

The Performance of Distributed News Aggregators

Lead Researcher: Wolf Richter

Collaborative Partner: Tobias Escher, David Bray

Reviewer: Bill Dutton

Introduction

The production of news has always relied on employing cutting edge communication infrastructure to provide the most up to date information to the readers or later viewers as quickly as possible. But recent advances in digital communication technology have not only enabled a significant improvement in the internal performance of established news organizations, but have also questioned the structure of the industry as such. Traditionally, the news industry has been dominated by the duality of the news agencies, which own the input channels through which information is gathered, and the publishers and networks, which own the distribution channels and filter relevant information from the continuous flow of data provided by the agencies.¹ The separation between these two groups has never been a strict one: the larger publishers and networks employ significant staff to gather news items themselves and subsequently follow up on the main stories provided by the agencies to bring their own perspective to their readers.

With the widespread availability of news-gathering devices, like digital cameras, laptops, and mobile phones, and public access to the global distribution infrastructure of the Internet for most citizens in developed countries, the exclusive power of the agencies and publishers to collect and disseminate “news” has received a significant blow. Mobile phone users can take digital pictures with the built-in cameras and distribute the information together with a short story on a blog or other publicly accessible news forum, which turns the mobile user into a Citizen Journalist.² The immediate gathering and the direct dissemination of information over the Internet enables a shift in the news production process away from industrial publishing companies to the users, readers, and consumers empowered by shared digital information spaces. This shift comes in two dimensions: (1) An Internet user has the possibility to gather and publish news items over the Internet and (2) the user can search and aggregate her own news from the abundance of freely available content on the Internet. While the first dimension turns every user into a Citizen Journalist, the second empowers Internet users to become their own news editors.

The first wave of online news media has focused on this second dimension, enabled by technologies like customizable homepages or subscription to special interest email lists. This mode of news production by customized filtering is also called the “Daily Me”, because it allows every user to create her newspaper based on the topics that are subjectively interesting.³ With the increasing popularity of blogs, wikis, and other shared information spaces, and the spread of camera phones, users increasingly also take over the first dimension of news gathering, resulting in the “Daily We”, in which users produce news for

¹ Shayne Bowman and Chris Willis: We Media - How audiences are shaping the future of news and information, The media center at the American Press Institute, Thinking Paper, 2003, p.10.

² See Dan Gillmor: We the Media: Grassroots Journalism by the People, for the People, 2004, Ch. 2.

³ See Bowman and Willis: *supra* fn. 1, p. 7.

other users without relying on the established news gathering and selection process and the underlying infrastructure.⁴

The rise of the Internet as a means for news production raises two sets of questions: First, what different modes of production and aggregation have emerged in the online news space and how does their performance compare to each other? Second, do the new news production paradigms threaten to replace or promise to complement the traditional mode of news production? Or put differently: Is the question to hire a professional news team or to rely on an online news aggregator a “make or buy” decision for a business, or is it a paradigmatic decision with respect to the outcome, i.e. the kind of news produced?

Several studies have already addressed the impact of the new possibilities of news production and consumption and the resulting shift in the balance of power in the industry. These studies have focused on comparing the priorities or the agenda set by traditional, i.e. non-distributed news publishers with those set by the so-called “blogosphere”, i.e. the direct creation and dissemination of information on the Internet, and found that both spheres follow different agendas.⁵ But the space for Internet news production is not dominated by one homogeneous production paradigm. On the contrary, a variety of different production paradigms have emerged within the space of online news production, with the Web presence of a traditional news production organization like the BBC on one end of the spectrum and with the decentralized blogosphere, where a majority of amateurs share their thoughts and opinions, on the other end. So far, little attention has been paid to systematically account the different paradigms of online news production between these two extremes and comparatively study the specific performance of each of the distinctively different paradigms of Internet news production.

Therefore, in this paper we will:

- Describe the problem solved by Internet news production
- Derive a categorization for the production paradigms found in the Internet news space
- Present one particular representative for each category and discuss its strength and weaknesses, using “embedded case studies” where appropriate
- Introduce a method to retrieve Internet news headlines from different platforms and quantitatively analyze their performance, respectively
- Show first results from a one-week experiment tracing the news agendas set by the different production paradigms
- Discuss the findings from the case studies and the quantitative analysis to understand what advantages and disadvantages distributed news production has over the traditional models

The business of news production

This paper will study the performance of a specific information market: news production. News is a relative good, because what information is “news” is determined by subjective and geo-temporal factors. What factors determine news-worthiness has been of long-term interest in media and communication studies.⁶ A number of factors have been identified such

⁴ See *Id.*

⁵ See “Web users diverge from traditional news agenda” in *Financial Times*, Sep 12, 2007; see PEJ project: The latest headline, your vote counts, available at: <http://www.journalism.org/node/7493>; see Tobias Escher: Analysis of the news agenda of bloggers vs. the traditional media available at: http://people.oii.ox.ac.uk/escher/wp-content/uploads/2007/09/Escher_Blog_Agenda_Setting.pdf

⁶ See J. Galtung, J. and M. H. Ruge (1965). “The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers.” *Journal of Peace Research* 2(1):

as closeness – in geographical terms as well as culturally or in terms of relevance to the reader – and status – e.g. are the people affected prominent, or is the country politically powerful. Of particular relevance for this research are two characteristics that Galtung and Ruge identified as *frequency*, i.e. sudden events which concur with the organization's times of working are more likely to be covered, for example a particular murder case rather than a long-term trend, and *continuity*, i.e. events that are already in the news have a higher chance of being selected as news again.

The news production market is not supply-constrained. On the contrary, the amount of information gathered and disseminated every day is ever growing and exceeds by far an individual's capacity to process.

This also poses a challenge for a single news organization maintaining situational awareness in today's global and complex environment. Even with globally distributed field reporters, the capacity of a single organization may in fact be near the limit of impossibility. There simply may be too much news being produced globally for a single agency to report on it all. Such problems for conventional news agencies stem from the accelerating trends of globalization, where increasing volume, velocity, volatility, and concerns over the veracity of news produced challenge the traditional approach to news production and reporting.

For the individual reader, news organizations have so far played an important role as gatekeepers to preprocess and select the news items before publication.⁷ This role is ambivalent: On the one hand they help their readers to manage the amount of information they can handle, and on the other, they take important decisions on what information the readers can access on their behalf.

In a digital environment, physical pages or broadcasting minutes are no longer the limiting factor. It is only the bandwidth of the readers or viewers to process, i.e. read or watch the information available on a given day. Online news production therefore also needs to address the challenge of information selection to avoid information overload. When exploring the solution scope we have found a variety of mechanisms to tackle this issue.

Non-distributed alternatives

There are three non-distributed alternatives to the production paradigms covered in this case study:

- Traditional offline media like newspapers
- Broadcasting media like radio or TV
- Online presences of centrally managed offline publications like bbc.co.uk

Of these three only the third alternative falls within the scope of the Internet news production space.

Nature of the problem

News production is a routine operation, which is performed on a daily, if not hourly basis by a large group of individuals – professionals, private persons, and all sorts of actors which fall into the middle grounds between these two categories.

The production process consists by and large of two different problems: The first is gathering or reporting of news items and the second is selecting relevant news items for publication.

64-90 and the more recent review by Harcup, T. and D. O'Neill (2001). "What Is News? Galtung and Ruge revisited." *Journalism Studies* 2: 261-280.

⁷ See e.g. David Manning White: *The "Gate Keeper". A case study in the selection of news* in: Lewis Anthony Dexter and David Manning White (Publisher): *People, society and mass communications* (1964).

News production strikes a delicate balance between an – at least in theory – objective account of the facts and a subjective evaluation of the facts and selection of the stories deemed news worthy.⁸ News publishers have played the role as gatekeepers, which decide on the allocation of bandwidth, traditionally printed pages or broadcast minutes, to certain topics or news items. The subjective bias and the balance between objective and subjective aspects constitute the profile of every news-producing organization. Traditional news publishers have carefully built their reputation in the market place over time and used the profile to assemble a loyal group of customers, which agree with the specific bias of the news producing organization. To maintain a professional standard of reporting and satisfy the expectations of the readers, news producers have traditionally required journalists to possess a certain level of understanding of the topics they cover. A Code of Conduct or Professional Ethics has emerged among professional journalists, which regulates the borders of journalistic activities, for example with regards to printing disturbing or unnecessarily intimate pictures or respecting off the record announcements.⁹

The production process of traditional media production follows “filter, then publish”.¹⁰ The editorial board of the publisher allocates the limited (physical) bandwidth to certain topics, which is filled by the items produced by news agencies and reporters and only those selected are published. On the Internet, a new production paradigm has appeared, which is “publish, then filter”. With almost unlimited physical bandwidth, almost every story and every accompanying piece of information can be stored and potentially published in digital format. The role of online news producers is to assist the users to filter the relevant information from all published information on the Internet, hence “publish, then filter”.

Based on this basic analysis of the news production process, we will propose the following categorization of the various instances of online news production studied in this paper. Within each of the two problems, news reporting and news selection, there are multiple points of control. Therefore it would be overly simplistic to categorize each mechanism as a central or distributed mechanism. We would like to postpone this discussion to the individual presentation of the mechanisms later in the paper.

Table 1: A functional analysis of news production organizations

News reporting	News production process	News selection	Examples
Professional reporters create original content	Filter then publish	Editorial Board	NYTimes.com, IHT.com, bbc.co.uk
Citizen reporters create original content	Filter then publish	Editorial Board	OhMyNews
Automatic algorithm collects content from all kinds of sources	Publish then filter	Algorithm	Google News, Yahoo News
Users submit content from other sources	Publish then filter	Ranking Mechanism	Digg, Reddit, del.icio.us

⁸ See Axel Bruns: Gatewatching, Not Gatekeeping: Collaborative Online News *available at* http://eprints.qut.edu.au/archive/00000189/01/Bruns_Gatewatching.PDF (last visited 22/11/2007).

⁹ See Gilmour: We the media, p. 138; see Bill Kovach and Tom Rosenstiel: The Elements of Journalism, London 2003, p. 12.

¹⁰ See Clay Shirky: The Music Business and the Big Flip *available at* http://www.shirky.com/writings/music_flip.html (last visited 15/11/2007).

Users creates original content or comments on content from other sources	Publish then filter	Social network	The Blogosphere, <i>Global Voices</i>
--	---------------------	----------------	---------------------------------------

Description of the cases

To map the Internet News aggregator space we have decided to choose one representative for each specific mode of production:

Digg News¹¹ is a proprietary platform that is open for unedited content submissions: it catalogs news stories submitted by its users as well as questions and comments, and invites users to rate the stories. The content with the most favorable votes is shown on Digg's front page. Digg's agenda is purely determined by the popular vote of its users – the wisdom of the crowd, and users have to rely on that what others find relevant or interesting, and what will also catch their interest.

Google News¹² is a proprietary platform that automatically collects news stories from news sites from all over the world, sorts and tags them using a proprietary classification technology, which makes the news stories searchable using Google's proprietary search technology. In its presentation Google news resembles most closely the look and feel of the online presence of a traditional news producer.

OhMyNews¹³ is an originally Korean citizen journalism platform that relies on original news contributions written and submitted by its members. Since 2001, OhMyNews also publishes an English language edition and has successfully established itself as the global market leader for citizen journalism. Members are volunteering their time. The articles are reviewed and selected by a professional board of editors that enforces the publicized style guidelines and chooses articles by their subjective importance.

bbc.co.uk news¹⁴ is the online presence of the BBC, which presents headlines and in-depth news coverage created by professional journalists on the Internet. Although bbc.co.uk employs dedicated online editorial staff, which is not integrated into the news production of the other BBC channels, radio and TV, and has adopted elements of citizen journalism like user forums to post comments or pictures taken by users, it is still rooted in the traditional, centrally managed approach to journalism. Therefore we will use bbc.co.uk to represent the traditional alternative to distributed online news production.

The Blogosphere cannot be reduced to one platform, which could represent this paradigm of online news production, which is made up by thousands of individual sources of information, the individual blogs. Blogs can be maintained by individuals or by groups, by amateurs or professionals, and cover all topics the bloggers find worthwhile writing about. Even aggregator sites like Technorati or Slashdot do not reflect the richness of the various opinions and degrees of reporting found in the myriads of private blogs. One approach to comparatively measure the performance of the blogosphere is to extract key words from the web site of a traditional news publisher and use a blog search engine, e.g. Google Blog search,¹⁵ to identify the number of blog entries covering this particular topic. This approach is burdensome and runs the risk of ignoring topics which are brought up by the blogosphere, and which are not covered by any other source. We will instead take a case study approach and closely examine a particular network of blogs called **Global Voices**,¹⁶ which is dedicated to international news reporting and aggregates a large number of disparate blogs

¹¹ <http://digg.com> (last visited 24/11/2007).

¹² <http://news.google.com> (last visited 24/11/2007).

¹³ <http://english.ohmynews.com> (last visited 24/11/2007).

¹⁴ <http://news.bbc.co.uk> (last visited 24/11/2007).

¹⁵ <http://blogsearch.google.com> (last visited 24/11/2007).

¹⁶ <http://www.globalvoicesonline.org> (last visited 22/11/2007).

in all parts of the world and should be representative for the mode of production and the range of topics covered in the blogosphere with regards to international news.

Wikinews relies on the Wiki technology, which also underlies the Wikipedia community, and allows major-scale collaborative production. Although the publication mode and in particular the possibilities of the collaborators to co-produce are significantly different from those of the blogosphere, we will not treat Wikinews as a structurally separate phenomenon, although this may be required for future analysis beyond the scope of this case study.

The Performance of Digg News

A history of Digg News

Digg News was developed by Kevin Rose, Owen Byrne, Ron Gorodetzky, and Jay Adelson in 2004 and was first launched on the 5th of December 2004. In an interview with ZDnews¹⁷ Kevin Rose explained that he never expected the site to take off, and called it a side project of his. But already in March 2005 a critical mass of users had been reached and more and more users joined the Digg community. Allegedly the popularity was triggered by a story about the hack of Paris Hilton's cell phone, through which the cell phone numbers of all of her contacts were made available to the general public. The story was first posted on Digg and then picked up by the major search engines like Google and Yahoo. As more and more people searched for the Paris Hilton story, Digg received a growing number of hits and became known as a platform for news.¹⁸

The estimates about how many users are actively using Digg vary widely. Digg reported 20 million users in October 2006, while conservative estimates from Comscore range at around 1.3 million.¹⁹ Digg was ranked the 76th most popular site on the Internet by Alexa and the 144th on Compete.com in 2006.²⁰

Other than its competitors del.icio.us and Reddit, which have been bought by Yahoo! and Conde Nast, respectively, Digg is still owned by the founders and Venture Capital firms Greylock Partners and Omidyar Network.²¹

A number of news sites, including the online version of the Wall Street Journal,²² have included buttons on their web sites, which allow users to “digg” the articles they have just read with one click. This marketing strategy is not only employed by Digg, but by many other social bookmarking tools like Reddit and del.icio.us, so readers often have a choice of platforms to submit an article directly from the site on which they have found the article. Aggregator services like Addthison.com²³ provide easy tools to web site operators, which allow to post an article to several social networking sites. Addthison currently features: Digg, Reddit, del.icio.us, Facebook, and Stumbleupon. The BBC uses addthison.com to provide

¹⁷ See Richard MacManus: Interview with Digg founder Kevin Rose *available at* <http://blogs.zdnet.com/web2explorer/index.php?p=108> (last visited 20/10/2007).

¹⁸ See id.

¹⁹ See Amit Chowdhry: Greylock Partners and Omidyar Network Plug Another \$8.5 Million into Digg *available at* <http://pulse2.com/category/omidyar-network/> (last visited 15/11/2007).

²⁰ See id.

²¹ See Michael Arrington: No Acquisition For Digg - Raise Series B Round Instead *available at* <http://www.techcrunch.com/2006/12/28/no-acquisition-for-digg-raise-series-b-round-instead/> (last visited 15/11/2007).

²² See Kevin Rose's blog post *available at* <http://blog.digg.com/?p=98> (last visited 14/11/2007).

²³ www.addthison.com (last visited 22/11/2007).

their users with the possibility to bookmark any article from bbc.co.uk on the social bookmarking site of their choice.²⁴

The nature of Digg News

Digg is a social bookmarking tool. Users can submit stories they find on the Internet to Digg. Newly submitted stories appear on Digg's web page as upcoming. Users can review the articles in the upcoming section and either vote in favor of them (digg them) if they think they are interesting, or decide to "bury" them, i.e. to reduce their ranking in the upcoming section. Once a submission has reached a certain amount of votes, it is moved to the popular section, which is also the frontpage of Digg. A quantitative study of this process has shown that because of the constant addition of new articles to the upcoming section and the focus of attention on new articles, the decision if an article will make it to the front page is made within the first couple of hours, after which attention and the likelihood of being digged decreases dramatically.²⁵ The same study also shows a high turnover of news stories on Digg's front page, with an average story remaining on the front page for 79 minutes.

Users have to register before starting to submit and vote for stories. During the registration process, users can choose a nickname and disclose some personal details.

Digg does not have a reputation system to allow users to rate the quality of the contributions of other users. Digg's reputation system focuses on stories and less on users.²⁶ Nevertheless, the influence of an individual user's votes (so called "diggs") seems to increase with every successful submission, i.e. one which is subsequently made popular by the community.²⁷ Every Digg user maintains a profile page, on which basic information about the user and his activity on Digg is displayed. A set of performance indicators provides details about a user's activity and his effectiveness in identifying popular stories. The indicators comprise among others the number of original submissions (first diggs), the number of overall votes (diggs), and the number of comments. The most important indicator of a user's effectiveness is the "popular" ratio, which indicates the amount of a user's original submissions which finally made it to the front page of Digg. Users can declare themselves as fans of a particular digger and receive notifications, whenever their idol contributes something.

The development of Digg over time has been well documented in numerous blogs, most of them maintained by active Digg users. These blogs also contain several stories, which will be used in the following to qualitatively describe the performance of Digg as a distributed problem solving network and outline some of the potential weaknesses of the model.

Critical account of Digg's performance

1. Groupthink produces mob behavior – the O'Reilly journalist incident

One of the aspirations of the "open submission – verification by 1000 eyeballs" approach taken by Digg is to get a more balanced and accurate view of what happened. But there are also concerns about what is called mob behavior or group thinking, i.e. that once a certain threshold of users have verified a story by digging it, others are prone to believe the story and may amplify the perceived reliability of the story by also digging it. If a rumor has been certified by a sufficient number of users, it turns into reality as the following incident involving an editor at O'Reilly demonstrates:

²⁴ See BBC: Social bookmarking links available at <http://news.bbc.co.uk/1/hi/help/6915817.stm> (last visited 24/11/2007).

²⁵ See Fang Wu and Bernardo A. Huberman: Novelty and collective attention in PNAS Vol. 104 No. 45 p. 17599ff (November 2007).

²⁶ See MacManus *supra* fn 17.

²⁷ See discussion on gaming of Digg below at p. TBD.

Steve Mallet was accused in a story submitted to Digg to have stolen some code from Digg's CSS page. The story became popular very quickly and subsequently appeared on the front page of Digg. When investigating the story, it turned out that the editor had indeed taken some code for his homepage, but from an Open Source Project called Pligg, which aims at producing an Open Source version of the voting backends and content management tools behind social bookmarking sites like Digg. A Spanish social bookmarking site called Meneame had contributed significant amounts of their code to the Open Source project, but as it turned out had itself copied code from Digg instead of writing their own code. The O'Reilly editor Steve had taken the code and incorporated it in his own site, trusting that it would be Open Source code. So while the code was indeed identical to Digg's CSS page code, it was hard to maintain that Steve had actually stolen the code from Digg. While the sensationalist news of an O'Reilly editor stealing Digg's code immediately became a popular story, the insight perspective of how the code actually made it from Digg to his home page remained buried in less popular comments and blog posts.²⁸

It would be too fast to conclude from this incident that popular vote mechanisms favor short, sensationalist, and catchy stories over well-founded and thoroughly researched pieces, but there certainly is the risk of such a bias unless appropriate checks and balances are in place.

2. Tipping point effect makes Digg prone to gaming – the “photographs of crowds” blog incident

The existence of a tipping point, a threshold after which a self-amplification or auto-feedback effect starts to push a story without judging the quality of the submission, opened the door to all sorts of gaming. In an experiment published by Wired magazine editor Annalee Newitz, she created a new blog intended to be as random and boring as possible showing random pictures of crowds obtained from the Creative Commons area on Flickr.²⁹ She submitted the story to Digg, but nobody else seemed to be fascinated by her random collection of pictures of groups of people accompanied by little insightful comments. Then Annalee paid roughly USD 100 to a Digg-gaming service, which started digging her blog post. After getting roughly 40 paid hits, verified by the fact that most of the diggers causing these first 40 Diggs also digged the same two stories on a Photoshop tutorial and a temperature sensor and were obviously paid by the same Digg-gaming service, the story started to gain momentum and organic diggs started to come in. With 121 Diggs the story finally made it to Digg's front page.

The story got finally removed from the front page, when some Digg users bothered to investigate the blog and found that the stories were bogus and decided to bury it. What can be learnt from this case is that the popularity curve of a story at Digg is characterized by a tipping point, and users seem to vote in favor of stories which have crossed the tipping point without carefully reviewing their content. But it also shows that once a story has made it to the front page, it undergoes a more thorough review, which finally led to the story's removal. After an initial period of collective failure, the self-correction mechanisms seem to have worked in this instance.

3. Platform is kept hostage by its users – the Blu-Ray code incident

Digg gained widespread popularity when its users published the secret code to remove the technical protection mechanism of a brand-new HD-DVD standard called the Blu-Ray format.³⁰ Lawyers representing the Blu-Ray Association, a consortium of hardware manufacturers who own the intellectual property of the Blu-Ray discs, contacted Digg and asked them to take down the posts referring to the code. But Digg's users submitted

²⁸ See Nat Torkington: Digging the madness of the crowd *available at* http://radar.oreilly.com/archives/2006/01/digging_the_madness_of_crowds.html (last visited 15/11/2007).

²⁹ See Annalee Newitz: I bought Votes on Digg *available at* <http://www.wired.com/techbiz/people/news/2007/03/72832> (last visited 15/11/2007). Wired is owned by CondeNet, which also owns Digg's competitor Reddit.

³⁰ See Ian Douglas: Digg out on a limb for DVD hackers *available at* <http://blogs.telegraph.co.uk/technology/iandouglas/may2007/diggdvd.htm> (last visited: 15/11/2007).

reference after reference to the secret code. Wikipedia, afraid of being sued, deleted the key immediately and froze all suspicious pages to prevent the code from reappearing. The decentralized architecture of Digg made it quite hard for the editors to shut down the post and ever more creative variants of the story were posted to Digg. Also the community of Digg seemed to be determined to decide what appears on Digg and what should be prevented. In numerous comments and blog posts they expressed their anger about the attempts of Digg to suppress the publication of the key. Finally Kevin Gower realized that no matter what decision they were taking, Digg would face severe consequences: Either it alienated its users / members by imposing the ban against their will or risked of being sued by the industry consortium. In a blog post he decided to join Digg's users and go down fighting, rather than bow down to a bigger company.³¹

This incident shows that intermediaries and their users are entering in a symbiotic relationship and are not in full control of the uses made by the problem solving network they are facilitating.

4. Power to shape agenda in the hands of a few power diggers – the Jesusphreak incident

One of the promises of the distributed news submission and filtering concept employed by Digg is to leverage the wisdom of the crowd and the '1000 eyeballs' principle to select the news which the majority of Internet users or at least a representative amount of Digg users consider news-worthy. A recent study of the Project for excellence in Journalism comparing the agendas set by Digg and its competitors Reddit and Del.icio.us has found that the community of each particular online news aggregator introduces a strong bias with regards to what is perceived as news-worthy.³² We have also already seen that Digg's 1000 eyeball principle is prone to herd-mentality. A closely related question is how many eyeballs are actually determining the popularity of a news story, i.e. how large is the group of people that are shaping the news agenda of Digg news? Digg published until February 2007 a list of Top Diggers, which also contained the number of submissions and their popularity. An analysis of the Top Diggers list published by Digg found that the top 100 diggers determined 55% of the content shown on Digg's front page.³³ The distribution of power to shape the news agenda on Digg's front page seemed to be highly unequal. In the wake of this finding, the Digg community started a major discussion about how democratic Digg actually was.³⁴ Rumors about the existence of Digging circles, which exclusively digged each other's stories to increase their popularity ratio and hence their potential to shape Digg's front page agenda, arose.³⁵ At the end, Digg founder Kevin Rose announced changes to the Digg algorithm and the removal of the list of Top Diggers.³⁶

Digg finally removed the Top Diggers list on 1 February 2007, allegedly to avoid gaming, as some (unnamed) companies were accused of recruiting members of the Top Diggers list in order to draw attention to their sites by getting a favorable Digg from an influential top

³¹ See Kevin Rose: Digg This: 09-f9-11-02-9d-74-e3-5b-d8-41-56-c5-63-56-88-c0 available at <http://blog.digg.com/?p=74> (last visited 15/11/2007).

³² See Project for Excellence in Journalism: The latest news headlines, your vote counts available at <http://www.journalism.org/node/7493> (last visited 24/11/2007).

³³ See Randfish's blog post: Top 100 Digg Users Control 56% of Digg's Homepage Content available at <http://www.seomoz.org/blog/top-100-digg-users-control-56-of-diggs-homepage-content> (last visited 15/11/2007).

³⁴ See Jesusphreak's blog post available at jesusphreak.infogami.com/blog/what_happened_to_digg (last visited)

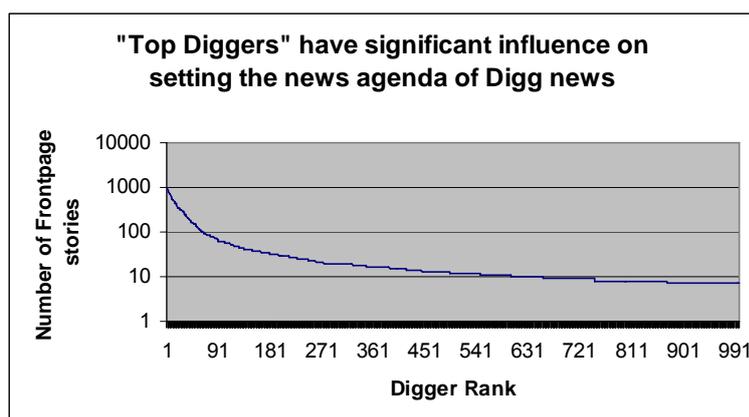
³⁵ See Dave's blog post: Digg Corrupted: Editor's Playground, not User-Driven Website available at http://forevergeek.com/news/digg_corrupted_editors_playground_not_userdriven_website.php (last visited 15/11/2007).

³⁶ See Kevin Rose's blog post available at <http://diggtheblog.blogspot.com/2006/09/digg-friends.html> (last visited 15/11/2007).

digger.³⁷ Since then, the Top Diggers list has been recreated by a private party outside of Digg, who regularly crawls Digg's popular section and counts the number of contributions for Digg's top 1000 users.³⁸ We have analysed the distribution of power based on the data published on the unofficial Top Diggers list and found that the distribution has not changed, despite the announced changes: Of the 36,917 front page stories submitted by the top-1000 users, 12,042 (or 32.6%) had been submitted by 20 users, and 22,965 (or 62.5%) by a group of 100 users. An analysis of the popularity ratios shows a similar picture: eight of the top-1000 users have popularity ratio of greater than 75%, i.e. three out of four stories they submit make it to the front page, and 41 users have popularity ratios of greater than 50%. Not surprisingly, Kevin Rose leads the popularity ratio with 98.3%. There is no obvious correlation between the number of submissions and the number of stories finally featured on the front page. The 41 users with a popularity ratio of higher than 50% submitted anything between 8 and 1484 stories, resulting in anything between 7 and 752 stories appearing on the front page. Still, the content shown on the front page of Digg is determined by a small group of diggers, who exercise significant influence over Digg's news agenda. There is some evidence that del.icio.us and Reddit show very similar patterns.³⁹

The following long-tail curve was derived from the data published by the inofficial top diggers list:⁴⁰

Table 2: The distribution of power at Digg



Conclusion

- The potential point of failure described in the first three embedded case studies do not reflect the normal performance of Digg. On the contrary, they demonstrate exceptional events, which are nevertheless important to understand the weaknesses and vulnerabilities of the model. These exceptions point to risk factors involved with this mode of production.
- The last embedded case, though, points to a structural imbalance inherent in the normal performance of Digg, and, if the additional evidence turns out to be reliable, also at other instances of Digg-style news aggregators. The power law distribution of influence is a strong indicator that the news agenda of these aggregators is not shaped by a popular

³⁷ See Kevin Rose's blog post available at <http://blog.digg.com/?p=60> (last visited 14/11/2007); see Christian Finke's blog available at <http://www.chrisfinke.com/2007/02/01/top-diggers-no-longer-recognized/> (last visited 14/11/2007).

³⁸ See Christian Finke's blog post available at <http://www.chrisfinke.com/2007/04/12/top-1000-diggers-20070411/> (last visited 14/11/2007).

³⁹ See Randfish *supra* fn. 33; see Christian Wallenta and Mohamed Ahmed: Analysing Systems with User Generated Content (working title), unpublished working paper, draft on file with authors.

⁴⁰ See Christian Finke's blog *supra* fn. 38 (last visited 14/11/2007).

vote of its users. Instead, the gathering and selection process is similarly concentrated at a few Power Diggers, which replace the gatekeeper function of journalists and editors in the traditional paradigm.

The performance of OhMyNews

A short history of OhMyNews

The Korean entrepreneur Mr Oh Yeonho started OhMyNews in 2000 as a Korean news site with a budget of USD 85,000, four professional editors and an initial group of 700 registered volunteers acting as Citizen journalists. The number of volunteers had increased to 14,000 by October 2001, to 20,000 by September 2002 and by July 2004, more than 34,000 citizen journalists had registered.⁴¹ In 2007, the community has grown to over 60,000 registered contributors.⁴² In 2001, OhMyNews started the English-language platform *OhMyNews International*, which is currently (2007) supported by 1,900 registered citizen journalists reporting from 100 countries all over the world.⁴³

The nature of OhMyNews

OhMyNews currently employs a permanent staff of 90 people, of which 30 are administrators and technicians and 60 editors, who review and select the articles submitted by the citizen journalists. The permanent staff is mainly financed through the sales of advertising space on the OhMyNews homepage.⁴⁴ Further revenue is generated by sale of news pieces, press reviews and other value-adding services like news alerts.⁴⁵

The citizen reporters, who have to register first to become accredited OhMyNews reporters, write stories based on first hand experience about their surroundings. The staff reporters and editors review the submissions and select those deemed worthy for publication. OhMyNews pays citizen journalists a small amount for articles, which finally make it to the front pages of one of the sections of OhMyNews. OhMyNews also publishes 'rare species' articles without remuneration. These latter contributions have not been selected for prominent placement, but still have been found worthwhile to be published somehow.⁴⁶ Another interesting form of appreciating the contribution of an individual reporter is the possibility for OhMyNews readers to tip her. Allegedly, one citizen reporter received almost USD 30,000 in only three days from readers in response to an article criticizing a conservative political group in Korea.⁴⁷ In another case, a citizen journalist, who documented the difficult situation of immigrants in Korea, raised over USD 30,000 from readers through tipping to help said immigrants.⁴⁸

41 See Eun-Gyoo Kim and James W. Hamilton: Capitulaton to capital? OhmyNews as alternative media *in* Media Culture Society 2006; 28; 541ff, 565.

42 See Cynthia Yoo: 'This Medal Belongs to OhmyNews Citizen Reporters and Staff' available at http://english.ohmynews.com/articleview/article_view.asp?article_class=8&no=380650&rel_no=1 (last visited: 14/11/2007).

43 See Todd Thacker: OhMyNews celebrates 7 years of Citizen Journalism available at http://english.ohmynews.com/articleview/article_view.asp?article_class=8&no=346674&rel_no=1 (last visited: 14/11/2007).

44 See Kim and Hamilton *supra* fn. 41 at 544.

45 See Judith Siers-Poisson: A Visit to OhmyNews available at: <http://www.prwatch.org/node/6407> (last visited: 14/11/2007).

46 See Kim and Hamilton *supra* fn. 41 at 545.

47 See Siers-Poisson *supra* fn 45.

48 See *id.*

OhMyNews is having significant political influence in Korea. Already in 2001, Oh was named as one of the 55 most influential people in Korea by the Korean newspaper *KyungHyang*. The weekly Korean magazine *Sisa* ranked OhMyNews as the sixth most influential media in Korea in 2003 and 2004.⁴⁹

The tremendous success of OhMyNews must be considered in the socio-political context of Korea: In the late nineties Korea had witnessed a change in generations, when the generation socialized after the end of the military dictatorship in 1987 started to gain influence in society.⁵⁰ Still at that time, the mainstream press widely ignored progressive ideas and openly supported only the conservative candidates during the general elections. The younger generation learned to use other channels of communication, which “matche[d] their openness, responsiveness and level of engagement.”⁵¹ OhMyNews satisfied the demand for an online platform to communicate and report the news, which the mainstream press, in the view of many younger citizens, did not cover sufficiently.

The performance of Global Voices

A short history of Global Voices

Global voices was founded in December 2004 by the activist Ethan Zuckerman and former CNN manager Rebecca MacKinnon at the Berkman Center at Harvard Law School to address the perceived need of better international collaboration between bloggers. They built Global Voices to help them find each other and engage in global conversations. After receiving initial funding from the MacArthur foundation and Hivos, Global Voices recently received a major grant from the news agency Reuters, with whom Global Voices cooperates in exploring the interaction between professional journalism and the blogosphere.

The nature of Global Voices

Global Voices acts as a news aggregator of 80 author-blogs from all over the world. A team of 17 volunteers is organized as an editorial team. Some of the blogs Global Voices links to are written by anonymous bloggers (most of them for political reasons), but Global Voices strives to link only to blogs of persons known to the regional editors, to verify its authenticity. The editorial team decides to include a blog if they are updated frequently, represent a point of view rarely seen in mainstream North American/European media, are well written and are written in a language one of Global Voices' editors and volunteer authors can understand (currently English, Chinese, Spanish, French, Arabic, Russian or Portuguese).⁵²

The existence of an editorial team differentiates Global Voices somehow from the pure blogosphere, where attention and selection is only created by an emergent and somehow random network of links between individual bloggers and their blogs. In Global Voices every blogger is free to publish whatever she deems newsworthy in her own blog. The editorial team reviews and features individual blog posts. In contrast to OhMyNews, the editorial team does not “decide” on what gets published, as long as the blogs they link to do not promote violence, terrorism, racism, and bigotry.⁵³ Their goal is rather to direct attention to individual topics and to present a variety of themes across geographies on the Global Voices homepage. The responsibility for the individual blog posts rests purely with the author of the

⁴⁹ See Kim and Hamilton *supra* fn. 41 at 551.

⁵⁰ See *id* at 553.

⁵¹ *Id* at 554.

⁵² See Global Voices FAQ *supra* fn. 53.

⁵³ See <http://www.globalvoicesonline.org/about/faq-frequently-asked-questions/> (last visited 15/11/2007).

blog post. The editorial team does not check the facts in the blog posts and encourages the readers to approach the blog posts with the usual amount of skepticism required when reading blogs. Another important function of the editorial team is to translate blogposts, if they are written in a language other than English to help the blogger to get international exposure. The multi-lingual dimension of Global Voices makes it special. While the majority of blogs are no longer written in English, for most readers there is little visibility of blogs outside their own culture and language. The volunteers at Global Voices translate several blog posts daily from foreign languages into English to make them accessible to the English-speaking world.

The amount of blog posts accessible through Global Voices is very large to process for an individual reader. One important function of the editorial team is to bundle blog posts by region and by topic and to create specific RSS feeds. Regional editors summarize the most important developments in the blogosphere of their region regularly and provide an overview over the most important developments. Users can tap into these resources and produce their own "Daily Global Me" with information from the blogosphere.⁵⁴

Users are encouraged to comment on the articles they read. Every blogpost can be commented on at the Global Voices homepage. The comments are moderated to avoid spam and hate speech, both dominant features in the blogosphere and a headache for every blogger. Users are also encouraged to suggest blogs to expand Global Voices' coverage of the world.

The core values of Global Voices are captured in the Global Voices Manifesto.⁵⁵ It contains a commitment to preserve and enable freedom of speech and universal access to information. To achieve this goal, Global Voices does not only offer a news platform, they are also active in developing tools to improve global communication, invest in training of bloggers, and engage in advocacy against censorship and freedom of press.⁵⁶

Wikinews and the Virginia Tech Shooting

Overview of Wikinews

Wikinews is an open source news Wiki established by the Wikimedia Foundation and launched in beta stage in December 2004. Founder Jimmy Wales has distinguished Wikipedia from Wikinews by stating "on Wikinews, each story is to be written as a news story as opposed to an encyclopedia article." In addition, Wikinews allows original work to be published under the form of original reporting and interviews and in contrast to OhmyNews and other citizen journalism efforts, Wikinews requires that a neutral point of view policy be maintained.⁵⁷

The idea for Wikinews originated in early 2003 as an anonymous post on Wikipedia's Metawiki, an auxiliary website for the coordination of all Wikimedia Foundation projects. The proposal was later redeveloped by freelance journalist and software developer Erik Möller, suggesting that a Wikinews effort could complement Wikipedia by providing "news on a wide variety of subjects, unbiased and in detail". By December 2004, a beta version of Wikinews was launched in both English and German. Later, in chronological order, editions in Dutch, French, Spanish, Swedish, Bulgarian, Polish, Portuguese, Romanian, Ukrainian, Italian,

⁵⁴ See above at p.2.

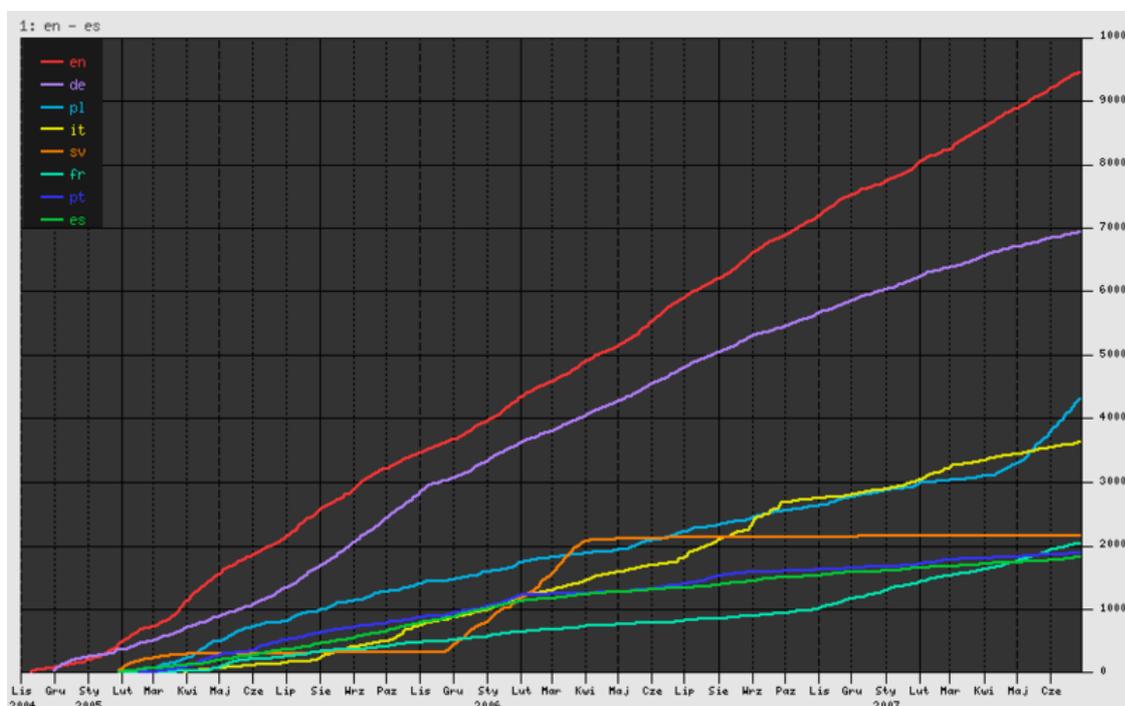
⁵⁵ See Global Voices Manifesto available at http://www.globalvoicesonline.org/wiki/article/Global_Voices_Draft_Manifesto (last visited 20/11/2007).

⁵⁶ See Global Voices FAQ *supra* fn. 53.

⁵⁷ See J. Glasner: Wikipedia Creators Move Into News *in*: WIRED, 29 November 2004, available at: <http://www.wired.com/culture/lifestyle/news/2004/11/65819> (last visited 08/01/2008).

Serbian, Japanese, Russian, Hebrew, Arabic, Thai, Norwegian, and Chinese were established. In September 2005, the Wikinews project moved to the Creative Commons Attribution 2.5 license. By September 2007, Wikinews published its 10,000 article in English, whereas the German version had approximately 6,300 articles.⁵⁸

Table 3: Growth of Wikinews over time in multiple languages from Nov 2004 to June 2007⁵⁹



While Wikinews focuses primarily on text articles, additional efforts are being launched to consider Audio Wikinews, which delivers Ogg Vorbis audio files, Wikinews Video 2.0, and a Wikinews daily print edition. However, like Wikipedia, Wikinews has been criticized for possible lapses in being neutral and including only verified and true information. Robert McHenry, former editor-in-chief of the Encyclopedia Britannica specifically objected to Wikinews, stating “the central question about the Wikinews effort is its credibility... making a newspaper is hard... someone who wants to do it but doesn't really know how hasn't solved the problem by gathering a lot of other people who don't know, either.”⁶⁰

One of the challenges with Wikinews involved discerning when an edit made to an evolving story is legitimate and when it is false or even vandalism. Even Jimmy Wales notes that “With a news story, the actual story has a limited lifespan. If it's not neutral, you've got to fix it quickly.” Another challenge with Wikinews involves maintaining a separate identity from Wikipedia. In 2007, Jonathan Dee of The New York Times, raised concerns that “So indistinct has the line between past and present become that Wikipedia has inadvertently all but strangled one of its sister projects, the three-year-old Wikinews.”⁶¹

⁵⁸ See A. Weiss: The Unassociated Press in: The New York Times, 10 February 2005, available at: http://www.nytimes.com/2005/02/10/technology/circuits/10wiki.html?_r=1&ex=1177300800&en=024e251d2c696137&ei=5070&oref=slogin (last visited 08/01/2008); see J. Glasner: All the News That's Fit to Wiki available at: WIRED, 04 April 2005, available at: <http://www.wired.com/techbiz/media/news/2005/04/67286> (last visited 08/01/2008).

⁵⁹ Source: http://en.wikinews.org/wiki/English_Wikinews_publishes_10000th_article (last visited 12/01/2008).

⁶⁰ See J. Glasner *supra* fn. 58; see J. Dee: All the News That's Fit to Print Out in: The New York Times, 01 July 2007 available at: <http://www.nytimes.com/2007/07/01/magazine/01WIKIPEDIA-t.html> (last visited: 08/01/2008).

⁶¹ See Dee *supra* fn. 60.

Wikipedia and the Virginia Tech Shootings

On 16 April 2007, a school shooting comprising two separate attacks occurred at the Virginia Tech in what was the deadliest school shooting in U.S. history to date. The perpetrator, Seung-Hui Cho, killed 32 people and wounded several others before committing suicide. While conventional media had challenges trying to piece together the chaotic events that occurred, a Wikipedia page was created by user Alangu that same day. That same day, more than 1500 edits were made to a specific Wikipedia page entitled “Virginia Tech massacre” by multiple users.⁶² The following day, more than 2000 additional edits were made to the page.⁶³

A time-lapse video was later posted on YouTube showing all the edits that were made to the Virginia Tech massacre Wikipedia page within 24 hours.⁶⁴ The Wikipedia page included a cell phone photo taken of French class students taking cover in Virginia Tech's Holden Hall, Room 212. Later contributions highlighted the student response and campus memorial for the 32 people killed, ultimately citing more than 120 online references as part of the Wikipedia entry.

What is particularly interesting about the Virginia Tech example is that individuals opted to use Wikipedia, not Wikinews, which also existed at this same time. While some Wikinews articles were created, by far the largest volume of activity (as indicated in the links above) occurred not on Wikinews by Wikipedia. This activity demonstrates two things. First, that the comments by Jonathan Dee and Jimmy Wales are right, it is hard to distinguish Wikipedia from Wikinews (in fact Wikipedia may be strangling Wikinews). Second, that a large number of motivated people could collaborate over a period of 48 hours and make more than 3500 distributed updates to a chaotic and evolving news story.

When Trends in News Influence Intelligence Agencies

Related to the theme of chaotic and evolving news environments, a group of national security experts met in Glasgow, Scotland in November 2007 to discuss new approaches to energy and environmental security efforts. This group specifically identified that understanding and anticipating security problems that stem from energy and environmental issues will be a powerful tool for government and business-sector decision makers. However, the group also noted that decision makers often lack the concrete data and pragmatic assessments necessary to understand fully how changes in the energy sectors or in environmental systems will affect other elements of either sector or system, national economies, global institutions, local cultures, or regional rivalries. Intelligence reports frequently focus in-depth on a single issue, thus ignoring many of the inter-relationships among diverse issues and key stakeholder groups. Decision-makers also may lack sufficient knowledge regarding the sensitivity of key relationships to regional and global stability and the fragility of local interactions to environmental perturbations.

Given these concerns, the group decided as one of their future action items, they would “demonstrate and prototype knowledge cultivation activities, specifically applying the 'lens' of national and international security implications.” The group decided that what happened with the Virginia Tech shooting, and trends in social media such as Digg.com, OhmyNews, and Wikipedia – cumulatively represented trends that might help address the challenges and complexity of gathering intelligence on energy and environmental issues to inform government and business-sector decision makers. Whether this trend will transform not only

⁶² See

http://en.wikipedia.org/w/index.php?title=Virginia_Tech_massacre&dir=prev&limit=500&action=history (last visited 12/01/2008).

⁶³ See *id.*

⁶⁴ See <http://www.youtube.com/watch?v=zrCQ9dUsfqU> (last visited 12/01/2008).

how government intelligence agencies collect and distribute information, but also how news and citizen journalists interact with intelligence agencies is unclear.⁶⁵

News at bbc.co.uk

A short history of bbc.co.uk

www.bbc.co.uk is the news website of the British Broadcasting Company (BBC). It was established in November 1997 by merging several entertainment and educational websites of the BBC and publishes content written by a dedicated team of online journalists based in London. It draws from the BBC's regional and international infrastructure and their local expertise. BBC specialists and correspondents also frequently write on particular topics. bbc.co.uk uses the BBC World Service as another source of news and has contracts with a large number of news agencies like Press Association, Associated Press, Reuters and Agence France-Presse.⁶⁶ bbc.co.uk is funded by the money collected from the UK's compulsory TV license and is free of advertisement within the UK. In 2007 the BBC announced plans to introduce banner ads for viewers from outside the UK.⁶⁷

Google News

A short history of Google News

Google News was launched in September 2002 for English-language news. In 2006, Google offered 22 regional editions in 10 languages. Google News relies on Google's proprietary Page Rank algorithm to identify news items and assess their relevance. Google claims that the selection of news items is adjusted to the taste of their users based on the stories they have read previously – provided they decide to identify themselves by logging in. While the exact workings of the Page Rank mechanism constitute a trade secret, some basic principles are known: Google most likely manually selects a set of highly regarded news sources, which receive a very high score. Pages which these sources link to also receive a high score, although lower than the original source. As a result of the distance to the initial set of sources and the number of links and link paths to get there, every news source receives a score, which influences their rank in the search catalog. It is therefore difficult to categorize the algorithm as centralized or decentralized. The design of the algorithm is determined centrally by Google's engineers. Nothing is known about the selection of the trusted set of news sources, but again this is most likely a central decision. The link structure followed by Google's algorithm has been created by the editors of the news sources, and therefore decentrally. We will for the purposes of this analysis treat the news selection algorithm as central, because the users only have very limited influence on changing it by their interaction with the platform. But it is important to note that there are several points of control, on which a categorization of the solution mechanisms could be built.

⁶⁵ See The Glasgow Group: Enabling Strategic Intelligence on Energy and Environmental Security Impacts and Consequences, Working Paper Released in November 2007 *available at*: <http://ssrn.com/abstract=1069129> (last visited: 08/01/2008).

⁶⁶ See <http://news.bbc.co.uk/1/hi/help/3281815.stm> (last visited 23/11/2007).

⁶⁷ See Richard Sambrook: Adverts on BBC.com *available at* http://www.bbc.co.uk/blogs/theeditors/2007/10/adverts_on_bccom_1.html (last visited 23/11/2007).

Roles in online news production – Problem Holders and Problem Solvers

In the traditional model of news production comprising news agencies and news publishers, the roles were clear cut: News agencies employed reporters, who gathered information, and news publishers employed editors, who selected the news for publication. Some larger publishing organizations employed their own reporters to investigate and their own column writers for opinion articles, which were clearly marked as such. Some writers also worked as reporters or as editors and vice versa, but by and large the distribution of roles was quite straightforward. The readers or viewers were a separate group, who got in contact with the editorial staff through written submissions or later call-in shows.

In the online news production space the distribution of roles to individuals is no longer clear cut. The line between citizen reporters as contributors of news and readers of the same news is blurry. Internet users, empowered by intermediary facilitated distributed problem solving networks, can act as reporters, the editor, and the reader, and can seamlessly switch between these roles. Even the classical “problem holder” – “problem solver” dichotomy is becoming blurry in view of the multi-faceted interaction between the users and the online news production platforms.

At **Digg News** the “problem holders” are users who want to know what is happening in their area of interest. They can browse the catalogue of popular and upcoming stories, search for a key word, or even ask a specific question. At the same time they can vote on the articles they have read, thereby switching into the role of the “problem solver” and supporting Digg in refining the relevance of the produced news results. They can also submit articles they found interesting to Digg News and thereby act as Problem Solvers for the News reporting problem.

Digg News: Problem holders = Problem solvers

At **OhMyNews** we need to distinguish between the news reporting and the news selection problem: In the news reporting problem, the problem holders are the users, who expect first-hand reports from people who have actually witnessed the news they are reporting. These users could at the same time also be problem solvers if they have registered as citizen journalists before and actively contribute by writing and submitting stories. The selection of the news items is left to OhMyNews' editorial staff, who evaluate and prioritize the news based on their professional judgment and OhMyNews' style guidelines.

OhMyNews: Problem holders \neq Problem solvers for the selection problem, and Problem holders \supset Problem solvers for the production problem

Problem holders at **Global Voices** are looking for access to information from regions or in languages which they otherwise would not get access to. The problem solved by Global Voices is a more complex form of the reporting problem: Global Voices facilitates and broadens access to networks, which the problem holders would not have been able to access on their own, being restrained by their lack of knowledge of the appropriate bridgeblogs or by being unable to understand the language. The problem solvers of the reporting problem are volunteers regularly capturing their impressions, comments, and opinions in their blogs. The problem solvers for the selection problem are motivated by philanthropic motives and believe in freedom of speech and global access to information.

The problem holders at **Google News** are users, who use the search engine to access the news items they are interested in. Google News also features ample news coverage on their home page, directing users' attention to currently popular stories – where popular means being ranked highly by Google's proprietary search algorithm. The problem solving in the selection problem at **Google News** is confined to a handful of Google engineers, who implement the selection logic in the Google search algorithm.

Google News: Problem holders \neq Problem solvers

On **bbc.co.uk**, users can navigate a pre-structured catalog and choose from the headlines or editorial categories like World, UK, Business, etc. to find stories which they are interested in. Problem holders and problem solvers are clear-cut roles and the groups do not overlap.

Users of news sources display differing levels of sophistication and expectation. While some problem holders have a very confined expectation of a particular topic they are interested in, others enjoy browsing a catalogue of options. All presented news production networks confront the problem holders with more solutions to their problem than they had initially looked for, and thereby aim to broaden awareness beyond the problem holders' initial expectations.

Some problem holders may be capable of selecting the information which is subjectively relevant for them from the general flow of information available on the Internet. The decision to turn to a problem-solving organization can be understood as a decision to save time and effort or to expand the range of information one can get hold off by oneself. The nature of news production requires the collaboration of a significant number of distributed problem solvers, which an individual problem holder cannot get access to.

Under all analyzed production paradigms, intermediaries (i.e. the news production platform) connect problem holders and problem solvers. It is the role of the intermediaries to bring together the required resources and facilitate the production process. The process underlying the problem solving in each case has been pre-structured by the designers of the platform.

Table 4: Degree of openness in selected news aggregators

	Digg	Global Voices	Google News	OhMyNews	bbc.co.uk
News reporting	Open	Open	Open	Semi-open	Closed
News selection	Open	Semi-open	Closed/ automatic	Closed	Closed

Motivation of problem solvers

In his analysis of the motivations for contributing to the production of public goods, Yochai Benkler found a “diversity of production incentives” as one of the destabilizing factors of the current paradigm of industrial information production and concludes that the intention to exclude others from the information produced underlies only three of nine possible modes of production.⁶⁸

In a comparative study of motivation to contribute to several different online communities, the motivation of bloggers ranged from “Collaboration”, “Reputation”, and “Self Expression” to “Egoism”, “Egotism”, and “Power”.⁶⁹ The bandwidth of motivations is as diverse as the range of topics covered and the tone of voices reflected in the individual blogs. The mission statements or values of the intermediaries Digg, OhMyNews and Global Voices provide a bit more insight, although it would be over-simplistic to reduce the motivation of all contributors to the values presented in these statements.

In Digg's Blu-Ray incident we have seen a strong feeling of identity with the Digg platform and the strong sense of the community to “own” the platform. At the same time the

⁶⁸ Yochai Benkler: The wealth of network, p. 43.

⁶⁹ Trevor D. Moore and Mark A. Serva: Understanding member motivation for contributing to different types of virtual communities: a proposed framework in Proceedings of the 2007 ACM SIGMIS CPR conference on 2007 computer personnel doctoral consortium and research conference: The global information technology workforce available at <http://portal.acm.org/citation.cfm?id=1235035> (last visited 20/11/2007).

community is very aware and critical of the unequal distribution of power to influence and shape the news agenda of Digg. Some of the ‘power Diggers’ may also be incentivized by monetary rewards, which stem from selling their influence to interested parties. The problem solvers at **OhMyNews** are motivated by a wide range of incentives ranging from the satisfaction of seeing their own articles published, participating in a community that is changing the market of news production to receiving a (nominal) financial compensation, which ranges between US\$2 and \$22 per published story.⁷⁰ Problem solvers of the news reporting problem at **Google News** are usually paid by other organizations. The problem solvers for the selection problem are paid engineers at Google and therefore “motivated” by their salaries or other goals of professional satisfaction.

Degree of professionalism required to participate in news reporting

Each of the problem-solving organizations analyzed in this case studies takes a different approach on the degree of professionalism required to participate in news reporting. **Digg News** is agnostic about the production of the news because it focuses on the selection mechanism. To participate in the selection process no special expertise is required. **Google News** is also agnostic about the mode of production, but has “encoded” stochastic expertise in its selection mechanism, which is designed and maintained by experts and which seems to favor professional sources over blogs.⁷¹ **bbc.co.uk** relies on professional staff to write and select the stories and adheres to the professional standards of news production.

OhMyNews requires their citizen reporters to register before starting to submit stories. During the registration process, citizen reporters agree to adhere to a set of rules widely resembling the code of conduct for professional journalists which requires, among others, an obligation to ‘truth’, the verification of all facts presented, independence from those they cover as well as the governing power, the principle of proportional coverage, and a fair treatment of diverging opinions and public criticism.⁷² In comparison to this, OhMyNews does not require or even encourage reporters to maintain an objective point of view – to quote OhMyNews founder Oh Yeon-Ho: “OhMyNews does not regard straight news articles as the standard. Articles including both facts and opinions are acceptable when they are good”.⁷³ Core principles of OhMyNews' journalistic ethics include the truthful presentation of facts, no spreading of rumors or false information, the disclosure of all interests, which could influence the intention of the article, e.g. work in marketing or public relations, refraining from obtaining information illegally, and several others. It is noticeable that there is no requirement to be detached and disinterested in the object of coverage.

Quantitative approaches to measuring performance

Performance indicators identified in the literature

The most comprehensive study so far on online news aggregators has been produced by the Project for Excellence in Journalism⁷⁴ (PEJ), comparing the agenda of Internet news aggregators of the Digg News type (specifically Digg News, del.icio.us, and Reddit) with the traditional news agenda (derived from the editor-moderated online platform Yahoo! News).

⁷⁰ See The end of 20th century journalism *supra* fn. 73.

⁷¹ See Google News Monitor *available at* <http://88.198.49.15/~search/gnews/reports/us/index.php> (last visited 24/11/2007).

⁷² Bill Kovach and Tom Rosenstiel: The Elements of Journalism, London 2003, p. 12f.

⁷³ See The end of 20th century journalism: Every citizen is a reporter *available at* <http://www.slideshare.net/marketingfacts/case-ohmynews%20> (last visited 15/11/2007).

⁷⁴ www.journalism.org; see also above at fn.x.

The key focus of this study was on the sources of the stories and on the agenda set by the different paradigms of news production. To compare the agendas, they categorized the news items made popular during one week in July 2007 according to their own indexing system into different categories and quantitatively compared the number of indexed stories in each category.

The key findings of the PEJ study can be summarized as:

- The news agendas of Digg-type aggregators diverge significantly from the traditional news agenda: The topics covered on the news aggregator sites tend to be more focused on domestic topics, celebrity news, and technology
- Topics at the aggregator sites have a high volatility, e.g. some stories only appear once and are not followed up by another story
- There is some, but limited variety between the preferred topics at each the three aggregator sites studied
- When looking at the most emailed and most viewed stories at Yahoo! News, the stories chosen by the readers follow different priorities than the stories chosen by the editors.

The source of the content found at the online aggregator sites by the PEJ study was mainly blog posts (40%), followed by non-news items like YouTube videos (31%), and traditional news items from sites like the BBC. This is an interesting finding, because it implies that items considered not-newsworthy by traditional news organizations, like the trivial content found at YouTube, is turned into something newsworthy by reporting it on a news aggregator site.

The methodology used in the PEJ study would have allowed for further investigation of performance indicators on the nature of the news agenda, like the average time an individual story or a topic is featured on the front page or the fluctuation of stories in the headlines. But in the PEJ study, only anecdotal evidence is given to support the thesis that the attention span of “user-driven” news aggregator sites is shorter than for traditional news producers that rely on editors, and which follow and cover specific broad topics over a certain amount of time.

A different approach to quantify the performance of news aggregators uses Webmetric analysis. Webmetric analysis is a quantitative approach that maps the hypertext links between different websites and thereby shows the structure of the analyzed network. A Webmetric analysis of the news aggregator and traditional news producer space has been done by Rob Ackland, including the sites analyzed in this paper, but also several other English-speaking news sites.⁷⁵ In the Webmetric analysis the number of inlinks and outlinks between a fixed sample of news sites has been measured using a Yahoo tool. Although this tool is limited to 1000 links and the analysis is restricted to the selected sample of news sites, the analysis clearly shows a high degree of interlinks between the traditional news producers and the distributed news aggregators. From the Webmetric analysis we can learn that the Blogosphere does not exist as a hermetic and independent news space, but interacts with the traditional news space through linking to their article as well as being linked to by the traditional news producers.

The approach taken in this paper goes beyond the approach taken by the PEJ report and the Webmetric analysis in two major aspects: First it builds on an analytic understanding of the different news production paradigms in the Internet space and chooses representatives of each paradigm for the analysis. Unlike the PEJ report, not only user-driven *selection*, but also user-driven *creation* of news items has been included in our analysis. Second, we focus not primarily on the topics constituting the agendas of the different news production paradigms or the structure of the network feeding the network, but on the analysis of the dynamic dimensions of news production, i.e. how fast news items become popular and how

⁷⁵ See Robert Ackland: Draft paper for OII-MTI Distributed Problem Solving Networks Project (2007).

long they remain so. As a result of this analysis we will get a better understanding of the continuity of the agendas constructed by the different paradigms.

Digg News and Google News provide publicly accessible tools to allow tracing of the published news stories over a period of time. Although the tools provided in particular by Digg News are useful for visualizing the development of a news agenda, they do not allow in-depth analysis because they do not provide access to the underlying data.

We therefore decided to use a neutral and platform-independent solution to investigate the performance of the news sites by analyzing their dynamic output. We used a modified version of the agenda analysis tool developed by Tobias Escher to compare the agenda set by the Blogosphere with that of traditional news publishers,⁷⁶ and extracted the most popular stories from the RSS feeds of some of the news platforms portrayed in this paper. From this data we extracted the number of new stories in the headlines of each site per day, the time a particular story stayed in the headlines, the number of stories picking up the same topic on subsequent days, and the average time a particular topic remained in the headlines in the follow-up stories.

Tools and methods

We analyzed the RSS feeds of the news sites Digg, OhMyNews, Global Voices, bbc.co.uk, and iht.com. We subscribed to the “Headlines” or “World News” RSS feed or the equivalent and also to the feed on technology news. The modified RSS reader obtained an update of the RSS feed every three hours over a period of 14 days in late November and early December (November 24 until December 7). The RSS feed was stored in a basic database, which allowed for easy comparison between the content of two feeds to analyze the changes that occurred between the two extractions. In the second stage, Yahoo!'s key word extractor provided us with a list of key words in the article. We stemmed the key words to improve the comparison across stories. Using the stemmed key words we compared the news stories of each feed over time to see how topics developed or were followed up by subsequent articles over time. To normalize across the different time zones, local times were converted to GMT.

Following the performance framework described by Galtung and Ruge⁷⁷ we focussed our analysis on the frequency and the continuity of news production

Early findings

Frequency

In the first analysis, we looked at the news turn-over of the different sites and measured the percentage of new news stories that appeared per 3 hour cycle in the feed. The perhaps surprising result shown in table 5 is that the traditional news editors BBC and IHT had a far higher turn-over than the distributed alternatives Digg and Global Voices. OhMyNews also scored far below the traditional platforms. bbc.co.uk, iht.com and OhMyNews are characterized by clearly identifiable “update cycles” in which 50-60% ([bbc](http://bbc.co.uk), [iht](http://iht.com)) of the news feed gets replaced by new stories, while Digg and Global Voices continuously update 10-15% of their news per cycle. The big spike for iht.com in cycle 34 is an artifact, caused by technical problems with its RSS feed in cycle 33: in the next cycle all articles were therefore recognized as “new”.

What is striking is the existence of a “weekend hole” across all analyzed platforms. News production at iht.com is very quiet on Saturday and only picks up on Sunday afternoon, when staff begins to prepare the Monday morning paper. BBC News is a bit more active on Saturday, although clearly reduced, and only picks up at full speed on Monday afternoon. OhMyNews and Global Voices also show a slow down over the weekend, although it is hard

⁷⁶ See Tobias Escher *supra* fn.5.

⁷⁷ Above at fn. 6.

to speculate about the reasons. It could either be that their editorial boards are not working at full capacity over the weekend or it could be that citizen journalists are also taking weekends off. Without further investigation it is hard to draw a definite conclusion from the data. It is quite interesting to see that Digg's news production also slows down over the weekend, which could be explained by a large number of contributors using their PC at work to contribute to Digg. Another possible explanation could be that since many "Diggs" refer to articles published on other platforms (as Rob Ackland has shown in his Webmetric analysis of the news space) the slowing down of news production over the weekend also affects the digging activity of news on Digg.

Table 5: New news stories per collection cycle (Main News)

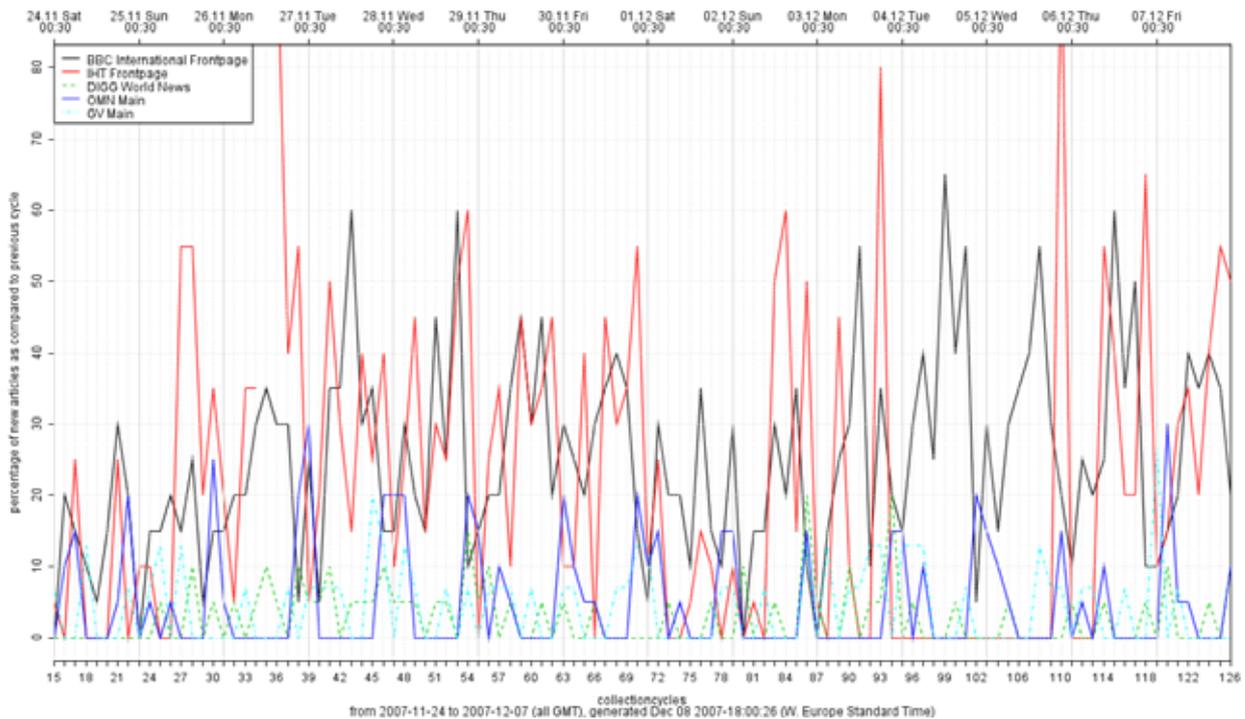
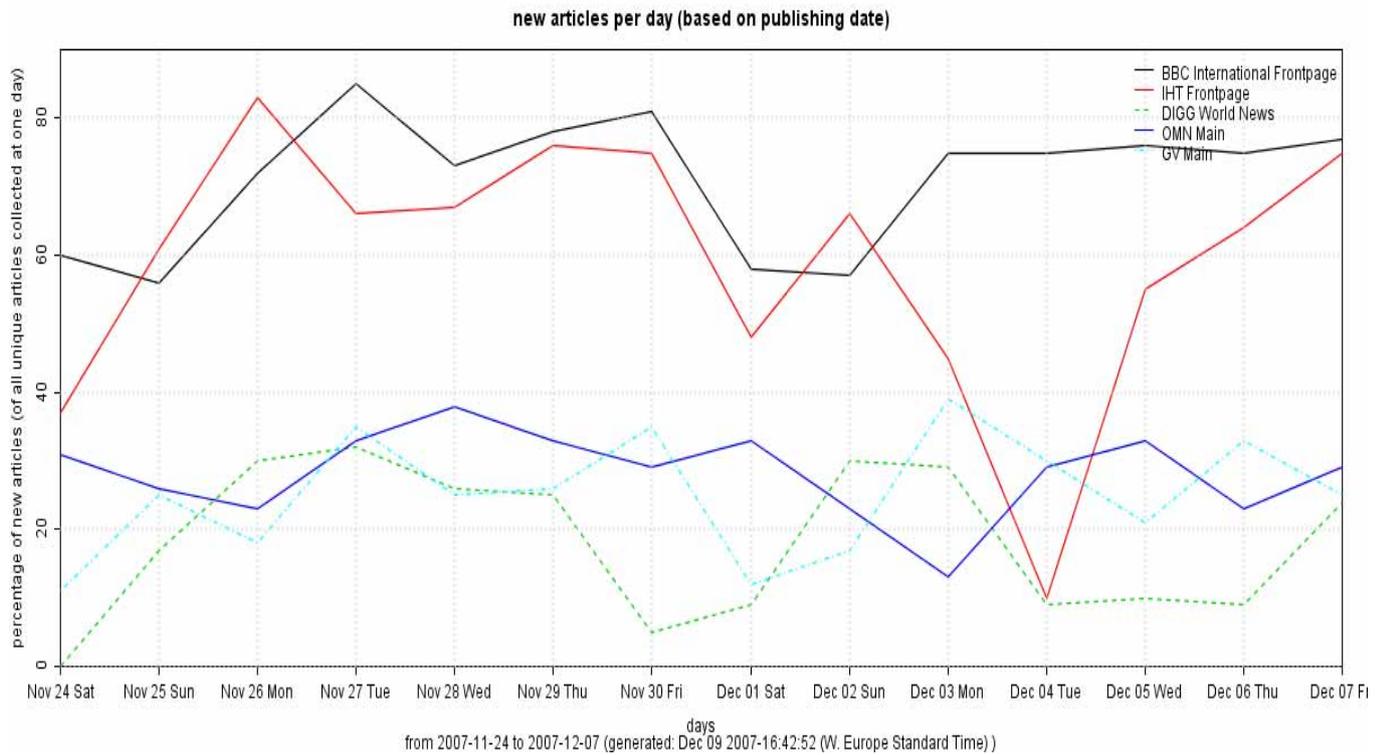


Table 6: New news stories per day (Main News)



The aggregated version of the news turn-over analysis shows a similar picture: We counted the new stories per day per platform to check if the continuous updating mode of Digg and Global Voices resulted in a similarly high news turn-over as the “step-change” approach taken by bbc.co.uk and iht.com. But also in the aggregate version, the traditional news aggregators have a much higher percentage of new stories in the news feed per day than the distributed alternatives.

The PEJ study pointed out that technology news was a domain of high interest among the distributed news aggregators like Digg news. We therefore subscribed separately to the technology news feed and analyzed the results accordingly. Indeed, the turn-over of news at Digg outnumbers the other platforms by far. Between 20% and 50% of the news feed at Digg was replaced by new stories for most analyzed news cycles. The BBC updated their news section roughly once every day with 10-20% of new stories per update, while OhMyNews and Global Voices reported on technology only sporadically.

Table 7: New news stories per collection cycle (Technology News)

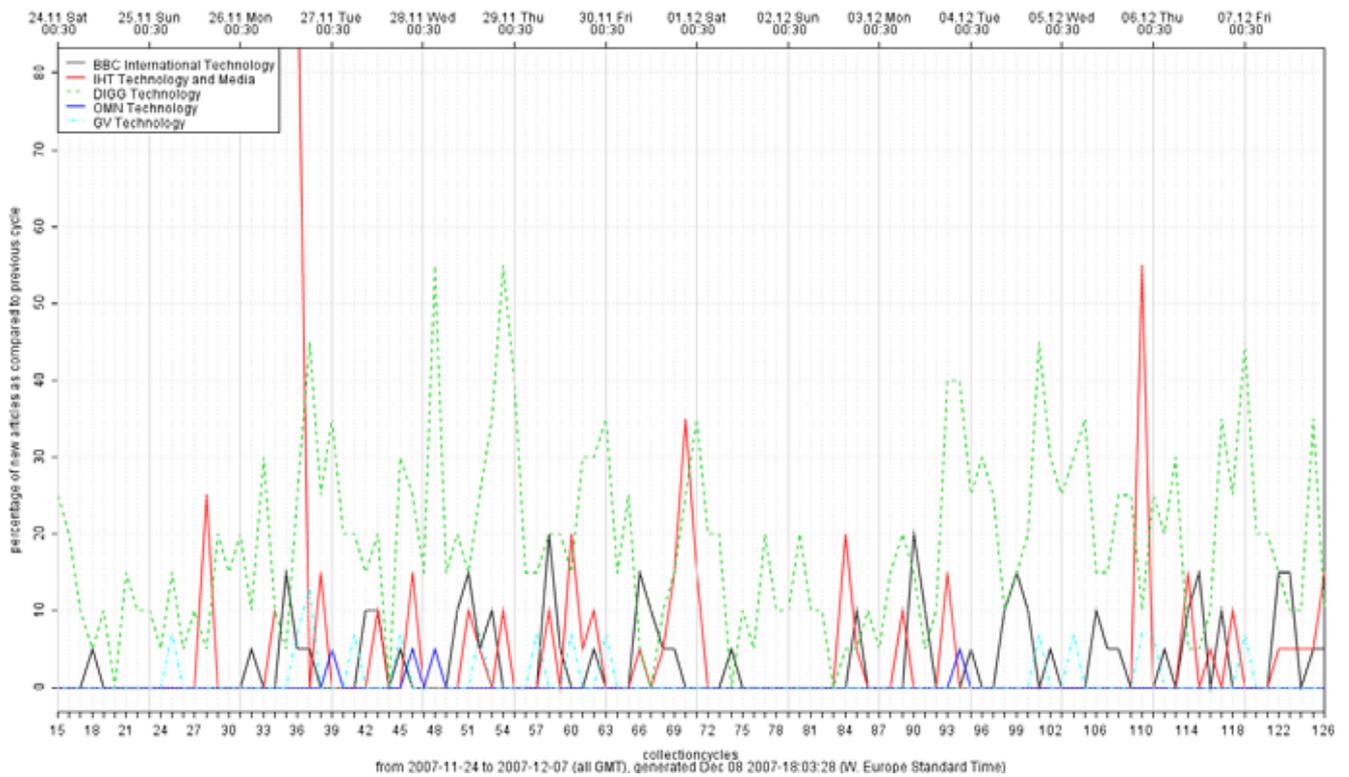
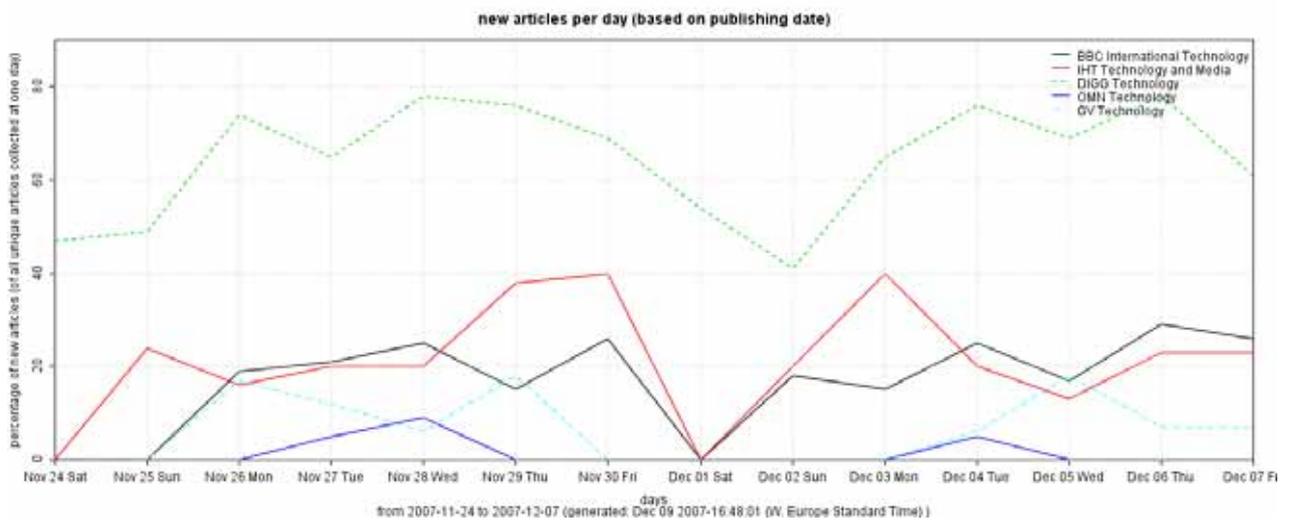


Table 8: New news stories per day (Technology News)



Continuity

To analyze the continuity of news production we have measured the average time an article stays on the front page of each news platform. In the second analysis of this part we will attempt to compare the key words in the published news items over time. The latter analysis is at the time of the writing of this paper still ongoing and has not delivered significant results as of now. We will therefore focus on the first aspect of continuity: the life-span of a news item on the front page.

Table 9: Life-span of a news items in hours (Main news)

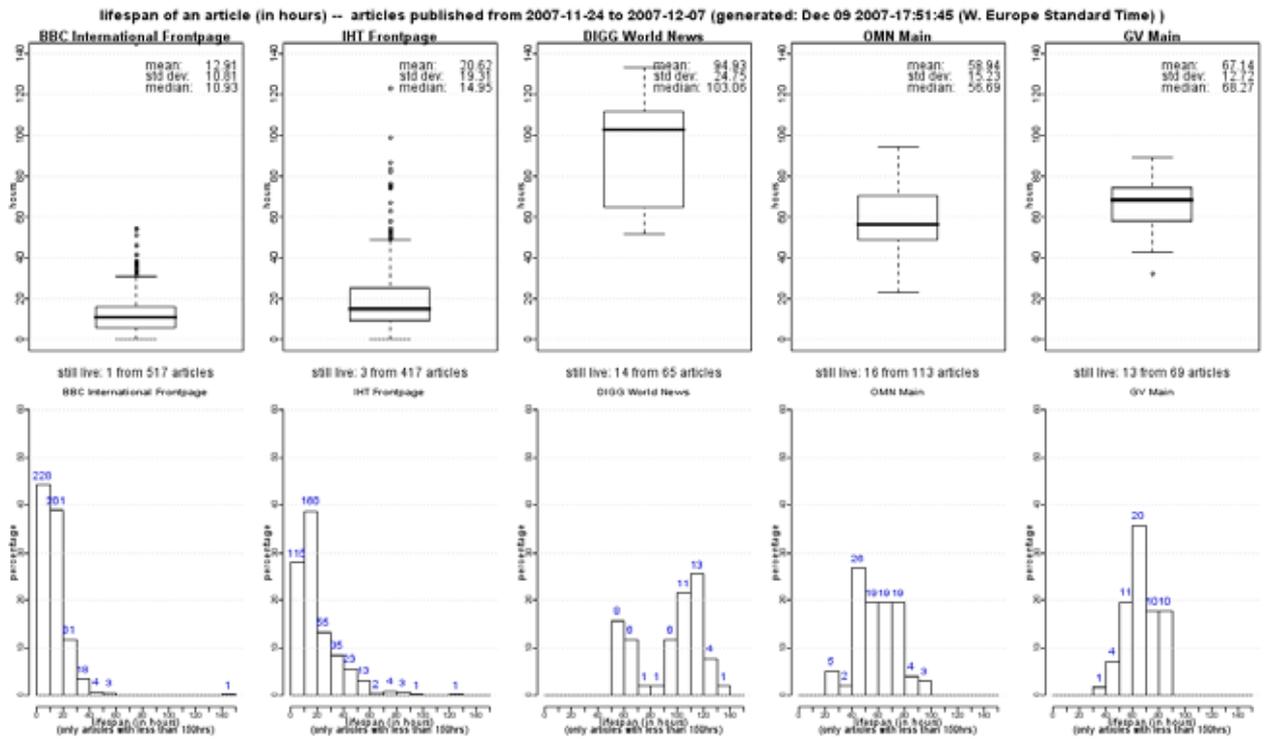
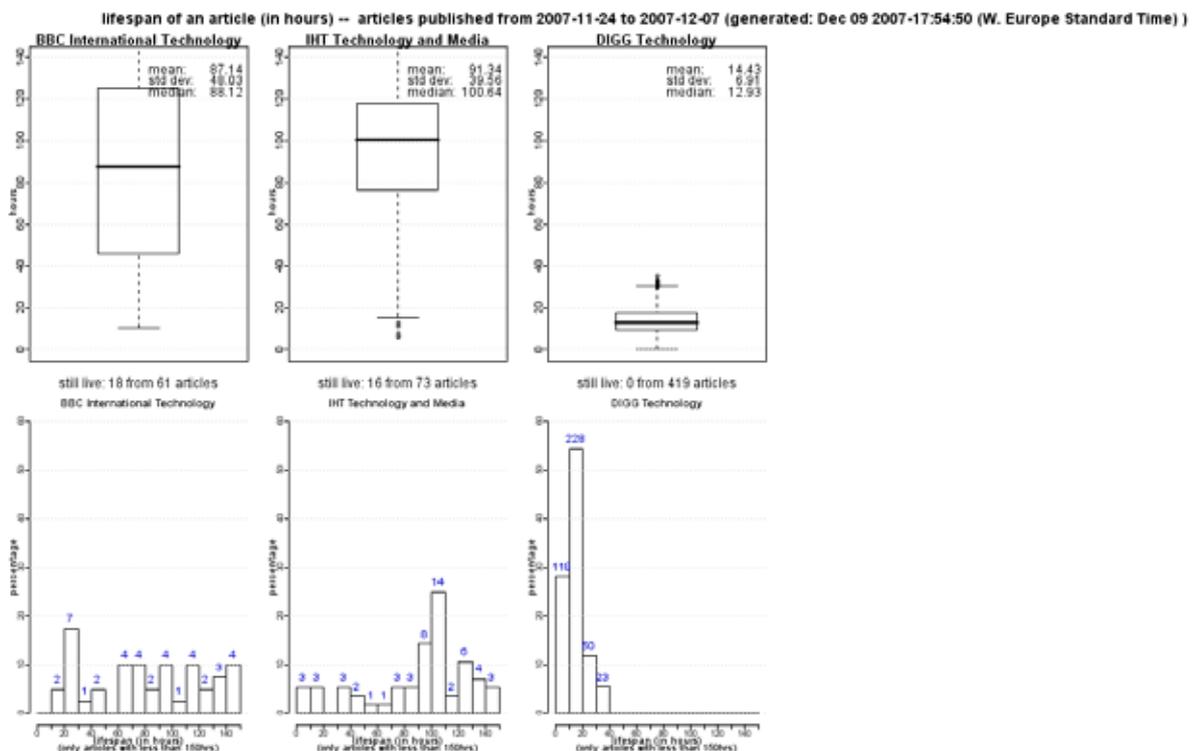


Table 10: Life-span of a news items in hours (Technology news)



In the main news feed the life-time of an article is shortest at iht.com and bbc.co.uk. Both show a sharply descending pattern with most articles being very recent and only a few articles staying on the frontpage for more than 2 days. Digg's curve shows two fields: The first cluster of news items has been around for 2-3 days, while the other has been active for 5-6 days. Further observation is required to check if this is a characteristic result of a systematic bias or a random artifact. Global Voices and OhMyNews show life-span curves,

which resemble more a Normal Distribution with a cluster of articles around the mean of 60 and 70h respectively.

The picture looks fundamentally different when analyzing the Technology feed: At Digg, the average article in the Technology feed has a life-span of 14.4 hours compared to 87 and 91 hours and *bbc.co.uk* and *iht.com*, respectively. While *iht.com*'s and *bbc.co.uk*'s life-span curves look almost linear, with a high deviation around the mean, the curve at Digg shows a deep decline after 20h, reflecting the short average life-span and the high turn-over of stories.

In a similar study of Digg News' front page undertaken by Wu and Huberman,⁷⁸ they observed that “a story usually lives on the front page for a period between 1 and 2 hours.” The front page is made up of the most popular stories from all categories. If we assume that all eight categories of Digg News contribute equally to the front page, this finding is in line with our findings. Wu and Hubermann also analyzed the distribution of the collective attention received by a particular article over time. In particular they looked at how the increase of popularity caused by the positive re-enforcement effect of more people reading an article was outweighed by the decline in novelty. They found that the growth of attention halved after 69 minutes.⁷⁹

Key findings of analysis

The edited news producers *bbc.co.uk*, *iht.com*, and *OhMyNews* are characterized by clearly identifiable updating cycles: they possess a frequency as described by Galtung and Ruge. Digg and *Global Voices* show a more continuous update pattern. Their frequency is created by the course of events, not the other way round. All platforms are more or less affected by a reduction of activity over the weekend.

The traditional news publishers changed the articles in the analyzed news feeds more often and covered more topics per period than the analyzed distributed platforms. This somehow counterintuitive finding is highly dependent on the analyzed news feeds. When looking at technology news, the articles on Digg were replaced four times faster than on *bbc.co.uk* and *iht.com*. This finding leads to the assumption that the differences in continuity between the production paradigms of Digg and the traditional publishers are less structural, but more related to the attention the shapers of the agenda devote to the topic. The technological community at Digg produces technology information more quickly and keeps it more up to date, while technology is only a sideline for the traditional media. The traditional platforms devote a fair amount of their energy to keeping the world affairs section up to date and following up with new stories. To translate the life-span into a proper indicator for the continuity of the news agenda, we need to complement it with the key word analysis to verify if the higher turn-over of stories at the traditional publishers is offset by the coverage of the same topics in related articles, which further expand the information provided in the original article.

Towards a conclusion

What do distributed news production platforms do “better”?

From our analysis we conclude that the distributed platforms seem to be particularly strong in shaping a news agenda that is in line with their users' interests and thereby expand the

⁷⁸ See Fang Wu and Bernardo A. Huberman: Novelty and collective attention *in* PNAS vol. 104, no. 45, p. 17599–17601 (November 2007)

⁷⁹ See *id.*

range of information available to the general public. Digg's performance in technology news is stronger than that of its competitors. Global Voices and OhMyNews have created a news profile which selects news that would not have been available to the readers through the traditional channels. The distributed platforms seem to be particularly strong in shaping a news agenda that is in line with their users' interests and thereby expanding the range of information available to the general public.

While the broadening of the range of news constitutes clearly "better" performance, the 'alignedness' with special interests is controversial: The ethics of traditional journalism have avoided focusing the agenda on satisfying public demand for information. At least in theory, media have a duty to broadly inform the public and raise and keep attention on unpopular, but important topics.⁸⁰ Cass Sunstein has raised the concern that the Daily Me created by tapping into the blogs that only reflect the reader's own opinion constituted an "Echo Chamber", which reiterates the point of view of the reader and makes it easier to avoid being exposed to information the reader does not like.⁸¹

It may be that traditional news agencies, even with globally distributed field reporters, are out-of-date for today's complex news environment. When the campus shootings occurred in the United States at Virginia Tech, Wikipedia actually had the most-up-to-date information compared to traditional news agencies. In addition, several intelligence agencies in the U.S. have recently re-examined their own traditional approaches to gathering intelligence, and in at least one instance a set of experts concluded that single agency-approaches to gathering intelligence were incapable of dealing with the interdependent events of the modern day. We may very well be approaching a point where there may be simply too much news being produced globally for a single organization's capacity to report on it all. The accelerating trends of globalization, where increasing volume, velocity, volatility, and concerns over the veracity of news produced challenge the traditional approach to both intelligence gathering and news production and reporting. Having reviewed several distributed modes of news production in this paper, we suggest that web technologies may represent viable solutions to addressing these challenges, and represent the next stage of development in the evolution of intelligence and news gathering activities.

Who captures the benefits?

- **Readers** may perceive the pre-selected access to information that reflect their opinion and meet their interests as a benefit, which saves them the time spent skimming through information on topics they are not interested in. On the other hand, to avoid the "Echo Chamber" effect, a reader needs to review several sources of information on the Internet to assemble a complete and balanced picture of the news.
- **New intermediaries**, which position their tools of getting access to and selecting information against the traditional media, capture the benefits of drawing significant traffic to their pages. Digg News and OhMyNews monetize the attention by advertising space on their sites. In addition, OhMyNews is selling added-value information services to paying customers.
- **Bloggers** capture the benefit of increased credibility by being featured on a news aggregator site. The "peer reviewed" of the public vote or an editorial board indicates that they are not alone with their view of the world. We have seen in our study that the peer review can also produce erroneous results, but that after some time errors will be corrected and false information be identified as such through self-correction mechanisms.
- **Traditional news reporters** capture benefits by reviewing news aggregators to find topics which may have so far evaded their attention. They thereby increase their network of informants. Traditional media have in recent years reduced the number of permanent

⁸⁰ See Gillmor *supra* fn.2. at p.5.

⁸¹ See Cass Sunstein: Republic.com 2.0.

staff reporters. As a result, the range of topics or regions covered by an individual reporter has increased. Many reporters also work as free-lancers and are hired to cover particular stories only. Blogs can act as a point of entry to start an investigation. Unscrupulous reporters will even capture more benefits by misappropriating content originally published in a blog and save the time and effort of further investigation by watching the reactions of others on a news aggregator site.

Emerging topics

- Distributed news aggregators provide solutions to the same two problems traditional news publishers have solved so far: News reporting and selection
- Traditional news publishers strike a delicate balance between the interests of the readers / viewers and their professional responsibility. Gillmor has documented that as a result of increasing cost pressure, traditional news producers have already tilted their balance towards the interests of the readers, which seems to favor local and celebrity news.⁸²
- Each of the analyzed news aggregators approach this balance in the selection of news differently: While Digg News' public voting mechanisms favor an agenda reflecting the particular interests like technology news and geek trivia, the editorial board at Global Voices tries to build an open space for promoting unheard "voices", and the editorial board at OhMyNews even tries to mirror the continuous news agenda of the traditional media.
- An important topic for the solution to the news reporting problem is identity as a means of motivation. In all studied news producing networks, the news reporters feel a strong sense of community and identification with the network they contribute to. The embedded case study on the publication of the Blu-Ray code shows that a platform can even be held hostage by its contributors, who enforce the principles they believe in even against the founders of the platform. The contributors to Global Voices and OhMyNews believe in building an alternative source of information, which more closely resembles their perception of the world. The "public good" motivation seems to be a strong unifying force. At the same time, the focus of attention of the contributors is put – following the self-selection principle – more on niche topics which the contributors deem subjectively relevant and underrepresented in the traditional media.
- Motivation and identity, in all aggregators with human contribution, stems from complementing or adjusting the agenda set by the mainstream (read traditional) media. The agendas do not seem to be totally detached, but rather complementary, as many topics brought up in Digg or commented on in the blogosphere relate to content produced by traditional media. The ratings and comments can direct attention to underrepresented topics or reevaluate the opinions expressed by the traditional media.
- It therefore seems highly unlikely that the news production industry can currently rely on the "wisdom of the crowds" in a "make or buy"-decision manner.
- At the same we have a strong hypothesis that the traditional news organization is no longer capable of coping with the complexity of covering the full range of global events and that distributed news gathering is a viable way of going forward.
- Popular voting mechanisms for news selection seem biased towards particular interest topics and prone to errors, although a strong community like Digg is capable of correcting mistakes after some time. Still, popular votes do not seem like a reliable mechanism that traditional news publishers could rely on to solve the news selection problem. It also seems unlikely that the sense of community required would emerge if a traditional news publisher attempted to create one.

⁸² See Gillmor *supra* fn.2 at p.5.

- The question for traditional media platforms is therefore less how to substitute their traditional model with a distributed model, but how to open their existing model to integrate elements of distributed news gathering and citizen journalism into the news production process. Given the experience of the editorial boards, it seems plausible to leverage the power of Citizen Journalism and other forms of distributed news gathering for the solution of the reporting problem without compromising on the quality standards and professional journalistic responsibility.
- It is too early to estimate the market effect on traditional news publishers. So far, it looks like distributed news aggregators are acting not as a substitute but as a complement, which in one way or another references or rejects the agenda and the material published by traditional news publishers, and hence distributed news aggregators cannot (yet) exist as an information space on their own.⁸³

⁸³ Further insights into the interaction between distributed news aggregators and traditional media portals could be derived from the Webmetric method developed by Robert Ackland et. al, see R. Ackland,, M. O'Neil, R. Standish. and M. Buchhorn: VOSON: A Web Services Approach for Facilitating Research into Online Networks, paper presented at the Second International Conference on e-Social Science, 28-30 June 2006, University of Manchester *available at* <http://voson.anu.edu.au/papers/ncess-conf06-full-paper-final.pdf> (last visited 20/11/2007).