



HAL
open science

Communication virale dans la publicité au sein des espaces numériques : jeux de données

Ugo Roux

► **To cite this version:**

Ugo Roux. Communication virale dans la publicité au sein des espaces numériques : jeux de données. Revue française des sciences de l'information et de la communication, 2019, 18, 10.4000/rfsic.7784 . hal-02951386

HAL Id: hal-02951386

<https://hal.science/hal-02951386v1>

Submitted on 29 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Communication virale dans la publicité au sein des espaces numériques : jeux de données

Ugo Roux



Édition électronique

URL : <http://journals.openedition.org/rfsic/7784>

DOI : 10.4000/rfsic.7784

ISSN : 2263-0856

Éditeur

Société Française de Sciences de l'Information et de la Communication

Ce document vous est offert par Aix-Marseille Université (AMU)



Référence électronique

Ugo Roux, « Communication virale dans la publicité au sein des espaces numériques : jeux de données », *Revue française des sciences de l'information et de la communication* [En ligne], 18 | 2019, mis en ligne le 01 décembre 2019, consulté le 29 septembre 2020. URL : <http://journals.openedition.org/rfsic/7784> ; DOI : <https://doi.org/10.4000/rfsic.7784>

Ce document a été généré automatiquement le 29 septembre 2020.



Les contenus de la *Revue française des sciences de l'information et de la communication* sont mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Communication virale dans la publicité au sein des espaces numériques : jeux de données

Ugo Roux

- 1 La viralité est un phénomène communicationnel majeur qui s'est grandement développé au sein du creuset fertile et créatif que sont Internet et le Web 2.0. Dresser une « théorie de la viralité » (Beauvisage et *al.*, 2011) serait aujourd'hui difficile, voire impossible. En effet, la viralité est la résultante d'une multiplicité de facteurs potentiellement à son origine, qui lui sont favorables et/ou qui l'orientent (p. ex. Beuscart et *al.*, 2009 ; Beuscart et Couronné, 2009). Mais de plus en plus de recherches qui s'intéressent aux phénomènes viraux mettent aujourd'hui en avant des facteurs qui jouent un rôle majeur dans le développement de ces événements communicationnels. Ainsi, les dispositifs éditoriaux qui peuvent mettre en avant un objet plutôt qu'un autre par recommandation (p. ex. Cha et *al.*, 2008), la structure du réseau et des interactions sociales qui l'animent (p. ex. Centola, 2010), les éléments créatifs de l'objet (p. ex. Porter et Golan, 2006), le profil et les motivations des usagers de médias sociaux (p. ex. Borges-Tiago et *al.*, 2019) ou encore les émotions et l'activation physiologique provoquées par l'objet (p. ex. Berger, 2011) sont autant de facteurs qui, parmi d'autres, sont reconnus pour avoir potentiellement une influence notable sur le devenir d'un objet culturel évoluant dans les espaces socionumériques – par exemple une vidéo.¹ Cependant, aucun chercheur ne s'était penché sur l'influence de facteurs techniques tels que la qualité de la définition sur la diffusion en ligne d'une vidéo (publicitaire) jusqu'à ce que nous le fassions dans notre thèse et dans des travaux postérieurs (Roux, 2016, 2018). Sur ce point l'approche théorique semblait donc limitée sur le plan épistémologique. L'étude d'un tel facteur pourrait sembler négligeable et ses conclusions évidentes, mais il n'en est rien. En effet, la viralité est une intrication factorielle tellement complexe, erratique et labile qu'il convient de s'attarder sur le moindre détail.

- 2 Le présent article nourrit deux enjeux majeurs. Tout d'abord, et nous rejoignons en cela l'objectif suivi par cette nouvelle rubrique *Data Papers* de la *Revue Française des Sciences de l'Information et de la Communication* (RFSIC), nous nous positionnons dans une logique d'accessibilité et d'intégrité du savoir scientifique (Le Deuff, 2018). Nos travaux ont produit de nombreuses données que nous souhaitons désormais partager et diffuser auprès de la communauté scientifique, en particulier celle des sciences de l'information et de la communication, afin qu'elle puisse les exploiter à des fins heuristiques. Ces données n'ont pas encore révélé leur plein potentiel, car nombre d'entre elles n'ont été que peu, voire pas du tout, exploitées. Enfin, nous rejoignons également l'objectif de valorisation de la variété des données obtenues et par extension des méthodes qui ont permis leur obtention. Ainsi donc, nous nous positionnons dans cette démarche de promotion d'un pluralisme méthodologique en sciences de l'information et de la communication pour mettre en avant l'intérêt de l'emploi de la méthode expérimentale – méthode que nous avons employée – dans notre (inter-)discipline, et ce, à travers les données que nous partageons ici. Il semblerait que cette méthode ne soit que très peu développée et exploitée dans les sciences de l'information et de la communication malgré la mise en œuvre d'une diversité méthodologique par les chercheurs dans ce champ de recherche (Bernard et Joule, 2005). La faute incomberait, selon Matalon (1988), à son côté artificialisant qui ne serait pas adapté aux situations réelles et naturelles du social. Pourtant, la méthode expérimentale permettrait, d'un point de vue « quantitatif », d'élargir le nombre des objets et des problématiques possibles dans un contexte de complexité, d'élargir les possibilités de définition des concepts et, d'un point de vue « qualitatif », de donner du sens et d'expliquer des phénomènes communicationnels, de stimuler la réflexion épistémologique ou encore d'aider à la solidification de la construction des connaissances par l'apport de nouvelles méthodes et de leur croisement (Courbet, 2013). Pour reprendre les mots d'Olivier Le Deuff, « sur l'ensemble de ces points énoncés, les SIC cherchent à être moteurs » (2018).

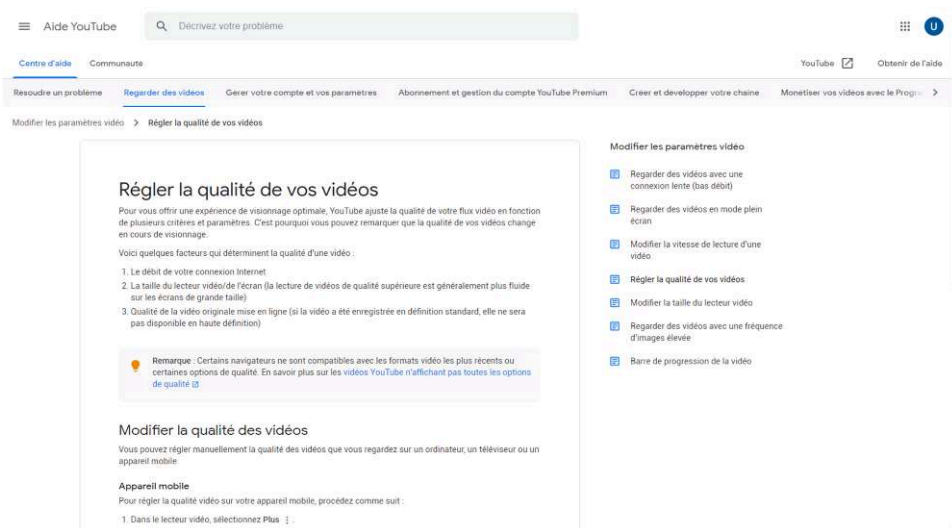
Expérience

- 3 Avant de présenter nos jeux de données, il nous semble judicieux de dérouler le fil de notre pensée et de nos protocoles qui ont guidé nos travaux afin de donner au lecteur une vision claire sur l'ensemble de ceux-ci et sur la manière dont ces jeux de données ont été produits et exploités. C'est pourquoi nous décrivons en détail dans cette partie le déroulement de notre expérience. Nous proposerons également une synthèse de quelques-uns de nos résultats afin de donner au lecteur un aperçu de l'usage que nous avons pu faire de nos jeux de données. Les conclusions tirées de ces résultats ne seront toutefois pas commentées ici. Aussi, nous invitons le lecteur à consulter nos travaux (Roux, 2016, 2018) s'il souhaite connaître les résultats complets, détaillés et commentés de notre étude.
- 4 Savoir ce que nous voulions démontrer était essentiel pour structurer l'ensemble de notre recherche pour atteindre un quelconque résultat. En effet, « dans une recherche bien menée, c'est l'aval (les objectifs visés et les analyses envisagées) qui doit déterminer l'amont (la nature et l'organisation des données nécessaires) » (Matalon, 1988, p. 11). C'est ainsi que nous nous sommes posé la question de recherche suivante : pourquoi certaines vidéos sont-elles partagées plus que d'autres sur Internet ? Autrement dit, nous voulions savoir quels éléments créent, favorisent et/ou orientent

un événement viral. Si nos lectures avaient alors déjà mis en évidence certaines des raisons qui font que des objets connaissent plus de succès que d'autres, nos observations ont suscité chez nous des interrogations sur l'influence de la qualité de la définition sur le potentiel viral d'une vidéo en ligne.

- 5 Les variations de la qualité de la définition d'une vidéo sont un phénomène « habituel » lorsque l'on visionne des vidéos hébergées sur les plateformes de *streaming* – en particulier *YouTube* dans le cas de notre expérience. Ces variations peuvent être la conséquence de plusieurs facteurs. Par exemple, *YouTube* prévoit d'adapter automatiquement la qualité de la définition des vidéos qu'elle donne à visionner, et ce, en fonction du débit de la connexion à Internet de l'utilisateur qui les visionne. Aussi, *YouTube* permet à ses usagers de modifier eux-mêmes, dans la limite du possible, les paramètres de lecture des vidéos que la plateforme héberge. En effet, *YouTube* (et les autres plateformes de *streaming* également) ne permet de modifier la qualité de la définition de la vidéo visionnée qu'à hauteur et en deçà de la qualité de la définition du fichier vidéo hébergé. Enfin, la qualité de la définition peut varier en fonction de la taille du lecteur vidéo ou de l'écran où est diffusée la vidéo et aussi en fonction du navigateur où celle-ci est visionnée. Le centre d'aide *YouTube* rappelle ces différents facteurs (figure 1). Cette variable – de la qualité de la définition – a ceci d'intéressant qu'elle est aisément manipulable en laboratoire, ce qui rend d'autant plus évident et pertinent l'usage de la méthode expérimentale qui est « considérée comme le mode le plus sûr d'administration de la preuve » (Matalon, 1988, p. 10). En nous basant sur un traitement statistique, les conclusions tirées à partir des résultats obtenus par cette méthode nous semblent ainsi d'autant plus robustes qu'ils ne laissent que peu de place à la subjectivité dans l'interprétation des données récoltées. Autrement dit, la méthode expérimentale nous permet de mieux assurer la validité de notre compréhension de phénomènes et de situations complexes comme la viralité.

Figure 1 – Régler la qualité de la définition sur YouTube



source : <https://support.google.com/youtube/answer/91449?hl=fr>

Hypothèses

- 6 Comme nous avons pu le voir, les variations de la qualité de la définition d'une vidéo sont un phénomène que l'on peut régulièrement observer lorsque l'on visionne des vidéos hébergées sur les plateformes de *streaming* comme *YouTube*. Ce constat a fait émerger la question du lien entre variations de la qualité de la définition d'une vidéo, appréciation et partages de celle-ci. C'est sur cette base empirico-inductive qui a animé notre recherche doctorale – et celles qui ont suivi – que nous avons émis les hypothèses principales suivantes :
- 7 H1 : L'appréciation (variable dépendante) que l'on a du contenu d'une vidéo est influencée par sa qualité de la définition (variable indépendante). Plus précisément, nous estimons que le contenu d'une vidéo est plus apprécié visionné en haute définition (HD) plutôt qu'en définition standard (SD) et inversement.
- 8 H2 : Par corollaire, les intentions de partage (variable dépendante) que l'on peut exprimer vis-à-vis d'une vidéo sont influencées par sa qualité de la définition (variable indépendante). Plus précisément, nous estimons qu'une vidéo visionnée en haute définition est plus volontiers partagée que la même vidéo visionnée en définition standard et inversement.

Matériel et méthodes

Procédures expérimentales

- 9 Notre expérience s'est déroulée en cinq étapes (ou phases). L'expérience a débuté par la phase 0 (manipulation exploratoire) puis la phase 1.0. Cette dernière étape a été répétée trois autres fois (phases 1.1, 1.2, 1.3) en faisant varier certains paramètres. En voici le descriptif :
1. La phase 0 était une manipulation exploratoire qui avait pour objectif de vérifier la faisabilité de notre expérience en testant notre protocole expérimental ; elle a été réalisée sur 3 sujets. Les résultats s'étant montrés concluants, nous les avons rajoutés à ceux de la phase 1.0 qui suit. Cette manipulation a été réalisée avec le corpus de 36 vidéos (corpus 36).
 2. La phase 1.0 fut menée en groupe (12 sujets) dans une salle informatique tandis que les phases suivantes furent organisées en séances individuelles pour des raisons logistiques et pratiques. Cette phase a été réalisée avec le corpus de 36 vidéos (corpus 36).
 3. La phase 1.1 (16 sujets) était un test de reproductibilité de la phase 1.0. Cette manipulation a donc aussi été réalisée avec le corpus de 36 vidéos (corpus 36).
 4. La phase 1.2 (13 sujets) était un test de reproductibilité avec un corpus réduit à 24 vidéos (corpus 24).
 5. La phase 1.3 (12 sujets) était un test de reproductibilité de la phase 1.2 qui proposait une même liste de lecture aux sujets pour vérifier s'il existait des effets d'ordre et/ou de proximité. Cette manipulation a donc aussi été réalisée avec le corpus de 24 vidéos (corpus 24).
- 10 Pour chacune des phases précédemment décrites, nous avons appliqué les procédures expérimentales qui suivent :
1. Installation des sujets devant les ordinateurs (avec casques audio) ;
 2. Lecture individuelle des consignes sur les différents questionnaires distribués à chaque sujet ;

3. Présentation des outils et du déroulement de l'expérience avec possibilité de précisions ;
4. Remplissage de la fiche de renseignement distribuée à chaque sujet ;
5. Visionnage individuel du corpus en respectant l'ordre indiqué, grâce à un code unique à chaque vidéo, sur la liste de lecture remise à chaque sujet (ordre établi par un tirage aléatoire différent pour chacun des sujets) ;
6. Pause individuelle après chaque vidéo pour répondre par écrit aux questions lui correspondant sur le questionnaire prévu à cet effet.

Sélection du corpus vidéo

- 11 L'expérience repose sur deux variables principales ; à savoir, la qualité de la définition (variable indépendante intra-sujet provoquée [contrôlée]) qui se décline en deux modalités (haute définition [HD] et définition standard [SD]) et la viralité (variable indépendante intra-sujet invoquée) qui se décline aussi en deux modalités (virale [V] et non-virale [NV]). Le corpus retenu est composé de vidéos publicitaires en ligne (récoltées sur *YouTube*) sélectionnées en fonction de ces variables et modalités. Autrement dit, le corpus est composé de publicités virales et de publicités non-virales (corpus témoin), chacune d'entre elles étant présentée en deux définitions (HD [720p] et SD [144p]). Enfin, les vidéos sont réparties de manière égale selon trois thématiques ; à savoir, alimentaire, automobile et sport. Chaque vidéo s'est vu attribuer un code d'authentification composé de trois lettres et d'un chiffre. La première lettre correspond au caractère viral de la vidéo (V pour viral ou N pour non-viral), la seconde à la qualité de sa définition (H pour haute définition ou S pour définition standard) et la dernière à sa thématique (A pour alimentaire, V pour automobile ou S pour sport). Le chiffre distingue les vidéos qui conjuguent les mêmes variables (1, 2 ou 3 pour le corpus 36 et 1 ou 2 pour le corpus 24). Le détail de la sélection et de la répartition des vidéos est précisé dans les jeux de données et leurs annexes.
- 12 Nous avons sélectionné les vidéos publicitaires virales qui composent notre corpus grâce au site Internet de l'agence *Advertising Age*. *Advertising Age* dresse en effet un classement hebdomadaire des vidéos publicitaires les plus virales. Quelques études récentes ont proposé de caractériser et de modéliser les dynamiques virales (p. ex. France et al., 2016 ; Boppolige et Gurtoo, 2017) ; il ressort de ces travaux que les vidéos virales ont en commun le fait qu'elles comptent un très grand nombre de vues et de partages en des laps de temps très courts, et ce, peu de temps après leur diffusion. En revanche, nous n'avons retenu que des vidéos publicitaires avec un nombre de vues inférieur à 20 000 un mois après leur mise en ligne pour constituer le corpus non-viral. Il s'agit d'un choix arbitraire qui offre une marge de sécurité très confortable qui permet d'affirmer raisonnablement que ces vidéos ne sont pas virales.

Sélection des sujets

- 13 L'échantillon des 53 sujets uniques – qui n'ont participé qu'une seule fois à notre expérience – était composé comme suit :
 - 71,43 % d'étudiants de l'Université de Toulon ;
 - 49,2 % d'hommes et 50,8 % de femmes ;
 - 84,1 % de sujets âgés entre 18 et 25 ans et 15,9 % de sujets âgés de plus de 25 ans.

- 14 Même si notre étude permet de dégager des tendances, il convient de déterminer le caractère généralisable de celles-ci à d'autres populations et/ou d'autres zones géographiques.

Mesures

- 15 Trois questionnaires ont été utilisés comme outils de mesure et de récolte des données :
1. Le premier questionnaire est une fiche de renseignements sur le sujet. Ce questionnaire est divisé en trois parties ; à savoir, des renseignements sociodémographiques, des renseignements sur les activités physiques et sportives et des renseignements en lien avec la consommation de contenus vidéo en ligne.
 2. Le deuxième questionnaire pose trois mêmes questions pour chaque vidéo ; à savoir, « À quel point appréciez-vous la qualité de la définition de cette vidéo ? » (Q1) (variable dépendante), « À quel point appréciez-vous le contenu de cette vidéo ? » (Q2) (variable dépendante), « Partageriez-vous cette vidéo ? » (Q3) (variable dépendante). Des échelles graduées en dix points (type « Likert » allant de (1) à (10) [variable quantitative discrète] ; (1) correspond au degré le plus bas d'appréciation et (10) correspond au degré le plus élevé d'appréciation) permettent aux sujets de répondre aux deux premières questions (Q1 et Q2). La troisième question (Q3) a pour modalités de réponse : « oui », « non », « peut-être » (échelle et variables qualitatives – ou catégorielles – nominales). Ce questionnaire propose également une quatrième unique question ouverte (Q4) commune à l'ensemble des vidéos : « Pour quelles raisons partageriez-vous ou ne partageriez-vous pas certaines vidéos ? ». Ce questionnaire a pour particularité d'exister en deux versions, une pour 36 vidéos et une pour 24 vidéos.
 3. Le troisième questionnaire n'a été distribué qu'à partir de l'expérience 1.1. Il s'agit d'un questionnaire sur les valeurs motivationnelles de l'athlète (Albouza, 2014 ; Albouza *et al.*, 2014) – nous voulions alors établir des liens avec les activités physiques et sportives des sujets. Ce questionnaire contient 20 items auxquels on répond avec une échelle graduée en six points (type « Likert » allant de (1) à (6) [variable quantitative discrète] ; (1) correspond au degré le plus bas de ressemblance et (6) correspond au degré le plus élevé de ressemblance). Chaque item renvoie à une proposition où il est demandé au sujet d'évaluer sa ressemblance avec un profil d'athlète. Ces propositions et ces profils correspondent à l'une des cinq valeurs motivationnelles de l'athlète ; à savoir, affirmation du statut, dépassement de soi, plaisir, respect des règles et morale (il existe des corrélations significatives entre ces valeurs et les valeurs motivationnelles non sportives équivalentes).

Données et exploitation²

Présentation des données

- 16 L'ensemble des données récoltées lors de notre expérience, les différents outils de mesure et de récolte de ces données ainsi que des exemples de fiches d'ordre de lecture sont accessibles en *open access* selon les termes prévus dans la licence *Creative Commons Attribution 4.0 International*. Ces fichiers sont disponibles à l'adresse suivante : <https://zenodo.org/record/3350486> (Roux, 2019). Précisons que les données présentées ont été récoltées dans la zone géographique environnant l'Université de Toulon sur une période de temps qui s'étend de mars 2014 à janvier 2015.

- 17 Les données sont réparties dans quatre fichiers qui respectent les recommandations relatives aux jeux de données formulées par *Open Data France*. Ces jeux de données sont donc enregistrés sous le format *.csv* encodé en *UTF-8* :
1. Le premier d'entre eux contient les données récoltées avec les questionnaires des expériences (20191025_roux-2016-questionnaire.csv). Ce jeu de données présente deux types d'échelles et de variables ; à savoir, qualitatives – ou catégorielles – ordinales (réponses aux Q1 et Q2), qualitatives – ou catégorielles – nominales (réponses aux Q3). De plus, ce jeu de données contient aussi les réponses faites à la question ouverte Q4).
 2. Le deuxième fichier contient les données récoltées avec les fiches de renseignements (20191025_roux-2016-fiche-renseignements.csv). Ce jeu de données présente trois types d'échelles et de variables ; à savoir, qualitatives – ou catégorielles – nominales (par exemple le sexe ou la profession), quantitatives de ratios – ou de rapports – discrètes (par exemple l'âge) et quantitatives de ratios – ou de rapports – continues (par exemple le temps passé sur Internet).
 3. Le troisième fichier contient les données récoltées avec le questionnaire des valeurs motivationnelles de l'athlète (20191025_roux-2016-qvmh.csv). Ce jeu de données présente un seul type d'échelles et de variables ; à savoir, qualitative – ou catégorielle – ordinale.
 4. Le quatrième fichier contient les ordres de lecture spécifiques à chaque sujet (20191025_roux-2016-liste-lecture.csv).

Tests de reproductibilité

- 18 Avant d'envisager de traiter, d'analyser ou d'interpréter les données récoltées, il nous a fallu vérifier la reproductibilité³ de celle-ci et la fiabilité de notre protocole. Cette étape fut essentielle pour nous assurer de la robustesse des résultats obtenus et pour ainsi tirer des conclusions valides. C'est pourquoi nous avons divisé notre expérience en quatre phases principales dont nous avons comparé les résultats. Pour ce faire, nous avons comparé en inter-groupes les moyennes des résultats obtenus aux mêmes questions d'une même catégorie de vidéos (les réponses à la Q1 ont été comparées entre elles en fonction des discriminants HD ou SD tandis que les réponses à la Q2 ont été comparées entre elles en fonction des discriminants V ou NV). En d'autres termes, nous comparons par exemple les moyennes des résultats HD obtenus à la Q1 de la manipulation 1.0 aux moyennes des résultats HD obtenus à la Q1 de la manipulation 1.1 et ainsi de suite. Pour cela nous avons réalisé des tests *F de Fisher* (comparaison de variances) puis des tests *t de Student* non-appariés avec seuil à .05 avec correction de *Bonferroni* (comparaison de moyennes). Enfin, nous avons calculé les moyennes des moyennes de ces résultats – et leurs écarts-types – pour faciliter la comparaison des résultats (la figure 2 et le tableau 1 proposent une synthèse et une représentation graphique de ces comparaisons). Lorsque deux phases se révélaient être reproductibles, nous en associons les résultats pour les comparer avec la phase suivante. Autrement dit, une fois la reproductibilité des phases 1.0 et 1.1 démontrée, nous avons réuni leurs résultats (pour en faire la phase 1CC) pour les comparer à ceux de la phase 1.2 et ainsi de suite jusqu'à la phase 1.3. Chaque phase s'étant révélée reproductible nous avons réuni l'ensemble des données obtenues en un seul jeu. C'est sur ce jeu de données unifié que nous avons réalisé nos analyses et interprétations afin d'en tirer des conclusions.

Figure 2 – Tests de reproductibilité des différentes phases de l'expérience

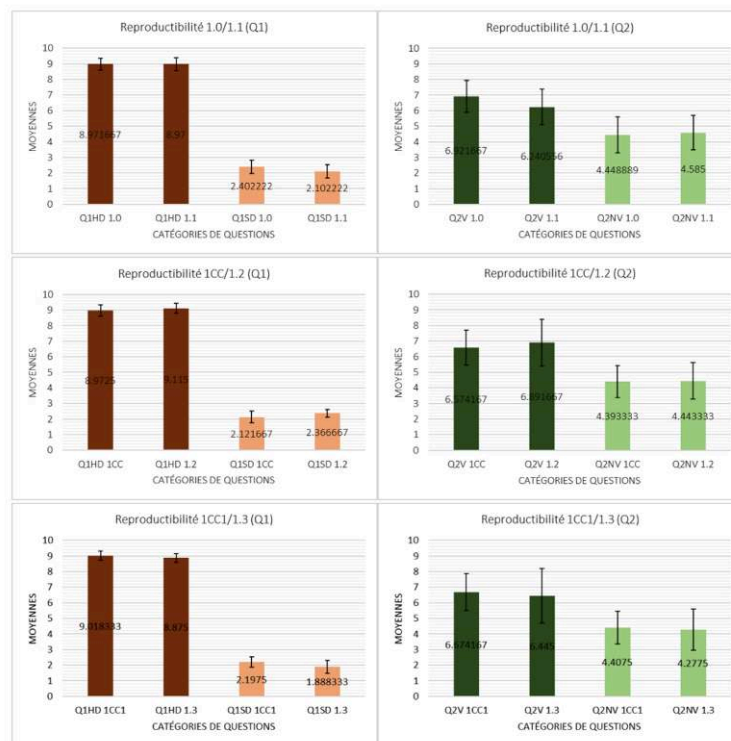


Tableau 1. Reproductibilité des différentes phases de l'expérience

Phases	Comparaison de variances (Test <i>F de Fisher</i>)		Comparaison de moyennes (Test <i>t de Student</i>)			
	F	p	m1	m2	t	p
1.0/1.1 Q1HD	(17, 17) .8578	.7555	8.971667	8.97	(33.802) .0126	.99
1.0/1.1 Q1SD	(17, 17) 1.0103	.9834	2.402222	2.102222	(33.999) 2.0974	.04347
1.0/1.1 Q2V	(17, 17) .8088	.6667	6.921667	6.240556	(33.624) 1.887	.06782
1.0/1.1 Q2NV	(17, 17) 1.0642	.8994	4.448889	4.585	(33.967) -0.3638	.7183
1CC/1.2 Q1HD	(11, 11) 1.371	.6097	8.9725	9.115	(21.474) -1.0701	.2965
1CC/1.2 Q1SD	(11, 11) 2.5235	.1401	2.121667	2.366667	(18.535) -1.909	.07187
1CC/1.2 Q2V	(11, 11) .5367	.3168	6.574167	6.891667	(20.167) -0.5875	.5634
1CC/1.2 Q2NV	(11, 11) .7723	.6757	4.393333	4.443333	(21.653) -0.1119	.9119
1CC1/1.3 Q1HD	(11, 11) 1.1483	.8227	9.018333	8.875	(21.896) 1.1933	.2455
1CC1/1.3 Q1SD	(11, 11) .5919	.3979	2.1975	1.888333	(20.644) 2.0622	.052
1CC1/1.3 Q2V	(11, 11) .4703	.2266	6.674167	6.445	(19.473) .3767	.7105
1CC1/1.3 Q2NV	(11, 11) .6431	.476	4.4075	4.2775	(21.009) 0.2691	.7905

Comparaison de variances et comparaison de moyennes

- 19 Nous avons déjà évoqué dans la précédente partie les comparaisons de variances ainsi que les comparaisons de moyennes. Si nous revenons sur ces deux tests, c'est qu'ils ont démontré leur utilité au-delà de la validation de la reproductibilité de notre expérience. En effet, ces tests nous ont notamment permis de comparer, entre autres, les réponses données aux différentes catégories de questions et ainsi de nous faire une première idée des tendances d'appréciation – et de leurs écarts – selon les différentes variables (la figure 3 et le tableau 2 proposent une synthèse et une représentation graphique de ces comparaisons). Par exemple, lorsque nous avons comparé en intragroupe les réponses aux Q1HD aux réponses aux Q1SD, nous avons pu apprécier

l'écart très significatif d'appréciation de la qualité de la définition qu'ont les sujets entre les vidéos HD et les vidéos SD (ce qui est dû au caractère objectif de l'appréciation de la qualité de la définition) ; *a contrario*, nous avons remarqué que l'écart se réduisait lorsque nous comparions les réponses aux Q1V aux réponses aux Q1NV (ce qui est dû au caractère subjectif de l'appréciation du contenu vidéo).

Figure 3 – Comparaisons réponses Q1HD/Q1SD et Q2V/Q2NV

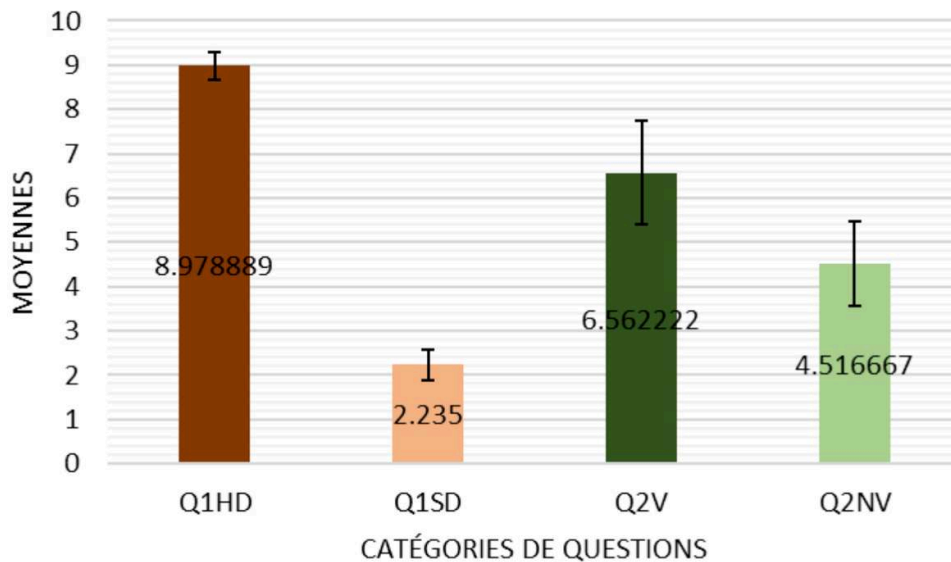


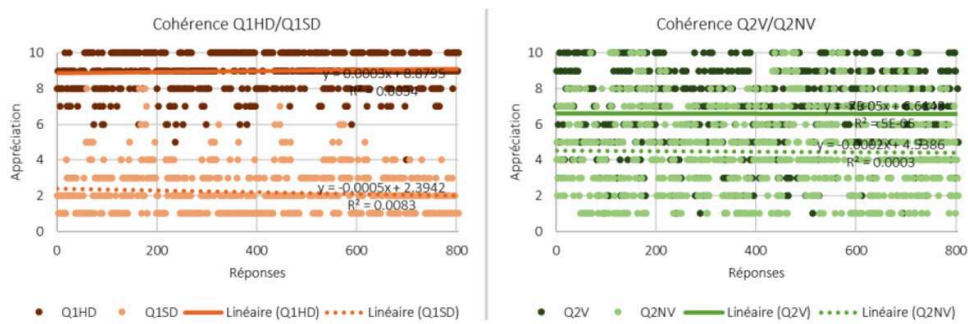
Tableau 2 – Comparaisons réponses Q1HD/Q1SD et Q2V/Q2NV (test t de Student)

Comparaisons	m1	m2	t	p
Q1HD/Q1SD	8.978889	2.235	(33.702) 60.0138	< .001
Q2V/Q2NV	6.562222	4.416667	(32.861) 5.7275	< .001

1.3.3. Cohérence

- 20 Les tests de cohérence ont cet intérêt qu'ils permettent une visualisation graphique et rapide de la tendance des réponses – et de leur cohérence – données par tel ou tel groupe d'individus par rapport à telle ou telle catégorie de vidéos, et ce, sans écraser les réponses en une moyenne. Il est également possible de comparer plusieurs tendances en superposant différents tests de cohérence. Les réponses aux Q1 et aux Q2 ont été exploitées pour réaliser ces tests de cohérence (la figure 4 propose une représentation graphique de ces tests qui éclairent les résultats présentés dans la figure 3 et le tableau 2 ; on remarquera aussi les courbes de tendance qui marquent bien les écarts d'appréciation).

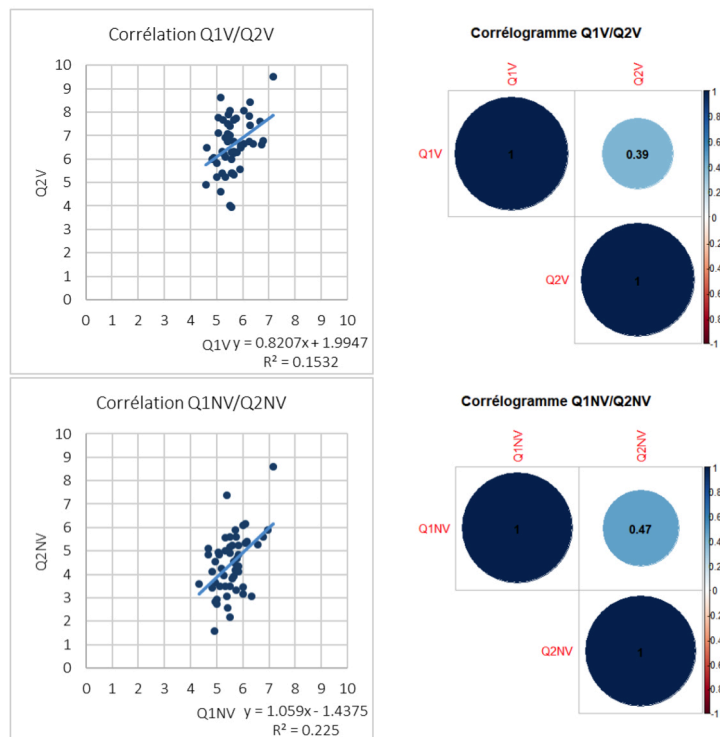
Figure 4 – Cohérences Q1 HD/SD et Q2 V/NV



Corrélations

- 21 La corrélation mesure la relation entre plusieurs variables aléatoires ou statistiques. Dans notre cas, les corrélations nous ont servi à mettre en évidence un lien entre l'appréciation de la qualité de la définition des vidéos et l'appréciation de leur contenu. Pour établir la significativité des corrélations, nous nous sommes appuyés sur les seuils de signification en sciences humaines (Fiske, 2008). Les réponses aux Q1 opposées aux réponses aux Q2 ont été exploitées pour réaliser ces corrélations (la figure 5 propose deux représentations graphiques de ces corrélations). Les résultats obtenus ont démontré une corrélation entre appréciation de la qualité de la définition et appréciation du contenu vidéo. Normalement, une corrélation ne permet pas de dire quelle variable en influence une autre. Toutefois, on peut ici affirmer que la variable « qualité de la définition » influence la variable « viralité » puisque l'inverse ne peut être possible ou, tout du moins, fortement improbable.

Figure 5 – Corrélations et corrélogrammes



1.3.5. Répartition des intentions de partages

- 22 Comparer les répartitions des intentions de partage se révèle utile pour comparer le potentiel viral de chacune des catégories de vidéos que nous avons établies selon les deux variables principales croisées (virale HD [VHD], virale SD [VSD], non-virale HD [NVHD] et non-virale SD [NVSD]). Pour réaliser cette comparaison, ce sont les réponses aux Q3 qui ont été exploitées. Plus précisément, nous avons calculé les répartitions en pourcentages des « oui », « non » et « peut-être » pour chaque catégorie de vidéos et nous les avons comparées entre elles avec des tests du χ^2 (la figure 6 et les tableaux 3 et 4 proposent une synthèse et des représentations graphiques des différentes répartitions des intentions de partage ainsi que de leurs comparaisons). On constate notamment que les vidéos virales sont celles qui obtiennent le plus d'intentions positives de partage lorsqu'elles sont regardées en HD – elles voient cependant leur potentiel viral divisé presque par quatre si elles sont regardées en SD.

Figure 6 – Différentes représentations graphiques des répartitions des intentions de partage (%) (Q3)

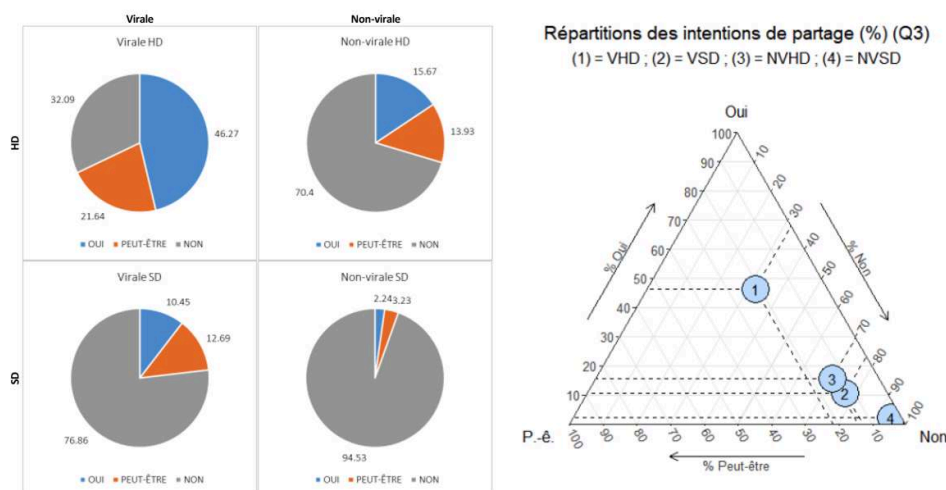


Tableau 3 – Répartitions des intentions de partage (Q3)

Modalités de réponse	VIRALE HD	VIRALE SD	NON-VIRALE HD	NON-VIRALE SD
OUI	186 (46,27 %)	42 (10,45 %)	63 (15,67 %)	9 (2,24 %)
PEUT-ÊTRE	87 (21,64 %)	51 (12,69 %)	56 (13,93 %)	13 (3,23 %)
NON	129 (32,09 %)	309 (76,86 %)	283 (70,4 %)	380 (94,53 %)
TOTAL	402	402	402	402

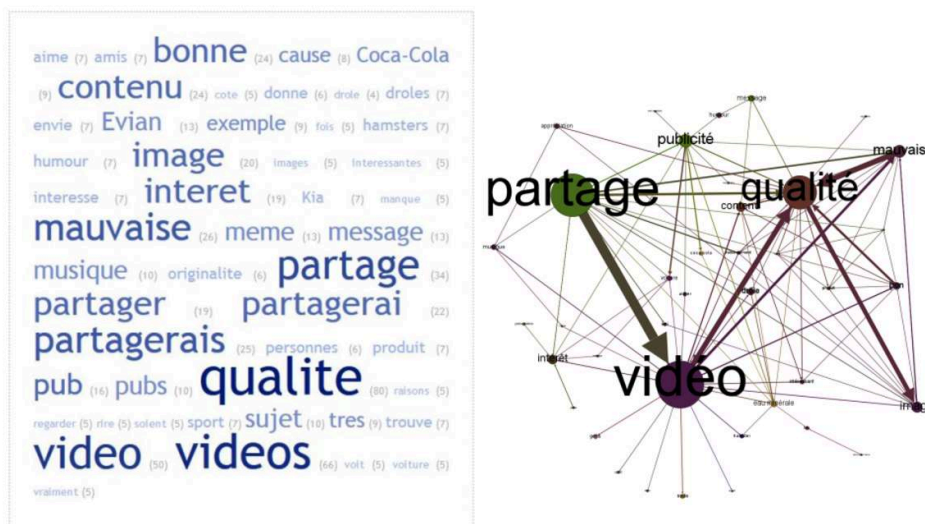
Tableau 4. Comparaisons (χ^2) des répartitions des intentions de partage (Q3)

Variables croisées	VHD	VSD	NVHD	NVSD
VHD	-	$\chi^2 (2) = 43.3515$ $p < .001$	$\chi^2 (2) = 31.1084$ $p < .001$	$\chi^2 (2) = 84.3827$ $p < .001$
VSD	$\chi^2 (2) = 43.3515$ $p < .001$	-	$\chi^2 (2) = 1.3843$ $p = .5005$	$\chi^2 (2) = 12.7547$ $p = .0017$
NVHD	$\chi^2 (2) = 31.1084$ $p < .001$	$\chi^2 (2) = 1.3843$ $p = .5005$	-	$\chi^2 (2) = 20.2729$ $p < .001$
NVSD	$\chi^2 (2) = 84.3827$ $p < .001$	$\chi^2 (2) = 12.7547$ $p = .0017$	$\chi^2 (2) = 20.2729$ $p < .001$	-

Analyse sémantique

23 L'analyse sémantique réalisée nous a permis d'affiner l'interprétation des résultats statistiques obtenus en posant des mots sur les chiffres. La Q4 a fourni la matière pour réaliser cette analyse. Les réponses à cette question sont riches de sens et leur lecture nous a permis de dégager les motivations des individus dans leurs intentions de partage ou de non-partage (la figure 7 propose deux représentations graphiques des lemmes – mais aussi de leurs occurrences et de leurs relations – les plus employés par les sujets dans leur réponse). On remarque dans les déclarations des sujets un lien très marqué entre intentions de partage et qualité de la définition de la vidéo – ce qui corrobore les précédents résultats.

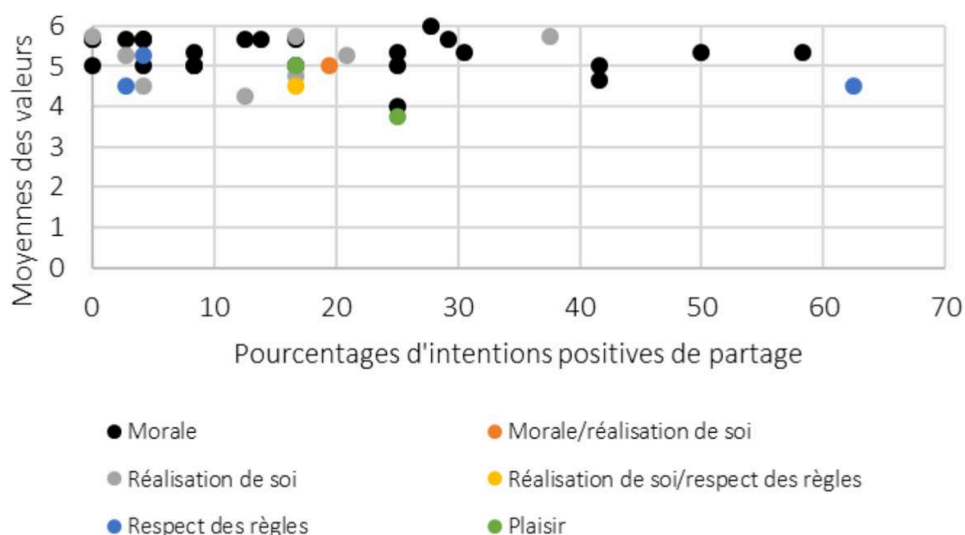
Figure 7 – Nuage de mots et relations terminologiques des termes les plus utilisés



Clustering

24 Le *clustering* permet de dégager des groupes homogènes de données, d’individus ou autres selon des caractéristiques communes. Dans notre étude nous souhaitions vérifier si des groupes d’individus se dégageaient en fonction de leurs valeurs motivationnelles dominantes et de leurs intentions de partage. Les réponses à la Q3 ainsi que les réponses au questionnaire des valeurs motivationnelles de l’athlète ont été exploitées pour réaliser ce *clustering* (figure 8) ; pour déterminer la valeur motivationnelle dominante d’un individu, nous avons fait la moyenne des réponses aux différentes propositions correspondant à une même valeur motivationnelle (nous avons également calculé les écarts-types). Le *clustering* n’a pas permis de dégager un profil de valeurs motivationnelles qui serait susceptible de plus partager qu’un autre.

Figure 8 – *Clustering* des valeurs motivationnelles par rapport aux intentions positives de partage



Conclusion

- 25 Comme nous le disions en introduction, nous nous positionnons dans une logique d'accessibilité et d'intégrité du savoir scientifique. C'est pourquoi nous souhaitons partager et diffuser en *open access* les jeux de données que nous présentons ici auprès de la communauté scientifique, et plus spécifiquement celle des sciences de l'information et de la communication. Les cadres potentiels de réexploitation de ces données peuvent être multiples (tant qu'ils sont en accord avec les termes prévus dans la licence *Creative Commons Attribution 4.0 International*). Celles-ci peuvent déjà être utilisées pour poursuivre nos travaux sur la viralité puisqu'elles n'ont pas toutes été exploitées et croisées (par exemple, nombre de données de la fiche de renseignements n'ont pas encore été utilisées). On peut aussi envisager que ces jeux de données soient utilisés à des fins pédagogiques et de formation aux statistiques (en sciences sociales ou autres) ou à des logiciels de traitements statistiques comme R.

BIBLIOGRAPHIE

- ALBOUZA Youssef, *Le système des valeurs des athlètes et leur rôle dans la régulation de leur agressivité lors des pratiques sportives compétitives*, thèse de doctorat en sciences et techniques des activités physiques et sportives, 2014, 282 p.
- ALBOUZA Youssef, CHAZAUD Pierre, WACH Monique, « Validation d'un questionnaire des valeurs motivationnelles appliquées au sport de compétition », *5^e Congrès internationales de la Société Française de Psychologie du Sport (SFPS)* (pp. 87), Nice, 2014.
- BEAUVISAGE Thomas, BEUSCART Jean-Samuel, COURONNÉ Thomas et MELLET Kévin, « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité », *Tracés*, 2011, 21, n° 2, p. 151-166.
- BERNARD Françoise et JOULE Robert-Vincent, « Le pluralisme méthodologique en sciences de l'information et de la communication à l'épreuve de la "communication engageante" », *Questions de communication*, 2005, 1, n° 7, p. 185-207.
- BERGER Jonah, « Arousal increases social transmission of information », *Psychological Science*, 2011, 22, n° 7, pp. 891-893.
- BEUSCART Jean-Samuel, CARDON Dominique, PISSARD Nicolas et PRIEUR Christophe, « Pourquoi partager mes photos de vacances avec des inconnus ? Les usages de Flickr », *Réseaux*, 2009, 27, n° 154, p. 91-129.
- BEUSCART Jean-Samuel et COURONNÉ Thomas, « La distribution de la notoriété en ligne. Une analyse quantitative de MySpace », *Terrains et travaux*, 2009, n° 15, p. 147-170.
- BOPPOLIGE Archana Anand et GURTOO Anjula, « Growth of a viral phenomenon : Development and testing of a new methodological framework », *Australasian Marketing Journal*, 2017, 25, n° 1, p. 57-66, doi : 10.1016/j.ausmj.2017.01.002.

- BORGES-TIAGO Maria Teresa, TIAGO Flavio et COSME Carla, « Exploring users' motivations to participate in viral communication on social media », *Journal of Business Research*, 2019, 101 ; p. 574-582, doi : 10.1016/j.jbusres.2018.11.011.
- CENTOLA Damon, « The spread of behavior in an online social network experiment », *Science*, 2010, 329, n° 5996, p. 1194-1197.
- CHA Meeyoung, MISLOVE Alan, ADAMS Ben et GUMMADI Krishna P., « Characterizing Social Cascades in Flickr », dans Faloutsos Christos, Karagiannis Thomas et Rodriguez Pablo (sous la dir. de), *Proceedings of the first workshop on Online social networks*, New York, ACM, 2008, p. 13-18.
- COURBET Didier, « Pour un développement raisonné des méthodes expérimentales en SIC. Quelques intérêts épistémologiques du pluralisme méthodologique », *ESSACHESS*, 2013, 6, n° 1(11), p. 15-34.
- FISKE Susan, *Psychologie sociale*, Paris, De Boeck, 2008, 759 p.
- FRANCE Stephen L., VAGHEFI Mahyar Sharif et ZHAO Huimin, « Characterizing viral videos : Methodology and applications », *Electronic Commerce Research and Applications*, 2016, 19, p. 19-32, doi : 10.1016/j.elerap.2016.07.002.
- LE DEUFF Olivier, « Une nouvelle rubrique pour la RFSIC : *Le Data Paper* », *Revue française des sciences de l'information et de la communication*, 2018, n° 15.
- MATALON Benjamin, *Décrire, expliquer, prévoir : Démarches expérimentales et terrain*, Paris, Armand Colin, 1988, 272 p.
- PORTER Lance et GOLAN Guy J., « From Subservient Chickens to Brawny Men : A Comparison of Viral Advertising to Television Advertising », *Journal of Interactive Advertising*, 2006, 6, n° 2, p. 30-38.
- ROUX Ugo, *Communication virale dans la publicité au sein des espaces numériques : approche critique et expérimentale du phénomène*, thèse en sciences de l'information et de la communication, 2016, 432 p.
- ROUX Ugo, « Communication virale dans la publicité au sein des espaces numériques : jeux de données », *Zenodo*, 2019, doi : 10.5281/zenodo.3350486
- ROUX Ugo, « Communication virale dans les espaces numériques : effet de la définition de l'image sur la diffusion d'une vidéo publicitaire en ligne », *Interrogations ?*, 2018, n° 26.
- THOMPSON Clive, Fast Company, [En ligne], 2008 [Page consultée le 23 juillet 2019]. Disponibilité et accès <https://www.fastcompany.com/641124/tipping-point-toast>
- WILSON Ralph F., PracticalEcommerce, [En ligne], 2018 [Page consultée le 23 juillet 2019]. Disponibilité et accès <http://www.practicalecommerce.com/viral-principles>

NOTES

1. L'intégration de ces contenus est un choix éditorial humain, c'est leur implémentation qui est une résultante algorithmique, qu'il s'agisse de données sélectionnées de manière « sociale » ou « centrée sur la donnée (ou les logs d'usage) ».
2. La sauvegarde, le traitement, l'analyse, l'interprétation et la présentation des données quantitatives récoltées ont été réalisés avec l'assistance du logiciel de statistiques R (et RStudio) et du logiciel tableur Microsoft Excel. La sauvegarde, le traitement, l'analyse, l'interprétation et la présentation des données qualitatives récoltées ont été réalisés avec l'assistance du logiciel

d'analyse sémantique *Tropes*, du logiciel tableur *Microsoft Excel*, du logiciel d'analyse et de visualisation de réseaux *Gephi* et de l'application en ligne pour créer des nuages de mots *TagCrowd*.

3. Au-delà de vérifier la validité de notre protocole expérimental, l'intérêt de tester la reproductibilité de nos manipulations est de révéler une loi générale dans le comportement des individus interrogés.

RÉSUMÉS

Le présent article revient sur nos différents travaux pour présenter les données qui ont servi à en dresser les conclusions (Roux, 2016, 2018, 2019). Nos travaux ont produit de nombreuses données que nous souhaitons désormais partager et diffuser auprès de la communauté scientifique, en particulier celle des sciences de l'information et de la communication, afin qu'elle puisse les exploiter à des fins heuristiques. Ces données n'ont pas encore révélé leur plein potentiel, car nombre d'entre elles n'ont été que peu, voire pas du tout, exploitées.

This article comes back to our various works to present the data that helped to draw their conclusions (Roux, 2016, 2018, 2019). Our works have produced many data that we want to share with the scientific community, more specifically the information and communication sciences community, in order to let it use these in a heuristic way. These data have not revealed their full potential yet. Indeed, many of these have not been used yet.

INDEX

Mots-clés : communication virale, viralité, publicité, vidéo, espaces numériques, jeux de données

Keywords : viral communication, virality, advertising, video, digital spaces, data sets

AUTEUR

UGO ROUX

Ugo Roux est docteur en sciences de l'information et de la communication. Il est actuellement attaché temporaire d'enseignement et de recherche (ATER) au sein du département information communication option métiers du livre et du patrimoine de l'IUT d'Aix-Marseille Université. Il est également membre de l'Institut Méditerranéen des Sciences de l'Information et de la Communication – EA 7492 (IMSIC). ugo.roux(at)gmail.com