¿ Interrogations ?
revue pluridisciplinaire de sciences humaines et sociales

Segault Antonin, Tajariol Federico, Domenget Jean-Claude, Roxin Ioan

Communication de crise sur les médias sociaux : outils pour l'étude des publics

Pour citer l'article

Segault Antonin, Tajariol Federico, Domenget Jean-Claude, Roxin Ioan, « Communication de crise sur les médias sociaux : outils pour l'étude des publics », dans *revue ¿ Interrogations ?*, N°24. Public, non-public : questions de méthodologie, juin 2017 [en ligne], http://www.revue-interrogations.org/Communication-de-crise-sur-les (Consulté le 8 septembre 2026).

ISSN 1778-3747

Tous les textes et documents disponibles sur ce site sont, sauf mention contraire, protégés par la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 3.0 France.

Résumé

Cet article présente des outils permettant l'étude des publics qui communiquent *via* les médias sociaux lors d'une situation de crise. Il décrit les techniques de collecte des contenus créés par les utilisateurs sur ces plates-formes, propose des techniques de stockage des corpus ainsi constitués, et identifie des métriques pour leur analyse. Il s'adresse tant aux chercheurs s'intéressant aux publics en situation de communication de crise et à leurs échanges, qu'aux concepteurs de plates-formes d'aide à la gestion de crise intéressés à y intégrer les données issues des médias sociaux.

Mots clés : communication de crise, médias sociaux, collecte de données, fouille de données, web sémantique.

Abstract

Crisis communication on social media : tools for the study of publics

This paper presents some tools to study publics communicating through social media during a crisis. It describes methods to capture user-generated contents on these platforms, proposes a system to store the corpora, and identifies metrics for their analysis. It is aimed at researchers interested in analyzing how emerging publics communicate during a crisis event, as well as designers of crisis management platforms willing to integrate several data from social networks.

Keywords :crisis communication, social media, data capture, data mining, semantic web.

Introduction

Au cours de ces dernières années, différents publics ont eu fréquemment recours à des services Web tels que les médias sociaux pour communiquer en situation de crise. Ces publics ont créé de nouveaux usages, tirant parti de la large diffusion de ces services, pour transmettre des informations liées à la crise et à leur état matériel et psychologique. L'étude des échanges sur ces plates-formes peut permettre, d'une part, de mieux comprendre les usages émergents en situation de crise, et par conséquent de préconiser des moyens de les soutenir. D'autre part, elle permet d'analyser les obstacles méthodologiques auxquels sont confrontés les chercheurs. Les contenus générés par les publics sont notamment difficiles d'accès dans les médias sociaux en raison du manque d'interopérabilité entre les différentes plates-formes et de l'importance des quantités de données à analyser.

Dans cet article, nous présenterons des outils et des techniques permettant de surmonter ces obstacles. Nous donnerons dans un premier temps un aperçu de la communication de crise et de son extension *via* les médias sociaux. Nous verrons quels outils peuvent être utilisés pour constituer des *corpora* de contenus générés par les utilisateurs sur les différentes plates-formes, et quelles contraintes compliquent cette collecte. Nous proposerons ensuite une solution basée sur les technologies du Web Sémantique pour stocker ces données. Enfin, nous présenterons différentes pistes pour l'analyse de ces corpus, et nous soulignerons les limites de ces outils et les enjeux d'une telle démarche méthodologique

Le contexte

La communication de crise

Le terme de crise recouvre une grande variété d'événements, allant des catastrophes naturelles aux crises organisationnelles (Sellnow, Seeger, 2013). Une crise est la concrétisation d'un risque qui n'a pas pu être maîtrisé (Sellnow, Seeger, 2013, citant Heath, 1995). Elle peut être définie comme la perception d'un événement imprévisible qui menace les attentes des parties prenantes (Coombs, 2010, citant Coombs, 1999). L'évaluation d'une crise ne saurait cependant être objective, car sa gravité relève des perceptions individuelles et collectives (Sellnow, Seeger, 2013).

Nous nous focaliserons dans cet article sur les crises constituant des urgences à grande échelle affectant les membres du public (Palen *et al.*, 2009). Cette définition comprend les catastrophes naturelles mais aussi les

accidents industriels ou encore les actes de terrorisme : la gestion de ces types de situations est assurée par le gouvernement, la protection civile et les organisations humanitaires (Seeger *et al*., 1998), et prend en compte non seulement les aspects matériels mais aussi les besoins informationnels et psychologiques. Dans ce but, la communication de crise cherche à éviter ou limiter les conséquences négatives de la crise (Reynolds, Seeger, 2005, citant Coombs, 1999). Elle est l'œuvre de nombreux acteurs (*e.g*. victimes, secouristes, témoins, volontaires, journalistes) dont l'activité évolue selon les phases de la crise (Bruns, Liang, 2012), et exige un haut niveau de coopération (Sellnow, Seeger, 2013). Ainsi, les crises sont des situations complexes, pouvant comporter des cascades d'événements (Provitolo *et al*., 2011) dans des périmètres spatiaux et temporels différents qui touchent des populations différentes. Ces populations peuvent constituer des objets d'étude en tant que publics (Proulx, 1998 ; Esquenazi, 2006).

La communication de crise *via* les médias sociaux

Les médias sociaux sont des services Web, appelés aussi plates-formes, permettant la création de profils publics ou semi-publics ainsi que la gestion de connexions avec d'autres utilisateurs (Boyd, Ellison, 2008). On peut notamment citer les plates-formes Facebook (créé en 2004), Flickr (2004), Youtube (2005) et Twitter (2006). La large diffusion et la diversité des médias sociaux ont facilité leur appropriation par les publics lors d'une situation de crise (Vieweg *et al.*, 2010). Par exemple, lors de la fusillade de Virginia Tech (USA) en 2007, les médias sociaux ont été utilisés pour identifier les victimes. Tandis que les autorités et la presse étaient contraintes au silence par la loi et l'éthique, les étudiants et leurs proches ont mené un travail collaboratif rigoureux grâce aux médias sociaux. En collectant et en croisant les témoignages et les traces d'usage en ligne (connexion sur une messagerie instantanée, contact téléphonique), ils ont pu établir la liste complète des victimes avant que les autorités ne diffusent l'information officielle (Palen *et al.*, 2009).

Durant les crises, la nature des échanges sur ces plates-formes évolue. Les utilisateurs privilégient notamment le partage de liens et la rediffusion d'informations plutôt que les messages personnels (Hui *et al.,* 2012, citant Hughes, Palen, 2009 ; Thomson *et al.*, 2012). Des conventions émergent entre les publics frappés par des catastrophes pour favoriser la transmission et la vérification des informations. De cette manière les publics ont choisi des mots-clés (*hashtags*) facilitant ainsi l'indexation des contenus dans Flickr (Liu *et al.*, 2008) et l'agrégation et l'exploitation des informations sur Twitter (Bruns, Burgess, 2012). En créant une communication officieuse (en ang., *backchannel communication*) à grande échelle entre les membres du public, ces outils favorisent la gestion collaborative des informations liées à la crise (Sutton, Palen, Shklovski, 2008), faisant émerger une intelligence hautement parallèle et distribuée (Palen *et al.*, 2010).

Cependant, les organisations officielles telles que la protection civile et les organisations non gouvernementales (ONG) peinent à intégrer ces données dans leurs systèmes d'information opérationnels, estimant que les échanges sur les médias sociaux ne sont pas assez fiables, précis et légitimes au regard des procédures standardisées (Tapia *et al.*, 2011). Néanmoins, les équipes de gestion de crise ont conscience de la valeur de ces outils pour l'évaluation en temps réel de la situation, et commencent à les utiliser pour obtenir des informations complémentaires (non cruciales), notamment en mobilisant leurs réseaux personnels (Tapia *et al.*, 2013). Les médias sociaux rendent la communication de crise persistante, permettant la constitution de *corpora* représentatifs dont la construction ouvre des problèmes méthodologiques importants (Palen *et al.*, 2009 ; Bruns, Liang, 2012).

Dans les sections suivantes, nous décrivons les méthodes de collecte des contenus créés par les publics *via* les médias sociaux, puis les techniques de stockage et d'analyse de ces données.

La collecte de données

Différents logiciels permettent d'automatiser la collecte des données et des contenus générés *via* les médias sociaux en situation de crise, et peuvent être exploités par les chercheurs et par les gestionnaires des situations de crise.

Les interfaces de programmation des médias sociaux

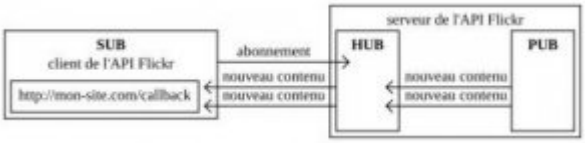
Les médias sociaux mettent à la disposition des développeurs des *Application Programming Interface* (*APIs*, interfaces de programmation), qui permettent à tout chercheur disposant d'une machine connectée à Internet (client) d'interroger les plates-formes de médias sociaux (serveurs) pour y récupérer des données concernant

les contenus publiés. Ces interfaces comportent de nombreux outils pour collecter les données relatives aux utilisateurs (identité, relations, centres d'intérêt) et aux contenus partagés (date de création, commentaires et rediffusions, liens vers d'autres contenus). Certaines APIs permettent en outre l'ajout ou la modification de ces informations par le client, fonctionnalités qui requièrent souvent l'authentification de l'auteur de la requête, nécessitant parfois des procédures complexes. Les plates-formes fournissent généralement une ou plusieurs implémentations de leur API sous forme d'une bibliothèque logicielle dans un langage de programmation courant (*i.e.* PHP, Java, Python, Ruby), et la communauté des développeurs élargit ensuite cette offre. La collecte est le plus souvent effectuée dans une configuration de type *pull* : le client, pour obtenir des données, envoie une requête au serveur de l'API (*i.e.* Facebook, Twitter, *etc*.) ; le serveur lui envoie les données demandées.

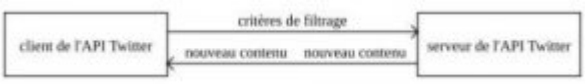
Collecte de données en temps réel

Des données peuvent être collectées en temps réel lorsque le *client* interroge régulièrement le *serveur* de l'API pour obtenir les derniers contenus publiés par les publics. Cependant, si l'on cherche à obtenir un *corpus* d'échanges exhaustif, il faut envoyer ces requêtes plus fréquemment que nécessaire, afin de ne perdre aucun message (par exemple, en augmentant graduellement la fréquence jusqu'à ce qu'un même message apparaisse deux fois). Cette marge de sécurité entraîne un usage superflu des ressources (bande passante, puissance de calcul) du client comme du serveur. Pour pallier ces difficultés, un certain nombre de plates-formes (Facebook, Twitter, Youtube, Flickr, Tumblr et Instagram) ont doté leurs APIs de fonctionnalités dédiées à la capture de données en temps réel. Bien qu'utilisant des techniques différentes, elles reposent toutes sur le même principe : c'est le serveur de l'API qui notifie au client la publication d'un nouveau contenu correspondant à la requête (on parle alors de méthode *push*).

La majorité de ces systèmes de notification sont des implémentations du protocole PubSubHubHub [1] (ou PuSH), qui définit un écosystème d'éditeurs (*Publisher – Pub*), d'abonnés (*Subscriber – Sub*), et de relais (*Hub*). Pour s'abonner à un contenu publié par un éditeur, il suffit d'envoyer au relais une requête indiquant le contenu en question, et l'adresse Internet (*i.e. callback*) à laquelle envoyer les mises à jour. L'éditeur envoie une notification d'ajout ou de modification du contenu, et celui-ci transmet l'information aux abonnés de ce contenu [2] (voir Figure 1). Le nombre de requêtes et de ressources, notamment la bande passante, est ainsi fortement limité.



Ce mode de fonctionnement est généralisé avec quelques exceptions. Par exemple, l'API Streaming [3] de Twitter fonctionne d'une manière un peu différente (voir Figure 2). Le chercheur, *via* son application client, interroge le serveur. Celui-ci ouvre une connexion permanente, à travers laquelle les nouveaux contenus sont diffusés comme un flux de données en continu : cette diffusion ne s'arrête que lorsque le client ferme la connexion, mettant ainsi fin à la requête des données [4]. En revanche, quand le volume de données envoyées par le serveur dépasse 1 % des tweets publiés à un instant précis, le serveur cesse la diffusion en continu et envoie seulement un échantillon jugé représentatif [5] (Morstatter *et al.*, 2013). Cet échantillonnage peut influencer les résultats de la requête (*e.g.* en fragmentant ou en supprimant des chaînes de réponses ou de rediffusions) et ainsi avoir un impact sur les objectifs attendus par le chercheur. Pour pallier cette limite dans la collecte des données, le serveur notifie au client [6] le dépassement du 1 % des tweets publiés, permettant au chercheur de tenir compte de l'échantillonnage, ou d'envisager une requête plus précise.



Le chercheur devra aussi tenir compte des limitations imposées par les différentes APIs concernant les contenus publiés par les utilisateurs. Plus précisément, les APIs de Facebook, Tumblr et Youtube permettent la capture en temps réel uniquement des contenus publiés par un utilisateur ou un groupe d'utilisateurs

spécifiques (voir Tableau 1). En revanche, Les APIs de Twitter, Flickr [7] et Instagram [8], permettent d'effectuer des requêtes portant sur les données publiées par l'ensemble des utilisateurs de la plate-forme.

Ce court panorama montre que ces dernières APIs sont les seules adaptées à la capture en temps réel des contenus produits *via* les médias sociaux pendant une situation de crise, dont les publics nombreux et variés ne peuvent pas forcement être définis *a priori*.

Critères de collecte

Ces trois APIs en temps réel permettent de collecter les contenus publiés en fonction de différents critères (voir Tableau 1).

Tableau 1 : Fonctionnalités des APIs temps réel des principaux médias sociaux

Plate-forme		**Facebook**	**Twitter**	**Youtube**	**Flickr**	**Instagram**	**Tumblr**
Portée des requêtes		Un utilisateur, une page	Globale	Un utilisateur, un groupe	Globale	Globale	Un blog
Critères	Tag		Oui		Oui	Oui	
	Texte		Oui		Non	Non	
	Localisation		Oui (faible proportion)		Oui	Oui	
	Opérateurs logiques		ET/OU		OU	Non	

Ces critères portent principalement sur les données textuelles et les métadonnées associées [9]. Les plates-formes Flickr et Instagram permettent la collecte uniquement *via* les mots-clés (*e.g. tags*) associés à une image, tandis que Twitter accepte d'y inclure le texte du tweet, ses mots-clés et les URLs, mais aussi le nom de l'utilisateur. Globalement, les APIs gèrent de manière limitée les requêtes complexes (*e.g.* combinant plusieurs opérateurs booléens). L'API de Twitter permet de composer des requêtes associant plusieurs mots-clés à l'aide des opérateurs logiques ET/OU (représentés par l'espace et la virgule) [10], tandis que l'API de Flickr n'accepte qu'une liste de mots-clés traités comme une disjonction (OU logique) [11] et l'API d'Instagram n'effectue la recherche que sur un seul mot-clef à la fois [12]. Il est néanmoins possible d'envoyer plusieurs requêtes simples et d'effectuer les opérations logiques sur les résultats reçus.

Ces APIs proposent également des méthodes pour collecter les contenus en fonction des métadonnées géographiques associées, qui peuvent être de simples coordonnées ou d'identifiants de lieux. Les images ainsi géolocalisées sont nombreuses sur Flickr (Breslin, Passant, Decker, 2009) et plus encore sur Instagram, dont les photographies proviennent exclusivement de téléphones mobiles, car la majorité des modèles récents intègre le service de localisation GPS. Sur Twitter, en revanche, une très faible proportion des contenus échangés portent des métadonnées géographiques (Bruns, Liang, 2012). Ainsi, Twitter semble être la plate-forme la plus limitée pour constituer des corpus représentatifs sur la base d'informations géographiques [13].

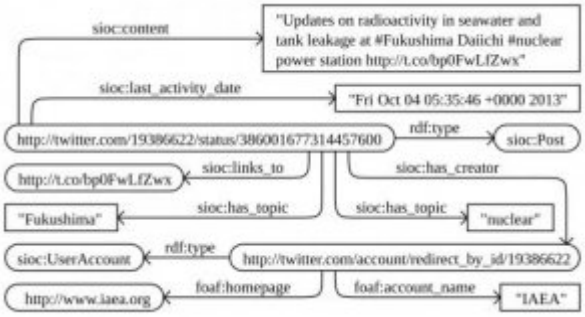
Le stockage des données

Le stockage de données provenant de multiples plates-formes est également complexe. En effet, les réponses des APIs utilisées par les médias sociaux diffèrent par la structure et le format des données (*e.g.* Atom/XML pour Flickr, JSON pour Twitter et Instagram). Pour exploiter ces données conjointement, il est donc nécessaire d'unifier leur représentation et d'assurer un niveau de formalisation suffisant pour en permettre leur traitement automatisé. Dans cette direction, les concepts du Web Sémantique (Berners-Lee, Hendler, Lassila, 2001) peuvent être exploités pour construire une représentation commune des données issues des médias sociaux (Breslin, Passant, Decker, 2009). Pour ce faire, il est nécessaire d'identifier des formalismes permettant de décrire ces données, leur structure et les relations possibles. La description est assurée par le modèle RDF (Resource Description Framework), permettant la description des données sous forme de triplets *sujet, prédicat, objet*. La structure des données est décrite sous forme d'une ontologie, qui est « *la spécification de la conceptualisation d'un domaine* » (Provitolo *et al.*, 2011, citant Gruber, 1995). Une ontologie définit les concepts, leurs relations, ainsi que les règles et contraintes les régissant (Breslin *et al.*, 2009). Plusieurs

ontologies existantes et reconnues peuvent être mobilisées pour formaliser les données issues des médias sociaux (Breslin *et al.*, 2009), comme SIOC (*Semantically-Interlinked Online Communities*), une ontologie pour modéliser les messages échangés dans les communautés en ligne, et FOAF [14] (*friend of a friend*), qui décrit les personnes, leurs caractéristiques et leurs relations. On peut aisément combiner les vocabulaires de ces deux ontologies (voir Tableau 2) afin de représenter les données capturées grâce aux outils précédemment évoqués [15]. Par exemple, le tweet [16] de la Figure 3 peut être représenté par le graphe de la Figure 4.

Tableau 2 : Description de données issues des médias sociaux à l'aide d'ontologies existantes

Sujet	Prédicat	Objet		
		Twitter	Instagram	Flickr
Identifiant de la publication	rdf:type	sioc:Post		
	sioc:title			Titre de la publication
	sioc:content	Texte du tweet	Description de la publication	
	sioc:attachment		Image	
	sioc:has_topic	Mot clef		
	sioc:links_to	URL dans le tweet		
	sioc:references	Message retweeté		
	sioc:addressed_to	Utilisateur mentionné		
	sioc:reply_of	Message initial		
	sioc:last_activity_date	Date et heure de création		
	sioc:has_creator	Identifiant de l'utilisateur		
Identifiant de l'utilisateur	rdf:type	sioc:UserAccount		
	foaf:acountName	Nom de l'utilisateur		
	foaf:homepage	Page Web de l'utilisateur		



Les triplets *sujet, prédicat, objet* ainsi constitués (par exemple, , *rdf:type, sioc:Post*) peuvent ensuite être stockés dans une base de données optimisée pour RDF, appelée *triple-store*. Au-delà de la grande diversité des bases de données de type *triple-store* [17], toutes permettent la récupération de triplets grâce au langage de

requête dédié SPARQL (Sparql protocol and RDF query language). Ce langage de requêtes rend possible l'analyse des données collectées *via* les médias sociaux.

L'analyse des données

Métriques

Grâce à SPARQL, le chercheur peut calculer diverses métriques à partir des métadonnées associées aux contenus publiés *via* les médias sociaux, puis formalisées selon le modèle RDF et enregistrées dans des *triple-store*. Par exemple, à l'aide de requêtes SPARQL, il est aisé de déterminer les utilisateurs les plus actifs durant la période de capture (voir Figure 5) ou encore de calculer les fréquences de réponse ou de rediffusion des contenus. Ces deux dernières métriques ont été étudiées par Amanda Lee Hughes et Leysa Palen (2009), révélant que, sur Twitter, les situations de crise sont caractérisées par une augmentation des rediffusions (*retweets*) et une diminution des réponses.

```
SELECT ?name (COUNT(?post) AS ?nb)
WHERE _ ?user a sioc:UserAccount . _ ?user
foaf:accountName ?name . _ ?post sioc:has_creator ?user .
_ ?post a sioc:Post .

GROUP BY ?user
ORDER BY DESC(?nb)
LIMIT 10
```

Concernant l'analyse du contenu des messages, le langage SPARQL permet aussi de rechercher les mots-clés ou encore les URLs les plus fréquents. Ainsi, la comparaison des pourcentages de retweets et de tweets contenant des URLs au cours de plus de 40 événements a montré des différences d'usages entre les événements sensationnels (catastrophes naturelles, scandales, caractérisés par une forte part de retweets et d'URL) et les autres événements médiatiques (sport, émissions télévisées) (Bruns, Stieglitz, 2012). Des outils d'analyse de texte peuvent être utilisés pour extraire les principaux thèmes du corpus (Guille *et al.*, 2013). Ces programmes permettent en outre la classification automatique des sentiments exprimés dans les messages (positifs ou négatifs), ainsi que l'analyse des émotions (peur, joie, tristesse, *etc*.), bien que cette dernière puisse encore être perfectionnée (Johansson *et al.,* 2012).

Afin de décrire et comprendre le déroulement d'une crise sur les médias sociaux, il est nécessaire d'analyser les variations temporelles de ces différentes métriques. Ainsi, l'on pourra calculer l'évolution du nombre d'utilisateurs actifs et de messages publiés au cours du temps, ou encore identifier les mots-clés, les URLs et les utilisateurs les plus fréquents au cours des différentes phases (Bruns, Liang, 2012).

Analyse structurale des réseaux

L'analyse structurale des réseaux peut être mobilisée pour explorer les données issues des médias sociaux. Cette méthodologie quantitative s'appuie sur la théorie des graphes et l'algèbre linéaire afin de représenter des réseaux de relations et d'en mesurer différentes propriétés (Mercklé, 2011). Le système de collecte précédemment décrit ciblant l'ensemble des utilisateurs des plates-formes de médias sociaux (dans les limites émanant des critères de collecte et des outils eux-mêmes), les réseaux ainsi constitués se rapprochent de réseaux complets (plutôt que de réseaux personnels) (Forsé, Degenne, 2004). Différents types de réseaux peuvent être considérés : homogènes, comme les interactions entre les usagers (réponses, rediffusion), ou encore les co-occurrences de mots-clés au sein des messages ; *hétérogènes*, comme les mots-clés ou les URLs partagés par les utilisateurs (Bruns, Liang, 2012).

Les métriques de la théorie des graphes peuvent être distinguées selon leur portée, locale ou globale. Les métriques locales, telles que les différents degrés de centralité, sont propres à un nœud, et permettent d'évaluer sa place dans le réseau. Les métriques globales, comme la densité et la connexité, portent sur l'ensemble du graphe, ou sur une portion de celui-ci. Un grand nombre d'outils logiciels dédiés au calcul de ces métriques sont déjà disponibles [18]. Nous avons notamment utilisé la librairie Python NetworkX, qui propose un grand nombre d'algorithmes pour l'analyse des graphes [19]. Une fonction d'import de triplets RDF est en

cours d'intégration [20], permettant de travailler directement sur les données collectées avec les outils précédemment évoqués. De même, le projet Statnet [21] propose les paquets Network et SNA pour l'analyse structurale des réseaux avec le logiciel de statistiques R, dont le module Rrdf [22] permet la gestion de la structure RDF.

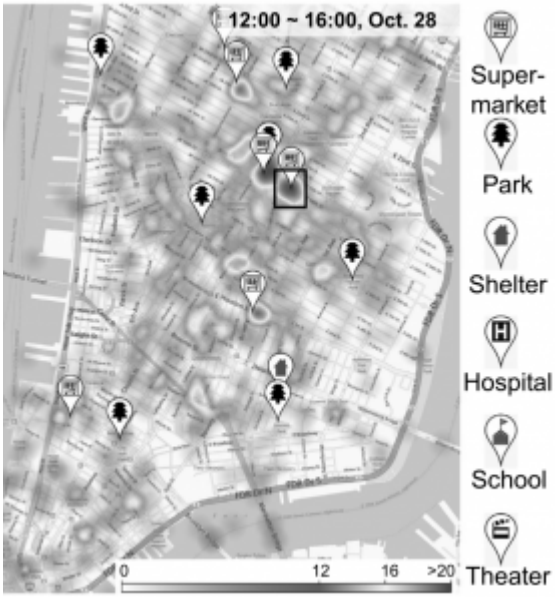
Un exemple d'utilisation de l'analyse structurale de réseaux est offert par (Hui *et al.*, 2012) avec l'étude des messages échangés sur Twitter en 2010, lors de la présence d'un homme armé sur un campus du Rensselaer Polytechnic Institute. Les cascades de retweets, modélisées sous formes de réseaux, ont permis de mettre en évidence la diffusion de l'alerte au cours du temps. Le degré de centralité entrant des nœuds y est interprété comme la crédibilité perçue des utilisateurs, et le degré de centralité sortant comme leur propension à partager l'information .

En outre, les méthodes de regroupement (*clustering*) basées sur des algorithmes par apprentissage (supervisé ou non) peuvent être mobilisées pour identifier différents groupes d'utilisateurs (Imran *et al.*, 2013) selon des similarités dans leurs profils ou encore selon les données qu'ils partagent. Cela permettra la compréhension de logiques d'interactions sociales intra-groupes et inter-groupes.

Visualisation des données

Des visualisations peuvent être créées pour faciliter la compréhension des larges corpora issus des médias sociaux (Malizia *et al.*, 2011 ; Chae *et al.*, 2014). De nombreux programmes permettent la représentation de graphes, notamment l'API GraphStream (Guille *et al.*, 2013), le logiciel Gephi (Bruns, 2012) ou encore les outils du projet GraphViz. Ces derniers reposent sur DOT [23], un langage pour représenter des graphes sous forme de fichiers textes simples, utilisé par de nombreux autres logiciels d'analyse ou de visualisation de réseaux. Ces visualisations peuvent en outre inclure des métriques issues de l'analyse structurale.

Plusieurs autres types de visualisation peuvent être mobilisés. La représentation des séries temporelles, par exemple avec E-DataViewer [24] (Vieweg *et al.*, 2010, citant Starbird, 2008) ou StreamGraph (Byron, Wattenberg, 2008), favorise la visualisation de l'évolution de différentes métriques au cours du temps, et l'identification d'anomalies (Terpstra *et al.*, 2012). Des outils de cartographies, comme ceux utilisés par les projets Ushahidi et Safecast [25], permettent également une meilleure compréhension des données géographiques (Terpstra *et al.*, 2012 ; Malizia *et al.*, 2011). Ainsi, dans la figure 6, la cartographie des utilisateurs de Twitter à l'annonce de l'arrivée de la tempête Sandy à Manhattan, a permis de constater que les habitants n'ont pas rejoint les refuges, mais se sont rendus dans les supermarchés, vraisemblablement pour s'y procurer des réserves alimentaires (Chae *et al.*, 2014).



Cadre d'interprétation

L'interprétation des résultats doit néanmoins tenir compte des limites des différentes méthodes utilisées pour

la collecte et l'analyse de ces données. Les critères de collectes de données définissent les limites du corpus qui sera constitué. Le choix des mots-clés, des fenêtres temporelles et spatiales exprime des hypothèses préalables quant à la distribution des données, dont dépendent la précision et l'exhaustivité du corpus (Gerlitz, C., Rieder, 2013). Les métriques d'analyse et les représentations visuelles ne sont également pas des outils neutres (Barats, 2013 ; Tufte, 1983) , mais le fruit de partis pris, généralement par souci de clarté ou d'esthétique (Mercklé, 2011) mais influençant néanmoins l'interprétation. Le chercheur ne doit pas utiliser ces outils comme des « *boîtes noires* » (Bruns, Liang, 2012 ; Barats, 2013) mais, au contraire, en étudier le fonctionnement afin de tenir compte des biais qu'ils induisent, et de définir les limites du cadre d'interprétation des résultats.

Conclusion

Dans cette contribution, nous avons donné un aperçu des enjeux de l'usage des médias sociaux en situation de crise, et des contraintes entravant son étude. La constitution de corpus exhaustifs nécessite la capture en temps réel des contenus générés par les utilisateurs, grâce aux APIs spécifiques fournies par certaines plates-formes. Les ontologies SIOC et FOAF permettent le stockage de ces données sous forme de triplets RDF, dont le langage de requête SPARQL permet l'analyse. Les métriques de l'analyse structurale des réseaux sociaux peuvent également être exploitées, et différentes représentations graphiques des résultats sont utilisables pour en favoriser la compréhension.

Cependant, ces techniques de collecte ne peuvent être appliquées à toutes les plates-formes de médias sociaux, et la complexité des requêtes reste limitée. Par ailleurs, le traitement de corpus de données conséquents nécessite d'importantes capacités de calcul et de stockage, et les outils de capture et d'analyse présentés dans cet article devront être adaptés pour ce passage à l'échelle. En outre, tant lors de la collecte que pour l'analyse, le chercheur se doit d'étudier précisément les outils qu'il choisit d'utiliser, afin d'être conscient des limites de l'interprétation et des éventuels biais liées à leur fonctionnement.

Les méthodes décrites dans cet article présentent un double enjeu. D'une part, elles peuvent être exploitées par les chercheurs pour l'étude des usages des médias sociaux en situation de crise, et notamment des publics impliqués. D'autre part, elles peuvent être mises à la disposition de ces mêmes publics, en permettant d'intégrer les données issues des médias sociaux au sein d'outils d'aide à la gestion de crise favorisant l'extraction, la vérification et la diffusion d'informations utiles en temps réel. Ce double enjeu se retrouve dans le projet SCOPANUM (Stratégies de COmmunication de crise en gestion Post-Accident NUcléaire *via* les Médias Sociaux), financé par le CSFRS [26], et mobilisant l'équipe OUN [27] du laboratoire ELLIADD [28] et le CEPN [29]. Ce projet a pour objectif d'étudier les usages des médias sociaux dans la phase post-accidentelle nucléaire, et de concevoir des outils favorisant le processus de résilience.

Bibliographie

Barats Christine (2013), « Le Web : outils de communication, objet de connaissance », pp. 155–174, dans *Sciences de L'information et de La Communication. Objets, Savoirs, Discipline*, Oliviesi S. (dir.), Grenoble, Presses universitaires de Grenoble.

Berners-Lee Tim, Hendler James, Lassila Ora (2001), « The semantic web », *Scientific American*, 284, pp. 28–37.

Boyd, Danah, Ellison, Nicole(2008), « Social Network Sites : Definition, History, and Scholarship », *Journal of Computer-Mediated Communication*, 13, pp. 210, [en ligne]. http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full (consulté le 6/05/2015)

Breslin John, Passant Alexandre, Decker Stefan (2009), *The Social Semantic Web,* Berlin, Springer..

Bruns Axel (2012), « How long is a tweet ? Mapping dynamic conversation networks on Twitter using Gawk and Gephi », *Information, Communication, Society*, 15, pp. 1323–1351, [en ligne] http://eprints.qut.edu.au/47819/ (consulté le 6/05/2015)

Bruns Axel, Burgess Jean (2012), « Local and global responses to disaster :# eqnz and the Christchurch earthquake », *Disaster and Emergency Management Conference, Conference Proceedings*, 2012, pp. 86–103,

[en ligne] http://snurb.info/files/2012/Local%20and%20Global%20Responses%20to%20Disaster.pdf (consulté le 6/05/2015).

Bruns Axel, Liang Y. (2012), « Tools and methods for capturing Twitter data during natural disasters », *First Monday*, 17, pp. 1–8, [en ligne] http://firstmonday.org/article/view/3937/3193 (consulté le 6/05/2015).

Bruns Axel, Stieglitz Stephan (2012), « Quantitative approaches to comparing communication patterns on Twitter », *Journal of Technology in Human Services*, 30, pp. 160–185, [en ligne] http://www.tandfonline.com/doi/abs/10.1080/15228835.2012.74424 (consulté le 6/05/2015).

Byron Lee, Wattenberg Martin (2008), « Stacked Graphs-Geometry, Aesthetics », *IEEE Trans. Vis. Comput. Graph.*, 14, pp. 1245–1252, [en ligne] http://leebyron.com/streamgraph/(consulté le 6/05/2015).

Chae Junghoon, Thom Dennis, Jang Yun, Kim SungYe, Ertl Thomas, Ebert David (2014), « Public behavior response analysis in disaster events utilizing visual analytics of microblog data », *Computers, Graphics*, 38, pp. 51–60, [en ligne] http://www.sciencedirect.com/science/article/pii/S0097849313001490 (consulté le 6/05/2015).

Coombs Thimoty (1999), *Ongoing Crisis Communication : Planning, Managing, and Responding*, Los Angeles, SAGE Publications.

Coombs Thimoty (2010), « Parameters for crisis communication », dans *The Handbook of Crisis Communication*, Coombs W. T. , Holladay S. J. (dirs), Oxford, Wiley-Blackwell, pp. 17–53.

Esquenazi Jean-Pierre (2006), « Les médias et leurs publics. Le processus de l'interprétation », dans *Sciences de L'information et de La Communication. Objets, Savoirs, Discipline*, Oliviesi S. (dir.), Grenoble, Presses universitaires de Grenoble, pp. 11-20.

Forsé Michel, Degenne Alain (2004), *Les réseaux sociaux*, Paris, Armand Colin.

Gerlitz Carolin, Rieder Bernhard (2013), « Mining One Percent of Twitter : Collections, Baselines, Sampling », *M/C Journal*, 16, [en ligne] http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/620 (consulté le 6/05/2015).

Gruber Thomas (1995), « Toward principles for the design of ontologies used for knowledge sharing », *International Journal of Human-Computer Studies*, 43, pp. 907–928.

Guille Adrien, Favre Cécile, Hacid Hakim, Zighed Djamel Abdelkader (2013), « SONDY : An Open Source Platform for Social Dynamics Mining and Analysis » dans *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, New York, ACM, pp. 1005–1008, [en ligne] http://mediamining.univ-lyon2.fr/people/guille/publications/sigmod13.pdf. (consulté le 6/05/2015).

Heath Robert (1995), « Corporate environmental risk communication : Cases and practices along the Texas Gulf Coast », *Communication Yearbook*, 18, pp. 255–277.

Hughes Amanda Lee, Palen Leysa (2009), « Twitter adoption and use in mass convergence and emergency events », *International Journal of Emergency Management*, 6, pp. 248–260, [en ligne] http://www.iscramlive.org/ISCRAM2009/papers/Contributions/211_Twitter%20Adoption%20and%20Use%20in%20Mass%20Convergence_Hughes2009.pdf, consulté le 6/05/2015

Hui Cindy, Wallace William., Magdon-Ismail Malik, Goldberg Mark (2012), « Information cascades in social media in response to a crisis : a preliminary model and a case study », *Proceedings of the 21st international conference companion on World Wide Web*, New York, ACM, pp. 653–656, [en ligne] http://www.cs.rpi.edu/magdon/ps/conference/InfoCascadesSWDMwww2012.pdf (consulté le 6/05/2015)

Imran Muhammad, Elbassuoni Shady, Castillo Carlos, Diaz Fernando., Meier Patrick (2013), « Extracting information nuggets from disaster-related messages in social media », *ISCRAM'13 : Proceedings of the 10th International ISCRAM Conference*, 26. Karlsruhe, KIT, pp. 791–800, [en ligne] http://qcri.org.qa/app/media/1843 (consulté le 6/05/2015)

Johansson Fredrik, Brynielsson Joel, Quijano Narganes Maribel (2012), « Estimating Citizen Alertness in Crises using Social Media Monitoring and Analysis », *Intelligence and Security Informatics Conference (EISIC)*,

Washington, DC, IEEE Computer Society, pp. 189–196.

Liu Sophia, Palen, Leysia, Sutton Jeannette, Hughes Amanda Lee, Vieweg Sarah (2008), « In search of the bigger picture : The emergent role of on-line photo sharing in times of disaster », *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM)*. Washington, DC, F. Fiedrich / B. Van de Walle (dirs), pp. 140–149, [en ligne] http://digital.cs.usu.edu/amanda/OnlinePhotoSharingISCRAM08.pdf (consulté le 6/05/2015)

Malizia Alessio, Bellucci Andrea, Diaz Paloma, Aedo Ignacio, Levialdi Stefano (2011), « eStorys : A visual storyboard system supporting back-channel communication for emergencies », *Journal of Visual Languages, Computing*, 22, pp. 150–169, [en ligne] http://www.sciencedirect.com/science/article/pii/S1045926X10000789,(consulté le 6/05/2015).

Mercklé Pierre (2011), *Sociologie des réseaux sociaux*, Paris, La découverte.

Morstatter Fred, Pfeffer Jürgen, Liu Huan, Carley Kathleen (2013), « Is the sample good enough ? Comparing data from Twitter's streaming API with Twitter's firehose », dans *Proceedings of ICWSM,* Palo Alto, AAAI, pp. 400–408, [en ligne] http://arxiv.org/abs/1306.5204 (consulté le 6/05/2015)

Palen Leysia, Anderson Kenneth Mark, Mark Gloria Martin James, Sicker Douglas, Palmer Martha, Grunwald Dick (2010), « A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters », dans *Proceedings of the 2010 ACM-BCS Visions of Computer Science Conference*, Swinton, British Computer Society, pp. 1-12, [en ligne] https://www.cs.colorado.edu/ palen/computingvisionspaper.pdf (consulté le 6/05/2015)

Palen Leysia, Vieweg Sarah, Liu Sophia, Hughes Amanda Lee (2009), « Crisis in a networked world features of computer-mediated communication in the April 16, 2007, Virginia Tech Event » dans *Social Science Computer Review*, 27, pp. 467–480.

Proulx Serge (1998), *Accusé de réception : le téléspectateur construit par les sciences sociales*, Québec, Presses de l'Université Laval.

Provitolo Damienne, Dubos-Paillard Edwige, Müller Jean-Pierre (2011), « Vers une ontologie des risques et des catastrophes : le modèle conceptuel », *XVI Èmes Rencontres de Rochebrune, Du 19 Au 23 Janvier 2009, Rencontres Interdisciplinaires Sur Les Systèmes Complexes Naturels et Artificiels*, pp. 1–16, [en ligne] https://halshs.archives-ouvertes.fr/halshs-00643597/document (consulté le 6/05/2015).

Reynolds Barbara, Seeger Matthew (2005), « Crisis and emergency risk communication as an integrative model », *Journal of Health Communication*, 10, pp. 43–55.

Seeger Matthew, Sellnow Timothy, Ulmer, Robert (1998), « Communication, organization, and crisis », *Communication Yearbook*, 21, pp. 231–276.

Sellnow Thimoty, Seeger Matthew (2013), *Theorizing Crisis Communication*, Chichester, John Wiley Sons.

Sutton Jeannette, Palen Leysia, Shklovski Irina (2008), « Backchannels on the front lines : Emergent uses of social media in the 2007 southern California wildfires », dans *Proceedings of the 5th International ISCRAM Conference*, Fiedrich Franck, Van de Walle Bartel (dirs), Washington, DC, pp. 624–632, [en ligne] https://www.cs.colorado.edu/ palen/Papers/iscram08/BackchannelsISCRAM08.pdf (consulté le 6/05/2015).

Tapia Andrea, Bajpai Kartikeya, Jansen Jim, Yen John., Giles Lee (2011), « Seeking the trustworthy tweet : Can microblogged data fit the information needs of disaster response and humanitarian relief organizations », dans *Proceedings of the 8th International ISCRAM Conference*, Santos Maria A., Sousa Luisa, Portela Eliane (dirs), Lisbonne, pp. 1–10.

Tapia Andrea, Moore Kathleen, Johnson Nicholas (2013), « Beyond the Trustworthy Tweet : A Deeper Understanding of Microblogged Data Use by Disaster Response and Humanitarian Relief Organizations », dans *Proceedings of the 10th International ISCRAM Conference*, Karlsruhe, KIT, pp. 770–779, [en ligne] http://www.iscramlive.org/ISCRAM2013/files/121.pdf (consulté le 6/05/2015)

Terpstra Teun, de Vries Arno, Stronkman Richard, Paradies Geerte (2012), « Towards a realtime Twitter analysis

during crises for operational crisis management », dans *ISCRAM'12 : Proceedings of the 9th International ISCRAM Conference*, Vancouver, Simon Fraser University, pp. 1-16, [en ligne] http://www.iscramlive.org/ISCRAM2012/proceedings/172.pdf (consulté le 6/05/2015)

Thomson Robert, Ito Naoya, Suda Hinako, Lin Fangyu, Liu Yafei., Hayasaka Ryo, Ryuzo Isochi, Wang Zian (2012), « Trusting Tweets : The Fukushima Disaster and Information Source Credibility on Twitter », dans *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management*, Vancouver, Simon Fraser University, pp. 1-10, [en ligne] http://www.iscramlive.org/ISCRAM2012/proceedings/112.pdf (consulté le 6/05/2015)

Tufte Edward (2001), *The visual display of quantitative information*, [1983], Cheshire, CT, Graphics press USA.

Vieweg Sarah, Hughes Amanda Lee., Starbird Kate, Palen Leysia (2010), « Microblogging during two natural hazards events : what twitter may contribute to situational awareness », dans *Proceedings of the 28th international conference on Human factors in computing systems*, New York, ACM, pp. 1079–1088, [en ligne] https://www.cs.colorado.edu/ palen/vieweg_1700_chi2010.pdf (consulté le 6/05/2015).

Notes

[1] Accès : https://pubsubhubbub.googlecode.com/git/pubsubhubbub-core-0.4.html. Consulté le 12/02/2014

[2] L'API d'Instagram, contrairement à celle de Flickr, n'envoie pas directement les nouveaux messages au *callback*. Elle le notifie seulement de l'existence de nouveaux contenus. Il faut donc ensuite interroger l'API standard pour obtenir les contenus en question (avec une requête comme /tags/[nom du tag]/media/recent)

[3] Accès : https://dev.twitter.com/docs/streaming-apis. Consulté le 12/02/2014

[4] Accès : https://dev.twitter.com/docs/streaming-apis/connecting#Overview. Consulté le 18/02/2014

[5] Accès : https://dev.twitter.com/docs/faq#6861. Consulté le 18/02/2014

[6] Accès : https://dev.twitter.com/streaming/overview/messages-types#limit_notices. Consulté le 29/10/2014

[7] Accès : http://www.flickr.com/services/api/. Consulté le 12/02/2014

[8] Accès : http://instagram.com/developer/realtime/. Consulté le 12/02/2014

[9] Une métadonnée est une donnée à propos d'une autre donnée. Par exemple les dates et auteurs d'un document.

[10] Accès : https://dev.twitter.com/docs/streaming-apis/parameters#track. Consulté le 20/02/2014

[11] Accès : http://www.flickr.com/services/api/flickr.push.subscribe.html. Consulté le 20/02/2014

[12] Accès : http://instagram.com/developer/endpoints/tags/#get_tags_media_recent. Consulté le 20/02/2014

[13] En outre, sur Twitter, le système d'identifiants de lieux pose problème, car certaines publications peuvent être associées à des lieux très étendus (un pays, par exemple). Une recherche portant sur une localisation précise peut ainsi retourner des résultats associés à une zone englobant cette localisation, mais provenant d'un lieu éloigné.

[14] Accès : http://xmlns.com/foaf/spec/. Consulté le 12/02/2014

[15] Basé également sur un script de Tom Morris, Accès : https://github.com/tommorris/twitter-rdf/blob/master/twitter-rdf.xsl. Consulté le 12/02/2014

[16] Accès : https://twitter.com/iaeaorg/status/386001677314457600. Consulté le 12/02/2014

[17] Accès : http://www.w3.org/wiki/RdfStoreBenchmarking. Consulté le 18/02/2014

[18] Accès : http://quanti.hypotheses.org/845/. Consulté le 20/02/2014

[19] Accès : http://networkx.github.io/documentation/latest/reference/algorithms.html. Consulté le 12/02/2014

[20] Accès : https://github.com/networkx/networkx/pull/855. Consulté le 12/02/2014

[21] Accès : http://statnet.org/. Consulté le 12/02/2014

[22] Accès : https://github.com/egonw/rrdf. Consulté le 20/02/2014

[23] Accès : http://graphviz.org/content/dot-language. Consulté le 12/02/2014

[24] Accès : http://www.cs.colorado.edu/ starbird/e-dataviewer.html. Consulté le 12/02/2014

[25] Accès : http://safecast.org/. Consulté le 20/04/2014

[26] Accès : http://www.csfrs.fr/. Consulté le 22/11/2014

[27] Accès : http://semlearn.pu-pm.univ-fcomte.fr/. Consulté le 20/06/2914

[28] Accès : http://elliadd.univ-fcomte.fr/. Consulté le 20/06/2014

[29] Accès : http://www.cepn.asso.fr/. Consulté le 20/06/2014