

Les mèmes Internet ont-ils un mode de propagation spécifique ? Analyses à partir de topogrammes de mèmes de l'Internet chinois

Clément Renaud, Post-Doctorant, Telecom ParisTech, UMR I3

Clement.Renaud@telecom-paristech.fr

Valérie Fernandez, Professeur, Telecom ParisTech, UMR I3

Valerie.Fernandez@telecom-paristech.fr

Gilles Puel, MdC HDR, Université de Toulouse, LEREPS

Gilles.Puel@univ-tlse1.fr

Résumé

Contenus humoristiques circulant très largement sur la Toile, les mèmes Internet représentent aujourd'hui un phénomène notable parmi les usages des réseaux sociaux en ligne. Sur le réseau social chinois Sina Weibo, les mèmes sont devenus une forme de discussion répandue aux particularités encore méconnues. Nous proposons ici d'observer la diffusion de différents types de contenus circulant sur Sina Weibo afin de mieux comprendre les spécificités des mèmes Internet. Pour ce faire, nous procédons à la visualisation des dynamiques sémantiques, sociales et spatiotemporelles d'après un échantillon de contenus : marketing, actualité, humour et scandale politique. Nous proposons un modèle appelé *topogramme* afin de structurer la lecture des résultats produits par les outils de visualisation et d'analyse des données. Nous montrons comment ce dispositif d'analyse peut être utile pour définir des modèles socio-sémantiques et spatiotemporels des contenus Internet à travers l'exemple particulier des mèmes Internet sur Sina Weibo.

Mots clés : mèmes Internet, outils d'analyse des réseaux sociaux, big data, graphes de visualisation, Sina Weibo, topogramme

Composante importante des échanges en ligne, les mèmes Internet sont un sujet de recherche foisonnant dans le domaine des sciences sociales. Ces petites blagues, images et messages partagés sur la Toile offrent en effet un objet intéressant pour comprendre notamment les parcours de la dissémination en ligne, mais également pour analyser les expressions et attitudes culturelles. Ce papier cherche à rendre compte de la diversité de leurs modes de propagation en Chine et plus spécifiquement des formes particulières qui accompagnent leur diffusion selon des modèles encore méconnus.

Nous introduisons la notion de *topogramme* pour décrire les dynamiques de diffusion de différents contenus en ligne. Défini dans le champ des sciences cognitives comme une carte heuristique, nous utilisons ici le topogramme à la fois comme outil d'observation et de représentation graphique. Dans notre dispositif, un topogramme donne à voir différentes dimensions de la diffusion des contenus sélectionnés : le réseau sémantique associé à une analyse lexicale, la dynamique conversationnelle des échanges entre les utilisateurs, la structure géographique de ces échanges ainsi que leurs évolutions temporelles.

La première partie de ce papier introduira le concept de même Internet, en présentant son histoire et ses usages les plus courants. Nous nous intéresserons ensuite au service de microblogging chinois Sina Weibo en présentant notamment le jeu de données du projet Weiboscope utilisé dans cette étude. Dans une troisième partie, nous décrirons le dispositif d'analyse et de visualisation développé pour ce travail en introduisant la notion de topogramme. La quatrième partie proposera une série d'exemples montrant la diffusion de différents contenus (marketing, publicité, faits divers, mèmes humoristiques). Après avoir discuté les particularités de la diffusion de chacun de ces contenus, nous concluons sur les limites et opportunités dans l'utilisation des topogrammes pour l'observation de contenus en ligne.

LES MÈMES INTERNET

Origines du concept de mème

Le concept de mème trouve son origine dans l'évolutionnisme culturel de Dawkins (1989) qui voit dans le mème une « *unité distincte de la pensée* » définissant une culture. Défini par analogie au *gène* biologique, le *mème* se

comporte selon lui comme un « *gène égoïste* » agissant de manière isolée et distincte, diffusant une “culture”. Souvent représenté par l’image du virus, ce “*réplicateur culturel*” (Blackmore, 2001) serait doté de la capacité de se reproduire lui-même, se propageant rapidement de cerveau en cerveau. Situé à l’avant-plan d’une lutte pour la survie et la fécondité des idées, il serait une extension à part entière du vivant au-delà du biologique (Bloom, 2012).

Si la construction théorique sur la transmission culturelle peut paraître attrayante, le concept de mème reste néanmoins incapable d’asseoir une réelle légitimité scientifique, notamment de par l’absence d’études de cas complètes et qui l’empêche de dépasser la simple analogie (Jouxtel, 2014). Largement marginal dans le domaine scientifique, le terme a néanmoins été adopté pour décrire un type particulier de contenus se diffusant très rapidement sur Internet. Les “*Internet memes*” ont ainsi fait leur entrée dans les dictionnaires anglo-saxons (*Oxford dictionary* et *Merriam-Webster* le classe parmi les 10 mots marquants de l’année 2012).

Les mèmes Internet

Un mème Internet apparaît bien souvent comme une forme humoristique au contenu comique, voire absurde. Parfois appelés “contenus viraux”, les mèmes se propagent sur le Web par vague rapide de popularité croissante, avec l’appui notamment des services de réseaux sociaux (Shifman, 2014). Le cycle de vie informationnel suit un schéma assez caractéristique des modes de diffusion en ligne (voir fig. 1).

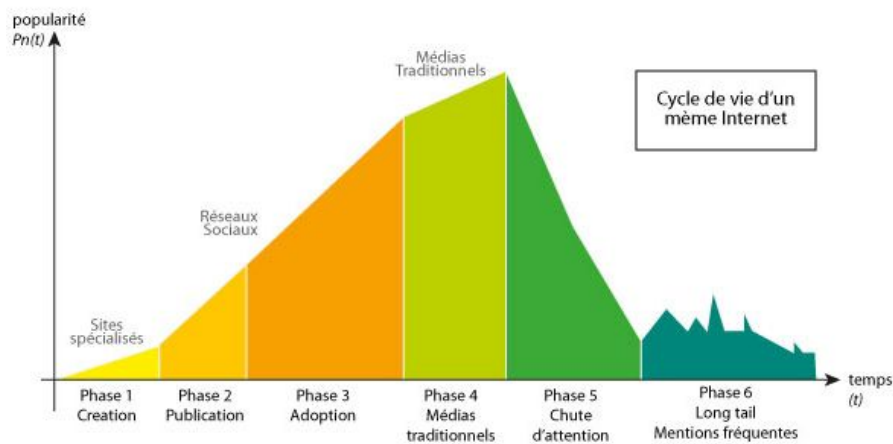


Fig 1 - Cycle de vie informationnel d'un mème Internet, réalisée d'après une compilation des études existantes dans la littérature. (source: auteur)

Similaire à une rumeur, les mèmes sont la plupart du temps mis en circulation de façon très localisée sur un petit nombre de sites spécialisés (phase 1), 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) (phase 2). Les réseaux sociaux agissent alors comme une chambre d'écho, qui détermine si le "proto-mème" (Ferrara, 2013) encore en devenir deviendra mème ou restera simple message isolé. Durant cette phase souvent nommée *adoption* (phase 3), le mème entre en concurrence avec d'autres informations sur les réseaux sociaux (Weng & al., 2012). 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) (phase 4). Ensuite, l'attention envers le mème décroît fortement et rapidement (phase 5). La présence épisodique de citations maintient l'existence du mème apparente dans les groupes définis lors de la phase d'adoption (phase 6).

L'évolution du volume de la diffusion est donc un des premiers indicateurs permettant de qualifier un mème Internet. Néanmoins, il est impossible de donner une estimation du volume minimum nécessaire pour devenir "mème" tant ce chiffre dépend de la population étudiée et des groupes sociaux qu'elle constitue. Les utilisateurs tendent plutôt à choisir les mèmes selon le réseau

social auquel ils appartiennent et le moment d'exposition, produisant ainsi une grande hétérogénéité des mèmes dans le réseau (Weng & al., 2012).

Typologie des mèmes

Knobel et Lankshear (2007) ont identifié trois caractéristiques communes à une centaine de mèmes parmi les plus largement diffusés d'Internet :

- **Humour** : Le mème doit posséder une dimension comique et accrocheuse
- **Intertextualité** : Le mème met en jeu un ou des renvois à d'autres éléments culturels ou textuels, souvent implicites.
- **Juxtaposition atypique** : C'est la mise en relation de plusieurs objets visuels ou sémantiques sans corrélations apparentes fait du même un objet intéressant.

L'étude porte seulement sur un type précis de mèmes à caractère plutôt comique, ignorant les discussions plus sérieuses, d'ordre politique notamment. Les différentes intentions des discours et actes d'énonciation à l'œuvre dans un mème peuvent nous aider à dresser une première typologie non-exclusive des mèmes (voir Tableau 1). Les différentes pratiques entourant les mèmes sont souvent la reconduction sous une forme numérique de formes de discours préexistantes (publicité, propagande politique, rumeurs, etc.). La forme et l'échelle de la diffusion ainsi que l'importance du caractère comique ou humoristique sont néanmoins les différences importantes à prendre en compte.

Contenus	Exemples célèbres et références
Absurdiste, humour	Les mèmes comiques sont la catégorie de mèmes la plus répandue avec notamment les vidéos ou images de chats (LOLcats) qui constituent un des contenus les plus visionnés de l'Internet (Shifman, 2014).
Actualité, satire, commentaire social	Souvent étudiés, ces "graffitis du web censuré" (Mina, 2012) mettent en œuvre le comique et la satire comme stratégie de contournement des interdits médiatiques pour discuter les faits divers et d'actualité (Zuckerman, 2010).

Publicité, marketing viral	Outil important des campagnes de publicité et de promotion (Bergstrom, 2012), en utilisant notamment les hashtags (Shi, 2011).
Marketing politique, soutien, petition	L'importance des réseaux sociaux dans les stratégies de communication politiques a fait des mèmes un outil de campagne, notamment pour Obama (Walker, 2012), ou de diffusion virale des pétitions (Adamic & al., 2013).
Fan clubs, adoration	Sina Weibo s'est historiquement développé par l'acquisition de blogs de stars chinoise et asiatique, puis plus récemment américaine.
Hoax, spam	Emails dits de fraude 4-1-9 ou Nigerian scam cherchant à extorquer de l'argent au destinataire a trouvé sa reproduction sur les réseaux sociaux.

Tableau 1 - Les types d'utilisations des mèmes Internet selon la littérature.

LES CONTENUS SUR SINA WEIBO

Les études consacrées aux mèmes sur l'Internet chinois et plus particulièrement sur le réseau social Sina Weibo se sont focalisées sur les usages commerciaux ou politiques des mèmes, considérées comme catalyseurs pour la mobilisation des publics (Wang, 2012). Le présent papier s'intéresse également aux spécificités de cette plate-forme, mais cherche davantage à proposer une démarche d'analyse des modèles de diffusion de différents contenus transposable à d'autres réseaux sociaux.

Sina Weibo : industrie de contenus de réseau social

Alors qu'un blocage officiel s'applique en Chine sur les plus célèbres services Internet américains, de nombreux outils se sont développés pour répondre aux besoins et intérêts des internautes chinois. Historiquement, l'interdiction de Twitter en Juin 2009 a fait de Sina Weibo un des services de microblog les plus utilisés. Favori de la classe moyenne montante et des jeunes travailleurs urbains, Sina Weibo contraste avec Tencent Weibo utilisé typiquement par des personnes aux revenus plus faibles accédant au web depuis leurs mobiles

(Sullivan, 2012). À l'origine similaire à Twitter (une messagerie de 140 caractères), le service a peu à peu évolué pour intégrer une vaste offre de contenus, centrée autour de nombreuses personnalités publiques célèbres (Gao, 2012).

Le jeu de données Weiboscope

La présente analyse s'appuie sur un jeu de données collecté par le *Journalism and Media Studies Center* de l'Université de Hong Kong (JMSC) lors de son projet *Weiboscope*. Il s'agit d'un échantillonnage aléatoire de messages effectué quotidiennement durant l'année 2012 sur un panel d'environ 350 000 utilisateurs ayant au moins 1000 *followers* (Fu & 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 - parfois sur ordre du gouvernement chinois. Ce jeu de données a été mis à disposition¹ sous une forme réduite et anonymisée. La collection des données se base sur une série d'utilisateurs (génération aléatoire d'identifiants dont l'existence est validée) afin de donner *“une image représentative des usages et utilisateurs de Sina Weibo [alors que] les études auparavant limitées à des analyses non-aléatoires (...) se cantonnaient aux utilisateurs les plus populaires”* (ibid, 2013). Ainsi, plutôt que de considérer uniquement les “stars” de Sina Weibo, cet échantillon s'attache à refléter également les pratiques des utilisateurs “lambda”.

Sélection d'un échantillon de différents contenus

Les hashtags associés à des activités commerciales, de loisir ou de divertissement (Renaud, 2014) ont été écartés en raison du manque de diversité dans leurs modèles de diffusion. Les usages majoritaires de Sina Weibo correspondent relativement bien à ceux des autres mass-médias, sociaux et traditionnels de par le monde (produits culturels, divertissement, publicité, etc.). Les discussions parlent également de la vie de tous les jours : les embouteillages dans les cités chinoises est un des sujets les plus discutés de l'année 2012.

Nous avons sélectionné un ensemble de contenus de nature différente : campagnes de marketing, faits d'actualité, scandales politiques et mêmes

¹ Disponible à l'adresse : <http://147.8.142.179/datazip/>

comiques (voir Tableau 2). En s'appuyant sur les littératures scientifique et secondaire ainsi que sur un relevé de terrain, nous avons sélectionné cinq évènements ayant connu une forte diffusion sur *Sina Weibo* en 2012, afin de comparer leurs modèles de diffusion. Pour chacun, nous avons déterminé une période d'activité de quatre semaines et un ensemble de mots-clés, utilisés comme critères pour une recherche plein-texte dans l'intégralité du corpus.

Type	Description
Absurdité, humour	Le mème « Dufu est très occupé » (杜甫很忙) détourne une calligraphie bien connue du poète Du Fu (杜甫) de la dynastie Tang (712-770) en le mettant en scène dans des situations de la vie au 21 ^{ème} siècle. (<i>Mots-clés</i> : 杜甫很忙, 李白不服气了)
Absurdiste, humour	La phrase « Yuanfang, qu'en penses-tu? » (元芳,你怎么看?) a été la source d'innombrables jeux de mots en Chine. Issu d'un dialogue d'une série policière, elle est devenue aussi célèbre que le "élémentaire" de Sherlock Holmes au Dr. Watson. (<i>Mots-clés</i> : 元芳,你怎么看)
Fan clubs, adoration	The Voice of China (中国好声音) désigne la première saison d'une émission de télé-crochet musical diffusée en Chine depuis le 13 juin 2012 sur Zhejiang Television. Chaque épisode a été visionné en moyenne par 7 millions de téléspectateurs et plus de 70 millions d'internautes. (<i>Mots-clés</i> : 中国好声音, 吴莫愁)
Actualité, satire, commentaire social	Sextape désigne ici un scandale né d'une vidéo intitulée " <i>Lei, the secretary who accepts sex bribes</i> " montrant Lei Zhengfu, secrétaire du Parti du district Beibei à Chongqing en compagnie d'une de ses maîtresses. A peine 63h après la publication de cette vidéo, Mr. Lei sera mis à pied de ses fonctions politiques. (<i>Mot-clé</i> : 雷政富)
Actualité, satire, commentaire social	Qiegao (切糕) est un fait divers devenu débat de société autour de la condition du peuple Uyghur. En Décembre 2012, une dispute concernant éclate entre un villageois du Hunan et un vendeur originaire de la région du Xinjiang. L'altercation dégénère en bagarre générale, occasionnant des dégâts estimés par la police à 160 000 yuans (environ 19 500€), somme jugée exorbitante par les internautes. (<i>Mot-clé</i> : 切糕)

Tableau 2 - Liste des mèmes sélectionnés pour l'étude

MÉTHODOLOGIE D'ANALYSE ET DE VISUALISATION DE DONNEES

Face à la diversité des contenus disponibles en ligne, plusieurs recherches se sont intéressées à comprendre les modèles de diffusion et de popularisation des mèmes (Leskovec & al., 2009; Weng, 2012). Notre travail utilise le *topogramme* comme outil d'observation pour décrire les similarités et différences des structures de diffusion en associant plusieurs dimensions : lexicale, sociale et spatio-temporelle.

Topogrammes: modélisation de la diffusion

À l'instar des géogrammes du paysage de Berque (1999) ou des chorèmes de l'espace de Brunet (1980), les topogrammes nous permettent ici de considérer sous un jour commun des faits et objets digitaux dissemblables. En introduisant ce concept, nous cherchons notamment à décrire plus précisément des lieux du Web en donnant à voir les caractéristiques sémantiques (graphe lexical), conversationnelles (graphe social) et spatio-temporelles (carte et ligne de temps) d'un fait sélectionné. Si les médias traditionnels ont cherché à définir les territoires du discours, l'enjeu stratégique des médias du web se situe davantage dans la définition d'espaces pour l'énonciation. Marketing, communication politique, journalisme ou activisme social, la fonction du média dans une économie de l'attention devenue hyper-compétitive (Weng & al., 2012) n'est plus communicative (“dire”) mais performative : “faire dire” ou plutôt “faire faire” (cliquer, liker, acheter...). Bâtir l'image d'une marque, d'une entreprise, d'une personne ou d'un fait public nécessite aujourd'hui la construction de réseaux sémantiques, conversationnels (sociaux) et territoriaux définissant les fondations d'un espace de “participation” où se déroule l'énonciation - l'individuation. Ainsi, la gestion stratégique d'un réseau de symboles, de personnes et de lieux devient l'enjeu du média pour un contrôle de la topologie des espaces d'expression. En représentant la circulation de différents contenus, les *topogrammes* nous aide à considérer la façon dont sont décrits et actualisés ces espaces particuliers du Web.

Réseaux de mots, d'utilisateurs et de lieux

Une pratique répandue dans l'analyse des contenus en ligne est la représentation des discussions sous la forme d'un graphe de diffusion dans le but d'identifier les structures particulières des conversations (Nettleton, 2013). Le modèle émetteur-récepteur (point-trait) décrit par ce type de visualisation de réseaux reste néanmoins une limite importante. Pour une visualisation intéressante, le rôle de l'espace perceptif dans la représentation doit être élaboré pour permettre une restitution des enjeux et tensions complexes présentes lors des actes de communication en ligne. Une telle perspective implique donc une méthodologie renouvelée et le développement d'outils adaptés permettant d'observer les actes de parole qui constitue la diffusion. Il ne s'agit plus simplement d'étudier un réseau d'échanges entre utilisateurs, mais plutôt de s'interroger sur les relations entre différents réseaux et objets, de nature souvent diverses.

Le *topogramme* propose donc d'observer trois aspects importants pour comprendre l'organisation et les motifs émergents de différents phénomènes de diffusions en ligne : (1) champ sémantique, (2) dynamique des discussions et (3) dimension spatio-temporelle de ces échanges.

- **(1) le champ sémantique** d'un thème ou contenu est constitué des mots qui sont prononcés lors de sa diffusion. Les associations et jeux de mots d'un thème dessinent des liens entre des signifiants souvent improbables. En représentant les co-occurrences de mots dans une même phrase, nous pouvons créer un graphe sémantique montrant les relations entre ces mots-clés sous la forme d'un réseau.

- **(2) les dynamiques conversationnelles** : au-delà des mots, un thème se constitue sous la forme d'une conversation où différents acteurs discutent, commentent et se saisissent des actions disponibles sur une plateforme web (like, retweets, etc.). Pour chaque thème, nous constituons un graphe dirigé des conversations où chaque utilisateur est un nœud lié aux autres par ses interactions (mentions, citations et retweets). Une fois l'ensemble du graphe constitué, nous utilisons l'algorithme dit de Louvain (Blondel & al, 2008) pour identifier des communautés dans la discussion.

- **(3) son existence spatio-temporelle** : ces discussions possèdent également une existence physique, que nous considérerons ici par celle des utilisateurs. Le jeu de données *Weiboscope* comprend les localités mentionnées par les utilisateurs dans leurs profils et nous permet ainsi de reconstituer un graphe

socio-géographique où les interactions entre utilisateurs sont retranscrites sous formes d'interactions entre les provinces renseignées dans leurs profils.

Choix technologiques

Pour réaliser cette cartographie particulière, il n'existe pas d'outils directement disponibles. Nous avons donc entrepris de développer entièrement un dispositif spécifique afin de pouvoir considérer l'ensemble de ces graphes et de leurs relations. Pour extraire ces connaissances de la masse des données étudiées, nous avons construit à la fois un système d'analyse de texte en langue chinoise et une interface de représentation propice à l'exploration et à la découverte des phénomènes². Cet espace perceptif nous a permis de mettre en relation des entités dissemblables comme des utilisateurs, des lieux, des mots, ainsi que d'observer leurs relations mutuelles.

La langue écrite chinoise n'utilisant pas d'espace pour séparer les mots d'une phrase, elle nécessite d'être segmentée en mots avant de pouvoir être traitée (un nom commun chinois prend souvent la forme d'un, deux ou trois caractères). Nous avons utilisé la librairie *Jieba*³ pour mener à bien cette tâche.

Pour le graphe lexical, chaque mot est un nœud dans le réseau. La co-occurrence de deux mots dans une même phrase définit une relation entre ces mots. L'ensemble de ces relations constitue le réseau sémantique. Nous limitons également le graphe aux 500 mots les plus utilisés (dont les occurrences sont les plus nombreuses) afin de préserver la lisibilité lors de l'étape suivante de visualisation.

Les actions effectuées par les utilisateurs lors des conversations (mentions, citations et retweets) nous permettent de reconstituer le graphe dirigé et pondéré des échanges. Chaque utilisateur y est un *node* et chaque interaction un *edge*. La taille du graphe est limitée aux utilisateurs ayant effectué au moins deux échanges avec seulement les 500 utilisateurs les plus actifs.

Le jeu de données *Weiboscope* donne également accès aux localités mentionnées par les utilisateurs dans leurs profils, nous permettant de retracer la géographie de la diffusion. Cette information doit néanmoins être utilisée

2. Ce logiciel réutilisable est disponible à l'adresse <http://github.com/topogram/topogram>, consulté le 1er Janvier 2015 à 16 :45

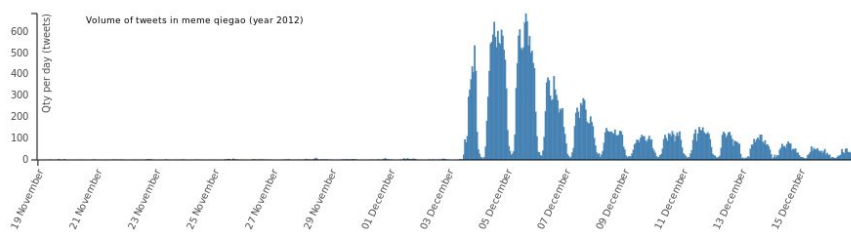
3. Jieba, <https://github.com/fxsjy/jieba>, consulté le 14 Mars 2014 à 18 :30

avec précaution : les utilisateurs peuvent indiquer ce que bon leur semble et mettent rarement à jour leurs informations de profil lors de leurs déplacements. L'heure d'envoi du message est utilisée pour organiser les données selon un axe temporel.

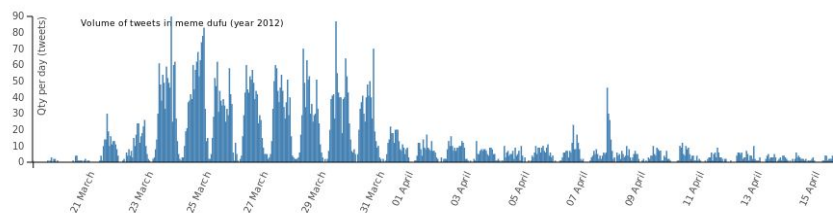
RÉSULTATS: GÉNÉRICITÉ ET SPÉCIFICITÉS DES MODES DE PROPAGATION

Structures temporelles

Dans un premier temps, nous avons choisi de considérer les différentes structures temporelles des mèmes. Ces graphes donnent à voir différentes formes représentatives : la “breaking news” du mème *qiegao* avec un départ abrupt de la courbe (fig. 2a) suivi d'une retombée rapide de l'attention.



2a - Qiegao



2b - Dufu

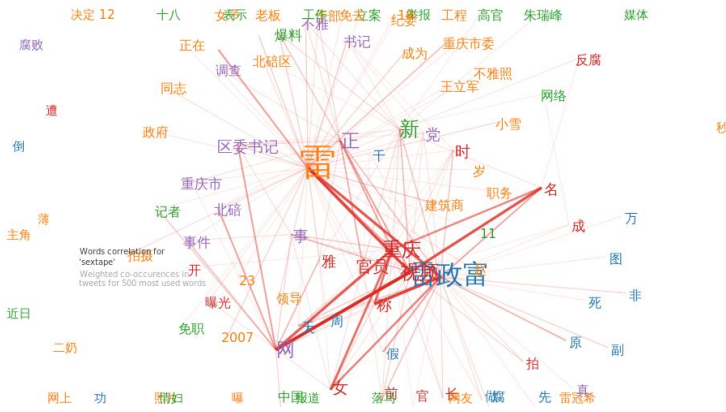
Fig 2 - Graphe temporel représentant le volume de messages échangés sur une durée de 4 semaines. La diminution régulière de la quantité de messages observés dans les graphes correspond à la baisse d'activité pendant les nuits. (source: auteur)

Le graphe du même absurdiste *dufu* (fig. 2b) monte quant à lui une lente montée d'attention puis un plateau de plusieurs jours. La blague semble durer et même connaître un regain d'intérêt une semaine plus tard. Cette tendance du même comique à durer se vérifie également avec *yuanfang*, qui mentionné très régulièrement sur une période de plusieurs mois. La discussion plus informative autour de la *sextape* croît doucement avant de retomber rapidement. *The Voice* suit le rythme des émissions avec un volume important lors des diffusions, dépassant très largement les trois autres mêmes présentés.

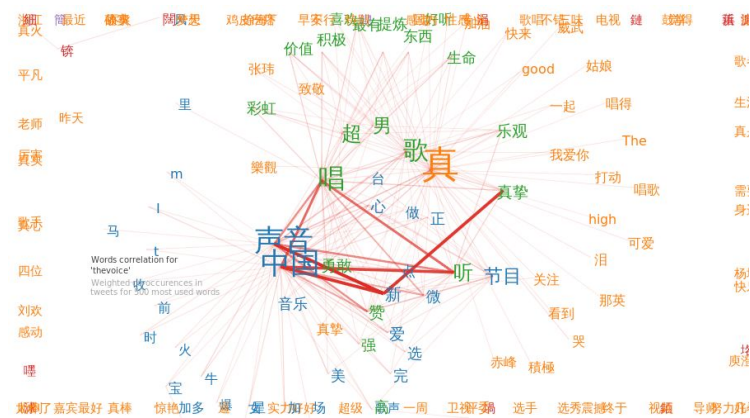
Structures lexicales

Nous avons choisi de représenter les réseaux de mots dessinés par ces échanges en ligne afin de comprendre comment l'activité symbolique se construit autour d'interactions entre signifiés particuliers⁴. Les graphes des mêmes absurdistes *dufu* et *yuanfang* correspondant aux structures des jeux de mots : peu de mots centraux entourés d'une myriade de déclinaisons faits de mots de plus faible importance. La discussion concernant le fait divers politique *sextape* est structurée en communautés de mots plutôt bien définies (fig 3a) : l'une décrit les détails de l'affaire (noms de lieux et de personnes, verbes d'actions), l'autre se compose surtout d'adjectifs transcrivant le caractère houleux de l'histoire, la dernière est faite de mots plus anecdotiques. La structure lexicale qui entoure *The Voice* (fig 2b) semble très architecturée, avec des associations plus convenues (voix-chant, trac-courage, etc).

⁴ Les graphes de mots sont construits grâce aux co-occurrences de mots dans les messages. Si deux mots sont utilisés dans un même message, un edge est ajouté entre eux. La taille des mots représente le nombre de fois où ils sont été cités dans l'ensemble du corpus. Les couleurs définissent des communautés de mots calculées grâce à l'algorithme de Louvain. L'épaisseur des traits représente l'intensité des relations (leur nombre de co-occurrences dans le corpus). La proximité des mots est calculée sur la base de l'intensité de leurs relations par l'algorithme force de d3js (Bostock & al, 2013).



(a) Graphe sémantique du scandale de la "sextape" d'un officiel chinois



(b) Graphe sémantique de l'émission de télé-crochet The Voice

Fig 3 - Graphe sémantique représentant les co-occurrences entre les 500 mots les plus courant du corpus (mots cités dans un même message) - source: auteur

Structures conversationnelles

Les graphes de conversation⁵ rendent compte de la structure et de la taille des conversations, très variables : les discussions autour d'aspects politiques et sociétaux sont plus larges, alors que les autres semblent moins importantes. Dans le cas de *The Voice*, la conversation est organisée autour de peu d'acteurs qui centralisent les débats, à l'inverse de la discussion de société *Qiegao* (fig. 4a) où un vaste foyer de discussions agrège de nombreux utilisateurs très actifs. Le même absurdiste (fig. 4b) semble être composé d'utilisateurs très distants discutant peu entre eux.

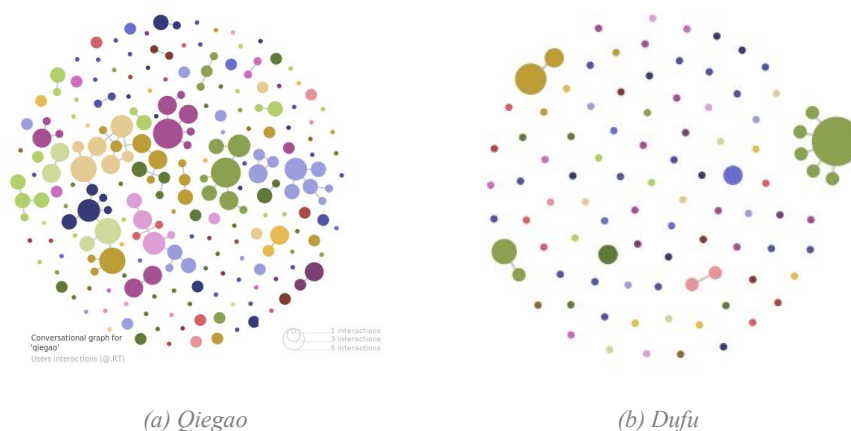


Fig 4 - Graphes conversationnels représentant les échanges en ligne (commentaires et citations) entre groupes d'utilisateurs - source: auteur

⁵ Chaque utilisateur est un point. Sa couleur montre son appartenance à une même communauté de conversation (algorithme de Louvain), les traits expriment des échanges supérieurs à 4 interactions. Enfin, la taille des points exprime le nombre de connexions entrantes et sortantes d'un utilisateur. La distance entre les groupes d'utilisateurs est calculée par l'algorithme force de d3.js (Bostock & al, 2013): plus ils sont proches dans la conversation, plus ils sont proches dans le graphe Les utilisateurs possédant moins de deux interactions ont été supprimés pour rendre le graphe lisible.

Structures géographiques

En associant chaque utilisateur à sa province d'origine⁶, nous pouvons reconstituer le réseau des interactions entre provinces sous la forme d'un graphe pondéré et dirigé. Afin d'éviter de reproduire simplement les foyers de population, les résultats ont été pondérés par le pourcentage de population totale représentée par chaque province dans le corpus utilisé. Pour des raisons de lisibilité, les liens entre les provinces sont dirigés non pas vers la capitale de la province mais vers le *centroid* (*barycentre*) de la forme géométrique qui la représente.

Dans le cas de *The Voice*, l'interaction s'origine dans la province du Zhejiang, où est localisé le compte officiel de la chaîne pour se diriger majoritairement vers Pékin et Shanghai. À l'inverse, les discussions entourant *Qiegao* semblent suivre des patterns beaucoup plus diversifiés. La ville de Guangzhou joue un rôle de diffuseur alors que les informations convergent vers Shanghai et Pékin. De nombreuses conversations se déroulent également dans l'ouest de la Chine, région d'origine des peuples Uyghur – sujets de la conversation.



Jour 1 (22/11/2012)
Faible propagation



Jour 2-5 (23-25/11/2012)
Diffusion depuis Pékin et Canton

⁶ Les données concernant les provinces d'origine des utilisateurs proviennent de leurs informations de profil. Elles permettent de représenter la dynamique des échanges. Les dynamiques mineures de contenus apparaissent car le volume d'informations pour chaque province. L'épaisseur des traits exprime le volume des échanges en pourcentage du total sur la période observée.



Jour 5+ 26/11-15/12/2012 - Large diffusion à travers le pays

Fig 5 - Évolution dans le temps des interactions entre utilisateurs (commentaires et citations), regroupés par province d'origine selon leur information de profil, pour le même "sextape" - source: auteur

Ce pattern particulier, qu'on retrouve dans d'autres faits de société comme *sextape* ou *biaoge*, illustre la plus grande liberté de ton des médias du sud de la Chine, instigateurs d'affaires importantes alors que les médias pékinois jouent souvent un rôle plus conservateur de diffuseur. Les cartes des mêmes absurdistes *dufu* et *yuanfang* montrent que si Pékin est bien présent, les échanges des premiers jours se font dans des provinces typiquement peu actives (respectivement Hubei et Yunnan). De nombreuses villes prennent part à la discussion et les échanges sont plutôt internes aux provinces.

DISCUSSION : SPÉCIFICITÉS DES MÈMES

La lecture des différents topogrammes permet d'ébaucher une première typologie des contenus en ligne selon les modalités de leurs diffusions. Néanmoins, cette étude expérimentale porte sur un nombre trop restreint de mèmes pour valider pleinement notre hypothèse méthodologique. Ces premières observations permettent toutefois de tirer des conclusions provisoires sur les spécificités des différents types de diffusion des contenus.

La diffusion des faits d'actualité est entourée de grands volumes de discussion, qui lui permettent de se propager rapidement. Dans le cas du fait divers, son démarrage brutal puis sa transformation par le commentaire en débat de société parfois brûlant se caractérise par un grand nombre de communautés très actives dans la discussion pendant un laps de temps plutôt court. La conversation s'oriente généralement autour de termes simples et structurés selon des chemins assez typiques. Les conversations autour de faits d'actualité sont au début composées de communautés diverses. L'enjeu pour le contrôle de la discussion est de s'approprier le cœur de la diffusion, souvent en centralisant et fédérant l'ensemble des conversations. Cette stratégie qui cherche à couper court à une diffusion non-programmée de l'information semble particulièrement observable en Chine.

Une des pratiques usuelles est notamment de relayer et de diffuser le maximum d'informations afin de recentrer la discussion autour de mots et de groupes d'individus bien définis. En effet, les discussions entourant une actualité politique créent des communautés importantes qui restent néanmoins relativement distantes. Discutant sans vraiment se rencontrer, les différentes communautés se livrent à des échanges très polarisés avec peu de dialogues.

La projection cartographique montre que Canton joue un rôle de précurseur dans les faits d'actualité et paraît "sortir" les affaires. Pékin agit plutôt comme diffuseur en disséminant largement l'information. Ce résultat renforce plutôt l'hypothèse d'une similarité entre réseaux sociaux et médias traditionnels, puisque Canton possède traditionnellement une presse plus portée sur l'investigation que celle plus officielle de Pékin. Les disparités entre Est et Ouest sont également flagrantes lors de la diffusion, avec une domination des grandes zones urbaines de l'Est chinois qui, même pondérées, restent omniprésentes sur les cartes de la diffusion. Les zones de l'Ouest souvent moins urbanisées semblent être moins associées aux discussions, sauf dans le cas où celles-ci les concernent directement (comme le Xinjiang pour *Qiegao*).

L'émission de télévision *The Voice* propose en quelque sorte un contre-exemple du même. Le réseau de conversations entourant cette émission s'organise autour de très peu de diffuseurs dialoguant peu, mais entourés de très nombreux fans. Les mots de moindre importance sont exclus des conversations, contraints à la marginalité dans le réseau sémantique. La diffusion de *The Voice* ne possède pas à proprement parler les spécificités qu'on attendrait d'une diffusion sur les réseaux sociaux avec peu de conversations. Son topogramme se caractérise par une très forte agrégation et une très faible modularité (peu de groupes fortement dominés par des

diffuseurs importants). Il ne se pose pas du tout comme un lieu commun du Web, ouvert et associatif, mais plutôt comme une zone d'échanges clairement identifiée et territorialisée.

L'observation des spécificités de diffusion des faits d'actualité ou des campagnes commerciales a permis de vérifier des choses déjà connues : l'actualité se propage selon des cycles courts, les médias de Canton sont plus avant-gardistes que ceux de Pékin, la communication télévisuelle est soigneusement préparée, etc. Comme souvent dans les études utilisant l'analyse de données, ces quelques observations viennent confirmer des connaissances préalables sans vraiment apporter de nouveautés.

Mais les modèles peu connus des mêmes absurdistes sont une des dimensions intéressantes mises à jour dans cette étude. En effet, les topogrammes des mêmes comiques possèdent des structures atypiques. Les graphes de conversation des mêmes *dufu* et *yuanfang* sont fragmentés en nombreux petits groupes de discussions très intégrés qui communiquent très peu entre eux. Tout se passe comme si nous n'assistions pas à une vaste conversation mais plutôt à de nombreuses petites conversations. Les origines géographiques diverses des utilisateurs montrent que la blague fonctionne localement. Le caractère anecdotique et humoristique semble être un fort vecteur de diffusion avec une présence beaucoup plus durable dans le temps que dans le cas des actualités. Leurs graphes sémantiques se composent autour de peu de mots-clés qui sont réutilisés de manière non-définitive, permettant ainsi une grande variation et une grande appropriation par des utilisateurs même isolés. Cette structuration des conversations en groupes lexicaux ouverts aux associations peu communes semble favoriser la diffusion des mêmes et leur permettre de durer dans le temps. Les utilisateurs de Taiwan sont présents dans la diffusion des mêmes absurdistes, alors qu'ils sont absents des discussions politiques. Ainsi, les mêmes possèdent des topogrammes plus associatifs qui encouragent une participation durable.

Parmi les éléments observés dans notre échantillon, seuls *dufu* et *yuanfang* semblent réellement correspondre à la définition de "même Internet". Les débats de société et les faits d'actualité s'apparentent davantage à des "événements web", reconduction de phénomènes pré-existants à l'Internet - les faits divers ont toujours été débattus, même si l'Internet en change radicalement l'échelle. Les modèles issus des médias traditionnels ont également été reconduits comme le montre l'exemple de *The Voice*.

Les mêmes comiques présentent le modèle le plus atypique de cet ensemble. L'humour, s'il pré-existe heureusement à l'Internet, procède d'un régime éminemment conversationnel et prend parfois la forme particulière des mêmes sur Internet. Le jeu de mots, le détournement d'images ou de slogans sont des pratiques communes qui se voient reconduites sous des formes nouvelles en ligne. L'étude de ces formes humoristiques permet d'accéder à des conversations pas nécessairement planifiées et possédant un caractère plus informel. Ce type d'échanges presque spontané semble pourtant suivre des modèles identifiables qu'il est possible d'étudier plus en détails grâce à l'usage des topogrammes.

CONCLUSION

L'analyse des échantillons de contenus issus de Sina Weibo, permet de vérifier notre hypothèse et de valider notre outil méthodologique, le *topogramme*, en identifiant des particularités parmi les mêmes sélectionnés. Par exemple, l'humour semble être le plus fort vecteur pour qu'un même se propage dans la durée et auprès d'un plus grand nombre de communautés diverses. Egalement, un même absurdiste ou humoristique se compose autour de peu de mots-clés qui sont réutilisés et ré-agencés permettant ainsi une grande variation dans sa déclinaison. A l'inverse un débat sur un sujet de société brûlant, démarrant de manière brutale, centralise les conversations et met en relation des communautés diverses, provoquant des regroupements de mots et d'individus. Les discussions qui entourent une actualité politique créent des communautés importantes de discussions en ligne qui restent néanmoins relativement distantes, discutant sans se rencontrer.

Enfin, le cas de l'émission de télévision *The Voice* donne l'exemple d'une diffusion utilisant le réseau social uniquement comme une chambre d'écho. Ses conversations sont centralisées autour de peu de diffuseurs importants entourés de fans, autour d'un discours sémantiquement très établi qui inclut difficilement d'autres discussions ou éléments annexes. Nous pouvons donc dire que l'exemple de *The Voice* est en fait un contre-exemple et ne doit pas être compris comme un même Internet, puisqu'il ne crée pas de conversations et ne possède pas de spécificités de diffusion propres aux réseaux sociaux.

La taille réduite de notre corpus ne nous permet pas d'affirmer que l'absence de pattern clairement observable est une caractéristique représentative des mêmes Internet. Egalement, des mesures de centralité pourrait faire apparaître plus clairement le rôle sous-jacent des provinces dans la dynamique des

conversations. Notre étude permet néanmoins de montrer comment les analyses conversationnelles classiques des discussions peuvent être complétées par les dimensions temporelles, géographiques et temporelles pour mieux comprendre les phénomènes à l'œuvre lors de la diffusion des contenus en ligne. Cette lecture croisée des différentes dimensions permet notamment de concevoir ou de comprendre stratégiquement les ressorts de faits public ou de communication dans leur contexte.

Après avoir exploré ici les postulats méthodologiques de cette recherche, celle-ci doit maintenant être poursuivie sur un plus vaste corpus qui permettra de valider ou de découvrir des caractéristiques de la diffusion de contenus en ligne par recouplement des différents aspects observés. Cette observation permettra notamment de dégager des lectures stratégiques et analytiques pour une meilleure appréhension de la diffusion des actes de communication d'organismes tant privés que publics. Egaleme nt, nous pourrons poursuivre cette étude dans deux directions complémentaires : 1) d'une part, la possibilité de généraliser les leçons tirées de l'observation de ces patterns de propagation, et 2) d'autre part, la nécessité de considérer leur instanciation en situation d'usage pour comprendre la dimension pragmatique des mêmes. Le lien esquissé dans ce texte entre topoï géographiques et rhétoriques pourra notamment servir de base à la construction d'une démarche analytique pour cette observation.

BIBLIOGRAPHIE

- ADAMIC, L., LENTO, T., ADAR, E., & NG, P. (2014), « *Information Evolution in Social Networks* » in arXiv Preprint arXiv:1402.6792, 1–19.
- BAUCKHAGE, C. (2011), « Insights into internet memes » in Proc. *ICWSM2011*, p. 42–49.
- BERQUE, A. (1999), « Géogrammes, pour une ontologie des faits géographiques. » *Espace Géographique*, vol. 28(4), p.320–326.
- BOSTOCK, M., OGIEVETSKY, V., & HEER, J. (2011), « D3: Data-Driven Documents » in *IEEE Transactions on Visualization and Computer Graphics*, vol. 17(12), 2301–9. doi:10.1109/TVCG.2011.185
- BRUNET, R. (1980), « La composition des modèles dans l'analyse spatiale » in *Espace Géographique*, p. 253–265.
- BERGSTROM, M. (2012), *All Eyes East: Lessons from the Front Lines of Marketing to China's Youth*, Palgrave Macmillan.
- BLACKMORE, S. (2001). « Evolution and Memes: the Human Brain As a Selective Imitation Device » in *Cybernetics and Systems*, vol. 32(1-2), p. 225–255. doi:10.1080/019697201300001867
- BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R., & LEFEBVRE, E. (2008), « Fast unfolding of communities in large networks » in *Journal of Statistical Mechanics: Theory and Experiment*, doi:10.1088/1742-5468/2008/10
- BLOOM, P. (2012), « Religion, morality, evolution » in *Annual Review of Psychology*, vol. 63, p.179–99. doi:10.1146/annurev-psych-120710-100334
- DAWKINS, R. (1989). *The selfish gene*, Oxford University Press.
- FERRARA, E., JAFARI ASBAGH, M., VAROL, O., QAZVINIAN, V., MENCZER, F., & FLAMMINI, A. (2013), « Clustering Memes in Social Media » in *IEEE/ACM ASONAM'13*, doi:10.1145/2492517.2492530
- FU, K., & CHAU, M. (2013), « Reality check for the Chinese microblog space: a random sampling approach » in *PloS One*, vol. 8(3), p. 58-56. doi:10.1371/journal.pone.0058356
- GAO, Q., ABEL, F., HOUBEN, G., & Yu, Y. (2012), « A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter » in *20th international conference on User Modeling, Adaptation, and Personalization*, p.88-101
- JOUXTEL, P. (2013). « La mémétique, une science à l'état sauvage » in *Hermès, La Revue*, vol. 67, p.50-56
- KNOBEL, M., & LANKSHEAR, C. (2007). *A new literacies sampler*, Peter Lang Publishing, Inc, New York City, USA.
- LESKOVEC, J., BACKSTROM, L., & KLEINBERG, J. (2009), « Meme-tracking and the dynamics of the news cycle » in *Proceedings of the KDD '09*. doi:10.1145/1557019.1557077

NETTLETON, D. F. (2013). « Data mining of social networks represented as graphs » in *Computer Science Review*, vol.7, p. 1–34. doi:10.1016/j.cosrev.2012.12.001

RENAUD, C. et al. (2014). « Mêmes internet et réseaux sociaux chinois : état des lieux et perspectives d'analyse » in *AIM2014*, p. 1–25.

SHI, M. (2012). *Brand Management via China's Sina Weibo : Case Studies on Durex and VANCL*. Faculty of the American University.

SHIFMAN, L. (2014), *Memes in digital culture*, MIT Press, Cambridge, MA, USA.

SULLIVAN, J. (2012). « A tale of two microblogs in China” in *Media, Culture & Society*, vol. 34(6), p. 773–783. doi:10.1177/0163443712448951

WALKER, H. (2012). « Obama Campaign Deploys Cat Meme to Get Out the Vote in Ohio » in *Politicker*. Retrieved May 20, 2013.

WANG, S. (2012). « China's Internet lexicon: Symbolic meaning and commoditization of Grass Mud Horse in the harmonious society » in *First Monday*, vol. 17(1).

WENG, L., FLAMMINI, A., VESPIGNANI, A., & MENCZER, F. (2012). « Competition among memes in a world with limited attention » in *Scientific Reports*, vol. 2, p.335. doi:10.1038/srep00335

ZUCKERMAN, E. (2008). « The Cute Cat Theory Talk at ETech », in *O'Reilly E-tech conference 2*, O'Reilly Conferences.