Accès à l'information dans les réseaux sociaux : quelles formes de collaboration ?

Lynda Tamine, Laure Soulier, Lamjed Ben Jabeur, Frédéric Amblard, Chihab Hanachi, Gilles Hubert, Camille Roth

Accès à l'information dans les réseaux sociaux : quelles formes de collaboration ?

Lynda Tamine[*], Laure Soulier[*], Lamjed Ben Jabeur[*], Frédéric Amblard[**], Chihab Hanachi[**], Gilles Hubert[*], Camille Roth[***]

[*] *Université Paul Sabatier - IRIT. 118 route de Narbonne. 31062 Toulouse Cedex 9*
tamine, soulier, jabeur, hubert@irit.fr
[**] *Université Capitole 1 - IRIT. 118 route de Narbonne. 31062 Toulouse Cedex 9*
frederic.amblard, chihab.hanachi@univ-tlse1.fr
[***] *Centre Marc Bloch Berlin (An-Institut der Humboldt Universität, UMIFRE CNRS-MAE), Berlin*
roth@ehess.fr

RÉSUMÉ. *Nous présentons dans cet article une analyse exploratoire dans l'objectif d'identifier les patrons de collaboration au sein des réseaux sociaux pour l'accès à l'information. Nous nous focalisons sur le cas de la gestion de crise au travers de la plateforme sociale Twitter. Nous nous appuyons ainsi sur deux types de crises : une catastrophe naturelle (l'ouragan Sandy) et une crise sanitaire (Ebola). Nous proposons une méthodologie en trois étapes qui permet 1) d'analyser les structures de collaboration, 2) d'identifier les thématiques de collaboration et 3) de présenter les réseaux latents de collaboration sur l'ensemble du réseau. Les résultats montrent paradoxalement que le périmètre de la collaboration est réduit, que les sujets sous-jacents sont diversifiés et impliquent des utilisateurs qui peuvent être à l'interface de différents sujets.*

ABSTRACT. *This paper presents an exploratory analysis attempting to highlight patterns of collaboration in social medias. Particularly, we focus on crisis management through the Twitter social platform. We rely on two types of crisis: a natural disaster (Sandy hurricane) and a sanitary crisis (Ebola). We proceed in a three step-methodology consisting in 1) analysing the collaboration structure, 2) then identifying topics of collaboration and 3) finally presenting an overview of the latent collaboration network. Results highlight that collaboration is generally performed in a small network while characterized by diversified topics.*

MOTS-CLÉS : *Réseaux sociaux, Collaboration, Accès à l'information, Twitter*

KEYWORDS: *Social networks, Collaboration, Information access, Twitter*

1. Introduction

Des statistiques récentes sur les services des réseaux sociaux révèlent que près de 3,5 milliards d'acteurs y sont présents en 2015[1]. Des pans entiers de notre société évoluent dans leurs modes d'interaction, de production et d'accès à l'information (Lin et Lu, 2011). Cette évolution offre d'importantes opportunités de collaboration à large échelle et ce, pour diverses fins, dont la résolution d'un besoin en information (De Choudhury *et al.*, 2014) ou l'accomplissement de tâches (Dorn *et al.*, 2011). Dans le cadre de cet article, nous nous intéressons particulièrement à l'accès et au partage collaboratifs à l'information pour la résolution de besoins en information implicites ou explicites exprimés sur les plates-formes sociales. En effet, des travaux précédents ont montré que : *1*) les situations de recherche complexes, qualifiées d'exploratoires, gagneraient en efficacité si elles étaient réalisées de façon collaborative, et non individuelle, par un groupe d'utilisateurs (Shah, 2012 ; Soulier *et al.*, 2014), *2*) le besoin de recherche d'information collaborative est de plus en plus prépondérant dans les réseaux sociaux, aussi bien pour des besoins individuels (Morris, 2013) que pour des besoins collectifs comme dans le cas d'une gestion de crise (Heverin et Zach, 2012). Dans le but d'explorer la question de la collaboration dans les réseaux sociaux dans un objectif de partage et de réponse à un besoin en information, nous menons une étude exploratoire. Cette étude a pour objectifs de mieux comprendre les formes de collaboration associées, leur structures et leurs sémantiques, ainsi que d'identifier les verrous posés par une collaboration à une échelle aussi large que celle du Web 2.0.

L'organisation de cet article est la suivante : la section 2 introduit la notion de collaboration puis présente une synthèse des travaux sur l'accès collaboratif à l'information dans les plates-formes sociales. La section 3 détaille le cadre et les résultats de notre étude exploratoire. Enfin la section 4 dresse un bilan de notre étude puis cerne les principaux résultats ainsi que les pistes de travaux futurs qui en découlent.

2. Accès collaboratif à l'information et réseaux sociaux

2.1. *Notion de collaboration*

En lien avec les objectifs de notre étude exploratoire, nous déclinons la notion de collaboration sous l'angle de l'accès à l'information et de la gestion de crise.

– *Recherche d'information collaborative (RIC)*. Dans (Foster, 2006), les auteurs définissent la RIC comme l'étude des systèmes et des pratiques permettant à un groupe d'utilisateurs de collaborer durant une activité de recherche d'information pour répondre à un besoin en information commun. La collaboration permet ainsi de faire émerger une connaissance ou une intelligence collective qui permet la résolution de besoins informationnels complexes (Poltrock *et al.*, 2003). Néanmoins, la multiplicité des formes de collaboration rend les systèmes qui les supportent difficilement comparables. Pour pallier ce problème, une taxonomie de la collaboration en RIC a été proposée selon quatre dimensions (Gene Golovchinsky et Back, 2008) : *1*) l'*intention* (explicite vs. implicite) selon que le besoin en information est explicitement exprimé

1. http ://www.blogdumoderateur.com/chiffres-reseaux-sociaux/

ou implicitement perçu par les autres, *2*) la *médiation* selon que le pilotage de la coordination est assuré par les collaborateurs (via des interfaces généralement) ou par le système, *3*) la *concurrence* précisant l'aspect synchrone ou asynchrone de la collaboration, *4*) la *localisation* liée à l'aspect distribué ou centralisé du lieu de réalisation de la tâche de recherche d'information collaborative.

– *Collaboration pour la gestion de crise*. Dans les situations de crise (catastrophe naturelle ou industrielle, explosion de violence…), différents intervenants (hôpitaux, pompiers, police, samu, ONG…) sont amenés à agir de manière coopérative pour réduire la criticité et les impacts de la crise (Faure *et al.*, 2010). La coordination de ces intervenants est un point déterminant pour la maîtrise et la résolution de la crise. Fréquemment, cette coordination est assurée par une cellule de crise, composée des représentants des institutions impliquées dans la résolution de la crise. La cellule de crise met en œuvre des plans, qui peuvent être adaptés suivant l'évolution de la situation sur le terrain, des ressources disponibles et des échanges collaboratifs entre acteurs au sein de la cellule (collaboration horizontale) mais aussi entre les membres de la cellule et les acteurs sur le terrain (collaboration verticale). Il arrive de plus en plus que des associations ou des citoyens volontaires deviennent des acteurs à part entière de la gestion de crise. Ils peuvent aider à différents niveaux : en apportant des secours ou des ressources, en informant de la situation sur le terrain (vidéos, commentaires…). Ces volontaires se mobilisent sur les réseaux sociaux (Imran *et al.*, 2013) et quelques fois des organisations informelles émergent et s'auto-organisent (répartition des tâches (Roy *et al.*, 2013), répartition des secteurs géographiques…) pour secourir, financer l'aide, construire la meilleure représentation visuelle de la situation. Dans l'autre sens, les institutionnels sont de plus en plus à l'écoute des réseaux sociaux où ils découvrent des informations (avant leur arrivée par les circuits officiels) ou des collaborateurs, mesurent l'émotion, autant d'éléments à prendre en compte pour le pilotage de la crise.

2.2. *Partage et recherche d'information sur Twitter*

De nombreux travaux attestent l'usage de Twitter pour deux tâches dominantes qui induisent de la collaboration entre utilisateurs, qui sont, en l'occurrence, le partage et la recherche d'information (De Choudhury *et al.*, 2014 ; Morris et Teevan, 2012). Selon les dimensions évoquées précédemment, la collaboration sur Twitter peut être caractérisée par : *1*) une intention explicite (résolution de besoins en information) ou implicite (partage d'information), *2*) une médiation supportée par des protocoles de communication entre utilisateurs (ex. *hashtag*, *mention*, *retweet*, *reply*), *3*) un déroulement généralement asynchrone et *4*) une localisation distribuée. Dans (Morris et Teevan, 2012), les auteurs ont montré l'apport des plates-formes sociales comme Twitter et leur complémentarité avec les moteurs de recherche d'information. Plus précisément, les études montrent que les motivations envers l'usage des réseaux sociaux pour répondre à des besoins en information ponctuels (qui relèvent davantage des objectifs des moteurs de recherche d'information (Ben Jabeur *et al.*, 2012)) sont notamment : *1*) la création ou le renforcement de liens «sociaux», *2*) l'accès à des réponses personnalisées, produites par des utilisateurs perçus comme étant plus engagés

dans la collaboration, *3*) la crédibilité des réponses basée sur le principe de présence du lien «social» avec son auteur.

2.3. *Les réseaux sociaux comme support de collaboration pour la gestion de crises*

De nombreux médias sociaux conventionnels (Facebook, Twitter...) sont couramment utilisés dans la gestion de crise, mais des plates-formes dédiées ont également vu le jour (Sahana, Ushaidi, OneResponse, Google Crisis Response...). Twitter, Facebook et Ushahidi sont les plus utilisés (Wendling *et al.*, 2013) en situation de crises (notamment lors du tremblement de terre de Haiti en 2010 ou de l'accident nucléaire de Fukushima en 2011). Certains outils (ex. NYPA) offrent plusieurs fonctions tout en s'interfaçant avec les outils conventionnels et/ou des Systèmes d'Information Géographique. En termes d'usage, ces différents outils permettent d'une part aux citoyens de géo-localiser des éléments (événements, victimes, demandes, ressources...), d'être informés et actifs dans la résolution de la crise, de diffuser l'information dans la sphère publique et/ou de suivre l'évolution de la crise, et d'autre part, aux organisations parties prenantes, de collaborer, d'être plus efficaces et d'accélérer leur prise de décision. Malgré ce potentiel technologique, les volontés institutionnelles ne sont pas toujours au rendez-vous du point de vue organisationnel et d'autres difficultés demeurent. Parmi les freins principaux identifiés dans ce cadre là, la question de la participation des personnes concernées, victimes de la crise bien souvent, reste le plus problématique. L'apparition de zones aveugles au cours de la crise, dans lesquelles les communications sont perturbées, soit pour des raisons d'engorgement par suite d'une activité trop forte, soit parce que les infrastructures elles-mêmes ont pu subir des dommages suite à la crise, est également un phénomène fréquemment abordé et qu'il s'agit de prendre en compte dans ce type de support aux interventions (Kavanaugh *et al.*, 2013). Le traitement de toutes ces difficultés exige en amont de mieux connaître les structures de coopération existantes entre les différents acteurs de la crise.

3. Description de l'étude

Inscrite dans le cadre d'une gestion de crise, les objectifs de l'étude sont : *1*) la construction des réseaux latents de collaboration, *2*) la caractérisation des motifs récurrents de collaboration dans ces réseaux. À la différence de l'étude présentée dans Smith *et al.* (2014) où les auteurs s'intéressent à la forme des interactions sociales, nous proposons d'analyser de façon solidaire à la fois la structure et la thématique du réseau de conversations dans une situation de collaboration. Nous présentons dans ce qui suit les collections de données utilisées ainsi que les résultats de notre analyse.

3.1. *Collections de données*

La gestion de crise étant un contexte spécifique où émergent spontanément des groupes et des rôles sociaux qui répondent à des besoins généraux communs (secours, conseil...) (Hiltz *et al.*, 2011), nous avons fait le choix d'analyser deux collections de tweets, construites en utilisant l'API Streaming de Twitter, lors de deux événements de crises à grande échelle : *1*) l'ouragan Sandy qui a touché les côtes Est des États-Unis

en octobre 2012, causant 210 morts et un coût de 50 milliards de dollars[2] ; *2*) l'épidémie de virus Ebola qui a touché les pays d'Afrique de l'Ouest, apparue en septembre 2013 et ayant conduit à plus de 10 mille décès[3]. Dans le but d'améliorer la qualité des données, nous avons filtré les données dites «utiles» comme des données crédibles en lien avec la crise. Le filtrage a été basé sur un classifieur automatique entraîné sur une sous-collection de tweets (i.e., 1800 tweets sélectionnés aléatoirement pour chacune des collections Sandy et Ebola) annotée manuellement par dix volontaires. Afin d'apprendre le modèle de classification, nous avons extrait douze caractéristiques associées à trois catégories qui sont le contenu, la typographie et le lexique. Nous avons estimé le modèle de régression logistique pour la classification de tweets en utilisant les deux classes définies («Utile» et «Inutile»). La phase de validation du classifieur a montré un taux de 81%. Les tweets classés automatiquement dans la classe «Utile» constituent la collection finale utilisée dans la suite des analyses. Le Tableau 1 présente des statistiques de chaque collection.

Collection	Sandy	Ebola
Tweets	2 119 854	2 872 890
Microbloggers	1 258 473	750 829
Retweets	963 631	1 157 826
Mentions	1 473 498	1 826 059
Réponses	63 596	69 773
URLs	596 393	1 309 919
Photos	107 263	310 581

Tableau 1. *Statistiques descriptives des jeux de données nettoyés.*

3.2. *Résultats*

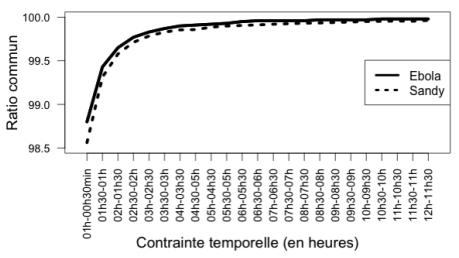
Notre étude exploratoire est basée sur une analyse en trois étapes permettant *1*) d'identifier les groupes de collaboration explicites sous-jacentes à des conversations (Section 3.3), *2*) de détecter les thématiques de collaboration (Section 3.4) et *3*) d'agréger les analyses structurelles et sémantiques afin de construire le réseau latent de collaboration (Section 3.5).

3.3. *Analyse structurelle des groupes de collaboration*

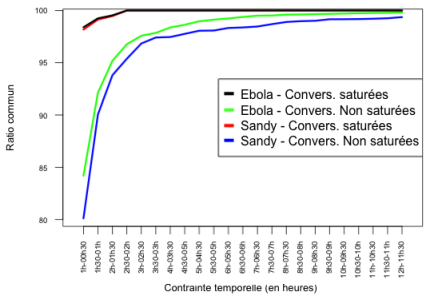
L'hypothèse qui a servi de base pour cette analyse est que les interactions (*reply*, *retweet*, *mention*) entre utilisateurs d'un réseau social sont des traits caractéristiques de la collaboration, matérialisée par des conversations. Afin de retracer la conversation autour d'un tweet, nous utilisons l'algorithme proposé par Cogan *et al.* (2012), qui dans un premier temps identifie le tweet racine de façon ascendante et retrace l'arborescence des interactions de façon descendante. À l'issue de cette étape, nous avons obtenu respectivement 240 991 et 196 005 conversations pour les collections Sandy et Ebola. Afin de pallier la problématique des conversations profondes, où la

2. https ://fr.wikipedia.org/wiki/Ouragan_Sandy
3. https ://fr.wikipedia.org/wiki/Virus_Ebola

thématique principale de la conversation peut dériver vers d'autres thématiques, nous ajoutons une contrainte temporelle qui limite la durée de publication entre le tweet racine et le dernier tweet considéré de la conversation. Cette contrainte temporelle consiste en une extraction d'une partie de la conversation et nécessite un paramétrage. Afin d'avoir un aperçu du ratio coût/bénéfice de cette troncature, nous avons représenté dans la Figure 1(a) le pourcentage moyen d'utilisateurs en commun pour une même conversation construite à partir de deux contraintes temporelles successives, noté «Ratio Commun». En d'autres termes, pour une troncature t donnée, il s'agit du ratio de recouvrement entre la population collectée jusqu'à t et celle jusqu'à t+30 minutes. L'axe des abscisses dans la Figure 1(a) représente les deux contraintes temporelles considérées pour l'estimation du «Ratio Commun». Compte tenu des courbes de la Figure 1, nous choisissons un intervalle de temps de 2h pour les deux collections qui permet d'avoir un taux de recouvrement entre deux intervalles de temps supérieur à 99%. Ce choix correspond par ailleurs à une densité raisonnable en termes de nombre de voisins dans le graphe social inter-utilisateurs induit. Afin d'avoir un meilleur aperçu quant à l'impact de cet intervalle sur les conversations extraites, nous avons séparé les conversations en deux groupes, les conversations «saturées» dont le ratio est égal à 100% au bout de 2h, et celles «non saturées» dont le ratio est inférieur à 100%. Nous avons ensuite estimé le ratio commun pour chacun de ces groupes. Les résultats sont présentés dans la Figure 1(b). Nous observons également que pour les conversations «non saturées», le ratio commun est supérieur à 95% pour un intervalle de temps de 2h, ce qui nous conforte dans notre paramétrage. Le Tableau 2, qui présente les statistiques détaillées associées aux collections, confirme la validité d'une contrainte temporelle de 2h (pourcentage de propriètés partagées $> 78\%$).



(a) Ratio d ?utilisateurs commun entre conversations



(b) Ratio d ?utilisateurs commun entre conversations en tenant compte du taux de saturation

Figure 1. *Paramétrage de la contrainte temporelle.*

Afin de mieux comprendre le comportement des utilisateurs, nous proposons d'illustrer les patrons de conversations les plus fréquents présentés dans le Tableau 3. Nous avons identifié ainsi les configurations en termes de nombre d'utilisateurs-nombre de relations basées sur les interactions sociales (notées y ;x dans le Tableau 3) à partir des conversations. Pour les deux jeux de données, nous observons que les patrons les plus fréquents sont identiques. L'analyse de ces patrons s'apparente à

Jeux de données	Sandy *2h*	base 100	Ebola *2h*	base 100
Nb de tweets	758 887	79,83	878 171	79,27
Nb d'utilisateurs	1 020 213	84,24	1 102 895	83,17
Nb de «Retweet»	702 227	78,75	825 642	78,77
Nb de «Reply»	56 682	96,28	52 624	88,06
Nb de «Mention»	90 370	83,51	157 457	80,23
Nb de conversations	240 991	100,00	196 005	100,00
Nb moyen de tweets par conversation	3,15	79,83	4,48	79,27
Nb moyen d'utilisateurs par conversation	4,23	84,24	5,62	83,17
Nb moyen de «Retweet» par conversation	2,91	78,75	4,21	78,77
Nb moyen de «Reply» par conversation	0,24	96,28	0,27	88,06
Nb moyen de «Mention» par conversation	0,37	83,51	0,80	80,23
Profondeur moyenne des conversations	1,10	99,68	1,23	98,83

Tableau 2. *Statistiques des jeux de données Ebola et Sandy pour la contrainte temporelle 2h. base 100 : comparaison avec la configuration sans contrainte temporelle*

l'étude des fréquences de motifs de cascades de diffusion dans les systèmes de blogs (Leskovec *et al.*, 2007 ; Franco et Kawai, 2010). Il s'agit par contre ici de motifs de conversations, c'est-à-dire de sous-graphes sociaux de mentions, retweets et reply afin d'observer, *in fine*, des rôles distincts (Eleta et Golbeck, 2014). Nous nous intéressons ainsi directement à la signification sociologique des patrons observés, dans un contexte de collaboration potentielle. On observe principalement deux types de patrons qui renvoient à deux types de dynamiques conversationnelles et de rôles distincts, à petite échelle : *1*) des réseaux en étoile (motifs 2 ;1, 3 ;2, 4 ;3, 6 ;5, etc.), avec une personne centrale et des individus qui s'y rapportent. Il s'agit principalement de sous-groupes de relais, où les individus périphériques citent l'utilisateur central et/ou en retransmettent le contenu. *2*) Des réseaux plus aplatis (motifs 4 ;4, 3 ;3, ou 5 ;6, représentant notamment un carré, un triangle ou un carré surmonté d'un triangle), qui correspondent davantage à des discussions collectives autour d'un ou plusieurs tiers.

3.4. *Analyse sémantique des sujets de collaboration*

Pour l'analyse des thématiques sous-jacentes aux conversations extraites lors de l'étape précédente, nous avons utilisé l'algorithme d'allocation latente de Dirichlet (LDA) (Blei *et al.*, 2003) en l'appliquant aux méta-documents formés des tweets associés à chacune des conversations. Afin d'identifier le nombre optimal de thèmes, nous avons utilisé la mesure de perplexité, comme proposé par Blei *et al.* (2003). Nous avons obtenu une valeur minimale de la mesure de perplexité pour 16 et 21 thèmes, respectivement pour les collections Ebola et Sandy. Trois juges humains ont effectué de façon indépendante une annotation manuelle et non supervisée des thèmes extraits automatiquement par la méthode LDA. En cas de désaccord quant aux libellés des thèmes, un consensus a été ensuite trouvé entre les trois annotateurs. Le Tableau 4 énonce les libellés obtenus pour chacune des collections. Nous pouvons observer que généralement les thématiques sont reliées avec des thématiques de gestion de crise, comme celles liées à la prévention, aux prières, aux victimes ou encore aux remercie-

Configuration	# Sandy	# Ebola	Patron	Configuration	# Sandy	# Ebola	Patron
2 ;1	157 687	96 573		3 ;2	36 929	35 694	
4 ;3	12 124	11 639		5 ;4	5 568	6 058	
4 ;4	2 767	4 342		6 ;5	3 394	3 855	
7 ;6	2 177	2 862		8 ;7	1 528	2 434	
5 ;6	1 446	2 322		3 ;3	750	2181	

Tableau 3. *Illustration des patrons de conversation les plus fréquents. # Sandy ou # Ebola : Nombre de conversations pour les jeux de données Sandy ou Ebola*

Collection	Thèmes
Sandy	*1*) État de la ville de New-York ; *2*) Pensées négatives ; *3*) Dons/aides ; *4*) Remerciements ; *5*) Explications ; *6*) Eau/Inondation ; *7*) Insultes ; *8*) Photos/Vidéos ; *9*) Morts/Décès ; *10*) Après sandy ; *11*) Dommages ; *12*) Gens perdus ; *13*) Prières ; *14*) Obama et l'attaque de Benghazi ; *15*) Alertes météos et nucléaires ; *16*) Humour ; *17*) Peur/Terreur ; *18*) Impact financier ; *19*) Rapport/État des lieux ; *20*) Outils de communication ; *21*) Informations via les médias
Ebola	*1*) Prévention ; *2*) Actions/Pensées envers les autres ; *3*) Rapports officiels ; *4*) Pensées personnelles ; *5*) Morts/Décès ; *6*) Urgences à l'échelle mondiale ; *7*) Exil ; *8*) Propagation ; *9*) Tests cliniques ; *10*) Recherche médicaments/Vaccins ; *11*) Soins ; *12*) Premier cas aux USA ; *13*) Maladie/Peur aux USA ; *14*) Victimes et quarantaine ; *15*) Plan d'action en Afrique ; *16*) Contrôle de la propagation

Tableau 4. *Thèmes extraits à partir des jeux de données.*

ments (Imran *et al.*, 2013). On note cependant quelques thèmes en marge, tels que les insultes (Sandy - 7) ou l'humour («Obama et l'attaque de Benghazi» Sandy - 14) qui sont liés aux taux d'erreurs de l'algorithme de classification (81%) lors du nettoyage des données.

Sandy				Ebola			
Thème	# Convers.	# Utilis.	# Tweets	Thème	# Convers.	# Utilis.	# Tweets
1	12 325	49 979	36 618	1	42 811	238 022	188 987
2	17 857	74 853	55 493	2	13 282	71 808	56 656
3	12 573	55 557	41 938	3	10 640	56 377	44 240
4	13 779	56 475	41 509	4	9 846	54 375	43 153
5	9 743	41 583	30 974	5	10 103	54 849	43 166
6	10 365	41 108	29 887	6	9 041	51 871	41 478
7	38 781	163 959	121 852	7	12 950	75 458	60 597
8	10 152	41 505	30 482	8	8 719	49 632	39 629
9	8 297	34 678	25 642	9	7 081	38 012	29 932
10	8 928	36 354	26 703	10	11 294	64 080	51 103
11	6 842	28 256	20 818	11	5 762	32 750	26 151
12	9 642	42 084	31 557	12	11 383	70 132	57 151
13	23 329	102 708	77 465	13	6 925	41 058	33 032
14	9 477	41 560	31 273	14	20 594	114 007	90 347
15	11 672	51 118	38 489	15	8 374	49 266	39 660
16	4 633	19 199	14 234	16	7 202	41 200	32 888
17	5 461	24 195	18 339				
18	8 278	35 337	26 313				
19	5 897	23 053	16 666				
20	6 817	29 855	22 496				
21	6 164	26 797	20 138				

Tableau 5. *Statistiques des regroupements des conversations par thème*

3.5. *Analyse des formes de collaboration*

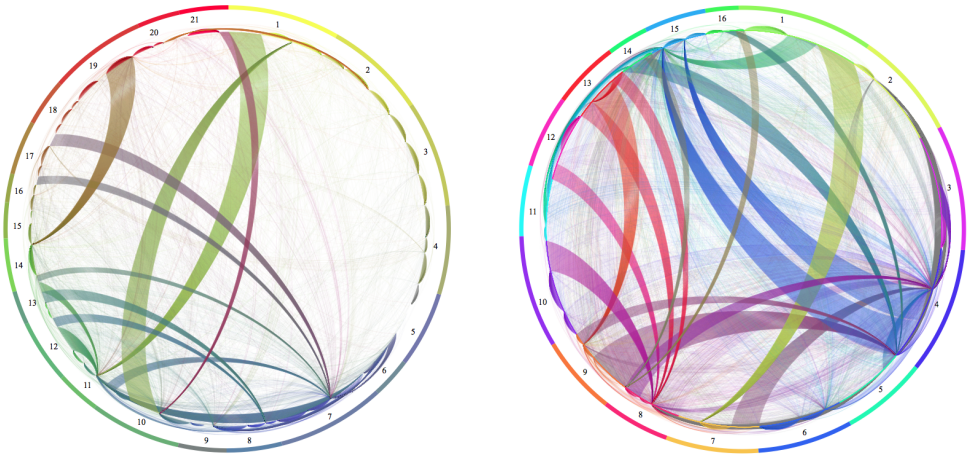
Dans le but de relier l'analyse structurelle à l'analyse sémantique des réseaux latents de collaboration, nous procédons en deux étapes :

1. Nous identifions pour chaque conversation la thématique la plus vraisemblable en retenant la paire conversation-thème ayant la probabilité maximale à partir de la distribution thème-document fournie par l'algorithme LDA. Le Tableau 5 présente quelques statistiques de ces regroupements par thématique. En observant les thèmes les plus représentés en termes de nombres de conversations, il en ressort les thématiques suivantes : *1*) pour Sandy : les prières, les pensées négatives et les remerciements et *2*) pour Ebola : la prévention, les victimes/quarantaine et les actions/pensées vers les autres.

2. Le réseau global est obtenu en représentant les conversations au niveau utilisateur et en les regroupant par classe de thèmes en fonction de l'assignation thème-document obtenue à l'étape précédente. Nous ajoutons également les relations entre tous les utilisateurs du réseau à partir des interactions sociales (*reply, retweet, mention*).

Les Figures 2(a) et 2(b) présentent respectivement les réseaux de collaboration latente obtenus à l'issue de cette étape pour les deux collections Sandy et Ebola. Les jeux de données étant très importants, nous avons représenté le réseau latent de collaboration en filtrant les trois conversations les plus peuplées, en termes de nombre

d'utilisateurs, pour chaque thème LDA. Les thèmes de collaboration sont dispersés autour du périmètre du cercle tandis que l'on distingue deux types de relations : *1*) les relations «intra», au sein même d'un thème, illustrées par des demi-cercles autour du périmètre et *2*) les relations «inter», entre deux thèmes différents, qui sont traversants. Nous observons que les relations «intra» sont plus prédominantes pour le jeu de données Sandy, alors que celui d'Ebola est caractérisé par un nombre plus important de liens «inter». Bien que ces figures représentent une représentation du sous-graphe global que l'on aurait obtenu à partir de l'ensemble des conversations, ces observations sont confirmées par la valeur de mesure de cohésion qui estime le ratio entre liens «intra» et liens «inter» (une valeur supérieure à 1 souligne la prépondérance des liens «intra»). Respectivement pour Sandy et Ebola, la mesure de cohésion est estimée à 1,43 et 0,63. Afin de comprendre ces résultats, nous analysons le graphe au niveau utilisateur. Les utilisateurs pouvant engager ou participer à plusieurs conversations, elles-mêmes assimilées à différentes thématiques, nous avons estimé le nombre moyen de thèmes associés à un utilisateur. Pour Ebola, nous avons obtenu une moyenne de 7 tandis que pour Sandy, un utilisateur est affilié en moyenne à 4 thèmes. Ce résultat peut expliquer les résultats de la mesure de cohésion dans la mesure où si un utilisateur est affilié à plusieurs thèmes, il a plus de chance de générer davantage de liens «inter» qu'un utilisateur associé à un seul thème. En outre, la construction du jeu de données pourrait être une explication, dans la mesure où le jeu de données Sandy a été collecté sur quelques jours, favorisant des interactions concentrées sur une même thématique (liens «intra»), alors que le jeu de données Ebola a été collecté sur un mois, plus favorable à des interactions couvrant plusieurs thématiques (liens «inter»).



(a) Jeu de données Sandy

(b) Jeu de données Ebola

Figure 2. *Réseaux latents de collaboration*

4. Conclusion

Nous avons présenté dans cet article une analyse exploratoire des formes de collaboration latente pouvant être extraites à partir des interactions sociales pour l'accès

à l'information. Un choix particulier a été porté sur la plate-forme Twitter, connue comme la plate-forme sociale la plus largement utilisée pour l'échange instantané d'informations, et deux cas de gestion de crise, reconnu comme un cadre qui a suscité dans le passé des collaborations spontanées entre citoyens lors de différentes crises sanitaires et/ou naturelles. Les résultats montrent paradoxalement que le périmètre de la collaboration est relativement réduit (taille des groupes entre 2 et 7 utilisateurs), que les sujets sous-jacents sont diversifiés et impliquent des utilisateurs qui peuvent être à l'interface de différents sujets. Notre étude est cependant limitée à l'analyse thématique en termes de sujets de conversations. Une analyse des actions (secourir, financer, apporter de l'information...) et rôles sous-jacents aux collaborations est nécessaire en vue de mieux comprendre la finalité de la collaboration et inférer son succès vs. son échec. Les résultats de notre étude permettent d'ouvrir des pistes de recherche atour de deux aspects complémentaires : *1*) la conception d'algorithmes et techniques qui facilitent la médiation entre utilisateurs d'un réseau social en vue d'une collaboration qui étend le périmètre de la collaboration sociale au-delà du voisinage structurel dans le réseau ; *2*) la définition d'une ontologie de collaboration qui allie sujet et action véhiculés dans l'interaction sociale sous-jacente à la collaboration.

5. Remerciements

Ce travail a été réalisé avec le support du projet EXPAC dans le cadre de l'appel PEPS[4] CNRS 2014-2015 .

6. Bibliographie

Ben Jabeur L., Tamine L., Boughanem M., « Uprising microblogs : A Bayesian network retrieval model for tweet search », *Proc. ACM Symposium on Applied Computing (SAC)*, ACM, p. 943-948, 2012.

Blei D. M., Ng A. Y., Jordan M. I., « Latent Dirichlet Allocation », *J. Mach. Learn. Res.*, vol. 3, p. 993-1022, 2003.

Cogan P., Andrews M., Bradonjic M., Kennedy W. S., Sala A., Tucci G., « Reconstruction and Analysis of Twitter Conversation Graphs », *Proc. of the First ACM Int. Workshop on Hot Topics on Interdisciplinary Social Networks Research*, HotSocial '12, ACM, p. 25-31, 2012.

De Choudhury M., Morris M. R., White R. W., « Seeking and Sharing Health Information Online : Comparing Search Engines and Social Media », *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, ACM, 2014.

Dorn C., Skopik F., Schall D., Dustdar S., « Interaction Mining and Skill-dependent Recommendations for Multi-objective Team Composition », *Data Knowl. Eng.*, vol. 70, n° 10, p. 866-891, October, 2011.

Eleta I., Golbeck J., « Multilingual use of Twitter : Social networks at the language frontier », *Computers in Human Behavior*, vol. 41, p. 424-432, 2014.

Faure C., Andonoff E., Hanachi C., Sibertin-Blanc C., Salatge N., « Flexibilité de processus de gestion de crise par intégration de protocoles d'interaction », *Ingénierie des Systèmes d'Information*, vol. 15, n° 3, p. 37-60, 2010.

4. Projet Exploratoire Premier Soutien

Foster J., « Collaborative information seeking and retrieval », *Annual Review of Information Science and Technology*, 2006.

Franco L., Kawai H., « News Detection in the Blogosphere : Two Approaches Based on Structure and Content Analysis », *Proc. 4th ICWSM AAAI Intl. Conf. on Weblogs and Social Media*, 2010.

Gene Golovchinsky J. P., Back M., « A taxonomy of collaboration in oline information seeking », *1st Intl Workshop on Collaborative Information Seeking*, 2008.

Heverin T., Zach L., « Use of Microblogging for Collective Sense-making During Violent Crises : A Study of Three Campus Shootings », *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, n° 1, p. 34-47, January, 2012.

Hiltz S. R., Diaz P., Mark G., « Introduction : Social Media and Collaborative Systems for Crisis Management », *ACM Trans. Comput.-Hum. Interact.*, vol. 18, n° 4, p. 18 :1-18 :6, December, 2011.

Imran M., Elbassuoni S., Castillo C., Diaz F., Meier P., « Practical Extraction of Disaster-relevant Information from Social Media », *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, p. 1021-1024, 2013.

Kavanaugh A., Sheetz S. D., Quek F., Kim B. J., « Cell phone use with social ties during crises : The case of the Virginia Tech tragedy. », *Using Social and Information Technologies for Disaster and Crisis Management*, 2013.

Leskovec J., McGlohon M., Faloutsos C., Glance N., Hurst M., « Cascading Behavior in Large Blog Graphs », 2007.

Lin K.-Y., Lu H.-P., « Why People Use Social Networking Sites : An Empirical Study Integrating Network Externalities and Motivation Theory », *Comput. Hum. Behav.*, vol. 27, n° 3, p. 1152-1161, May, 2011.

Morris M. R., « Collaborative Search Revisited », *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, ACM, p. 1181-1192, 2013.

Morris M. R., Teevan J., « Exploring the Complementary Roles of Social Networks and Search Engines », HCIC'12, ACM, 2012.

Poltrock S., Grudin J., Dumais S., Fidel R., Bruce H., Pejtersen A. M., « Information Seeking and Sharing in Design Teams », *SIGGROUP*, p. 239-247, 2003.

Roy S. B., Lykourentzou I., Thirumuruganathan S., Amer-Yahia S., Das G., « Crowds, not Drones : Modeling Human Factors in Interactive Crowdsourcing », *Proceedings of the First VLDB Workshop on Databases and Crowdsourcing, DBCrowd 2013, Riva del Garda, Trento, Italy, August 26, 2013*, p. 39-42, 2013.

Shah C., *Collaborative Information Seeking - The Art and Science of Making the Whole Greater than the Sum of All*, vol. 34 of *The information retrieval series*, Springer, 2012.

Smith M. A., Rainie L., Shneiderman B., Himelboim I., « Mapping Twitter Topic Networks : From Polarized Crowds to Community Clusters », 2014.

Soulier L., Tamine L., Bahsoun W., « On domain expertise-based roles in collaborative information retrieval », *Information Processing & Management*, vol. 50, n° 5, p. 752 - 774, 2014.

Wendling C., Radisch J., Jacobzone S., « The Use of Social Media in Risk and Crisis Communication », *OECD Working Papers on Public Governance*, 2013.