodes sur le graphe témoigne d'une proximité lors des conversations, c'est à dire de l'existence d'un échange entre eux (citations, commentaires ou retweets). Pour davantage de visibilité, certaines conversations sub-alternes représentant moins de 1% du total ont été effacées. Egalement, les nodes possédant un degré inférieur à 3 (moins d'un échange avec au moins trois autres nodes du graph) ne sont pas représentées.

Exemple 1 : Pluie torrentielle à Tianjing

Le premier même choisi parle d'une catastrophe naturelle, sous la forme d'une pluie diluvienne qui s'est abattue sur la ville de Tianjin durant la nuit du 21 au 22 Juillet 2012.

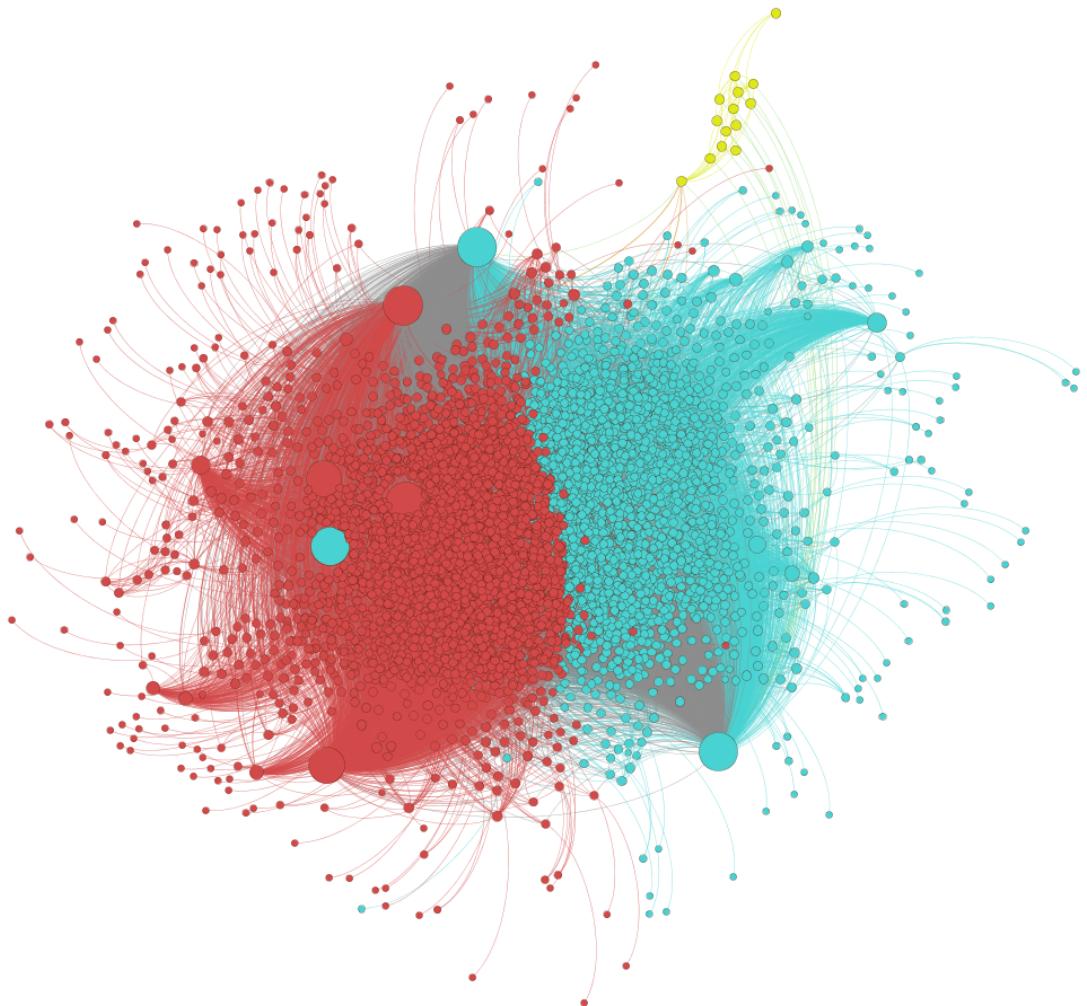
Ici 4 groupes composent 85% du graphe, constitués autour de gros diffuseurs (les nodes les plus gros). Plusieurs groupes semblent s'emparer de la conversation mais on voit peu d'activité entre les nodes alors que les échanges se déroulent autour de quelques utilisateurs très centraux. Cela traduit le fait que peu de personnes ont réellement discuté l'information et elles se sont simplement contentées de la relayer.

La diffusion d'un fait divers très local (il se passe à Tianjing) est entraînée par peu de sources très importantes (les quelques nodes de grande taille), vraisemblablement des journaux et médias locaux qui annoncent la nouvelle (presse, photos-choc d'innondations, etc). La conversation est peu active et très structurée, nous sommes en présence d'un modèle classique de diffusion de masse.

Exemple 2 : Veuve de l'enfant unique



Graphique 3.3.10 – Exemple 1 : Tianjin Baoyu



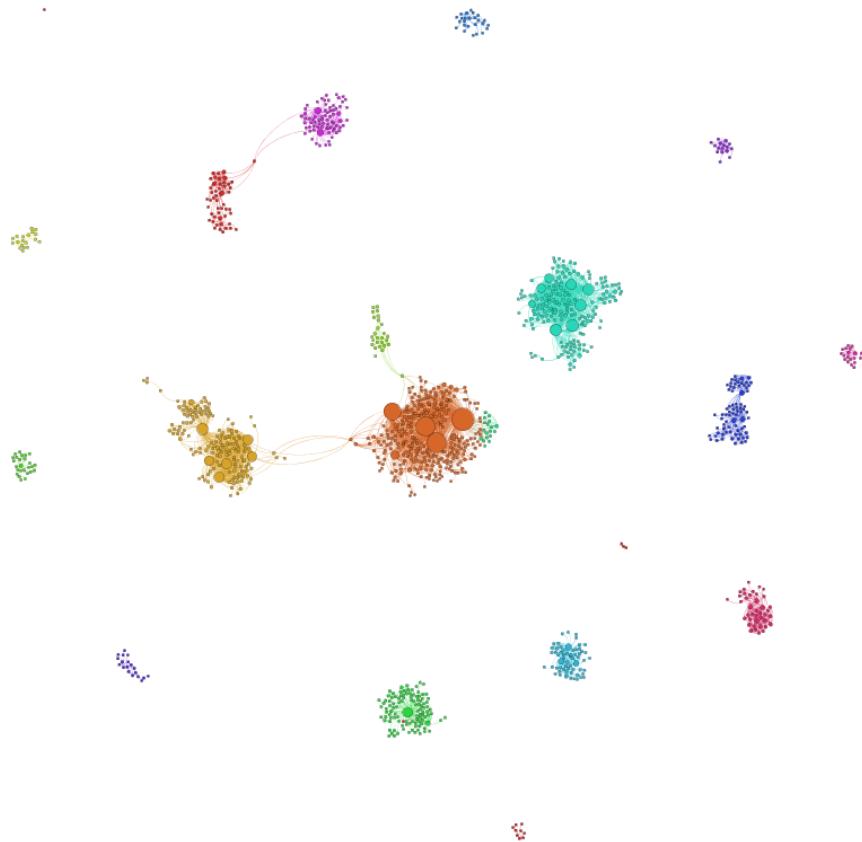
Graphique 3.3.11 – “Shidu Muqin”

Un autre sujet discuté par une très large quantité de personnes apparaît sous le terme "*shidu muqin*", forme contractée signifiant "*mère qui a perdu son enfant unique*" ([5931 ?][53BB ?][72EC ?][751F ?][5B50 ?][5973 ?][7684 ?][6BCD ?][4EB2 ?]). Ce hashtag désigne un phénomène de société bien connu en Chine où le deuil de la perte d'un enfant se double souvent pour une mère chinoise seule de l'absence de ressources pour vivre. En effet, l'absence de système de retraite fait porter aux enfants la responsabilité de la survie de la famille.

Le graphe ici est fait de deux grands groupes composant à eux deux près de 95% du graphe total. Les discussions sont très polarisées et menées par peu de participants (les nodes les plus gros sur le graphe). Peu de personnes très influentes concentrent les discussions autour d'eux, accompagnés ou suivis d'une foule de commentateurs.

Cet exemple donne à voir des groupes bien définis et très proches où plusieurs acteurs majeurs mènent la discussion. Les dynamiques d'échanges autour d'une question de société

(la loi de l'enfant unique en Chine et ses conséquences) s'articule en groupes distincts sans pour autant amener à des controverses importantes (qui se traduirait par des discussions longues et houleuses). Ici, les leaders d'opinion font la discussion et la diffusion se fait au travers d'eux.



Graphique 3.3.12 – “Piaowudong nuzui”

Exemple 3 : Abolition des lois sur la prostitution

Le hashtag “*Abolissons la loi piaowudong nuzui*” est l’expression d’une campagne pour l’abolition d’une législation scandaleuse sur la prostitution en Chine. Depuis les années 80, la loi chinoise interdit la prostitution et prévoit la condamnation des deux parties qui s’adonnent à un échange d’argent. Baptisé “*Piaowudong nuzui*”, cette loi a vu plusieurs cas absurdes impliquant des viols organisés sur mineurs se solder par la condamnation et l’emprisonnement des enfants incriminés. Relayés par les journalistes, les scandales à répétition ont éclaté à plusieurs reprises, impliquant parfois des officiels du Parti souvent blanchis alors que des enfants étaient eux emprisonnés. Le graphe des discussions autour de l’abolition de cette loi montre que de nombreux groupes discutent séparément de cette question puisque les premiers 50% du graphe sont déjà constitués de plus d’une quinzaine de clusters. Les groupes sont très éloignés entre eux, n’entretenant que peu de relations et connaissant une activité intense.

Cet exemple présente les caractéristiques d'une conversation très décentralisée dans laquelle de nombreux acteurs différents prennent part. L'émergence de ce type de discussion fragmentée témoigne d'un usage particulier de la discussion sur le réseau social et montre comment la discussion autour d'un sujet peut s'étendre sans afficher de relations directes dans le média lui-même. Ici un agent extérieur (en l'occurrence un article du journal *Nanfang Zhoumo* sur le sujet) fait naître la conversation sans pour autant l'accaparer et la centraliser. Plus difficilement détectable et contrôlable, cette dernière configuration est typique du même car elle se développe de façon large et durable entre des groupes à l'origine peu connectés.

Cette première visualisation des graphes conversationnels nous permet d'explorer quelques types précis d'échanges et d'en proposer une première lecture. Néanmoins, le procédé de visualisation reste rudimentaire et soulève plusieurs questions que nous nous donnons pour tâche de continuer à explorer. Premièrement, dans quel espace a lieu cette représentation ? En étalant ainsi ces graphes conversationnels, quelle action réalisons-nous réellement et quelle en est la valeur pour l'analyse ? À plus forte raison, quelle est la relation de cette espace du graphe conversationnel avec les autres formes d'espace, et plus notamment l'espace du réel géographique et l'espace de la représentation par le langage ?

3.4 Méthodologie de traitement et de visualisation des mèmes

L'étude des relations entre ces différents types d'espaces implique donc une méthodologie renouvelée et le développement d'outils adaptés. Il ne s'agit plus simplement d'étudier un réseau, mais plus précisément de s'interroger sur les relations entre différents réseaux, de nature souvent différentes. Au-delà du réseau multi-calques, nous sommes en présence de multiples réseaux possédant des relations communes.

Nous avons donc sélectionné trois aspects importants des mèmes que nous allons tenter de représenter au mieux afin d'en comprendre l'existence :

- ***langagier*** : le champ sémantique d'un mème est constitué des mots qui sont prononcés lors de sa diffusion. L'association de mots -souvent sous la forme du jeu de mots - est un des propres du mème et constitue ainsi une part importante de son existence. Ainsi, le mème produit à proprement parler des réseaux de mots en dessinant des liens entre des signifiants souvent improbables qui en font souvent le succès (Bauckhage, 2011).

- ***conversationnel*** : au-delà des mots, un mème se constitue sous la forme d'un échange, une conversation où les différents acteurs discutent, commentent et se saisissent des actions disponibles sur la plateforme web (like, retweets, etc.) pour converser. Comme nous l'avons vu précédemment, nous pouvons identifier et considérer un graphe conversationnel créé par le mème en se diffusant.

- ***réel*** : au-delà des échanges en ligne, ces discussions possèdent une existence physique, premièrement sous la forme de l'activité électrique des machines qui sont utilisées

lors de ce processus. Néanmoins, dans l'approche d'une géographie humaine des échanges numériques, nous considérerons ici l'existence physique des mèmes par celle des utilisateurs - de leurs corps - et non pas des machines.

Afin d'étudier chacun de ces aspects du même, nous allons donc procéder à la collection et la visualisation de données sur chacun de ces niveaux d'après un corpus de mèmes sélectionnés.

3.4.1 Sélection de mèmes pour l'étude

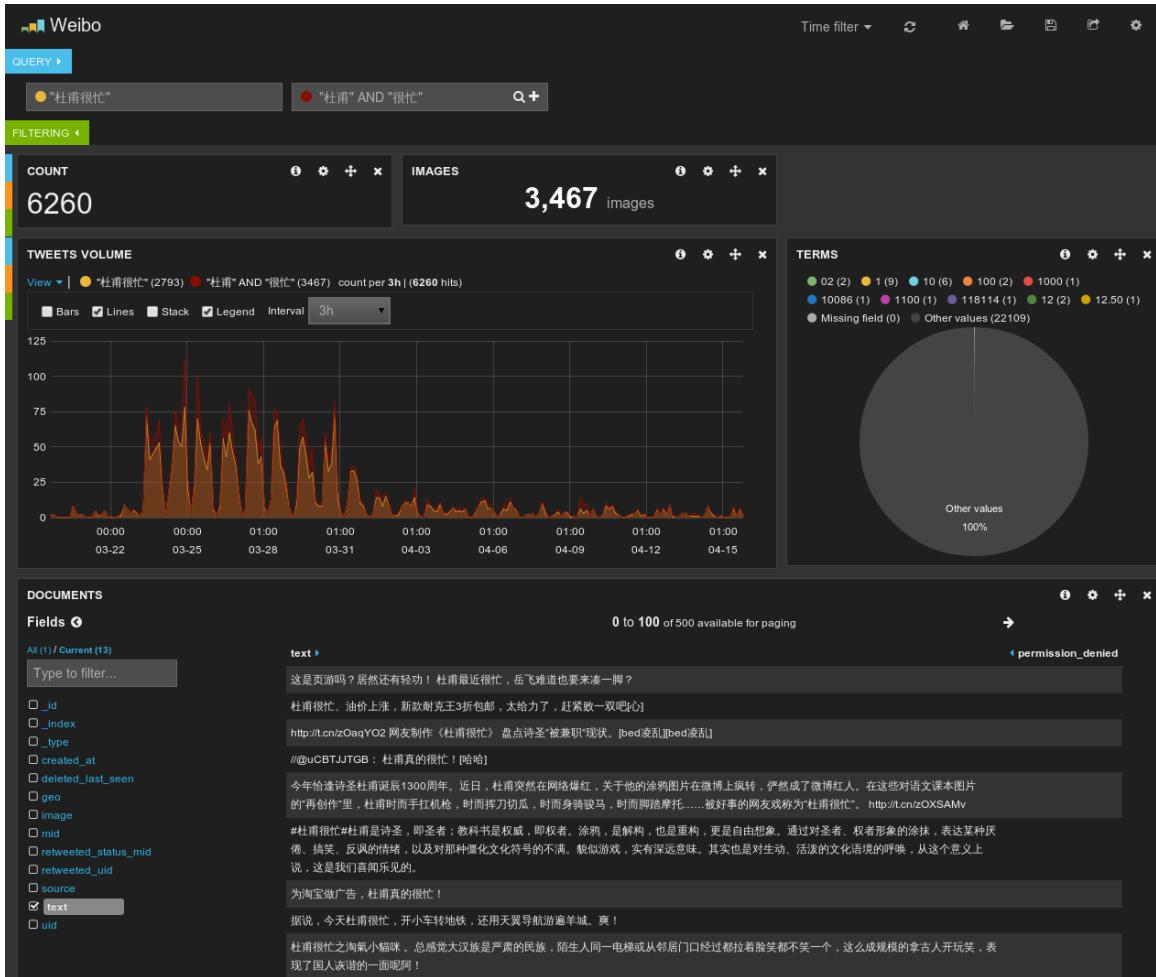
Choisir un ensemble de mèmes cohérents est une des étapes difficiles de notre recherche. En effet, dans le vaste corpus de données utilisées et plus généralement dans la multitude des échanges quotidiens sur les réseaux sociaux, trouver une prise pour l'étude n'est pas une tâche évidente. Afin de procéder à la sélection de mèmes, nous avons donc choisi d'utiliser la typologie des catégories de mèmes identifiées dans la littérature (voir partie 2) et d'en systématiser l'usage sur notre corpus. Ainsi, pour chaque catégorie, nous avons effectué une recherche mêlant des sources parlant des mèmes et de l'Internet chinois (journaux, blogs, encyclopédies en ligne, sites ressources), notre propre expérience du web chinois et notre corpus de données disponibles.

Dans un premier temps, nous avons donc identifié pour chaque catégorie de même plusieurs événements web de l'année 2012 sur Sina Weibo comme autant de candidats pour représenter l'ensemble de notre typologie. La classification de l'importance des événements web revêt une nature très différente selon les différentes sources. Une revue de la littérature web sur les "événements marquant du web social en 2012" montre que les contenus à caractère politique et polémique sont considérés comme plus importants sur les sites à audience majoritairement occidentale⁹ alors que les sites plus spécialisés sur la Chine¹⁰ prennent davantage en considération les phénomènes médiatiques commerciaux. Encore une fois, nous voyons comment les réseaux sociaux chinois sont représentés comme des phénomènes politiques, parfois au détriment de leur existence comme média à part entière. Après avoir comparé ces sources diverses, il nous fallait être sûr que nous disposions d'une quantité suffisante de données pour traiter le même choisi. Nous avons donc indexé les parties du corpus représentatives des différents événements afin de pouvoir effectuer des recherches plein-texte et ainsi vérifier le volume et la qualité des corpus mobilisables pour chaque même. Pour ce faire, nous avons mis en place un outil d'indexation et de recherche¹¹ qui nous permet de contrôler les différents paramètres et nous assurer de la

9. Global Voices, <http://globalvoicesonline.org/2012/12/07/top-10-chinese-internet-memes-of-2012/>, consulté le 22 Avril 2014 à 12 :10 ou WSJ <http://blogs.wsj.com/chinarealtime/2012/12/19/the-top-10-chinese-internet-memes-of-2012/>, consulté le 22 Avril 2014 à 12 :10

10. Danwei <http://www.danwei.com/chinas-hottest-styles-of-2012/> consulté le 22 Avril 2014 à 12 :12

11. Les technologies utilisées pour indexer le corpus sont *ElasticSearch* pour le moteur de recherche et



Graphique 3.4.1 – Ce tableau de bord permet de comparer la qualité de différentes requêtes dans le corpus. Capture d'écran réalisée le 23 Mars 2014 à 16h18

quantité (taille significative), des dimensions (dates correctes) et de la qualité du corpus (vérification du contenu d'un échantillon de 100 messages sélectionnés aléatoirement).

Cette étape nous a également permis d'identifier les mots-clés les plus appropriés pour définir chaque mème. En effet, la qualité des archives pour chaque mème dépend fortement de la méthode utilisée pour collecter les données. Ici, nous utilisons la recherche plein texte pour collecter les données et nous devons ainsi nous assurer que notre requête est bien construite, afin de minimiser le bruit dans chaque corpus (éviter les messages contenant les même mots-clés mais sans relations avec le mème).

Une fois assuré qu'un mème pouvait être correctement représenté dans un corpus de données, nous obtenons donc une liste réduite de mèmes à étudier pour chaque catégorie.

3.4.2 Extraction des graphes et traitement des données brutes

Une fois cette requête clairement identifiée nous procédons pour chaque mème à l'extraction d'un jeu de données contenant l'ensemble des messages correspondant à la

Kibana pour le tableau de bord. <http://www.elasticsearch.org> consulté le 22 Avril 2014 à 12 :23

requête définie. Ce jeu de données est ensuite complété par l'ensemble des messages mentionnés mais ne contenant pas le mot clé (commentaires, réponses, etc.) afin d'obtenir un ensemble de messages représentatifs pour chaque mème.

Nous procédons ensuite au traitement de chaque corpus selon une série de procédures définies comme suit :

Graphe langagière (langagier ?)

Le texte de chaque message est analysé de façon à ne conserver que les mots les plus importants. Cela s'effectue en cinq étapes successives :

i

1. La phrase est segmentée (analyse de la langue chinoise) pour obtenir un ensemble de mots du même.
2. A chacun des mots est associé l'ensemble des utilisateurs l'ayant mentionné.
3. Les mots les plus courants sont enlevés afin de réduire le bruit. La constitution de la liste des mots courants (*stopwords*) est une étape très importante. La liste que nous utilisons a été créée à partir de différents corpus lors d'expérimentations précédentes et augmentée de nouveaux mots au fur et à mesure.
4. La co-occurrence d'un mot dans une même phrase définit une relation entre ces deux mots.
5. Parmi l'ensemble de mots-clés ainsi obtenu, nous conservons les 500 mots les plus utilisés (dont les occurrences sont les plus nombreuses)
6. Nous reconstituons le graphe des relations entre ces 500 mots-clés d'après la série de leur co-occurrence : chaque mot est un node, chaque co-occurrence est un edge.
7. L'ensemble constitue le graphe sémantique du même, pondéré mais non dirigé.

3.4.2.2. Graphe conversationnel

Pour chaque mème, l'ensemble des interactions (mentions, citations et retweets) constituent le graphe conversationnel où un utilisateur est un node et une interaction est un edge. Le graphe est dirigé car les interactions vont depuis un utilisateur à un autre.

Une fois l'ensemble du graphe constitué, nous déterminons sa modularité et son coefficient moyen de clustering afin de posséder des informations sur sa structure. Nous calculons également la centralité intermédiaire (*betweenness centrality*) pour chaque utilisateur dans l'ensemble du graphe. Ensuite, nous identifions les différentes communautés en utilisant l'algorithme dit de Louvain et son implémentation par Blondel & al. (Blondel *et al.*, 2008)¹² qui nous permet d'attribuer à chaque utilisateur un groupe unique. Enfin, nous réduisons la taille du graphe final en ne conservant que les utilisateurs ayant effectué au moins 2 échanges - en supprimant les *edges* du graphe ayant un poids inférieur à 2.

Le graphe ainsi constitué correspond aux données conversationnelles du même.

12. <http://perso.crans.org/aynaud/communities/index.html> consulté le 22 Avril 2014 à 14 :24

Localisation des utilisateurs

Pour chaque mème, nous souhaitons également regrouper les informations de localisation des utilisateurs mentionnés ou actifs dans le même. Le jeu de données *Weiboscope* comprend les localités mentionnées par les utilisateurs dans leurs profils. Le nombre de ces localités est restreinte par l'interface de Sina Weibo elle-même à : l'ensemble des provinces de Chine continentale, Hong Kong, Macau, Taiwan, “à l'étranger” et “autres”. Ainsi, pour chaque utilisateur nous assignons l'information géographique mentionnée par l'utilisateur dans son profil.

3.4.3 Visualisation multi-graphes

Une fois que l'ensemble de ces données a été extrait et traité, nous pouvons considérer différents aspects au sein du même :

ii

- Graphe sémantique : mots-clés et relations entre les mots
- Graphe socio-sémantique : relations entre les mots et les utilisateurs
- Graphe conversationnel : relations et communautés d'utilisateurs
- Graphe socio-géographique : relations entre les utilisateurs et les provinces / villes chinoises

Nous avons donc plusieurs niveaux de lecture interdépendants qui nous permettent de considérer différents aspects de chaque mème. Néanmoins, les informations de graphe sont peu claires et difficilement exploitables sous forme de données, à l'exception d'une série de mesures indicatives sur leurs structures. Afin d'explorer ces différentes relations, il nous faut pouvoir visualiser différentes dimensions du mème afin d'observer les articulations et phénomènes possibles émergents de cette lecture.

Pour réaliser cette cartographie particulière, il n'existe pas d'outils disponibles permettant de mettre en relation différents types de réseaux. Nous avons donc choisi de développer une solution technologique adaptée à nos besoins permettant de considérer sous différents angles l'ensemble de ces graphes et de leurs relations. Les différents choix conceptuels et technologiques présidant au design de cet outil de visualisation sont détaillés ci-après.

Choix technologiques

Les choix technologiques effectués lors de toute cette recherche et plus particulièrement pour cet outil de visualisation ont vocation à faciliter l'interopérabilité et la publication des résultats (visualisation, code et données) en ligne. Egalement, l'ensemble de l'outil de visualisation se fonde sur HTML5, CSS3 et Javascript qui sont des langages ouverts issues d'un travail collectif de l'ingénierie du Web permettant la réutilisation partielle ou totale des travaux produits par d'autres et contribuant ainsi à davantage d'interopérabilité entre les différents design scientifiques de recherche. La mise en forme

des données utilise la librairie “Data-Driven Documents” (d3.js)¹³, les données elles-mêmes sont stockées en JSON pour les graphes et GeoJSON pour les données cartographiques.

Espace perceptif, plan projectif et choix d'un “layout”

Devant la disparité des données en présence, nous devons donc construire un espace de représentation ou plutôt une interface de représentation propice à l'exploration et la découverte des phénomènes dans les relations entre les différents graphes. Dans la définition de cet espace perceptif se joue la réconciliation d'entités dissemblables (des utilisateurs, des lieux, des mots, leurs relations mutuelles). La tâche consisterait ainsi à réconcilier dans le plan de l'espace graphique de la visualisation (ici celui de l'écran) les différents espaces où projeter ces différents graphes. A l'image du milieu numérique, l'existence même de ce plan réconciliant global, local, réel, symbolique et imaginaire pose question. A la simple question : “quelle doit être sa couleur ?”, une logique commune de lisibilité répondrait : “blanc ou noir”, mais quelle peut en être la justification logique dans l'espace de la visualisation ? Et que penser du côté de l'écran ? De quoi est-t-il le bord ? Dès lors que se pose la question de structurer un espace de perception, nous voilà devant plusieurs difficultés formelles majeures. L'exigence de rigueur scientifique et surtout l'imperméabilité des formats de publications scientifiques actuelles rajoutent à la complexité de cette entreprise.

Afin de pouvoir considérer les relations entre mots, communautés et territoires, nous devons dans l'espace graphique disponible, il est nécessaire de résoudre plusieurs problèmes :

ii

- Donner à voir chaque type d'entités de façon reconnaissable (provinces chinoises, communautés, mots)
- Donner à voir les relations entre les différentes entités (graphe multiples)
- Pallier l'augmentation de la complexité visuelle et préserver une lisibilité

Nous avons donc choisi de construire un espace sur plusieurs niveaux hiérarchisés.

Graphes et interactions

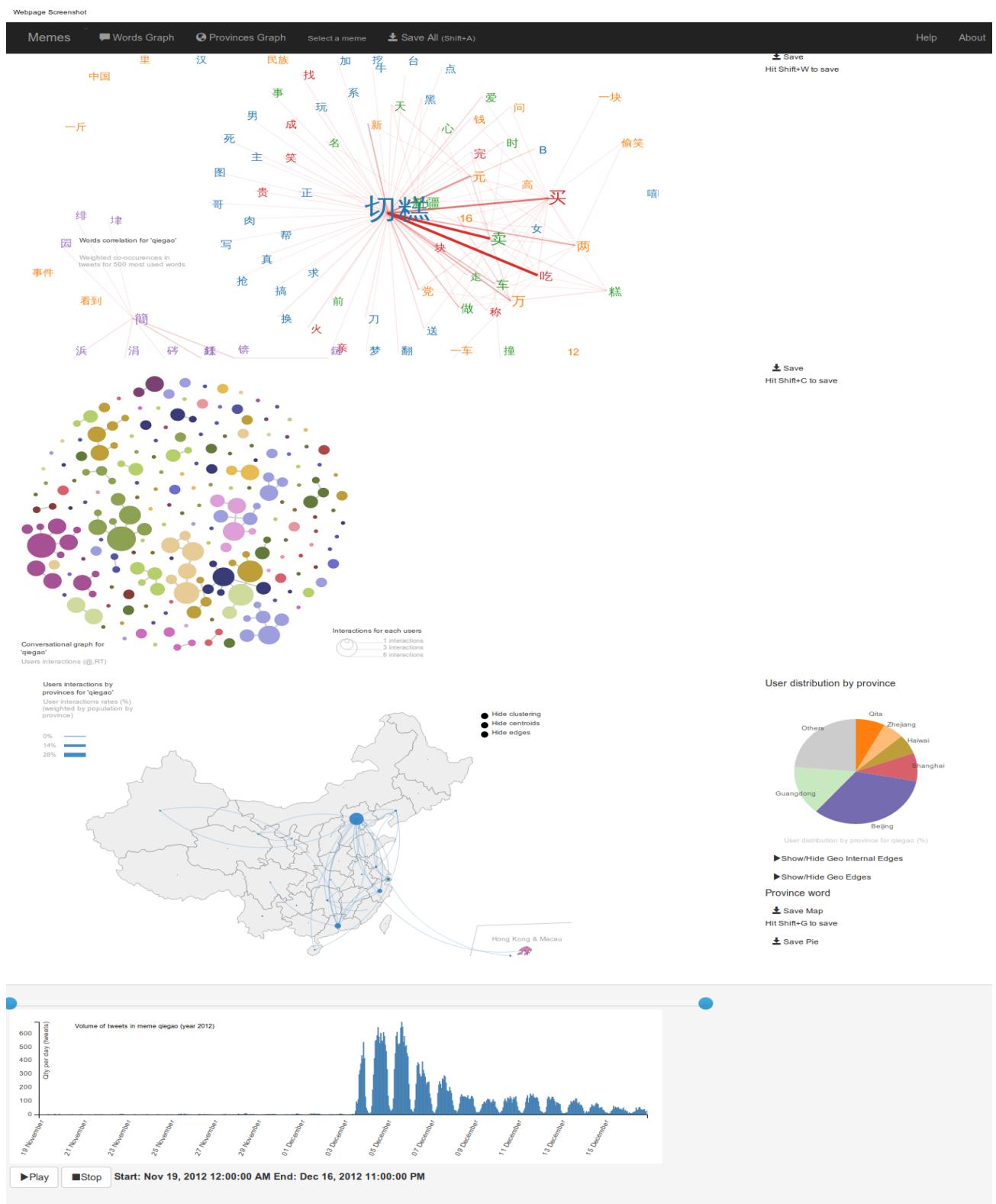
Le centre de la visualisation est occupé par les utilisateurs, reflétant à la fois l'intérêt central de cette étude pour l'étude des processus d'individuation et le rôle charnière des utilisateurs dans la structuration des données. Plus exactement, nous avons choisi de regrouper les utilisateurs par communautés afin de considérer les liens qui les unissaient entre elles et la structure générale de la discussion plutôt que de s'intéresser aux individus. Les individus sont symbolisés par des points au sein des communautés dont la couleur indique leur centralité dans le réseau global.

13. D3js at <http://d3js.org/>, consulté le 24 Avril 2014 à 14 :58

Notre première idée a été de représenter l'ensemble des utilisateurs sur une carte pour comprendre comment s'agencent spatialement chacune des discussions. La première étape a donc été de reconstituer une carte de la Chine comprenant Taiwan, Hong Kong et Macau. Chacun de ces territoires possèdent une influence médiatique certaine en Chine et méritent à ce titre d'être représenté. Néanmoins, le statut politique particulier de chacun de ces territoires nous a obligé à reconstituer une carte où ils figuraient tous. Nous avons choisi d'agrandir l'échelle de Hong Kong et Macau afin qu'il soit visible sur la carte. Egalement, nous avons choisi de ne pas restituer les données concernant les utilisateurs ayant choisi "autres" ou "reste du monde" comme leurs lieux de résidence. Si la carte se prête assez bien à la visualisation de quantité, les visualisations de graphe deviennent facilement confuses. Ainsi, nous avons choisi d'ajouter une option de visualisation des provinces sous la forme d'une liste, au moins toute aussi parlante. Cela nous permet en plus d'organiser cette liste selon différents critères afin d'étudier plusieurs aspects qui peuvent être intéressants.

Le graphe de mots revêt ici la forme la plus simple. Les relations entre les mots sont matérialisées par un graphe utilisant un algorithme force-based pour disposer les mots entre eux, recréant ainsi des groupes de mots intéressants d'après leur proximité dans les phrases lors des discussions. Un positionnement des mots selon leur importance (nombre de citations) est également disponible à loisir.

Afin d'explorer les différentes dimensions du graphes, nous avons également ajouté des interactions qui permettent de focaliser sur des zones ou des caractéristiques précises des graphes. ([Bostock et al., 2011](#))



Graphique 3.4.2 – Interface d'exploration et de visualisation des données

Table des matières

0.1	Introduction	1
1	Sina Weibo et le milieu numérique en Chine	3
1.1	L'Internet chinois : éléments de contexte	4
1.1.1	Petite histoire et évolution de l'Internet chinois	4
1.1.2	Censure sur l'Internet chinois	6
1.2	Médias sociaux en Chine : un paysage morcelé	9
1.2.1	Microblog en Chine et Sina Weibo	11
1.2.2	Sina Weibo, un usage plus ludique que <i>Twitter</i>	14
1.3	Code, langage et milieu(x) numérique(s)	17
1.3.1	Lieu, espace, territoire et technologies	18
1.3.2	Le milieu : richesse et désuétude	25
1.3.3	Cyberespace et milieu numérique	28
2	Les mèmes Internet, objets numériques culturels	31
2.1	Les mèmes : définitions et histoire	32
2.1.1	La mémétique : une éthologie culturelle teinté d'évolutionnisme	32
2.1.2	Mèmes Internet : définition, littérature et exemples	34
2.2	Mème, mémoire collective et culture	38
2.2.1	La mémoire comme trace	38
2.2.2	Diffusion de mèmes et structuration d'une mémoire collective	39
2.3	Textualité des mèmes et formes d'énonciations numériques	44
2.3.1	Le même comme figure rhétorique de l'écriture intertextuelle	44
2.3.2	Typologie des mèmes Internet	48
3	Méthodologie de recherche	57
3.1	Sciences sociales du réseau	57
3.1.1	Le réseau comme enjeu pour l'étude	58
3.1.2	Code, données et les nouveaux outils de l'écriture scientifique	60
3.1.3	Visualisation et espace perceptif pour l'information	63
3.2	Méthodes et outils d'analyse des réseaux sociaux	65

3.2.1	Anatomie d'un réseau social	65
3.2.2	Diffusion des discours dans les réseaux sociaux	69
3.2.3	Méthodes d'analyse de données des réseaux sociaux en ligne	71
3.3	Design de la recherche : Méthodologie choisie et schémas d'analyse retenus	72
3.3.1	Constitution et collection d'un corpus de données	72
3.3.2	Identification des mèmes en tant que “clusters”	76
3.3.3	L'identification des mèmes par hashtags	78
3.3.4	Visualisation du graphe conversationnel d'utilisateurs	82
3.3.5	Premiers éléments d'analyse : visualisation de mèmes	86
3.4	Méthodologie de traitement et de visualisation des mèmes	93
3.4.1	Sélection de mèmes pour l'étude	94
3.4.2	Extraction des graphes et traitement des données brutes	95
3.4.3	Visualisation multi-graphes	97
	Table des Matières	102
	Bibliographie	120
	Liste des Figures	122
	I Liste des mèmes	125

Bibliographie

- Jimi ADAMS, Katherine FAUST et Gina S. LOVASI : Capturing context : Integrating spatial and social network analyses. *Social Networks*, 34(1):1–5, janvier 2012. ISSN 03788733. URL <http://linkinghub.elsevier.com/retrieve/pii/S0378873311000608>.
- Muhammad A AHMAD et Ankur TEREDESAI : Modeling Spread of Ideas in Online Social Networks. *Reproduction*, 61:185–190, 2006. URL <http://dl.acm.org/citation.cfm?id=1273833>.
- Aris ANAGNOSTOPOULOS, Luca BECCHETTI, Carlos CASTILLO, Aristides GIONIS et Stefano LEONARDI : Online team formation in social networks. *Proceedings of the 21st international conference on World Wide Web - WWW '12*, page 839, 2012. URL <http://dl.acm.org/citation.cfm?doid=2187836.2187950>.
- Sinan ARAL, Lev MUCHNIK et Arun SUNDARARAJAN : Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21544–21549, 2009. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2799846/>&tool=pmcentrez&rendertype=abstract.
- François ASCHER : *L'âge des métapoles*. Éd. de l'Aube, 2009. ISBN 2815900173. URL <http://www.amazon.com/L%C3%A2ge-m%C3%A9ta%C3%A8res-Fran%C3%A7ois-Ascher/dp/2815900173>.
- J L AUSTIN : *How to do things with words*, volume 3 de *Syntax and Semantics*. Harvard University Press, 1975. ISBN 019824553X.
- Kent BACH : Meaning , Speech Acts , and Communication. In Robert M. HARNISH, éditeur : *Basic Topics in the Philosophy of Language*, pages 1–23. Prentice-Harll, 1994.
- Alain BADIOU : *L'Etre et l'événement : Tome 2, Logiques des mondes*. Seuil, 2006. ISBN 2020843242. URL <http://www.amazon.fr/LEtre-l%C3%A9v%C3%A9nement-Tome-Logiques-mondes/dp/2020843242>.
- Eytan BAKSHY, J. HOFMAN, W. MASON et D. WATTS : Identifying ‘Influencers’ on Twitter. In *Fourth ACM International Conference on Web Search and Data*

Mining (WSDM), Hong Kong, 2011. URL <http://kdpaine.blogs.com/files/twitterinfluencershofmanetal.pdf>.

David BAMMAN, Brendan O'CONNOR et Noah A SMITH : Censorship and deletion practices in Chinese social media. *First Monday*, 17(3-5):1–19, 2012. URL <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3943/3169> <http://brenocon.com/censorship.bamman+oconnor+smith.fm2012.pdf>.

K. BARBER : Orality, the media and new popular cultures in Africa. In Mena LAFKIOUI et Daniela MEROLLA, éditeurs : *Oralité et nouvelles dimensions de l'oralité Intersections théoriques et comparaisons des matériaux*, volume 00, pages 16–28, Paris, France, 2008. INALCO. ISBN 9782858311750.

Jp BARLOW : *Déclaration d'indépendance du Cyberspace*. L'Eclat, 2001.

F M BASS, T V KRISHNAN et D C JAIN : Why the Bass model fits without decision variables. *Marketing Science*, 13(3):203–223, 1994. ISSN 07322399. URL <http://www.jstor.org/stable/10.2307/183674>.

Michael BATTY : Virtual geography. *Futures*, 1997. URL <http://www.sciencedirect.com/science/article/pii/S0016328797000189>.

Christian BAUCKHAGE : Insights into internet memes. In *Proc. ICWSM2011*, pages 42–49, 2011. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/2757/3304>.

Mary BERGSTROM : *All Eyes East : Lessons from the Front Lines of Marketing to China's Youth*. Palgrave Macmillan, 2012. ISBN 0230120628. URL <http://www.amazon.com/All-Eyes-East-Lessons-Marketing/dp/0230120628>.

Augustin BERQUE : Géogrammes, pour une ontologie des faits géographiques. *Espace géographique*, 28(4):320–326, 1999. ISSN 0046-2497. URL http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1999_num_28_4_1275.

Jacques BERTIN : *Sémiologie graphique : Les diagrammes - Les réseaux - Les cartes*. Editions de l'Ecole des Hautes Etudes en Sciences, 1977. ISBN 2713212774.

Nick BINGHAM : Object-ions : from technological determinism towards geographies of relations, 1996. ISSN 0263-7758.

Susan BLACKMORE : The power of the meme meme. *Skeptic*, (2):43, 1997. ISSN 10639330. URL <http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=9711050421&site=ehost-live>.

Susan BLACKMORE : Evolution and Memes : the Human Brain As a Selective Imitation Device. *Cybernetics and Systems*, 32(1-2):225–255, 2001. ISSN 01969722. URL <http://www.informaworld.com/openurl?genre=article&doi=10.1080/019697201300001867&magic=crossref>.

Susan BLACKMORE : Why we need memetics. *Behavioral and Brain Sciences*, 29(04):349–350, 2006. ISSN 0140525X. URL <http://www.lib.ncsu.edu/cgi-bin/proxy.pl?server=http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=23334578&site=ehost-live&scope=site>.

Vincent D BLONDEL, Jean-Loup GUILLAUME, Renaud LAMBIOTTE et Etienne LEFEBVRE : Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment*, 2008(10):P10008, octobre 2008. ISSN 1742-5468. URL <http://arxiv.org/abs/0803.0476>.

Gillian BOLSOVER, William H. DUTTON, Ginette LAW et Soumitra DUTTA : Social Foundations of the Internet in China and the New Internet World : A Cross-National Comparative Perspective. *SSRN Electronic Journal*, juin 2013. ISSN 1556-5068. URL <http://papers.ssrn.com/abstract=2276482>.

Philippe BONNIN : Dispositifs et rituels du seuil. *Communications*, 70(1):65–92, 2000. ISSN 0588-8018. URL http://www.persee.fr/web/revues/home/prescript/article/comm_0588-8018_2000_num_70_1_2064.

Michael BOSTOCK, Vadim OGIEVETSKY et Jeffrey HEER : D³ : Data-Driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–9, décembre 2011. ISSN 1941-0506. URL <http://www.ncbi.nlm.nih.gov/pubmed/22034350>.

Danah BOYD et Kate CRAWFORD : Six Provocations for Big Data. In *SSRN Electronic Journal*, septembre 2011. URL <http://papers.ssrn.com/abstract=1926431>.

Ulrik BRANDES, Garry ROBINS, Ann MCCRANIE et Stanley WASSERMAN : What is network science ? *Network Science*, 1(01):1–15, avril 2013. ISSN 2050-1242. URL http://www.journals.cambridge.org/abstract_S2050124213000027.

Leo BREIMAN : Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 14787954. URL <http://www.springerlink.com/index/UOP06167N6173512.pdf>.

Roger BRUNET : La composition des modèles dans l’analyse spatiale. *Espace géographique*, pages 253–265, 1980. URL http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1980_num_9_4_3572.

Roger BRUNET : *La carte : mode d’emploi*. Fayard, 1987. ISBN 2213018480. URL <http://www.amazon.fr/La-carte-demploi-Roger-Brunet/dp/2213018480>.

Branislav BUCHEL : *Internet Memes as Means of Communication*. Master thesis, MARYK UNIVERSITY, 2012.

Simon BUCKINGHAM, Michael DAW, Roger SLACK et Ben JUBY : Memetic : From Meeting Memory to Virtual Ethnography & Distributed Video Analysis The Memetic toolset. In *2 nd International Conference on e-Social Science*, Manchester, UK, 2006.

Judith BUTLER : *Bodies that matter : on the discursive limits of "sex"*. Numéro Book, Whole. Routledge, 1993. ISBN 0415903661. URL <http://www.amazon.com/dp/0415903661>.

Alberto CAIRO : Forms and Functions : Visualization as a Technology An Information. In *The Functional Art*, chapitre 2, pages 25–44. New Riders, 2013. ISBN 978-0321834737. URL <http://www.thefunctionalart.com/>.

Georges CANGUILHEM : *La Connaissance De La Vie*. Vrin, 1965. ISBN 2711611329. URL http://books.google.fr/books/about/La_Connaissance_De_La_Vie.html?id=A7EUqvUHZsC&pgis=1.

Nicholas CARR : How the internet makes us stupid. *Sydney Morning Herald* 10 September 2010, pages 13–15, 2010.

Meeyoung CHA, Hamed HADDADI, F. BENEVENUTO et K.P. GUMMADI : Measuring user influence in Twitter : The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 10–17, Washington, 2010. AAAI. URL <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/download/1538/1826>.

Jun-Jun CHENG, Yun LIU, Bo SHEN et Wei-Guo YUAN : An epidemic model of rumor diffusion in online social networks. *The European Physical Journal B*, 86(1):29, janvier 2013. ISSN 1434-6028. URL <http://link.springer.com/10.1140/epjb/e2012-30483-5>.

Owen CHURCHES, Mike NICHOLLS, Myra THIESSEN, Mark KOHLER et Hannah KEAGE : Emoticons in mind : An event-related potential study Emoticons in mind : An event-related potential study. *Social Neuroscience*, 9(2):196–202, 2014. ISSN 1747-0919. URL <http://dx.doi.org/10.1080/17470919.2013.873737>.

Paul CLAVAL : L'évolution de quelques concepts de base de la géographie. In JF STASZAK, éditeur : *Les discours du géographe*, pages 89–118. L'Harmattan, Paris, France, histoire e édition, 1990.

Christopher CLIFTON, Deirdre MULLIGAN et Raghu RAMAKRISHNAN : Data Mining and Privacy : An Overview. *Privacy and Technologies of identity*, (2):191–208, 2006. URL http://dx.doi.org/10.1007/0-387-28222-X_11.

Antoine COMPAGNON : Théorie du lieu commun. *Cahiers de l'Association internationale des études francaises*, 49(1):23–37, 1997. ISSN 0571-5865. URL http://www.persee.fr/web/revues/home/prescript/article/caief_0571-5865_1997_num_49_1_1269.

Auguste COMTE : *Cours de philosophie positive : La philosophie chimique et la philosophie biologique, Volume 3.* Bachelier, 1838. URL <http://books.google.com/books?id=X3YrAAAYAAJ&pgis=1>.

Michael D CONOVER, Clayton DAVIS, Emilio FERRARA, Karissa MCKELVEY, Filippo MENCZER et Alessandro FLAMMINI : The Geospatial Characteristics of a Social Movement Communication Network. *PLoS ONE*, 8(3):e55957+, 2013. URL <http://dx.doi.org/10.1371/journal.pone.0055957>.

Michele COSCIA : Competition and Success in the Meme Pool : a Case Study on Quick-meme.com. avril 2013. URL <http://arxiv.org/abs/1304.1712>.

J W CRAMPTON : Cartography : maps 2.0. *Progress in Human Geography*, 33 (1):91–100, 2009. ISSN 03091325. URL <http://phg.sagepub.com/cgi/doi/10.1177/0309132508094074>.

Jeremy W. CRAMPTON, Mark GRAHAM, Ate POORTHUIS, Taylor SHELTON, Monica STEPHENS, Matthew W. WILSON et Matthew ZOOK : Beyond the geotag : situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, 40(2):130–139, mars 2013. ISSN 1523-0406. URL <http://dx.doi.org/10.1080/15230406.2013.777137>.

Tim CRESSWELL : *Place : A Short Introduction*, volume 58 de *Short introductions to geography*. Wiley-Blackwell, 2004. ISBN 1405106727. URL <http://www.amazon.co.uk/dp/1405106727>.

Olivier CREVOISIER : The Innovative Milieus Approach : Toward a Territorialized Understanding of the Economy ? *Economic Geography*, 80(4):367–379, 2004. ISSN 00130095. URL <http://onlinelibrary.wiley.com/doi/10.1111/j.1944-8287.2004.tb00243.x/abstract>.

Candice Tran DAI : L ’ Internet en République populaire de Chine : vers un outil de démocratisation ou de renforcement de l ’ État-parti ? Rapport technique, Fondation pour la Recherche Stratégique, 2007.

Hubert DAMISCH : *L'origine de la perspective*. Flammarion, champs art édition, 1999. ISBN 2080816055. URL <http://www.amazon.fr/gp/product/2080816055?ie=UTF8&tag=leproidi-21&linkCode=as2&camp=1642&creative=6746&creativeASIN=2080816055>.

G Elijah DANN et Neil HADDOW : Just Doing Business or Doing Just Business : Google, Microsoft, Yahoo ! and the Business of Censoring China's Internet. *Journal of Business Ethics*, 79:219–234, 2008.

M Bernard DEBARBIEUX : Non lieux. *Espace géographique*, 22(1):90–91, 1993.

Jérôme DENIS : Préface : Les nouveaux visages de la performativité. *Études de communication*, (29):8–24, décembre 2006. ISSN 1270-6841. URL <http://edc.revues.org/344>.

Daniel C DENNETT : *Breaking the Spell : Religion As a Natural Phenomenon*. Viking, 2006. ISBN 067003472X. URL <http://books.google.com/books?id=yWtwDDqR61QC&pgis=1>.

Cong DING, Yang CHEN et Xiaoming FU : Crowd Crawling : Towards Collaborative Data Collection for Large-scale Online Social Networks. pages 183–188, 2013.

Martin DODGE et Rob KITCHIN : Code , space and everyday life. 2004.

Martin DODGE et Rob KITCHIN : Virtual places, 2007.

Jean-François DOULET : Les nouveaux enjeux de la mobilité urbaine dans les villes chinoises - Rapport d ' une mission en Chine. Rapport technique, Institut pour la ville en mouvement (PSA Peugeot-Citroën), Paris, France, 2001.

F DOUZET : Les frontières chinoises de l'Internet. *Hérodote*, 125:127–142, 2007.

Gabriel DUPUY : Internet : une approche géographique à l'échelle mondiale. *Flux*, pages 5–19, 2004. URL <http://www.cairn.info/revue-flux-2004-4-page-5.htm>.

Umberto ECO : *Lector in fabula ou La Coopération interprétative dans les textes narratifs*. Grasset, 1985. ISBN 2246342619. URL <http://www.amazon.fr/Lector-fabula-Coopération-interprétative-narratifs/dp/2246342619>.

Bruce EDMONDS : Three Challenges for the Survival of Memetics. *Journal of Memetics Evolutionary Models of Information Transmission*, (6), 2002.

Jacob EISENSTEIN, Brendan O'CONNOR, Noah A. SMITH et Eric P. XING : Mapping the geographical diffusion of new words. octobre 2012. URL <http://arxiv.org/abs/1210.5268>.

Norbert ELIAS : *La Dynamique de l'Occident*. Poche Pocket, Paris, {FR}, 1975.

Sarah ELWOOD : Volunteered geographic information : future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal*, 72(3-4):173–183, 2008. ISSN 03432521. URL <http://www.springerlink.com/index/10.1007/s10708-008-9186-0>.

J M EPSTEIN et R AXTELL : *Growing artificial societies*. 1996. ISBN 0262550253.

Xiang FANG et David C YEN : Demographics and behavior of Internet users in China. *Technology in Society*, 28:363–387, 2006.

Valérie FERNANDEZ et Gilles PUEL : *La gouvernance territoriale des TIC*. Hermès Science/Lavoisier, Paris, mondialisa édition, 2010. URL <http://editions.lavoisier.fr/notice.asp?ouvrage=2139905>.

Emilio FERRARA, Mohsen JAFARIASBAGH, Onur VAROL, Vahed QAZVINIAN, Filippo MENCZER et Alessandro FLAMMINI : Clustering Memes in Social Media. octobre 2013. URL <http://arxiv.org/abs/1310.2665>.

Wolf FEUERHAHN : Du milieu à l’Umwelt : enjeux d’un changement terminologique. *Revue philosophique de la France et de l’étranger*, 134(4):419, 2009. ISSN 0035-3833. URL <http://www.cairn.info/revue-philosophique-2009-4-page-419.htm>.

Gustave FLAUBERT : *Le Dictionnaire des idées reçues*. Le Livre de Poche, 1997. ISBN 2253098361. URL <http://www.amazon.fr/Dictionnaire-idées-reçues-Gustave-Flaubert/dp/2253098361>.

Michel FOUCAULT : *Les Mots et les Choses : Une Archéologie du Savoir*. Gallimard, Paris, {FR}, 1996.

Michel FOUCAULT : *Sécurité, territoire, population. Cours au Collège de France (1977-78)*. Gallimard/Seuil (Collection « Hautes Études »), Paris, 2004.

King-wa FU, Chung-hong CHAN et Michael CHAU : Assessing Censorship on Microblogs in China : Discriminatory Keyword Analysis and the Real-Name Registration Policy. *IEEE Internet Computing*, 17(3):42–50, mai 2013. ISSN 1089-7801. URL <http://www.computer.org/csdl/mags/ic/2013/03/mic2013030042.html>.

Silvia Elena GALLAGHER et Timothy SAVAGE : Cross-cultural analysis in online community research : A literature review. *Computers in Human Behavior*, 29(3):1028–1038, mai 2013. ISSN 07475632. URL <http://linkinghub.elsevier.com/retrieve/pii/S0747563212002683>.

Qi GAO, Fabian ABEL, GJ Houben et Yong YU : A Comparative Study of Users’ Microblogging Behavior on Sina Weibo and Twitter. *User Modeling, Adaptation, and ...*, 2012. URL http://link.springer.com/chapter/10.1007/978-3-642-31454-4_8.

Minas GJOKA et Maciej KURANT : Practical recommendations on crawling online social networks. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Measurement of Internet Topologies*, 29(9):1872–1892, 2011. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6027868http://www.minasgjoka.com/papers/jsac11_crawling_osns.pdf.

Sharad GOEL, DJ WATTS et DG GOLDSTEIN : The structure of online diffusion networks. In *EC '12 Proceedings of the 13th ACM Conference on Electronic Commerce*, volume 1, 2012. ISBN 9781450314152. URL <http://dl.acm.org/citation.cfm?id=2229058>.

Michael GOLDHABER : The value of openness in an attention economy. *First Monday*, 11(6):3, 2006. URL <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1334/1254>.

Michele GRAFFIETI, Gaia SCAGNETTI, Donato RICCI, Luca MASUD et Mario PORPORA : Tell them anything but the truth : They will find their own. How we visualized the map of the future with respect to the audience of our story. *Leonardo*, 44(3):250–251, 2010.

Mark GRAHAM et Matthew ZOOK : Visualizing Global Cyberscapes : Mapping User-Generated Placemarks. *Journal of Urban Technology*, 18(1):115–132, janvier 2011. ISSN 1063-0732. URL <http://www.tandfonline.com/doi/abs/10.1080/10630732.2011.578412>.

S GRAHAM : The end of geography or the explosion of place ? Conceptualizing space, place and information technology. *Progress in Human Geography*, 22(2):165–185, 1998. ISSN 03091325. URL <http://phg.sagepub.com/cgi/doi/10.1191/030913298671334137>.

Éric GUICHARD : L'internet et les épistémologies des sciences humaines et sociales. *Revue Sciences/Lettres*, (2), octobre 2014. ISSN 2271-6246. URL <http://rsl.revues.org/389>.

Muki HAKLAY, Alex SINGLETON et Chris PARKER : Web Mapping 2.0 : The Neogeography of the GeoWeb. *Geography Compass*, 2(6):2011–2039, novembre 2008. ISSN 17498198. URL <http://doi.wiley.com/10.1111/j.1749-8198.2008.00167.x>.

Maurice HALBWACHS : *La mémoire collective et le temps*, volume 1945. UQAC, Les classiques des sciences sociales, 1947. URL http://www.psychanalyse.com/pdf/memoire_collective.pdf.

Harry HALPIN et Alexandre MONNIN, éditeurs. *Philosophical Engineering : Toward a Philosophy of the Web (Metaphilosophy)*. Wiley-Blackwell, 2014. ISBN 111870018X. URL <http://www.amazon.com/Philosophical-Engineering-Toward-Philosophy-Metaphilosophy/dp/111870018X>.

Frank HARARY : Graph theory 1736–1936, 1977. ISSN 03150860.

Douglas R HOFSTADTER : *Gödel, Escher, Bach*. Klett-Cotta, 1999. ISBN 360893037x.

Liangjie HONG et Brian D. DAVISON : Empirical study of topic modeling in Twitter. *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, pages 80–88, 2010. URL <http://portal.acm.org/citation.cfm?doid=1964858.1964870>.

Jeremy R L HOWELLS : Tacit Knowledge , Innovation and Economic Geography. *Urban Studies*, 39(5-6):871–884, 2002. ISSN 00420980. URL <https://webvpn.ethz.ch/http/0/www.informaworld.com/smpp/1814185133-11623592/content~content=a770985056>.

Yuk HUI : What is a digital object ? *Metaphilosophy*, 43(4):380–395, 2012.

Dino IENCO, Francesco BONCHI et Carlos CASTILLO : The Meme Ranking Problem : Maximizing Microblogging Virality. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 328–355, Sydney, NSW, 2010.

Sorin ISTRAIL et Solomon MARCUS : Alan Turing and John von Neumann - Their Brains and Their Computers. In Erzsébet CSUHAJ-VARJÚ, Marian GHEORGHE, Grzegorz ROZENBERG, Arto SALOMAA et György VASZIL, éditeurs : *13th International Conference on Membrane Computing CMC 2012*, volume 7762 de *Lecture Notes in Computer Science*, pages 26–35, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-36750-2. URL <http://link.springer.com/10.1007/978-3-642-36751-9>.

MO JACKSON et Leeat YARIV : Diffusion, strategic interaction, and social structure. in *Handbook of Social Economics*, pages 1–50, 2010. URL <http://citeseervx.ist.psu.edu/viewdoc/download?doi=10.1.1.207.4193&rep=rep1&type=pdf>.

Christian JACOB : *L'Empire des cartes : Approche théorique de la cartographie à travers l'histoire*. Albin Michel, 1992. ISBN 2226060839.

Christian JACOB : *Lieux de savoir : Tome 2, Les mains de l'intellect*. Editions Albin Michel, 2011. ISBN 978-2226187291. URL <http://www.amazon.fr/Lieux-savoir-Tome-mains-lintellect/dp/2226187294http://lieuxdesavoir.hypotheses.org/lieux-de-savoir-2-les-mains-de-lintellect-2>.

Roman JAKOBSON : *Essais de linguistique générale*. Numéro 1960. 1960.

Roman JAKOBSON et Morris HALLE : Phonemic patterning. *Fundamentals of language*, pages 37–44, 1956.

William JAMES : Essays in radical Empiricism. 1912.

Pascal JOUXTEL : La mémétique, une science à l'état sauvage. novembre 2014. ISSN 0767-9513. URL <http://documents.irevues.inist.fr/handle/2042/51885>.

KY KAMATH, James CAVERLEE, Kyumin LEE et Zhiyuan CHENG : Spatio-Temporal Dynamics of Online Memes : A Study of Geo-Tagged Tweets. In *WWW 2013*, Rio de Janeiro, Brazil, 2013. IW3C2. URL <http://students.cse.tamu.edu/kykamath/papers/www2013/fp113-kamath.pdf>.

Anand KARANDIKAR : Clustering short status messages : A topic model based approach. *Work*, page 55, 2010. URL <http://ebiquity.umbc.edu/get/a/publication/518.pdf>.

E KATZ et P F LAZARSFELD : *Personal influence : The part played by people in the flow of mass communications*, volume 21 de *Foundations of communications research*. Free Press, 1955. ISBN 9781412805070. URL <http://www.ncbi.nlm.nih.gov/pubmed/13409025>.

Christa KELLEHER et Thorsten WAGENER : Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling Software*, 26(6):822–827, 2011. ISSN 13648152. URL <http://linkinghub.elsevier.com/retrieve/pii/S1364815210003270>.

Rob. KITCHIN et Martin DODGE : *Code, Space : Software and Everyday Life*. MIT Press, 2011. ISBN 0262042487. URL <http://books.google.com/books?hl=en&lr=&id=ZHez2BXgIeQC&pgis=1http://mitpress.mit.edu/books/codespace>.

Michele KNOBEL et Colin LANKSHEAR : *A new literacies sampler*. Peter Lang Publishing, Inc., New York City, USA, 2007. ISBN 9780820495231. URL <http://books.google.com/books?hl=en&lr=&id=4Gjs8uT6dxIC&oi=fnd&pg=PR9&dq=A+New+Literacies+Sampler&ots=F14cYpvwjs&sig=4gyFVE88QSJvbC-jSnIUESPf-V8>.

Moo Nam KO, Gorrell P CHEEK, Mohamed SHEHAB et North CAROLINA : CONNECT SERVICES. *Computer*, 43(August):37–43, 2010. ISSN 00189162. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5551044.

Ravi KUMAR, Jasmine NOVAK et Andrew TOMKINS : Structure and evolution of online social networks. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining KDD 06*, 106(2):611, 2006. URL <http://portal.acm.org/citation.cfm?doid=1150402.1150476>.

Gerard KYLE et Garry CHICK : The Social Construction of a Sense of Place. *Leisure Sciences*, 29(3):209–225, 2007. ISSN 01490400. URL <http://www.tandfonline.com/doi/abs/10.1080/01490400701257922>.

Jacques LACAN : *Autres écrits*. Seuil, Paris, le champ f édition, 2001.

Bruno LATOUR : On actor-network theory. *Soziale Welt*, 47(4):369–381, 1996. URL <http://www.cours.fse.ulaval.ca/edc-65804/latour-clarifications.pdf>.

John LAURENT : A Note on the Origin of Memes & Mnemes. *Journal of Memetics Evolutionary Models of Information Transmission*, 3(1), 1999.

Alex LEAVITT, Evan BURCHARD, David FISHER et Sam GILBERT : The Influentials : New Approaches for Analyzing Influence on Twitter, 2009. URL <http://www.webecologyproject.org/2009/09/analyzing-influence-on-twitter/>.

Jure LESKOVEC, Lada A ADAMIC et Bernardo A HUBERMAN : The Dynamics of Viral Marketing. *ACM Transactions on the Web*, 1(1):5–es, 2005. URL <http://arxiv.org/abs/physics/0509039>.

Jure LESKOVEC, Lars BACKSTROM et Jon KLEINBERG : Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, page 497, New York, New York, USA, juin 2009. ACM Press. ISBN 9781605584959. URL <http://dl.acm.org/citation.cfm?id=1557019.1557077>.

Yuan LI, Haoyu GAO, Mingmin YANG, Wanqiu GUAN et H MA : What are Chinese Talking about in Hot Weibos. *CoRR*, 2(abs), 2013. URL <http://arxiv.org/abs/1304.4682>.

Nan LIN : Building a Network Theory of Social Capital. *Connections*, 22(1):28–51, 1999. ISSN 14691930. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.3792&rep=rep1&type=pdf>.

Michael R LISSACK : The Redefinition of Memes. *Journal of Memetics Evolutionary Models of Information Transmission*, (1):1–14, 2004. ISSN 13664786. URL <http://search.ebscohost.com/login.aspx?direct=true&db=afh&AN=15171172&site=ehost-live>.

Steve LOHR : Big Data's Impact in the World, février 2012. URL http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=1&pagewanted=1&ref=technology.

Gilad LOTAN, Erhardt GRAEFF, Mike ANANNY, Devin GAFFNEY, Ian PEARCE et Danah BOYD : The Arab Spring| The Revolutions Were Tweeted : Information Flows during the 2011 Tunisian and Egyptian Revolutions. *International Journal of Communication*, 5, 2011. URL <http://ijoc.org/ojs/index.php/ijoc/article/view/1246>.

Dave LYONS : China's Golden Shield Project. *In CIRC09*, 2009. URL <http://www.scribd.com/doc/15919071/Dave-Lyons-Chinas-Golden-Shield-Project>.

Rebecca MACKINNON : China's Censorship 2.0 : How companies censor bloggers. *First Monday*, 14(2-2), 2009. URL <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewarticle/2378/2089>.

Michael W MACY et Robert WILLER : From Factors to Actors : Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28(1):143–166, 2002. ISSN 03600572. URL <http://www.annualreviews.org/doi/abs/10.1146/annurev.soc.28.110601.141117>.

Lev MANOVICH : Media After Software. *Journal of Visual Culture*, 12(1):30–37, 2013.

Étienne Jules MAREY : *La méthode graphique dans les sciences expérimentales : et principalement en physiologie et en médecine*. G. Masson, 1885. URL <http://books.google.com/books?id=JnsAAAAAMAAJ&pgis=1>.

Frédéric MARTEL : *Mainstream*. Flammarion, Paris, 2010. ISBN 2081236176. URL <http://www.amazon.fr/Mainstream-Frédéric-Martel/dp/2081236176>.

P. R. MASANI : The Cybernetical Movement and von Neumann's Letter, 1946. *In Norbert Wiener 1894–1964*, pages 239–250. Birkhäuser Basel, Basel, 1990. ISBN 978-3-0348-9963-5. URL <http://www.springerlink.com/index/10.1007/978-3-0348-9252-0>.

Doreen MASSEY : Power-geometry and a progressive sense of place. *In J BIRD, B CURTIS, T PUTNAM, G ROBERTSON et L TICKNER, éditeurs : Mapping the futures Local cultures global change*, numéro 1984, chapitre 4, pages 59–69. Routledge, 1993. ISBN 041507018X. URL http://books.google.com/books?id=307_ioVON0gC.

Hugh MATTHEWS, Marilyn TAYLOR, Barry PERCY-SMITH et Melanie LIMB : The unacceptable flaneur : The shopping mall as a teenage hangout. *Childhood*, 7(3): 279–294, 2000. ISSN 09075682. URL <http://chd.sagepub.com/cgi/doi/10.1177/0907568200007003003>.

Richard MCCREADIE, Ian SOBOROFF, Jimmy LIN, Craig MACDONALD, Iadh OUNIS et Dean McCULLOUGH : On building a reusable Twitter corpus. *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '12*, page 1113, 2012. URL <http://dl.acm.org/citation.cfm?doid=2348283.2348495>.

Claire Cain MILLER : How Obama's Internet Campaign Changed Politics, 2008. URL <http://bits.blogs.nytimes.com/2008/11/07/how-obamas-internet-campaign-changed-politics/>.

Jacob Levi MORENO et Helen JENNINGS : Statistics of Social Configurations. *Sociometry*, 1(3/4):342–374, 1938. ISSN 00380431. URL <http://www.jstor.org/stable/10.2307/2785588>.

David F. NETTLETON : Data mining of social networks represented as graphs. *Computer Science Review*, 7:1–34, février 2013. ISSN 15740137. URL <http://linkinghub.elsevier.com/retrieve/pii/S1574013712000445>.

Jason Q. NG : *Blocked on Weibo : What Gets Suppressed on China’s Version of Twitter (And Why)*. New Press, The, 2013. ISBN 159558871X.

BD NYE : *Modeling memes : A memetic view of affordance learning*. Thèse de doctorat, 2011. URL <http://repository.upenn.edu/edissertations/336/>.

Jean-Marc OFFNER : Les « effets structurants » du transport : mythe politique, mystification scientifique. *Espace géographique*, 22(3):233–242, 1993. ISSN 0046-2497. URL http://www.persee.fr/web/revues/home/prescript/article/spgeo_0046-2497_1993_num_22_3_3209.

Ray OLDENBURG : *The Great Good Place*. Marlowe & Company, 1999. ISBN 1569246815. URL <http://books.google.com/books?id=0a0jHGdSKLMC>.

Pamela E OLIVER et Gerald MARWELL : Whatever Happened to Critical Mass Theory ? A Retrospective and Assessment. *Sociological Theory*, 19(3):292–311, 2001. ISSN 07352751. URL <http://doi.wiley.com/10.1111/0735-2751.00142>.

Jukka-Pekka ONNELA, Samuel ARBESMAN, Marta C GONZÁLEZ, Albert-László BARABÁSI et Nicholas A CHRISTAKIS : Geographic Constraints on Social Network Groups. *PLoS ONE*, 6(4):7, 2011. URL <http://arxiv.org/abs/1011.4859>.

OPENITP : A Snapshot of Chinese Internet Users Circumventing Censorship. Rapport technique April, Open Internet Tools Project, 2013.

Alain PAPAUX : De la nature au « milieu » : l’homme plongé dans l’environnement. *Revue interdisciplinaire d’études juridiques*, 60(1):29–57, 2008.

Stephen PETSCHULAT : Other people’s data. *Communications of the ACM*, 53(1):53, 2010. ISSN 00010782. URL <http://portal.acm.org/citation.cfm?doid=1629175.1629196>.

Christian PLANTIN : *Lieux communs, topoi, stéréotypes, clichés*. Editions Kimé, 1993. ISBN 2908212765. URL http://books.google.fr/books/about/Lieux_communst_sterotypes_clich.html?hl=fr&id=718UAAAIAAJ&pgis=1.

Ate POORTHUIS et Matthew ZOOK : Spaces of Volunteered Geographic Information. *SSRN Electronic Journal*, mai 2013. ISSN 1556-5068. URL <http://papers.ssrn.com/abstract=2259845>.

G PUEL : Géographie des lieux d'accès à Internet. Les conditions de l'accès public et les modèles d'usages dans les grandes villes de Chine. *Espace géographique*, 38:17–29, 2009.

Gilles PUEL, Blandine PONS et Jin XIAOTING : New Forms of Mobility and Social Practices : The Starbucks Cafés in Beijing. *China Perspectives*, (April), 2007. URL <http://hal.archives-ouvertes.fr/hal-00125873/>.

Xipeng QIU, Qi ZHANG et Xuanjing HUANG : FudanNLP : A toolkit for Chinese natural language processing. *Proceedings of ACL*, pages 49–54, 2013. URL <https://www.aclweb.org/anthology/P/P13/P13-4009.pdf>.

Anand RAJARAMAN et Jeffrey David ULLMAN : *Mining of Massive Datasets*. Cambridge University Press, Cambridge, 2011. ISBN 9781139058452. URL <http://ebooks.cambridge.org/ref/id/CB09781139058452>.

Dominique RAYNAUD : L ' émergence de l ' espace perspectif : Effets de croyance et de connaissance. In Odile JACOB, éditeur : *Les Espaces de l'Homme, Symposium annuel du Collège de France*,, 2005.

Clément RENAUD : Facebook face aux réseaux sociaux chinois, 2011. URL http://www.lemonde.fr/technologies/article/2011/04/27/facebook-face-aux-reseaux-sociaux-chinois_1513168_651865.html.

Bernhard RIEDER : Networked control : Search engines and the symmetry of confidence. *International Review of Information Ethics*, 3:26–32, 2005. URL http://www.i-r-i-e.net/inhalt/003/003_full.pdf#page=28.

Garry ROBINS : A tutorial on methods for the modeling and analysis of social network data. *Journal of Mathematical Psychology*, pages 1–14, mars 2013. ISSN 00222496. URL <http://linkinghub.elsevier.com/retrieve/pii/S0022249613000126>.

Daniel M. ROMERO, Brendan MEEDER et Jon KLEINBERG : Differences in the Mechanics of Information Diffusion Across Topics : Idioms, Political Hashtags, and Complex Contagion on Twitter. In *Proceedings of the 20th ACM International World Wide Web Conference*, Hyderabad, 2011. ACM. URL <http://www2011india.com/proceeding/proceedings/p695.pdf>.

Nicole RONALD, Virginia Dignum, Catholijn JONKER, Theo ARENTZE et Harry TIMERMANS : On the engineering of agent-based simulations of social activities with social

networks. *Information and Software Technology*, 54(6):625–638, 2012. ISSN 09505849.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0950584911002485>.

Richard B ROTHENBERG : Commentary : Sampling in Social Networks. *Connections*, 18 (1):104–110, 1995.

Matthew A RUSSEL : *Mining the Social Web*. O'Reilly Media, 2011.

G. SALTON, A. WONG et C. S. YANG : A vector space model for automatic indexing, 1975. ISSN 00010782.

John SCOTT : Social Network Analysis. *Sociology*, 22(1):109–127, 1988. ISSN 0038-0385, 1469-8684. URL [http://soc.sagepub.com/content/22/1/109\\$delimiter\\$026E30F\\$nhttp://soc.sagepub.com/content/22/1/109.full.pdf\\$delimiter\\$026E30F\\$nhttp://soc.sagepub.com/content/22/1/109.short](http://soc.sagepub.com/content/22/1/109$delimiter$026E30F$nhttp://soc.sagepub.com/content/22/1/109.full.pdf$delimiter$026E30F$nhttp://soc.sagepub.com/content/22/1/109.short).

Sebastian SEUNG : *Connectome : How the Brain's Wiring Makes Us Who We Are*. Houghton Mifflin Harcourt, 2012. ISBN 0547508174. URL <http://books.google.com/books?id=GXwEuoYl3wQC&pgis=1>.

Jonathan SILVERTOWN : A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9):467–471, 2009. URL <http://www.ncbi.nlm.nih.gov/pubmed/19586682>.

Gilbert SIMONDON : *Du mode d'existence des objets techniques*. Aubier philosophie, ISSN 1248-8658. Aubier, 1989. ISBN 2700734149. URL <http://www.sudoc.abes.fr/DB=2.1/SRCH?IKT=12&TRM=056398972>.

James SOMERS : The Man Who Would Teach Machines to Think. *The Atlantic*, 2013. URL <http://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/>.

Pascal SOUCY et Guy W MINEAU : Beyond TFIDF Weighting for Text Categorization in the Vector Space Model. *Corpus*, 19(1):1130–1135, 2005. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.4957&rep=rep1&type=pdf>.

C SOUKUP : Computer-mediated communication as a virtual third place : building Oldenburg's great good places on the world wide web. *New Media & Society*, 8 (3):421–440, 2006. ISSN 14614448. URL <http://nms.sagepub.com/cgi/doi/10.1177/1461444806061953>.

Isabelle STENGERS : Penser les sciences par leur milieu. *Rue Descartes*, 41(3):41, août 2003. ISSN 1144-0821. URL http://www.cairn.info/article.php?ID_ARTICLE=RDES_041_0041.

Isabelle STENGERS, Brian MASSUMI et Erin MANNING : Histoire du milieus : entre macro et mésopolitique. *Inflexions*, (3):1–17, 2009.

Alexandre STEYER, Renaud GARCIA-BARDIDIA et Pascale QUESTER : Online Discussion Groups as Social Networks : An Empirical Investigation of Word-of-Mouth on the Internet. *Journal of Interactive Advertising*, 6(2):51–60, 2006. ISSN 15252019. URL <http://www.jiad.org/article80>.

Bernard STIEGLER : Temps et individuation technique, psychique, et collective dans l'oeuvre de Simondon. *Intellectica*, 1-2(26-27):241–256, 1998.

Bernard STIEGLER : The Aufklärung in the Age of Philosophical Engineering. In *WWW12 Conference Proceedings*, 2012.

Ferhat TAYLAN : Centralité et modicabilité. *Le Portique*, 25(25), août 2010. ISSN 1283-8594. URL <http://leportique.revues.org/index2475.html>.

Nigel THRIFT et Shaun FRENCH : The automatic production of space. *Transactions of the Institute of British Geographers*, 27(3):309–335, septembre 2002. ISSN 0020-2754. URL <http://www.blackwell-synergy.com/links/doi/10.1111/1475-5661.00057>.

Paul M TORRENS : Geography and computational social science. *GeoJournal*, 75(2):133–148, 2010. ISSN 03432521. URL <http://www.springerlink.com/index/10.1007/s10708-010-9361-y>.

Lokman TSUI : An Inadequate Metaphor : The Great Firewall and Chinese Internet Censorship. *Global Dialogue*, 9(1-2), avril 2007. URL <http://www.worlddialogue.org/content.php?id=400>.

Paola TUBARO et Antonio A CASILLI : An ethnographic seduction” : how qualitative research and Agent-based models can benefit each other. *Bulletin de Méthodologie Sociologique*, 106(1):59–74, 2010. URL http://dx.doi.org/10.1177/0759106309360111http://hal.archives-ouvertes.fr/docs/00/66/14/25/PDF/EthnoSeductionBMSvol106n1pp116_draft.pdf.

Paul VIRILIO : *La bombe informatique*. Galilée, 1998. ISBN 2718605073. URL <http://www.amazon.fr/La-bombe-informatique-Virilio-P/dp/2718605073>.

Hunter WALKER : Obama Campaign Deploys Cat Meme to Get Out the Vote in Ohio, 2012. URL <http://politicker.com/2012/11/obama-campaign-deploys-cat-meme-to-get-out-the-vote-in-ohio/>.

Cara WALLIS : *Technomobility in China : Young Migrant Women and Mobile Phones (Critical Cultural Communication)*. NYU Press, 2013. ISBN 0814795269. URL <http://>

//www.amazon.com/Technomobility-China-Critical-Cultural-Communication/dp/0814795269.

Greg WALTON : China ' s Golden Shield : Corporations and the Development of Surveillance Technology in China. Rapport technique, International Centre for Human Rights and Democratic Development, 2001.

Lin WANG et Brendan C. WOOD : An epidemiological approach to model the viral propagation of memes. *Applied Mathematical Modelling*, 35(11):5442–5447, novembre 2011. ISSN 0307904X. URL <http://linkinghub.elsevier.com/retrieve/pii/S0307904X11002824>.

Colin WARE : *Information Visualization : Perception for Design*, volume 22 de *The Morgan Kaufmann series in interactive technologies / ed. Stuart Card. - San Francisco : Morgan Kaufmann*. Morgan Kaufmann, 2004. ISBN 1558608192. URL http://books.google.com/books?id=ZmG_FiqqyqgC&pgis=1.

Tetsurô WATSUJI, Augustin BERQUE, Pauline COUTEAU et Kurodamm AKINOBU : *Fûdo : Le milieu humain*. CNRS, Paris, 2011. ISBN 2271071372.

L. WENG, A. FLAMMINI, A. VESPIGNANI et F. MENCZER : Competition among memes in a world with limited attention. *Scientific reports*, 2:335, janvier 2012. ISSN 2045-2322. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3315179/>&tool=pmcentrez&rendertype=abstract.

Al B. WESOLOWSKY et Edward R. TUFTE : The Visual Display of Quantitative Information. *Journal of Field Archaeology*, 19(1):121, 1992. ISSN 00934690. URL <http://www.jstor.org/stable/530384?origin=crossref>.

Michel WIEVIORKA : *L'impératif numérique ou La nouvelle ère des sciences humaines et sociales ?* CNRS, 2013. ISBN 978-2-271-07981-7. URL <http://lectures.revues.org/12677>.

Philipp WINTER et Stefan LINDSKOG : How China Is Blocking Tor. *Arxiv preprint arXiv12040447*, (March), 2012. URL <http://arxiv.org/abs/1204.0447>.

An xiao MINA : A Curated History of the Grass Mud Horse Song, 2012. URL <http://www.88-bar.com/2012/02/a-curated-history-of-the-grass-mud-horse-song/>.

Guobin YANG : Activism beyond Virtual borders : Chinese Internet-mediated Networks and information politics in China. *Command Lines : The Emergence of Governance in Global Cyberspace*, 7, 2008.

Louis YU, Sitaram ASUR et Bernardo A. HUBERMAN : What Trends in Chinese Social Media. *In The 5th SNA-KDD Workshop'11 (SNA-KDD'11)*, San Diego CA USA, juillet 2011. ACM. ISBN 9781450302258. URL <http://www.ssrn.com/abstract=1888779><http://arxiv.org/abs/1107.3522>.

Louis YU et Valerie KING : Guanxi in the Chinese Web-A Study of Mutual Linking. *In WWW2008*, volume 30, pages 35–41, Beijing, China, 2008. A postdoctoral researcher in the Social Computing Lab at HP Labs., ACM. URL <http://portal.acm.org/citation.cfm?doid=1367497.1367706>.

Xun ZHAO, Feida ZHU, Weining QIAN et Aoying ZHOU : Impact of Multimedia in Sina Weibo : Popularity and Life Span. *In The Joint Conference of the Sixth Chinese Semantic Web Symposium and the First Chinese Web Science Conference (CSWS & CWSC '12)*, pages 1–11, 2012. URL <http://www.mysmu.edu/faculty/fdzhu/paper/CWSC'12.pdf>.

M. a. ZOOK : Old Hierarchies or New Networks of Centrality ? : The Global Geography of the Internet Content Market. *American Behavioral Scientist*, 44(10):1679–1696, juin 2001. ISSN 0002-7642. URL <http://abs.sagepub.com/cgi/doi/10.1177/00027640121958113>.

MA ZOOK et Mark GRAHAM : From cyberspace to DigiPlace : Visibility in an age of information and mobility. *Societies and cities in the age of instant access*, pages 241–254, 2007. URL http://link.springer.com/content/pdf/10.1007/1-4020-5427-0_16.pdf.

Arkaitz ZUBIAGA, Damiano SPINA et Raquel MARTÍNEZ : Classifying Trending Topics : A Typology of Conversation Triggers on Twitter. *Evaluation*, pages 8–11, 2011. URL <http://nlp.uned.es/~damiano/pdf/zubiaga2011trendingtopics.pdf>.

Table des figures

1.1.1 Jingjing et Chacha, les policiers de l'Internet chinois	8
1.2.1 Capture d'écran de Sina Weibo	12
1.2.2 sondage Weibo concernant l'accident de train de Wenzhou	13
1.2.3 Cours de l'action SINA au Nasdaq entre 2009 et Février 2014	14
1.2.4 Commentaires supprimés par Sina	17
1.3.1 Hexie	23
2.1.1 la première mention du smiley par Scott Fahlman	35
2.1.2 Un jeu d'émoticônes de Weibo	36
2.1.3 Cycle de vie d'un même Internet	37
2.1.4 Exemples de mèmes internet	38
2.2.1 Exemples de réponses à la question " <i>What are some quintessential Indian memes ?</i> " D'après http://www.quora.com/India/What-are-some-quintessential-Indian-memes Consulté le 12/08/2013 à 0 :41 GMT+8	41
2.2.2 " <i>High Expectations Indian Father</i> " d'après Quora.com	41
2.2.3 " <i>High Expectations Asian Father</i> " d'après Quora.com	42
2.3.1 Typologie des mèmes	49
2.3.2 <i>Caonima</i> : un animal mythique du web chinois	51
2.3.3 Lolcat utilisé lors la campagne d'Obama	53
2.3.4 Détournement de tracts hongkongais anti-chinois	54
2.3.5 La star du X japonaise Sola Aoi sur Sina Weibo	55
2.3.6 Extrait d'un spam du type <i>Nigerian Scam</i>	56
3.1.1 Triptyque des pratiques scientifiques	61
3.2.1 Représentation d'un réseau sous forme de graphes.	66
3.2.2 réseau social multi-couches	67
3.2.3 Tableau récapitulatif de méthodes d'analyse de données de réseaux sociaux en ligne.	71
3.3.1 Algorithme de reconnaissance de même (clustering) (<i>Ferrara et al., 2013</i>) . .	77
3.3.2 Volume de tweets et hashtags pour la semaine 1	80

3.3.3 Distribution des 429 hashtags sélectionnés	81
3.3.4 3 modèles de réseau	83
3.3.5 Graphe de diffusion de hashtags sur Twitter d'après (Weng <i>et al.</i> , 2012) . .	84
3.3.6 Visualisation simple du hashtag “WeicoPlus”	85
3.3.7 Visualisation simple des conversations autour du hashtag Veuve d'enfant unique	86
3.3.8 Visualisation des réseaux de conversation	87
3.3.9 Photo de Tianjin durant la pluie torrentielle en Juillet 2012	88
3.3.10 Exemple 1 : Tianjin Baoyu	89
3.3.11 “Shidu Muqin”	90
3.3.12 “Piaowudong nuzui”	92
3.4.1 Tableau de bords requêtes par mots-clés	95
3.4.2 Interface d'exploration et de visualisation des données	100

Annexes

Première partie

Liste des mèmes

