Republique du cameroun
****
Paix-Travail-Patrie
****
Universite de Ngaoundere

Facultes des sciences

Departement de maths-info

Republique of cameroun
****
Peace-Work-FatherLand
****
University of Ngaoundere

Faculty of sciences

Department of Computer science

Student personnal work

Topic: HOAX

Mention:MasterI SLED

Academic year : 2017-2018

Mentor: **Dr Tchakounte**

April 27, 2018

liste des menbres du groupes

Noms et Prenoms	Matricule
-chassem kamdem priva	14B052FS
-lonfo fofe blondo	16B013FS
-eloudou mvogo victor	14A249FS
-djimtor aladoum serge	14B046FS
-kemogne tchuinte clovis	14A366FS
-tchouke ntongna aristo	13A022FS
-houmgui samngadi	16A742FS
-fotso kuate franck	16B088FS
-bagaza derrick methode	14A888FS
-nadjilem adolphe	16B410FS
-bakou ndjèlémé	16B087FS
-djimbeye ludovic	09B586FS

Contents

Introduction

the web 2.0 designates all the techniques, uses and functionalities that followed the original form of the web characterized by: more simplicity and interactivity. It concerns in particular interfaces and exchanges allowing Internet users with little knowledge to appropriate new functionality of the web. Internet users can, on the one hand, contribute to the exchange of information and interact (share, exchange, etc.) in a simple way, both in terms of the content and the structure of the pages and secondly between them. , creating the social web. the user becomes thanks to the tools at his disposal, an active person on the web. his during the information they relay are not sure or not all simply from sources not sure.It is therefore likely to relay false information or inked either a hoax hoax (false information expire or unverifiable propagate spontaneously by Internet users) with the arrival of social networks including Facebook (2.5 million users in Cameroon), instagram, whassapp the dissemination of information has become easy on the web and offers the possibility of choosing the target audience because faced with naive users the spread of hoaxes can reach cruising speeds.

Faced with the naiveter of most users of web 2.0 it is important to distinguish between hoax and information because they play a role and not least on all aspects of society (politics: bad information can affect the stability of a country, economic: sabotage of products; ...) it is in this momentum that researchers like Marin vukovic, kresimir pripuzic and hrvoje belani from the University of Zagreb worked on an intelligent automatic Hoax detection system. during their work they showed that it was possible to a certain extent to distinguish and classify the true information of hoax but if a false post appeared having no similarity with those contained in the system it would be impossible to classify it.

Cedric Maigrot, Ewa Kijak and Vincent Claveau have been working on "the detection of false information in social networks: the usefulness of knowledge mergers", Eugenio Tachinni, Gabriele Ballarin, Marco L, Marco Della Vedova Stefano Moret and Lucia de Alfaro have is working on the theme "Automated fake news detection in social networks". It is in the same vein that this article is written, the main purpose of which is to present the hoax in their generalities as well as the methods of detection of evaluation and dissemination of these on social media this article is structured as follows and in this order: the means of communication (notament the social media, the mailing ....), the hoax (definition the objectives, the classification of the information, the characteristics , the life cycle, the types) the origin of the information, the evaluation, the

Related work

With regard to the work related to hoax, we put more emphasis on the detection false information in social networks: The analysis of the veracity of information is a line of research is studied in several projects, including work on **the fact checking** to verify the differences in the nature of the information (source, mode of dissemination), and of finality (help for journalism) involve methods different from those used for hoax.
The European project **PHEME** (**Derczynski et al.** , 2015) focuses on the detection of rumors about social media or online media. Several works of this project study the links between messages on social networks. This work focuses on responses and reactions to tweets to decide the truth.
In the same vein , the European project **Reveal Project** (Middleton, 2015b), aims to objective of developing information verification tools and services in social networks, from a journalistic and professional perspective. Different media such as the image (Zampoglou et al. , 2015), the video (Middleton, 2015a), and the text are analyzed. However, it is a matter of developing non-automatic tools but help for journalists. The works also make extensive use of external resources (Gottron et al. , 2014).
Finally, InVid 1 is a European project that focuses on automatic detection fake videos, working on images from videos analyzed (Foteini et al. , 2016).
Another angle of study of social networks in the literature is the analysis sources of messages and relationships between members. (Golbeck and Hendler, 2006) propose a measure of trust between users of social networks, which characterizes the trust of a relationship between two users. This relationship of trust can thus serve as an index to judge the reliability of the information transmitted. Several approaches have been proposed to determine the credibility of a source. . (Gupta et al. , 2012) propose an application of the PageRank algorithm (Page et al. , 1999) on a graph representing the relationships between tweets, the authors of these tweets and the events associated with these tweets.
These approaches, however, require extensive knowledge of the network that makes it difficult for them to applicable in practice for commercial and consumer social networks.
In regards to the analysis of images circulating on social networks to determine their veracity, the problem is multiple. An image may have been intentionally modified

(falsified), or used to illustrate a statement with which it has no relation (diversion). Two categories approaches to address these problems exist. On the one hand, those based on an analysis image statistics to detect changes. For example, in the case of images in JPEG format, it is possible to spot a double compression (Bianchi and Piva, 2012) and thus a partial modification of the image. (Goljan et al. , 2011) is base on the knowledge of the imprint of the camera capturing an image, which will be modified in the case of a modification of the image. The other category of approach uses information external to the image to determine its integrity. It is in this case to search in a database (or the web) for similar images or identical to determine if the image has been modified or diverted. The problem looking for similar images is an active domain whose latest works are base on deep convolutional neural networks to describe and compare images (Wan et al. , 2014). like other articles one has the research of automatic hoax detection interest in the last decade, but with the partial results and solutions that are based on different approaches, eg heuristics, traffic analysis, etc. Authors of developed a service that receivesand evaluates e-mail messages that users forward when they suspect hoax. Their approach and results are interesting, but not applicable for real-time detection, in which we are interested.
Furthermore, it is more difficult to detect if an unsuspected message is hoax.

Chapter un

Background

1.1 Communication means

Second generation of the World Wide Web, especially the movement away from static web pages to dynamic and shareable content and social networking. Web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web. There are over 100 million worldwide views per day on YouTube There are more Facebook users (350 million) than there are U.S. residents (309 million) Twitter will process about 10 billion "tweets" this year Fastest-growing Facebook segment: ages 35-54 "Force multiplier" through sharing The way business is being done. Nearly 3 in 4 people participate in at least one online community or social network. Facebook (58%) YouTube (31%) Myspace (24%) Twitter (15%) Groups (12%) LinkedIn (9%) Forums (9%) Blog (7%) Flicker (6%) Yelp (2%) Foursquare (2%).The majority (82 percent) participates in social media at least once a week, with nearly half participating every day or nearly every day. Every day or almost every day (48%) A few times a week (25%) Once a week (9%) A few times a month (10%) Less than once a month (8%)

1.2 Social media

1.2.1 Social media concept

Kubey (1998) opined that media education has yet to obtain popular support. He postulated that far more parents, for example, will say that they want their children to be computer literate than will say they want their children to be media literate. Parents believe that computer expertise can equal a leg up in the job market. Joint (2011) has wondered aloud if the new technology is making us stupid. Writing in the journal "ANTAEUS" he asked the question as the title of his research: If Google makes you stupid, what should librarians do about it? Social media as a concept is so new that it has not taken root in the lexicon of the code of ethics of Social Work organizations. Yet, the point cannot be overstated that cutting edge Social work

practice of the future will of necessity involve and include extensive use of social media and web based activities.

1.2.2 How social media Works

Now let`s take a look at each of the main types of social media, and how they work. These explanations are intentionally very general, because with social media every rule seems to have an exception. In fact, among the defining characteristics of social media are the blurring of definitions, rapid innovation, reinvention and mash-ups. Each explanation also has a section on how to try out that form of social media yourself, with pointers on both how to find social media that`s relevant to you and how you might go about creating it. If you want to really understand how social media works, there`s no better way than to take part in it. Mash-ups the combination of two or more pieces of content (or software, or websites) is one of the phenomena in social media that make it at once so exciting, fast-moving and sometimes bewildering. Mash-ups are possible because of the openness of social media many websites and software developers encourage people to play with their services and reinvent them. There are literally hundreds of mash-ups of the Google Earth service, where people have attached information to parts of the maps. A popular type of mash-up cannibalizes different pieces of content, typically videos and music. Popular videos on YouTube can spawn hundreds of imitations, homages and (frequently) comic reinterpretations. In communities like this, the number of mash ups a piece of content spawns is often an indicator of its popularity. Social networks on the web are like contained versions of the sprawling blog network. People joining a social network usually create a profile and then build a network by connecting to friends and contacts in the network, or by inviting real world contacts and friends to join the social network. These communities retain the interest of their members by being useful to them and providing services that are entertaining or help them to expand their networks. MySpace, for instance, allows members to create vivid, chaotic home pages (they`ve been likened to the walls of a teenager`s bedroom) to which they can upload images, videos and music. MySpace has built a lot of its popularity around its music services. There are said to be over three million bands and musicians registered on it, trying to attract a fan base from the 200 million registered accounts. According to Hitwise, in September 2006 MySpace was the 8th largest referrer of traffic to HMV.co.uk, more even than the MSN search engine. In 2007, Facebook, a social network that originated in US colleges, became available for public use in the UK. Its popularity quickly rocketed. Part of Facebook`s success is its creators` decision to ` open up and allow anyone to develop applications and run them on Facebook without charging them. This has seen Facebook users able to play each other at Scrabble and Chess, compare each others`

tastes and send "virtual gifts", among any number of new ideas vying for attention.

1.3 MAILLING

E-mailing has the double advantage of being both efficient and inexpensive. This support has many advantages and seduces more and more users. It allows the constitution of a database by qualifying the customer data as well as possible. Thus you will be able to select your targets using very fine criteria and often nonexistent in traditional databases. It allows the customization of messages. The insertion of personalization fields can be done both in the object and in the content of the message. It ensures a high reactivity of Internet users. Indeed, while for a traditional postal mailing operation, it will take weeks to measure the first returns, an e-mailing campaign will generate nearly 80% of returns in 3 days. It is inexpensive. Indeed, e-mailing is cheaper than a mail merge. It allows a measurability of the different actions of the user. While on other media, only some information is available, with e-mailing, you can quickly know the following indicators, including the number of emails sent, the number of unsuccessful emails, the number of e-mails received, the number of e-mails opened, the number of forward, the number of clicks and the number of subscriptions.

1.4 OTHERS

1.4.1 wikis

The name "Wiki" was chosen by Ward Cunningham the creator of the first Wiki. It is a shortened form of "wiki- wiki", the Hawaiian word for quick. A wiki is a web site that is generally editable by anyone with a computer, a web browser, and an internet connection. Wikis use a quick and easy syntax to allow users to apply formatting to text and create links between pages. This simple formatting syntax means that authors no longer need to learn the complexities of HTML to create content on the web. The main strength of a wiki is that it gives people the ability to work collaboratively on the same document. The only software you need is an Internet browser. Consequently, wikis are used for a variety of purposes. If you make a mistake, it`s easy to revert back to an earlier version of the document.

1.4.2 web blog

A weblog often has the quality of being a kind of "log of our times" from a particular point-of-view. Generally, weblogs are devoted to one or several subjects or themes, usually of topical interest, and, in general, can be thought of as developing commentaries, individual or collective on their particular themes. A weblog may consist of the

recorded ideas of an individual (a sort of diary) or be a complex collaboration open to anyone. Most of the latter are moderated discussions.

1.4.3 micro blogging

Microblogging is a combination of blogging and instant messaging that allows users to create short messages to be posted and shared with an audience online. Social platforms like Twitter have become extremely popular forms of this new type of blogging, especially on the mobile web – making it much more convenient to communicate with people compared to the days when desktop web browsing and interaction was the norm .These short messages can come in the form of a variety of content formats including text, images, video, audio, and hyperlinks. The trend evolved around the later end the Web 2.0 era after social media and traditional blogging merged to create a way that was easier and faster to communicate with people online and keep them informed about relevant, shareable information at the same time. Popular Examples of Microblogging Platforms : You may be using a microblogging website already without even knowing it. As it turns out, short but frequent social posting online is exactly what most people want, given that so many of us browse the web from our mobile devices when we`re out on the go and our attention spans are shorter than ever.

- Twitter

  Twitter is one of the oldest and most well-known social platforms to be put under the "microblogging" category. While the 280-character limit still exists today, you can now also share videos, articles links, photos, GIFs, sound clips, and more through Twitter Cards in addition to regular text.

- Tumblr

  Tumblr takes inspiration from Twitter but has fewer limitations and more features. You can certainly post a lengthy blog post if you want, but most users enjoy posting lots and lots of individual posts of visual content like photosets and GIFs.

- Instagram

  Instagram is like a photo journal for wherever you go. Rather than uploading multiple photos to an album the way we used to do via the desktop web on Facebook or Flickr, Instagram lets you post one photo at a time to show where you are and what you're doing.

1.4.4 podcast

A good starting point, is to think of a podcast as "Internet Radio On-Demand." It`s similar in that you can usually listen to it on your computer â but it`s more than

that. [However, and not to confuse the issue, podcasting isn`t confined to just audio but can be video as well].

With the amount of content that podcasting provides, regular Broadcast Radio, or "Terrestrial Radio" as they call it simply can never compete. The AM and FM radio band only has so many channels. Consequently, radio stations "Broadcast" their content meaning that they attempt to appeal to as broad of an audience as possible. Because, after all, this is what advertisers are looking for. But podcasting, by contrast, is not necessarily hamstrung to advertising revenue like its broadcasting cousin. With its specific and specialized content, it is able to "narrowcast" to only those who choose to listen. So while a particular podcast`s audience may be considerably smaller than the audience of a broadcast, one could argue that the podcast`s audience is a much more targeted and interested in the content being delivered. So, in a way, Satellite Radio, with its ability to provide more channels than Broadcast Radio, takes a step towards podcasting but still does not come close. Podcasts are "On Demand" and can be listened to on your schedule not when a Radio Station decides to air it. So, "t`s kind of like TiVo. Each podcast typically has a website where show episodes can be listened to or downloaded for future listening. With downloaded media, you can either listen to it on your computer or take it with you by transferring it to a portable digital media player or using a podcast app on your phone. So, in this way, it`s kind of like a small paperback book.

But what truly makes a podcast unique, and what gives a podcast its "casting" ability, is how it is able to immediately deliver itself to multiple podcast distribution points (such as iTunes and Sticker Radio) or pod catcher applications through a process of syndication known as RSS (Real Simple Syndication). Listeners can easily "subscribe" to podcasts (most are free) by clicking on its RSS icon or subscription button. The listener is then walked through how to add that podcast´s syndication "feed" to a pod catching application of their choosing. So, when a podcaster releases a new episode, subscribers are automatically notified without having to constantly check back with the podcast`s website to see if a new show has been produced. And, with the podcatching software, episodes of their favorite podcasts can be automatically downloaded all without having to lift a finger. So, in this way, podcasts are like magazine subscriptions. The differential aspect in "casting" is major to where podcasts can have a global audience reach as where tradition radio has a limitation of their broadcasting signal strength.

Chapter Deux

Hoax

Most poignant, shocking or alarming messages circulating on the Internet are hoaxs. Do not forward a received message because its contents "could be true", but step back and if necessary validate information from a reliable source, especially as all hoaxes are not without dangers, including the security of your computer. In this part, we will present the various elements characterizing and penetration trend hoaxs Inthe world, Africa and more specifically in Cameroon

2.1 Definitions and objectives

A hoax is a hoax or false information, outdated or unverifiable spontaneously propagated by the Internet. It may cover any subject that could trigger a positive or negative emotion to the user: virus alert, missing children, the promise of happiness, petition, etc. There are spam because the hoax is distributed manually and voluntarily by people being in good faith but who have not been in that spot the hoax. The aim of the hoax is to spread false information on a massive scale, the democratization of the internet has greatly helped his progress

2.2 Classification of hoaxes

The authors identify hoaxes using classification techniques:

Logistic Regression is a widely used method because it allows modeling or binary variables are binary variables. It is widely used in medicine (healing or not a patient) in sociology, epidemiology, quantitative marketing (or not buying products or services following a share) and finance to model risk (scoring) .The logistic regression and regression lin`eaire belong to the same family modelesGLM (Generalized Linear Models): in both cases one reads a èvènement a combination linèairedes explanatory variables. For règression lin`eaire, the variable dependent follows a normal distribution N (, s) and is a linear function of the explanatory variables. For logistic regression, the dependent variable, also called variable answers following a Bernoulli parameter p (p being the average probability of the event occurring) when the experiment is

repeated once or a binomial distribution when the experiment is repeated n times (eg the same dose is tested on n insects). The probability parameter p here is a function of linear combination of explicatives.Les variables most commonly used functions to link the probability p explanatory variables are the logistic function (called logit) and function of distributions of the normal distribution (this is called probit model). These two functions are symmetric and sigmoid.

The crowdsourcing algorithm is a generic term for a variety of approaches that exploit the potential of large crowds of people in èmettant contribution Appelsa for particular tasks. Although approachescrowd-sourcing can take many different forms, today it is made increasingly via the Web, which allows interaction with a plurality of contributors from around the world. Several approaches crowdsourcingincluent platforms on the Web. This practice allows companies to large-scale human à l`intelligence call to provide solutions to the problems they want to outsource.

2.3 Features hoaxes

The hoaxs can also take the form of a call for help, to organ donation, blood, spinal cord ... playing on the sensitive fiber of users, these messages, even if they were based on a true story, continuing on the canvas, also surfing the success of social media like Facebook and twitter.

These messages differ in some reccurent caracteristics . They advertise all be carriers of a message (hyper) important. This may be the quick win money, help has brought urgently a child in later life, the signature has been made to allow the neighbors not to be expelled or warning against danger imminent. The text is transmitted wants always dramatic or written emphatically. Emotions such as greed, fear, anger or compassion are always titillèes.

Other features, if you receive a hoax, you will be asked of the relayed quickly, most of the time is a maximum of knowledge (sometimes you just need to return to 10 person). It responds like a virus, but unlike these, the intentions of the sender are rarely wrong. In itself, a hoax can not cause any dommagea your computer. However, it may be the cause of the problem by the simple fact of being so frequently transmitted. He indeed uses bandwidth unnecessarily and cause you to lose energy and money foolishly.

NB: Acotè of hoaxs demanding money, those who warn against a computer virus to respond as a wildfire in the mail. Viruses are so feared by the computer owners

that they lend more attention to this type of message. But these hoaxs based on any existing viruses. They announce the loss of all your data stored on the hard drive or the imminent explosion of your computer. Pay attention to these kind of messages and delete them upon receipt. Do not forget either prevent your contact about the movement of such mails.

2.4 Life Cycle of a hoax

A hoax life cycle from its creation to its identification. At first, a hoaxer constructed the message based on the information collected and according to the analysis of the current situation. Then the hoax is widely distributed to across the network peers. They continue to use the hoax for a long time, and then redistribute to their own networks. The hoax can be changed (increased or decreased) in During replication. The root of the hoax can have several copies, including double and different. The hoax will create a negative impact for some time and will be captured at some time t by the victims. This will redistribute its networks additional information on the hoax of his malevolence. The hoax dies when victims have sufficiently recognized malevolence.

2.5 Hoaxes are of different types:

Hoax ad: The hoax spreading false messages announcing that an organization is launching an activity such as recruitment, training, competitions etc ... The messages invite the target to submit nominations before a deadline and mentioned ways (email, number phone) and the name of the contact to contact. The hoaxer can use real advertisements in the past with the legal organization information with a current date.

promoting anti-hoax: hoax spreading false messages to tarnish the brand of a company or conspire against a political figure. The message may report a bad situation involving business and personality. The hoax can broadcast images showing the political person in an adulterous situation. The hoax recruited in the dark by a company A can proliferate information that concurrent B company is involved in a deadly tragedy. The hoaxer often uses images as a message in this case to incite people against the competing company. Another example may be a carrier vehicle a tragic accident picture killing 100 people, involving rival transport company B. Such hoaxes can be used to decrease the odds of a candidate during elections.

Antipolitics hoax: The hoaxer spreading false messages to divert people's government views. For example, an event occurs about the cancellation of meetings of activists

due to non-compliance with administrative rules. But the hoaxer sends images showing contradictory facts with police officers who beat people. Such hoaxes can be used during electoral processes by a candidate to appeal to the people by spreading false facts against the outgoing regime.

Hoax rumor: The hoax spreads a false rumor for some time about a specific situation to analyze the reactions and behavior of people. For example, the president`s death or the end of the manufacture of a product.

Good Fortune / Fortune Poor

The message tells us that it`s our lucky day, or if it does not return the message a great misfortune will befall us. Misinformation

The message usually involves very large groups and announces a scandal that would react any normal web surfing. It invites readers to disseminate up misinformation. In this case, the address of the sender and this is false Practice is not just a hoax but of defamation, which can lead to very serious smart sanctions. Humor

The message addresses a universal and uses professional language subject (doctor, professional, computer, ...) and quickly takes the subject into derision. It can evoke so many areas of activity.Dans this case hoax there is never a valid signature, no name is mentioned. However, the reader will feel touched and concernè.Ce hoax is the most dangerous according Hoaxbuster.com because "it is painless, colorless, odorless, ... Due to its very high degree of contamination, it spreads very rapidly to the brain. It has a flash incubation time. To date no effective remedy against the hoax of humor was found. "

2.6 Tendency of penetration hoax

In Cameroon

In Cameroon, the trend of hoaxes is as strong as in some African countries with the exponential development of new technologies poorly assimilated for a good safe use. These hoaxes, we can mention among others:

He recently which caused a stir in Cameroon. The magazine Jeune Afrique published

on Sunday on the occasion of April 1, an imaginary interview with Samuel Eto`o. In this false interview,the Cameroonian footballer announced his fake candidacy for president of his country in October. But this "April Fool" does not laugh sport. Samuel Eto`o threatened to sue for libel.

Announced to last three days, the mood movement ended abruptly. The movement mood of traders in the city of Bafoussam, began April 3, 2017, has been likened to a "April Fool" bad taste. For the stuffing, came too late in the night of April 2 while the event was obsolete, is soon died of his own death. The mood is again normal. On site languages â that are loosened believe that it is a reckoning, "Some cybercriminals and unidentified wanted to create a psychosis through leaf lets ventilated by night, Bafoussam issue closer to other cities who live in a harmful situation" said a trader who requested anonymity. Market A, for example, large populations convergence pole, serenity returned. All shops have the doors wide open, next to stalls hosting large world. Traders quietly go about their business. Trade and other bystanders are not left behind in this canularesque movement.

The food company U-Fresh said to be defamed as a result of widely reported by the press publications that the quality of its products. Only, no evidence to the contrary has been made. The Director-General Shen Qiang has termed "rumor" and "defamation" these informations. For him, these were made available to the press by his opponents who would fear competition. Only, no evidence of these allegations or guarantees on the quality of U-Fresh products have been distributed to journalists. This suggests that the direction of the U-fresh company had more ambition as she manipulated the press and public opinion, without actually remove the ambiguity about the dangers of its products. Antic estimated at several hundred million CFA francs loss due to intrusions into the information systems of some companies. Multibillion losses FCFA are due to fraud Simbox (electronic box used to be charged the international telephone traffic to the price of domestic tariff by redirecting and manipulating the international Internet traffic, according Minette Libom Li Likeng. Millions of Hoax (false information disseminated by mass on the Internet, including social networks) and spoofing (identity theft) are not far behind. to all this must be added, according to Ms. Li Likeng, that 51% of volume of domestic Internet traffic is related to illegal downloads.

In Africa

There would be worse than a scare to die from Ebola viral Also the disease, "hoax"

has started to be denied that when moviegoers have recognized, the retouched picture, a character from the zombie film " World War Z ". Credulity? This alert is not the only cause trouble in the media. On ABC aired a video of a man moving an arm when it was thought the death of Ebola. In African newspapers "All Africa" âand "The New Dawn", articles evoked two separate cases of people who have died of haemorrhagic fever and would then come back to life in the Liberian Nimba Country. There would be worse than a scare to die from Ebola apprehension to resurrect.

World

In the whole world Guillaume Brossard, founder of false **www.hoaxbuster.com** tracker estimates the number of hoax between 10 to 20% of the total traffic of online messages. For example, a study of IBM Research Labs (Delhi), 29% of the content tweeted following the Boston bombings in 2013 related rumors, while the rest consisted of 20% information and 50% opinions.

Previously, most related hoax fake computer virus and they were mainly distributed by email. Some were found to be mechanisms by hackers to take profit, to collect mail from listings or to install malicious software, others were mere hoaxes.

Today the situation has changed: the hoax further spread via social networks, with a resurgence of ideological messages, especially the extreme right. Over 50% of verification requests are racist. Volunteers sites that fight against the spread of the hoax (mainly **www.hoaxbuster.com** and **www.hoaxkiller.fr** in Francophonie) are over whel med and do not have the means to treat all content. They therefore call for maximum autonomy and vigilance on the part of users

Reach the end of this part, or it is the question for us to present the different concepts and the hoaxs the penetration trend in the world, Africa and more specifically in Cameroon. It enabled us to understand well what is a hoax and that these different types. In the section that follows, we will present the origin of these.

Chapter Trois

Information Provenance

An information source is a person, thing, or place from which information comes, arises or is,obtained. That source might then inform a person about something or provide knowledge about it.

In social media,information is often transmitted and retransmitted from one user to other users and from one social media site to other social media sites.Social media propagates breaking news and disinformations alike fast and on an unsurpassed scale. It is a conglomerate of different types of social media sites,including social networking (e.g., Facebook, LinkedIn, etc.), blogging (e.g., Huffington Post, Business Insider,Engadget, etc.), micro-blogging (e.g., Twitter, Tumblr, Plurk, etc.), wikis (e.g., Wikipedia,Wikitravel, Wikihow, etc.), social news (e.g., Digg, Slashdot, Reddit, etc.),social bookmarking(e.g, Delicious, StumbleUpon, etc.), media sharing (e.g, Youtube, Flickr, UstreamTV, etc.),opinion, reviews and ratings (e.g., Epinions, Yelp, Cnet, etc.), and community QA (e.g.,Yahoo Answers, WikiAnswers, etc.).

In addition to providing a popular means to connect with friends, associates, and family members, social media has practical applications that are benefiting society as a whole. Social media has been used for gathering information about large-scale events such as fires, earthquakes, and other disasters, all of which impact government and non-government organizations at local, national. Individuals also use social media to find reliable information about what is going on around them and thus are able to leverage new information as quickly as possible.

One characteristic of social media is its low entry barrier enabling its wide use and explosive growth. Users simply need access to the Internet to participate in social media today.

With the ubiquitous availability of computational resources and Internet access, people produce a variety of content and interact with many others directly through social media. This is vastly different from the traditional media such as radio, television, and printed publications that dominated social communications in the past. These traditional media mechanisms are available to individuals or organizations with sufficient financial means to purchase "air time" or space on a printed page. Traditional media convey messages in a one-to-many fashion.

3.1 Diffusion Methods

The actors:
The followers: The followers are the users having made the choice to subscribe to the account of another (and thus to follow it) celebrity, company, brand out simply autension, this tespread ers to followers of social platforms such as Instagram, Soundcloud or Ask.However, followers play a very important role in spreading false information on the internet by sharing a false post with their friends, who in turn share with theirs.

Spreaders:
Spreads are followers who share a fake post with friends or followers. Friends tend to form groups. So, for example, as Alice knows Bob and Clive, the latter two are likely to know each other too, and therefore share the same opinions on a lot of things. Most of us often see some memes (the term usedfor characterize a link, video, phrase, or other online information unit.) many times, which increases the likelihood that we will share them as well. In addition, thecontagious nature of a meme - unlike viruses - depends on its number of shares, which makes things worse. In 2006, sociologist Matthew Salganik of Columbia University studied the behavior of 14,000 individuals who were much more likely to download a song if they knew other users liked it.

Channels:
Channels are the different voices by which to circulate false posts on the Internet The investigation of the"World" on the misleading articles circulating on the social network shows that their diffusion is amplified by unscrupulous pages. Behind the flow of false information often dries a well-oiled mechanics. The data collected by Le Mond Water over the last few months show, in fact, that Facebook pages with a high audience are responsible for a good deal of the circulation of false content on the platform. By mishandling the mechanisms that govern the ocial network, they manage to massively diffuse false content that would have remained much more confidential without their intervention. On Twitter, false information circulates better than accurate information Three researchers analyzed the fate of 126,000 children relaying information. They established that the real information had a much more confidential distribution than the fake ones, because of the interest of the users for the novelty

Models:
A new simple model describing how information spreads on social networks indicates

that any content can become viral very quickly. In the virtual world, the risk is that the beauty of a photograph or the persuasiveness of an article facilitates the diffusion of a "meme": the term used by Menczer and his colleagues to characterize a link, a video, a phrase or any other online information unit. However, researchers show that three factors are enough to explain why communities of social network users are unable to detect the true from the fake among memes, even if individuals can. These three factors are the enormous amount of information available, the limited time and attention that individuals can devote to browsing their newsfeed and choosing information to share, and finally the structure of socialnetworks.

The mathematical models used to study the distribution of false posts on social networks are so-called "multi-agent" models, where autonomous entities: the agents, which represent the users here interact. These models come from a class of older simulations that are used to study the spread of disease in a population. Imagine a graph in which each agent (each person) is represented by a point, or node. This node is connected by an edge to other nodes, representing friends or "followers". If, for example, Alice is "infected" by the flu virus or by false news, she can pass it on to her friends Bob and Clive, to whom she is connected, shaking their hand, or sharing the meme with them, respectively. Bob and Clive can, in turn, contaminate their contacts and so on. Based on this simple structure of the model and expanding it, researchers are trying to determine how far and under what conditions a false post can spread However, "information is not a virus," tempers Kristina Lerman, a computer scientist at the University of Southern California, who did not participate in the development of this new model. While weusually have to deal with one strain of influenza at a time, the number of memes that can infect us is huge. Researchers integrate this abundance into their model by imagining that each person has a screen on which they see the false posts arrive. The model gives a value to the probability that Alice creates or shares a new false pos for example, a video of a dancing cockatoos and does the same forany new possible false posts from all users. As new counterfeits increase the amount of total information in the system,these values,ââmeasure the amount of information each person feels behind their screen.

Another parameter in the model is counting the number of false posts that Alice sees in her newsfeed before choosing one to share with her contacts, instead of creating a new one. This parameter accounts for the attention span of the information Alice has focusedon. Once Alice posts a message, it appears on the screen of Bob, Clive and others, who, in turn, choose to create theirown fake posts or forward one from their newsfeed. The false posts of both modelsfollow globally what is called a power law , which means that the probability that a false post is shared a certain number of times decreases as a reverse power of that number. For example, on Twitter, a meme

is four times less likely to be shared twice than once. "If you look at thedistribution of images on Flickr, articles on Facebook or hashtags on Twitter based on the number of shares, all are governed by laws of power," says Menczer.

3.2 Identification of hoaxes

3.2.1 Prevention schemes

- Most hoaxes are transmitted by a well-known person (friend, colleague, client, etc.). The sender of the message should not be trusted a priori, because if a hoax arrives in your mailbox, it is because he has successfully passed the "natural selection" It will be likely to abuse any surfer hurry or uninformed.

- Most hoaxes are so effective that they are often relayed by a large number of people, even reproduced on a large number of websites. We must not believe information because it comes from many different sources: we must first of all ensure the competence of sources to judge the veracity of information.

- Most of the poignant, revolting or alarming messages circulating spontaneously on the Internet are hoax. We must not forward a similar message because its content "could be true", but take a step back and if necessary validate the information from a trusted source, because the doubt will always benefit rumors and hoaxes.

- Most hoax are well-constructed, well-argued messages, whose credibility seems to be beyond doubt. We must not stop at the presence of truthful elements, but be interested in the doubtful elements or for which no proof is advanced, because the hoax most often mix the true and the false. Even an apparently explicit photograph or video may have been rigged or partially selected to mislead users and accredit false or questionable information.

Classification techniques

The architecture of a rumour classification system can have slight variations depending on the specific use case. Here we define a typical architecture for a rumour classification system, which includes all the components needed for a complete system; however, as we point out in the descriptions below, depending on requirements, some of these components can be omitted. A rumour classification system usually begins with identifying that a piece of information is not confirmed (i.e., rumour detection)

and ends by determining the estimated veracity value of that piece of information (i.e., veracity classification). The entire process from rumour detection to veracity classification is performed through the following four components:

1. **Rumour detection:**
   In the first instance, a rumour classification system has to identify whether a piece of information constitutes a rumour. A typical input to a rumour detection component can be a stream of social media posts, whereupon a binary classifier has to determine if each post is deemed a rumour or a non-rumour. The output of this component is the stream of posts, where each post is labelled as rumour or non-rumour. This component is useful for identifying emerging rumours; however, it is not necessary when one needs to deal with rumours that are known a priori.

2. **Rumour tracking:**
   Once a rumour is identified, either because it is known a priori or because it is identified by the rumour detection component, the rumour tracking component collects and filters posts discussing the rumour. Having a rumour as input, which can be a post or a sentence describing it, or a set of keywords, this component monitors social media to find posts discussing the rumour, while eliminating irrelevant posts. The output of this component is a collection of posts discussing the rumour.

3. **Stance classification:**
   While the rumour tracking component retrieves posts related to a rumour, the stance classification component determines how each post is orienting to the rumour´s veracity. Having a set of posts associated with the same rumour as input, it outputs a label for each of those posts, where the labels are chosen from a generally predefined set of types of stances. This component can be useful to facilitate the task of the subsequent component dealing with veracity classification. However, it can be omitted where the stance of the public is not considered useful, e.g., cases solely relying on input from experts or validation from authoritative sources.

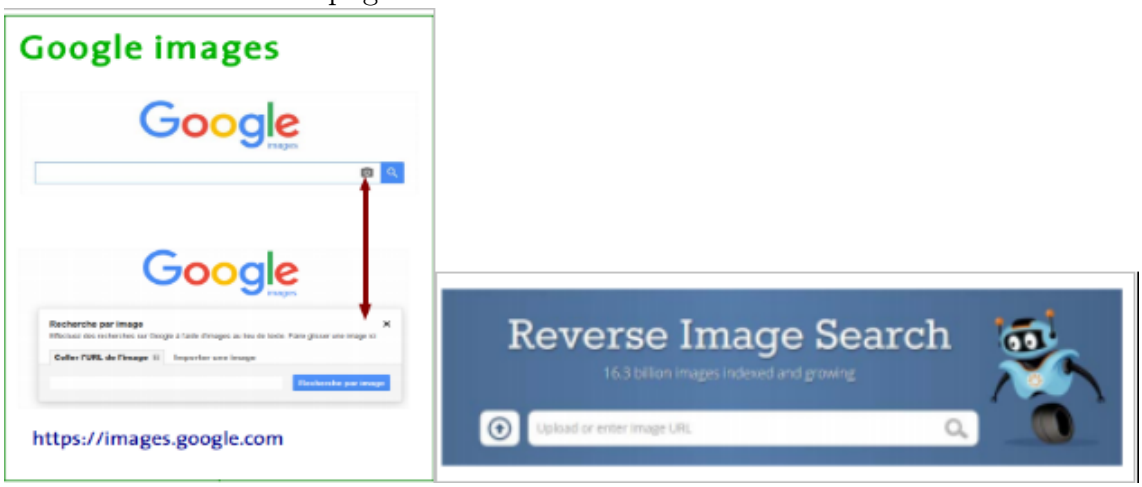
4. **Veracity classification:**
   The final, veracity classification component attempts to determine the actual truth value of the rumour. It can use as input the set of posts collected in the rumour tracking component, as well as the stance labels produced in the stance classification component. It can optionally try to collect additional data from

other sources such as news media, or other websites and databases.

3.3 Approaches based reliable web sources and images retrieval

Hoax identification approaches based on images and urls use the data sources in which images and urls related to hoax already exist; for example *https://www.tineye.com* allows you to find all the sites and dates where an image has been published as well as more complete versions if it comes from a crop. Easy to use, just enter the address of the image to check or simply download it, to start a search. It also offers a comparison tool to easily match the images found and the downloaded one.

du 2018-04-19 22-53-04.png



3.4 Detection schemes

There are several methods to detect Hoax. However according to the exigencies that one wants (Spreader, follower, content), one can categorize these methods in three big groups namely: Spreaders, follower and content.

- A follower is a person who has made the choice to subscribe to the account of another for the purpose of being notified about the activities of the latter as well, if a person is follower of another person then she is exposed to the risk of facing a hoax, when a hoax is published it is necessary to identify the persons exposed finally to evaluate the impact of this false position. The process of detecting the followers of a hoax is reluctant on the technique of "rumour tracking".

- A spreader can be seen as an active follower because it represents a relay in the spread of the fake post. To detect it you must first identify it as a follower then

evaluate if it relay this false post in turn. So use the techinique of the rumour tracking then that of the "stance classification".

- The detection of a false post based on the content is to compare a fake post with fake post already existing in databases. There are several sites that constitute a fake post database through which users can go to check if they are dealing with a false post; as an example we have: **hoaxkiller** and **haoxbuster**.

Thus in order to automate this process several techniques have been implemented. The figure below illustrates comparative elements of the existing techniques..

du 2018-04-19 23-07-56.png

Previous Work	**2002 Hoax Detector [3]**	**Croatian CERT (2004) [4]**	**2009 Automatic Hoax Detection System [2]**
Hoax Detection	Yes. Automated but not really intelligent. Expand to anti-virus software	Yes. Automated and rather intelligent (embed with a database)	Yes. Both automated and intelligent system integrated together
Target	More to Organizational level	Personal Based	Personal Based
Detection	Server based	Server Based	Server based
Form	Text	Text	Text

Moreover, it should be noted that the methods presented in this table are limited to the data source of the hoax only to be able to identify them whereas the method which is based on the technique of Levenshtein Distance is based on the data to identify the false post but in addition it allows if no elements exists in its base for the post to evaluate to define if this one can potentially represent a danger.

Conclusion

Arrived at the end of our work which related to the study and analyzes hoax fakes news or of the hoax,to measure their impact on our companies Il out of spring of our analysis that the hoax can touch a great number of Net surfers, who are especially hybrid of hoax and rumour, they draw their faculty to mislead the Net surfer by causing at his place sharp a emotion(por, compassion, revolt, hope, etc.) moreover they have a capacity to be propagated spontaneously within the community.This is why the hoax term remains most suitable to indicate the "cyber-rumours" or "hoax of the Web". However, Bruno Roy-Contancin, Guillaume Brossard and Pierre Roy-Contancin founded in 2000 HoaxBuster which is a Web site created with an aim of limiting the propagation of the data-processing hoaxes (hoax) and the nonfounded rumours circulating on Internet.Since its creation, its activity did not cease increasing, in particular with the emergence of the social networks, other sites then followed the concept, such as Hoaxkiller founded in 2001, Snopes is an anglophone Web site created with an aim of limiting the propagation of the data-processing hoaxes (hoax) and the unfounded rumours circulating on Internet It receives 300 000 visits per day Ces sites give the means of prevention, the means of fights and also sensitizes on the dangers to which represents the hoax.Nevertheless, in our company the hoax remains still a true threat and its impact influences on the community and according to our studies the means set up to fight against the i.e. (HoaxBuster, HoaxKiller, Snopes etc....) are ignored not visit by more by Net surfers in our company.However it will be judieux to find means more adequate compatible has our company to sensitize the Net surfers in connection with this plagues because the majority do not even know what it is the word "hoax". In Revenge, created a data-processing cell in partnership with the operators telephone (MTN, ORANGE, NEXTEL and CAMTEL) who will be able to sensitize through sms organize forums and to also make discover these sites (HoaxBuster, HoaxKiller, Snopes etc....) in order to reduce the rate of the hoax on our social networks.

Bibliography

[1] **Arkaitz Zubiaga.** *detection and resolution of rumours in social media : A survey.*

[2] **Pritam Gundecha Huan Liu .** *Mining Social Media : A Brief Introduction.*

[3] **Greoffrey Barbier , Zhuo Feng , Pritam Gundecha and Huan Liu .** *The Provenance Data in Social Media.*

[4] **Adrien Guille , and Al .** *Information in Online Social Networks A Survey.*

[5] **Victorial L. Rubin , and Al .** *Deception Detection for News:Three Types of Fakes .*