Towards a typology of social news apps from a Crowd Computing perspective

Daniel Schneider1, Jano de Souza1, Edson Mello Lucas1,2
1PESC/COPPE/UFRRJ, Graduate School of Engineering, Federal University of Rio de Janeiro, Brazil
2DEMAT/ICE/UFRRJ, School of Computer Science, Federal Rural University of Rio de Janeiro, Brazil
3IPRJ/UERJ, Polytechnic Institute, State University of Rio de Janeiro, Brazil
{schneider, jano, edmlucas}@cos.ufrj.br

Abstract— The Web is now both an information repository and a cyberspace supported by a participatory culture. In the domain of news and journalism, the Social Web has witnessed the emergence of Social News Websites (SNW), where users contribute for various reasons. Sites like Digg, Reddit and Storify, just to name a few, allow their users to discuss, comment, share, recirculate, tag and rate news from various sources. However, the social computing literature has described these sites so far through narrow definitions, e.g., as a synonym for social news aggregators, as specializations of social media or even as specialized forms of online social networks. In this research, we identify a broader landscape of social applications exploring citizen participation in the news value chain, and propose a broader definition and a typology for social news applications from the perspective of Crowd Computing.

Keywords—social news; typology; social aggregation; social curation; social production; piggyback systems; social push, backchannels.

I. INTRODUCTION

Since the early 2000s, when blogs appeared and emerged on the Web, social media has been gaining more impact on people's lives. Blogs were just one of the early pillars of a set of social media platforms that allowed people around the world, sometimes referred to as cyberspace, from connecting, sharing and discussing anything: personal experiences, photos, videos, events and thoughts. Social media quickly became a channel for the dissemination of news [1]. Technological advances in portable devices and WiFi networks facilitated this process even more, since people no longer need to sit at a computer desk to read news: they can access them anytime and anywhere through their smartphones and tablets.

According to Domingo and colleagues [2], academic publishing on social media has emphasized the shift in the role of the people "only" commenting and sharing news they read to a more prominent role of reporting instantly what happens around them - as local events that affect their lives, crimes and disasters. Leskovec et al. [3] complement arguing that this phenomenon sparked a symbiotic relationship between social media and news: much of what is discussed in social media is inspired by the news, and vice versa, the news instantly benefit from the content published on various topics in social media environments.

Ordinary citizens are gradually increasing their participation in the journalistic value chain by commenting on news stories through social media sites, posting reports or witnesses of events, checking information published by news agencies, or posting, sharing, commenting, tagging and evaluating news on social media sites. The concept of a production chain that does not end in the online journal has been extensively theorized in media studies through the paradigm of active audience, which emphasizes the role of audiences as producers of meaning - as they read, process, and discuss media texts in various ways, often rearranging them during this process [4,6].

Thus, the Web is now both an information repository and a social space supported by a participatory culture. In the domain of news and journalism, the Social Web has witnessed the emergence of social news websites (SNW), where users contribute for various reasons. Sites like Digg, Reddit and Storify, just to name a few, allow their users to discuss, comment, share, recirculate, tag and review news from various sources, practices that are known as meta-journalism in the new media and society literature. Moreover, the spread of ‘Web Journalism 2.0’ including blogs and other sites of citizen journalism has contributed to the sharp increase in the number of news broadcasted on the Internet. According to Monteiro et al. [7], readers run into some problems: "Where to read the best news? Where to read the news that matters the most?" Then followed a demand for recommender systems, or filtering news [7].

SNWs facilitate a participatory culture, eliminating the gatekeeping function exerted by news agencies and professional institutions, and allowing the public to decide what counts as "news". Social news sites also support democratic participation by allowing users to access the same information across geographic boundaries, and to respond to viewpoints and beliefs of other users, creating a virtual sphere in which anyone can contribute [24]. However, some theorists fear the changes in the pattern of news ranking and consumption, and see personalization as a risk to society, since the receipt of custom content tends to lead the reader to the loss of global vision of the events that happen in the world [35]. Others see it as a necessity to survive in the face of the avalanche of information that the Internet has allowed, and hold that the only way to reduce search costs is to allow the
reader to create custom electronic editions, containing news of interest.

Despite the complex network of participatory practices related to news production and consumption, the social computing literature has described social news sites so far through narrow definitions, e.g., as a synonym for social news aggregation, as specializations of social media or even as specialized forms of online social networks. In this research, we identify a broader landscape of social applications exploring citizen participation in the news value chain, and propose a broader definition for SNWs from the perspective of Crowd Computing. The main goal of this paper is to advance the understanding of the ‘social news’ concept by building a typology to differentiate among different social news apps. We believe this is the groundwork for a deeper understanding of this novel form of participatory culture. The proposed typology should also provide researchers with a useful framework to effectively understand and design novel forms of social news applications.

The rest of the paper is organized as follows. Section II provides a background for the concept of crowd computing in the context of news; in section III we advance in understanding the concept of social news and also present a definition for social news apps; section IV presents the typology and discusses the challenges and research issues associated with each of the proposed categories; and section V presents the conclusions of this research work.

II. CROWD COMPUTING IN THE NEWS DOMAIN

While there is no consensus definition for Crowd Computing, our research group makes use of a definition presented in similar way as in [8] as an overarching term which defines “the myriad of tools that enable idea sharing, non-hierarchical decision making and the full utilization of the world’s massive cognitive surplus”. Because it is a broad concept, it encompasses other terms of Computation presented in Figure 1 as sets of a Venn diagram whose intersections were defined based mainly on the discussions in [8][9][10][45].

![Diagram of Crowd Computing in the News Domain](Image)

Fig. 1. Crowd Computing in the news domain

It is worth mentioning that the literature describes similar approaches, perhaps more narrow than the Crowd Computing, like the concepts of 'crowd-powered systems' and 'social machines'. For instance, crowd-powered systems combine machine and crowd intelligence, and open up a broad new class of software systems that solve problems neither approach could solve alone [11]. Bernstein [11] argues that while crowds are increasingly adept at straightforward parallel tasks, they struggle with complex work because participants vary in quality, well-intentioned contributions can introduce errors, and future participants amplify and propagate those errors. The solution is to design techniques that decompose complex tasks into simpler and verifiable steps [11]. Social machines, on the other hand, are systems in which human and machine components make complementary contributions with respect to the performance of some larger joint process [12]. In investigating social news sites from the perspective of social machines, a key goal is to identify the roles played by human and computational components in the process supported by the socio-technical system infrastructure.

We exemplified the elements of each of the sets of Crowd Computing applications with sites or applications focused on news and on the journalistic value chain. Below we will briefly discuss each of the sets depicted in Figure 1.

A. Crowdsourcing

Here we work with a broader definition of crowdsourcing (CS), provided in [43], which defines the term as “tapping the perceptual, cognitive or enactive abilities of many people to achieve a well-defined result such as solving a problem, classifying a data set, or producing a decision”.

In the context of journalism, ordinary citizens have been encouraged to act as photographers, reporters and data analyzers. Pavlik [13] reported that on May 30, 2013, the Chicago Sun-Times fired all twenty-eight of its photographers, and with them the knowledge and experience they had in covering the city: “with revenues in steep decline and a changing media landscape, editors opted to go with images from freelancers, citizen journalists, and almost ubiquitous security cameras” [13]. As sustained by the author, this story is a clear demonstration that crowdsourcing may work for news agencies and organizations.

In a survey of CS systems in the World Wide Web, Doan and colleagues [14] describe two kinds of crowdsourcing: the explicit and implicit. Explicit systems let users evaluate, share, network, build artifacts, and execute tasks. In the domain of news, these systems include Wikinews, Wikipedia and social aggregation news sites. Implicit systems let users collaborate implicitly to solve a problem of the system owners, and fall into two groups: standalone and piggyback. Piggyback crowdsourcing in the context of news will be discussed in section IV.

B. Human Computation

In Human Computation (HC), the computer uses humans to perform tasks, inverting the roles of the traditional model in which humans delegate tasks to the computer [10]. HC systems use the power of human processing to solve problems
that computers still cannot solve [9] [47]. HC systems in the context of news are available on the Internet, and among them we can mention the Mechanical Turk (mturk) that uses turkers judgment to analyze and filter large data sets for the purpose of creating or uncovering a news story. This approach to journalism, the so-called data-driven journalism, builds on older practices, most notably on computer-assisted reporting. Data-driven journalism is based on the increasing availability of open data that are freely available online and analyzed with open source tools. This approach helps consumers, managers and politicians to understand the patterns and make decisions based on the findings. As such, data driven journalism might help to put journalists into a relevant role for society in a new way [15].

As an example of this approach, mturk became an integral part of ProPublica's newsroom operations, a "non-profit newsroom that produces investigative journalism in the public interest" [16]. For one of their projects, the company was facing a crucial data integration problem: they wanted to extract data from hundreds of different city, country, and state databases, and building an integration system of such scale was difficult and beyond its reach. The turkers created an effective human-powered web crawler that was up and running in a couple of days [16].

And last but not least, social news aggregation websites also fit in this category by allowing citizens to recommend news to other readers. In these sites, citizens are also involved in agenda setting and decision making related to issues in the news value chain.

C. Audience-Computer Interaction

The Audience-Computer Interaction is an area of Human-Computer Interaction (HCI) which involves the use of Crowd Computing technologies to obtain useful feedback from the crowds (virtual or real crowds) in the context of audience [17]. This technology increases the possibilities of interaction of large audiences with speakers and performers [18]. It should be noted that the role of the audiences as producers of meaning, in the context of the news production chain, has been extensively theorized in media studies through the paradigm of active audience.

In the context of news, these technologies may enable journalists to have greater interaction with the readers, making it easier to get immediate feedback from the audience. Audience backchannels, for instance, allow users to create interaction channels with defined audiences on Twitter, Facebook and other services, as will be discussed in section IV. Examples of such tools include sites such as WhoTalking and WhereTweeting.

D. Social Computing

Social Computing is an area of Computer Science concerned with the intersection of social behavior and computational systems [19], including (in addition to Computer Science) fields such as Sociology, Education, Communication, Psychology, among others. The core is to maintain (create and recreate) links and social contexts with the support of software and technology, so that people have an active online social life.

Youtube is an example of a popular social networking service that fits this group and focuses on sharing videos. In the context of news and journalism, blogs, online social networks (like Facebook, Twitter and Instagram) and also citizen journalism sites are examples of current pillars of a set of social platforms that allows citizens to connect, share and discuss news stories. Finally, social news aggregation sites (such as Digg and Reddit) and wiki-based platforms (like Wikinews) also fit the Social Computing systems class.

III. ADVANCING THE UNDERSTANDING OF SOCIAL NEWS

To the best of our knowledge, Lerman [20] was the first to present the concept of social news as a specialization of the concept of social media - which refers to media produced by ordinary users in applications whose content is primarily written by the users, such as blogs, wikis, Flickr, among others. Virasoro and colleagues [21] define a social news website as a synonym for social news aggregation website where "the visibility of an item is determined democratically by the preferences of its members, usually by voting" [21]. On the other hand, Tang and co-authors describe social news sites as specialized forms of online social networks and argue that, by incorporating a variety of social facilities, these sites allow users to publish, discover and promote the most interesting content without requiring a group of editors of the site itself [22]. Other authors - as Leavitt & Clark [23] - also work with a close definition to the preceding, presenting the social news sites (citing Digg and Reddit) as environments that allow users to submit hyperlinks for a 'feed' of content that other users can vote positively or negatively. "As the stories gain more positive votes, the most popular ones eventually reach the top of the site (main page) based on the total score of the news relating to positive and negative votes" [23].

As evident in the aforementioned definitions, social news is associated with the concept of social media - particularly with social networks and social aggregation. Social media sites share four characteristics [20]: users create or contribute content of multiple media types; users annotate content with tags; users evaluate content actively (through voting) or passively (by consuming content); and users create social networks [46] by designating other users with similar interests as contacts or friends. In the process of using these sites, users add rich metadata - in the form of social networks, notes and ratings - that improve the cooperative problem solving via the so-called social information processing [20].

First of all, we must acknowledge that news stories are socially constructed artifacts, and this clearly predates the era of Social and Ubiquitous Computing. That said, it is possible to adopt a narrower definition for "social news" as the broader the scope of the definition, the more nebulous becomes the term. But here we advocate a broader approach which widely includes any application where news stories are socially constructed, promoted and consumed with the open participation of citizens. The reason is straightforward: the production of news involves a complex and multifaceted chain of communication and sense-making where events, issues and
ideas are subject to the influence of various filters or 'guardians' (sources, journalists, editors) before reaching their target audience [5]. What social news platforms yield are new possibilities for citizen participation (often in the role of cybernauts) at various points along these chains of sense-making that shape the news, even in the reader and 'prosumer' roles.

In view of this discussion, we suggest that social news apps can be defined as socio-technical tools that allow the full utilization of the crowd's ability (idea generation, agenda-setting, framing etc) to collaborate (actively or passively) in some point along the news value chain.

IV. THE TYPOLOGY

The academic community has been analyzing the environment of social news sites from various perspectives such as social dynamics [25], participation [26], user generated content [27], crowdsourcing [14], social aggregation [20], design of incentives in social computing [28], uses and gratifications [29] and, more recently, social machines [12] and social curation [30].

Figure 2 depicts the typology of social news applications proposed in this work. We began this study with the following question regarding the investigated apps and platforms: How this application might exploit the contribution of the crowd on the news value chain? Our typology groups current applications in the context of crowd computing in the news value chain and classifies them according to the main goal of using the crowd work and intelligence, which may be: social news production (storytelling); social news aggregation (agenda-setting and news recommendation); social curation (preserving news stories); social push (personalization); social news backchannel (audience-centered crowdsourcing); and piggyback news apps (crowdsourcing for news value and attention). It should be stressed that the proposed typology covers most existing social news apps and does not prevent the existence of new groups later.

![Fig. 2. A typology for social news applications](image)

A. Social News Production

Social news production sites involve large groups of people (citizen journalists) who lend their talents to the creation and storytelling of news articles. Currently the main examples found in this category involve the blogs, wiki-based environments - such as Wikinews – and, in general, the so-called open collaboration environments related to news production.

Fortes & Lampe [44] define an open collaboration system as an online environment that supports the collective production of an artifact through a technologically mediated collaboration platform that has a low barrier of entry and exit and supports the emergence of persistent social structures. In the context of the news value chain, as the news also started to circulate on these environments, communities focused on comments on news began to play an important role in creating and maintaining repositories of community interest and opinion [44]. Wikipedia itself has been studied as an environment of open collaboration for news storytelling, and recent studies have investigated the nature of high-tempo collaborations over news published on Wikipedia [31]. Recent literature also reveals the role of news APIs in offering a technology infrastructure to support the public in the task of redefining, reorganizing and giving new meaning to content and data from news organizations [40].

However, structuring the crowd collaboration in news storytelling is challenging. Too little structure leads to sprawling narratives, and too much structure stifles creativity [41]. We believe much of the socio-technical challenges for the development of applications in this category are common to the general challenges of open collaboration. Key questions around who are the people who join these communities, why they join and contribute to these projects, and how they work remain largely open issues.

B. Online Social Network

From the point of view of Crowd Computing, online social networks involve crowds that trade information through a shared communication system such as Facebook or Twitter. Users of social networking sites continuously produce a stream of shared status updates, photos, links (and contributions for the storytelling of news stories) and receive a united stream from all their friends or followees [32]. However, today a lot of active social media users complain that their streams have become too overloaded and hard to extract useful information from [32]. Grineva & Grinev [32] describe three main reasons for this overload: the general increase of usage and amounts of data shared every day on social networks; the growth of the number of connections in the social graph; and automatic updates coming from applications.

Social network users tend to read and get immersed on irrelevant information than what is necessary. Most people disregard social media content and talk only to a thin circle of friends [32]. There are also people who receive newsfeeds with thousands of pieces of content every day. These are some of the problems faced by users of these applications in the role of news consumers. And last but not least, Grineva & Grinev [32] identified two groups of approaches to solve these problems: the first is based on filtering social media streams (quite mature and successfully used), and the second group of approaches proposes different paradigms for information sharing and consumption rather than stream, such as passive information consumption via the push model [32].
C. Social News Aggregator

According to Tang and colleagues [22], social media aggregation sites represent specialized forms of online social networks (OSNs) - and begin to change the way people seek and consume information on the Internet. By incorporating a variety of social features, these websites allow their users to discover, publish and promote the most relevant content without the need for a group of editors to perform this task. Other authors define a news aggregator as a site that adds a social structure made up of individuals connected by friendship or similarity of values (social networks) and a recommendation system. So, in the proposed typology, we consider that this category inherits characteristics of the online social networks application class.

Most social news aggregators allow users to submit content, in some way. However, each site is different in how content is moderated. On Slashdot and Fark, for example, the site administrators decide which news will be selected to the first page. In the case of Reddit and Digg, the news that get the most votes from the community are selected to the first page. Most of these sites have a commenting system, where users can create a discussion on each story. Some of them have also applied their voting system to the comments, so that the most popular comments are displayed with priority (on top). Users can contribute in different ways, depending on the site, for instance: commenting, gathering stuff (photos, videos, audio etc), submitting stuff (photos, videos etc), submitting content (stories), associating news to broader stories, reading news (value that cannot be neglected!), and so on.

Much of the social news aggregators investigated in this work present in its structure a first page containing a news agenda in accordance with the user's interests and profile (entered by the user), or containing the news that were published, rated or read by the declared 'friends' or contacts. This second feature is known as social filtering in the literature that investigates the social media aggregation sites.

A key challenge for designing social aggregation sites is evaluating how 'democratic' they may be, or at the opposite extreme, inclined to an elite group of contributors. In Digg, a popular social news aggregator, Goode [5] holds that there exists a two-tier agenda-setting process in which a core of users do the initial 'digging' required to bring stories to the surface, whilst other users rate stories already having high visibility (the so-called 'snowball' effect), and casual visitors merely consume news. Another common problem reported in the literature concerns the secretive nature of the algorithms used in these sites to compute the score of a news story [5].

D. Social News Curation

In recent years, a number of content curation platforms have emerged, some of which turn specifically for journalistic curation. Social news curation sites support users in collecting and sharing news found on the Web. Some websites market themselves as spaces for exploring a common interest through different types of media related, while others are promoted as a means to create and share stories, or even produce custom newspapers [30]. These environments benefit individually the user responsible for curating by organizing the news that matter to him, but also benefit the community as a whole that has an interest in that topic. Users can select, collect, annotate, crop, organize and present content from multiple media types. A single piece of content may turn out to be used in many different contexts from that for which it was originally planned.

Guerrini [33] depicts four main technologies used to support news coverage based on curation: Liveblogs, Storify, Storyful and ScribbleLive. Liveblogs are convenient to narrate an event as it is unfolding, and is the default format for the online coverage of major breaking news, sports and news events scheduled entertainment [33]. They do not follow the traditional format of reporting of the inverted pyramid, being usually presented in reverse chronological order with the most recent update at the top. Narratives are shaped by a single post, composed of short micro-updates that may consist of text, embedded images, videos, links or other items (such as tweets), and constantly updated by one or more authors. They combine conventional reporting with curation, where journalists sift and prioritize information from secondary sources and present them to the public almost in real time, often incorporating their own comments.

The Storify platform is defined as a content management system focused on curating stories, where users can enter text between various embedded images, videos or links, adding some explanation to give context. It is considered one of the most reputable curated tools, thanks to its flexibility, ease of use, and the option to incorporate the story itself curated on external sites or blogs [33]. The platform is often used as a liveblog, although this is not its original mission.

While Storify is more about collecting stories, ScribbleLive is more focused on storytelling in real time [33]. The technology allows journalists to cover any event in real time, building narratives from building blocks that are available on the Web. Each block can then be commented on or shared via Twitter or Facebook. Posts can be updated either through the Web interface or by applications developed for the mobile platform (iPhones or Androids) as well as SMS and voice mail. All posts can use the actual content of the institution or user-generated content. Reuters, CBS, CNN and other news organizations are clients of ScribbleLive.

The Storyful platform is defined as an intermediary between the newsroom and the producers of user-generated content [33]. Therefore, unlike Storify, Storyful's environment was not designed to facilitate the process of telling stories through the audience. Registered users have access to an online control panel that consists of curated lists of Twitter and videos on current breaking news. Verification of user-generated content is a crucial aspect of the Storyful environment. Although the news discovery process is largely automated, a team of experienced journalists conducts various checks and investigations to ensure (with the greatest possible accuracy) that the content found in social media is genuine and not a forgery prepared by someone trying to promote their own agenda [33].

It should be stressed that one could also include Wikinews (and Wikipedia) in this category of the typology, arguing that
wikipedians do nothing more than preserve information, particularly with regard to news articles.

E. Social Push System

Social push news applications allow their users to customize their 'feeds' of news by selecting among hundreds of sources suggested by the company behind the application. At the same time, they allow users to monitor trends of interest (professional, for example) and to connect with others by sharing, rating and commenting on news of interest.

Customization, in the field of journalism, implies the indication of items of interest by the user, such as desired color or font of an interface, or editorials of which he wishes to receive news by email [35]. Deuze [36] classifies the forms of customization in online journalism as pull, push and hybrid. In pull customization, the user "pulls" content and services he wants - accessing links provided by the website, by news editors and other available services. In push customization, the newspaper automatically "pushes" news to the user according to his profile, after he indicates the editorials he would like to receive information from. And in hybrid customization, the user can create his own page, previously selecting their topics and news services.

Social push news applications have begun to emerge in recent years, especially on mobile platforms like Android and iPhone (iOS). Recently the company LinkedIn bought one of the core social push applications on the market, the Pulse.

F. Social News Backchannel

The idea of backchannels was born in linguistics, and the term was first used to describe responses of listeners in the form of meta-communication during conversations [37]. According to Harris [38], communication technology supported by public backchannels has been of great interest to researchers, teachers and professionals. Harry, Green & Donath [39], for example, developed the backchan.nl system which allows very detailed feedback during a presentation of lectures. Voting and moderation were incorporated into the system, and the main comments voted by users were presented to the general public.

With the fever of communication tools associated with the Social Web, including network-based communication technologies such as Twitter and Facebook, some sites are designed to function as backchannels of such services (e.g. backchannel.us and WhoTalking). They allow users to create interaction channels with defined audiences on Twitter, Facebook and other services - typically using keywords - and interact with them. The backchannels represent a form of audience-centered crowdsourcing according to the CS framework proposed by Erickson [42].

G. Location-based Backchannel

Location-based backchannels are similar to common backchannels, but the service allows users to search what people are talking about in a geographic radius from a defined point on the map. An example of a site that offers this service is WhereTweeting.

H. Piggyback News System

Doan et al. [14] argue that one way of implementing implicit crowdsourcing systems is making the 'piggy back' of other established systems, by exploiting traces that users leave in this system to solve a target problem, and this provides a constant flow of the users activity. But, as these authors emphasize, it is necessary to solve the difficult challenge of determining how the traces can be exploited for the intended purposes [14].

Many news reporting or social media websites automatically extract of their own posts or comments, published on news sites, the most important issues or topics (the so-called 'trending topics'). Google News, for example, uses the log of searches done on his website and analysis based on the clicks curve (the links of the news) to assemble his own list of main topics. By exploiting the log in his own search site, Google is making use of the piggyback form of implicit crowdsourcing. Twitter, on the other hand, mines their 'hot trends' from tweets, benefiting from sharing content (done by re-tweets) and publishes the results as hashtags.

V. Conclusion

This study was an initial research effort to understand and reflect on the social news phenomenon. We have discussed a broader landscape of social applications exploring citizen participation in the news value chain, and proposed a definition and typology for social news applications from the perspective of Crowd Computing. Our discussion shows that crowd computing technologies are currently applied to a wide variety of problems in the context of the news value chain, and that it raises some interesting socio-technical challenges. We expect that this emerging field will grow rapidly. In the near future, we expect new techniques will be developed to engage a broader range of citizens in crowd computing, and to enable them to make increasingly complex contributions in the news value chain.

In the context of the Social Web and Social Machines, one possible future direction would be towards a culture of participation involving citizens throughout the whole journalistic value chain. If on one hand the Social Web has produced a large collaborative encyclopedia, on the other, has not been able to produce a collaborative newspaper in emerging form, designed and framed by the crowd, considering various perspectives while catering to audiences with different profiles and interests.

REFERENCES


Smart, P. R., & Shadbolt, N. R. (2014). Social Machines.


