

Anders Olof Larsson* and Hallvard Moe

Bots or journalists? News sharing on Twitter

DOI 10.1515/commun-2015-0014

Abstract: Online news sites have become an internet ‘staple’, but we know little of the forces driving the popularity of such sites in relation to what could be understood as the latest iteration of the web – social media services. This research in brief article discusses empirical results regarding the uses of Twitter for news sharing. Specifically, we present a comparative analysis of links emanating from the service at hand to a series of media outlets over a six-month period in two countries; Sweden and Norway. Focusing on linking practices among highly active Twitter accounts, we problematize the assumption that online communication involves two or more humans by directing attention to more or less automated ‘bot’ accounts. In sum, it is suggested that such automated accounts need to be dealt with more explicitly by researchers as well as practitioners interested in the popularity of online news as expressed through social media activity.

Keywords: online news, Twitter, news sharing, Sweden, Norway, bots, online journalism

1 Introduction

How can we understand the sharing of media content online? Such a question is arguably important for media and communication scholars, whether they are interested in the political economy of social networking sites, new forms of audience behavior, economic models for news provision, or the technological affordances of new media. Recent developments within the media industry have pointed towards the necessity to utilize a range of different platforms for promotion and dissemination of media content, as well as for connecting with users. Indeed, online services like Facebook, YouTube, various blogging servi-

*Corresponding Author: Anders Olof Larsson, University of Oslo, Department of Media and Communication, E-mail: a.o.larsson@media.uio.no

Hallvard Moe, University of Bergen, Department of Information Science and Media Studies, E-mail: hallvard.moe@infomedia.uib.no

ces or Twitter are becoming integrated into the flow of news provision. To generate more attention, any publication of a story on a news provider's website would supposedly be accompanied by invitations for involvement from readers, and simultaneously be disseminated and promoted widely through an array of external channels. But exactly to what extent, by whom and in which ways is the content distributed through novel services such as social media?

This research in brief article addresses such questions by presenting results from a study of the uses of Twitter for news sharing. Specifically, we discuss findings from a comparative analysis of links emanating from Twitter to a series of media outlets over a six-month period in two countries, Sweden and Norway. Twitter, a so-called micro blog that allows users to send short messages of up to 140 characters to a network of followers, has been assessed in a number of different journalistic contexts, such as news-gathering (Hermida 2010) and audience interactions (Larsson 2013). Previous research on social media use in our case countries has indeed found that journalists tend to be notably active on Twitter – especially in comparison with other groups of citizens (Larsson and Moe 2012).

Here, we extend this interest by looking at sharing practices. In so doing, we challenge a common assumption in scholarly discussions of online communication: that what we are studying is communication between two or more humans, rather than automated information transactions between so-called 'bots'. A 'bot' is a piece of more or less automated computer software, programmed to mimic the behavior of human internet users. Bots can be understood as a part of a series of wider developments, where computer processes, often collected under the heading of algorithms (e.g., Bucher 2012), take on more visible roles in online communication. On services like Twitter, bots can be utilized to boost statistics regarding the number of followers of a specific user (Messias, Schmidt, Oliveira and Benevenuto 2013) – practices that have been discussed in branches of society like the entertainment industry or in politics (e.g., Null 2013). As such, the presence of bots is being felt in a number of online situations, making a mark on statistics regarding users, followers and the like.

Our aim, then, can be specified as such: To what degree are different kinds of 'bots', journalists and readers sharing news on Twitter? Utilizing state-of-the-art methods for data collection and analysis, this article provides insights into the changing relationships between established news media and social media outlets such as Twitter. In so doing, the results presented here not only enrich our knowledge about social media more generally – they also emphasize the need for communication and media research to pay attention to 'bots' when making sense of online media.

2 Research approach

Our analytical setup follows a ‘most similar cases’-strategy (e.g., Lijphart 1975), where comparison provides context for individual cases, facilitating a more nuanced understanding of the emerging phenomenon at hand. The case countries are Sweden (9.4 million inhabitants) and Norway (4.9 million). Both represent Nordic welfare states (e.g., Hilson 2008; Syvertsen et al. forthcoming), and have similar media policy regulations, including publicly funded broadcasters and press subsidies. The cases also rank high on penetration of information technologies, with around 90 percent of citizens in both countries enjoying internet access at home (Nordicom 2009). As for Twitter use, a 2010 Swedish survey (Facht and Hellingwerf 2011) indicated that about one percent of online Swedes made use of Twitter every day, while nearly four percent of online Norwegians made use of Twitter on a weekly basis (NRK/Ipsos 2011). As such, while Twitter has enjoyed attention in the mass media as well as among researchers, it cannot be considered a widely used service in either of the cases – despite the number of advanced internet users. In an effort to base our analysis on a wide variety of online media actors, we chose to include a series of different types of online news outlets from the two countries: licence fee-funded public service broadcasters (nrk.no, sr.se and svt.se); commercial free-to-air TV broadcasters (tv2.no, tv4.se), the largest nation-wide tabloid newspapers (vg.no, aftonbladet.se), and the largest nation-wide broadsheet newspapers (aftenposten.no, dn.se). The included media were specifically chosen because of their dominant positions in their respective home markets (e.g., Facht 2012; Vaage 2012). Together, they represent the leading online news sites in the two countries, covering tabloid and broadsheet papers, as well as publicly funded and commercial broadcasters.

While research on Twitter is arguably at an early stage, recent years have seen the emergence of various approaches to studying the service. Perhaps the most popular mode of approaching the study of Twitter has been based on the presence of hashtags. The term refers to specific keywords preceded by the # character added to tweets by their respective posters to make their messages searchable in specific contexts, such as elections (Burgess and Bruns 2012; Larsson and Moe 2013), news discussion (Brunns and Burgess 2012), or natural disasters (Brunns and Liang 2012). Given the focus of the study at hand on the sharing of online news, the presence of hyperlinks was employed as a selection criterion for tweets to be included (Moe and Larsson 2013). Utilizing the Your-TwapperKeeper service to gather tweets containing relevant links (e.g., Bruns and Liang 2012; Moe and Larsson 2012) connected to a selection of Swedish and Norwegian media outlets, data collection was initiated on November 1,

Table 1: User categories for classifying top users.

User category	Description
Self-reference	User is presented as an official account operated by the media outlet linked to, or as a journalist working for the media outlet linked to.
Journalist	User is presented as a journalist working for another media outlet than the one linked to.
Citizen	User is presented as not related to any media outlet.
Bot	User is presented as a ‘digest’ of Swedish or Norwegian news media, or as an automated retweeting service for unknown purposes.

2011 and terminated six months later on April 30, 2012, resulting in a data set of 1,124,106 tweets with links sent from a total of 216,089 accounts. With this in mind, while established tools for data collection were utilized, they are nonetheless limited in terms of their “gardenhose” (Lewis, Zamith and Hermida 2013: 37) access to the Twitter API (Moe and Larsson 2012). Yet, as the search terms employed here were not expected to yield results beyond what the employed tool could handle, we should be less wary of these limitations than if we were interested in some international, large-scale phenomenon.

The collected data were subjected to a series of analyses using statistical software (SPSS) and Gawk, a scripting language often employed for the processing of this type of data (Bruns 2011). As our analytical focus here lay on the top users – specifically, the top ten most active user accounts in sending tweets containing links to the media outlets under scrutiny –, the employed tools were helpful in filtering out these high-end users for identification (following the suggestion by Lewis et al. 2013). As such, the top ten accounts relating to each media outlet were noted down, leading to a list of 90 such extremely active accounts, ten for each of the nine websites. The profile page of each account was visited, providing information for classification according to an iterative scheme. Table 1 presents our four distinguished user types.

While the categorizations presented in the table could be seen as rather broad, we suggest they can help function as a first step on the way towards richer insights into the characteristics of high-end Twitter users – like those under study here.

3 Results and discussion

Figure one presents the distribution of the collected tweets and users by the specific media outlet linked to. The black line represents the former, while the grey bars denote the latter.

While considerable differences between the numbers of users and tweets for each media outlet can be observed, Figure 1 suggests that relatively few tweets were sent from relatively few accounts linked to television broadcasters' websites – a trend especially visible in the Swedish case, where svt.se and tv4.se receive rather few links. By comparison, nrk.no, the website of Norwegian public service broadcaster NRK, appears rather popular in this regard. Other than that, the Swedish case presents few surprises; the national leading broadsheet (dn.se) and the leading tabloid (aftenbladet.se) are long established in the on-line realm, while the popularity of linking to online radio (sr.se) can perhaps be explained by “Click and share”-type integration of content streaming services. As for Norway, nrk.no receives the most links, followed by the leading national tabloid (vg.no) and broadsheet (aftenposten.no).

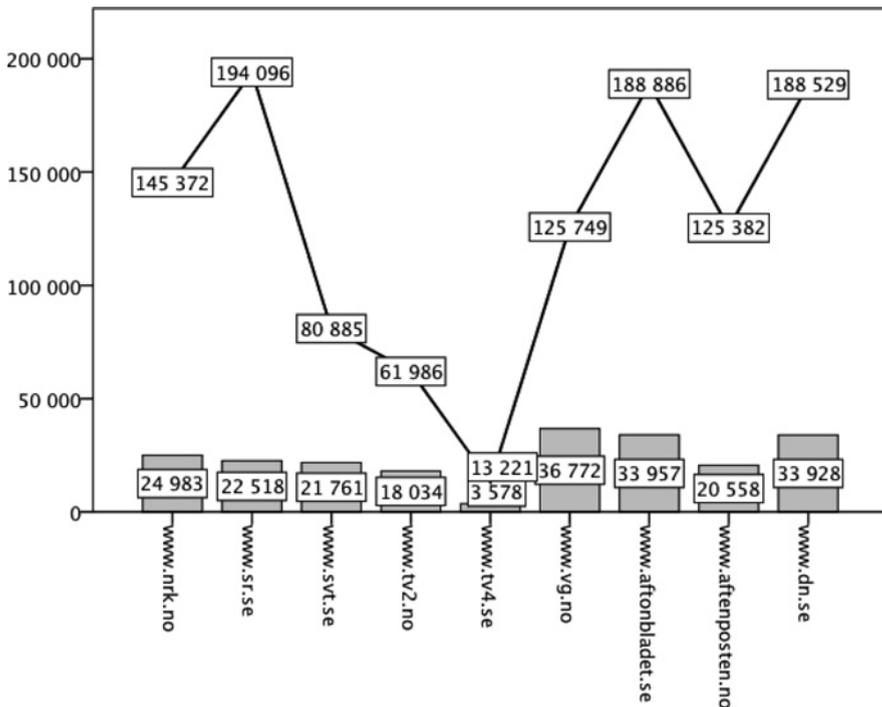


Figure 1: *N* of users (grey bars) and *N* of tweets containing links (black line).

While we need to keep the effect of the different sizes of the countries studied in mind, the results revealed in Figure 1 still provide important insights into the uses scrutinized here.

Figure 1 further suggests striking discrepancies in the ratio between users and tweets. While the users linking to the Swedish public service TV broadcaster (svt.se) on average sent 3.7 links during our 6-month data collection period (similar to the ratios of the commercial television broadcasters TV2 and TV4 as well as the Norwegian tabloid VG), the SVT's sister institution in radio (sr.se) had users sending on average 8.6 links. The Norwegian broadsheet Aftenposten ranks second in this regard, with a ratio of 6.1. Such variations point to different user patterns, or different kinds of users. With these tendencies in mind, we turn to analyzing the users.

Specifically, our focus here is placed on the very top link providers – on those high-end user accounts that appear among the top ten most active for each media outlet studied. Such a demarcation does not allow us to make certain claims about the large number of Twitter users sharing links to media outlets. Rather, what we are aiming for here is to look at the dominant minority, who through their tweeting activities are responsible for a considerable amount of online traffic directed towards specific news web sites.

Figure 2, then, provides the results of the classification process undertaken, as described above.

As previously mentioned, the 90 Twitter accounts categorized in Figure 2 were selected based on their belonging to the “top 10” most active accounts for each media outlet studied. Nevertheless, the degree to which the activity of these high-end users is tangible in the total amount of 1,124,106 collected tweets needs to be assessed. As such, the results show that while the *Bots* (129,605 tweets from 31 accounts) and *Self-reference* (128,962 tweets from 42 accounts) categories account for 11.5% of the total amount of tweets collected each, *Citizens* (61,754 tweets from 13 users) and *Journalists* (2,184 tweets from 4 users) account for 5.5 and 0.1% of the total number of tweets, respectively. Adding these percentages together, we can conclude that while these accounts are indeed rather few, these 90 users' (0.04% of the 216,089 users in total) combined activity represents 322,505 tweets – or almost a third (28.6%) of the total amount of tweets gathered. Furthermore, it is important to note that the two former categories make up the bulk of the activity studied, something which is not very surprising, given the automated character of these types of accounts. Nevertheless, the fact that the *Self-referencing* and *Bot* activities appear to account for such a large portion of the linking to these sites derived from Twitter has consequences for our understanding of news sharing online: As rather few *Bot* and *Self-referencing* accounts are responsible for a consider-

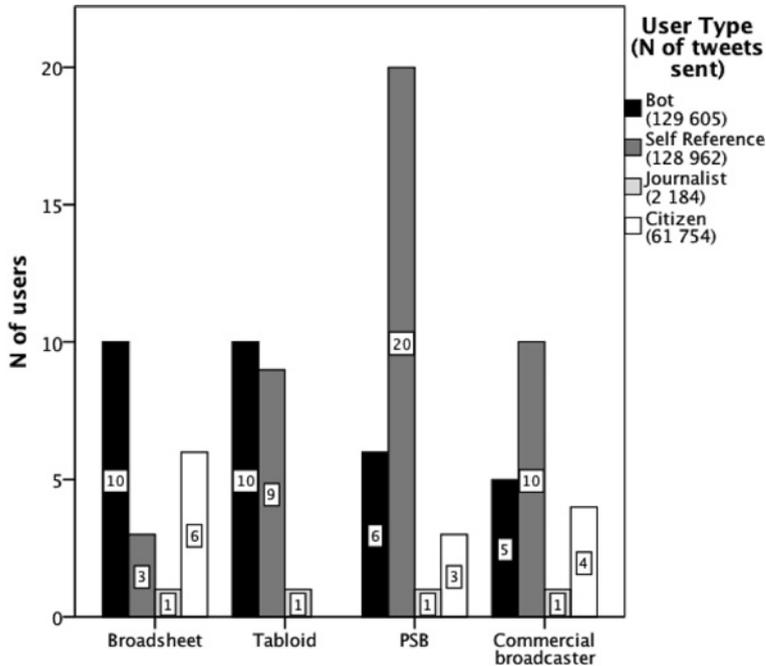


Figure 2: *N* of users classified into groups and *N* of tweets sent by each group.

able share of the total amount of traffic, this calls for further inroads into the activities of more or less automated accounts like these – and the repercussions which the activities of such accounts have on various online rankings.

Figure 2 further details the specifics of how the different user types tend to link to different varieties of media site as previously discussed. While the selection of high-end accounts employed here cannot be considered representative of the whole corpus of users, we can nonetheless distinguish three tendencies among these top users.

First, while no clear trend relating to the category of *Bot* accounts is apparent in Figure 2, these accounts seem to be somewhat more common when links to online newspapers (whether they be broadsheets or tabloids) are distributed. While it is hard to make any certain claims as to why this is, perhaps the longevity and popularity of online newspapers in both case countries has an influence here. As a mainstay of the media diets of many Swedes and Norwegians, these types of news sources enjoy considerable popularity online and could therefore be expected to be a suitable target for links.

Second, the *Self-reference* type accounts appear as more common when looking at public service broadcasters' sites. Relating to the previous category

dealt with, this could stem from the relative longevity of such media outlets in our case countries – a longevity that employees of these organizations apparently are attempting to uphold also in the online environment. Studies have also shown these institutions as willing and able to experiment with emerging online services (e.g., Burns and Brügger 2012; Moe 2013), which could rub off on the practice of Twitter uses as well.

Third, the smaller categories of *Journalists* and *Citizens* appear similarly spread among the categories of media types. As for the latter of these two categories, the account names in the *Citizen* category indicate that these accounts are integrated into the RSS readers of specific individuals – mentions of services like “Feedly” or “Flipboard” are common as parts of these user names, signaling that these are indeed Citizens rather than Bots, but that their active uses of RSS services contribute to a relative flooding of links. As for the former, self-identified *Journalist* accounts appear sparsely in our data. This could be related to the reported tendency for journalists to maintain a somewhat conservative stance towards the internet (Domingo et al. 2008; Larsson 2012; Vobic 2011), but perhaps also to patterns of competition moving online – traffic should be driven to your particular organization’s web site rather than to your competition. The sheer bulk of tweets from other types of users appears to almost drown out the activity of journalists. Indeed, this appears to hold true for the group of *Citizen* users as well. Further insights into the users and uses of links in this regard is needed, efforts that could perhaps take the categorizations suggested here as a starting point.

As has been noted throughout this brief research note, the approach presented is not without its limitations, but does nonetheless provide interesting insights into the flow of links from different types of Twitter users to various media outlets. As more and more media consumption takes place in various online environments, statistics regarding links posted to and from media sites are often used to gauge the popularity of different online actors. The results presented here, then, have indicated that we should be wary not to underestimate the role of high-end Twitter accounts when assessing such rankings. When one further considers that a not insignificant portion of these users appears to be professionally affiliated with the media outlet they are promoting online, the need for further insights into the digital strategies of media outlets becomes apparent.

References

- Bruns, A. 2011. How long is a tweet? Mapping dynamic conversation networks on Twitter using Gawk and Gephi. *Information, Communication & Society*, 15(9), 1–29.
- Bruns, A. & Burgess, J. 2012. Researching news discussion on Twitter. *Journalism Studies*, 13(5–6), 1–14.
- Bruns, A. & Liang, Y. E. 2012. Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4).
- Bucher, T. 2012. *Programmed sociality: A software studies perspective on social networking sites*. PhD dissertation, Faculty of Humanities, University of Oslo.
- Burgess, J. & Bruns, A. 2012. (Not) the Twitter election. *Journalism Practice*, 6(3), 384–402.
- Burns, M. & Brügger, N. (Eds.). 2012. *Histories of public service broadcasters on the web*. New York: Peter Lang.
- Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer J. B. & Vujnovic, M. 2008. Participatory journalism practices in the media and beyond. *Journalism Practice*, 2(3), 326–342.
- Facht, U. 2012. *Mediesverige 2012*. Gothenburg: Nordicom.
- Facht, U. & Hellingwerf, K. 2011. *Nordicom – Sveriges Internetbarometer 2010*. Gothenburg: Nordicom.
- Hermida, A. 2010. Twittering the news – The emergence of ambient journalism. *Journalism Practice*, 4(3), 297–308.
- Hilson, M. 2008. *The Nordic model. Scandinavia since 1945*. London: Reaktion books.
- Larsson, A. O. 2012. Interactivity on Swedish newspaper websites: What kind, how much and why? *Convergence: The International Journal of Research into New Media Technologies*, 18(2), 195–213.
- Larsson, A. O. 2013. Tweeting the Viewer – Use of Twitter in a Talk Show Context. *Journal of Broadcasting & Electronic Media*, 57(2), 135–152.
- Larsson, A. O. & Moe, H. 2012. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, 14(5), 729–747.
- Larsson, A. O., & Moe, H. 2013. Twitter in Politics and Elections – Insights from Scandinavia. In A. Bruns, J. Burgess, K. Weller, C. Puschmann & M. Mahrt (Eds.), *Twitter and Society* (pp. 319–330). New York: Peter Lang.
- Lewis, S. C., Zamith, R. & Hermida, A. 2013. Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52.
- Lijphart, A. 1975. The comparable-cases strategy in comparative research. *Comparative Political Studies*, 8(2), 158–176.
- Messias, J., Schmidt, L., Oliveira, R. & Benevenuto, F. 2013. You followed my bot! Transforming robots into influential users in Twitter. *First Monday*, 18(7).
- Moe, H. 2013. “Public Service Broadcasting and Social Networking Sites: The Norwegian Broadcasting Corporation on Facebook”. *Media International Australia*, 146: 114–122.
- Moe, H. & Larsson, A. O. 2012. Methodological and Ethical Challenges Associated with Large-scale Analyses of Online Political Communication. *Nordicom Review*, 33(1), 117–124.
- Moe, H. & Larsson, A. O. 2013. Untangling a Complex Media System. *Information, Communication & Society*, 16(5), 775–794.

- Nordicom. 2009. *Access to internet at home 1999–2009*. Retrieved March 20, 2013 from <http://www.nordicom.gu.se/?portal=mt&main=showStatTranslate.php&me=1&media=PC%20and%20Internet&type=media&translation>.
- NRK/Ipsos MMI. 2011. *Nettbrukerundersøkelse 2011*. Oslo: NRK.
- Null, C. 2013. Fake Twitter-follower market is adapting, growing, and getting ever cheaper. PC World, July 3, 2013. Retrieved August 27, 2013 from <http://www.pcworld.com/article/2043618/fake-twitter-follower-market-is-adapting-growing-and-getting-ever-cheaper.html>.
- Syvertsen, T. et al. forthcoming. *The media welfare state: Nordic media in the digital era*. Ann Arbor: The University of Michigan Press.
- Vaage, O. F. 2012. *Norsk mediebarometer 2011*. Oslo-Kongsvinger: Statistisk sentralbyrå (Statistics Norway).
- Vobic, I. 2011. Online multimedia news in print media: A lack of vision in Slovenia. *Journalism*, 12(8), 946–962.

Bionotes

Anders Olof Larsson

Anders Olof Larsson is postdoctoral fellow at the Department of Media and Communication, University of Oslo.

Hallvard Moe

Hallvard Moe is professor at the Department of Information Science and Media Studies, University of Bergen.