Conceptualizing and Measuring News Exposure as Network of Users and News Items

Trilling, D.

Publication date
2019

Document Version
Final published version

Published in
Measuring Media Use and Exposure

DAMIAN TRILLING

Conceptualizing and Measuring News Exposure as Network of Users and News Items

1. Introduction

Societally relevant information is no longer only transmitted via few distinct channels that are largely the same for everyone (a limited set of newspapers, TV stations, etc.) but also via social network sites, personalized media, and other channels. On social media, for instance, users read an individual item – and if they like this specific article, they can share it. Each user's click on an article might also serve as an input to an algorithm that – based on such clicks – shows that article to other users of the social network site or the news site in question or recommends another article to read next.

All of these developments are manifestations of what I call the *unbundling of news*: the development in which reading a full newspaper or watching a full TV news show decreases in favor of reading single articles or watching single clips. The trend towards unbundled news is not the same as the trend towards online news. Online, too, one can read an e-paper, which looks exactly as its offline counterpart, or watch an entire newscast via a video platform. Thus, while not all online news is unbundled, we can say that platforms such as social network sites, news aggregators such as Google News, and pay-per-article services like Blendle are built to enable the distribution of individual news items and thus lead to the unbundling of news.

Therefore, nowadays the entity in which news is spread often has become the single news item rather than a collection of items bundled in, for instance, a newspaper issue – very much like in the music industry,

where the ability to download or stream individual tracks has rendered the idea of an album antiquated for some. The solid-state entity of news has disappeared, and news, one might say, has become liquid (on the concept of liquidity, see BAUMAN 2007). For communication science research, this means that we have to re-think how to conceptualize news use. News unbundling makes it necessary to shift perspectives from survey questions like ›Did you watch the news yesterday?‹ towards the individual news item. This brings us to a dilemma. On the one hand, asking the broad questions we used to ask do not allow us to infer much about the content someone is exposed to. Asking ›How often do you read [Newspaper X]?‹ allows us to content-analyze the newspaper in question and infer what content someone is likely to have been exposed to (see also SCHARKOW/BACHL 2017), but asking ›How often do you read news on Facebook?‹ does not allow for such inferences. On the other hand, restricting ourselves to case-studies of individual news items (›Did you read the report on X?‹) makes it hard to abstract from the individual story and to generalize.

In this chapter, I therefore propose a network perspective that takes news unbundling into account and conceptualize news consumption in a way that allows for analyses on different levels of granularity. In doing so, I take mainly a methodological perspective. And, in fact, it is possible to read this chapter as a methodological guide for a more flexible way for storing news exposure data. But the reader can also go a step further and use the proposed methods to investigate new theoretical questions. In particular, I will argue that moving towards a network perspective on news exposure can help adequately measuring and theorizing current forms of news exposure.

2. Related research

Why news unbundling matters

The fact that news for a long time has been distributed in a bundled form of course has good practical reasons: It would be highly impractical to distribute single stories on paper rather than selling a complete newspaper that contains a variety of news items. This form of distribution has had a (beneficial) side-effect: Communication scholars have argued that it facilitates incidental exposure and incidental learning. Even if one is not particularly interested in a specific topic, one might be exposed to it anyway

while browsing the newspaper (e.g., DE WAAL/SCHOENBACH 2008). Similar arguments can be made for television news (e.g., SCHOENBACH/LAUF 2004). The Internet, in contrast, is often referred to as a pull-medium that requires the user to actively click and/or search, which could reduce incidental learning (see, e.g., SCHOENBACH/DE WAAL/LAUF 2005).

However, if we disregard badly lay-outed news sites and low screen resolutions of the early days of the Internet for a moment, the question arises if the difference between incidental exposure in online and offline news is really that strong. Skimming the homepage of a news site with its headlines and teasers should be roughly similar to browsing through a printed paper. After all, as eye-tracking studies show, people beeing offline do not read the paper in a linear fashion but focus on elements like headlines (see the literature review by LECKNER 2012). Accordingly, it has been shown that incidental exposure happens online, too (e.g., KIM/CHEN/GIL DE ZÚÑIGA 2013; LEE 2009; TEWKSBURY/WEAVER/MADDEX 2001).

Furthermore, the classic online news site is not fully unbundled. Continuous updates and a 24/7 news cycle shift the focus from ›yesterday's newspaper‹ to the individual item, but these items are still bundled and presented within one news site. It is therefore not surprising that browsing a news site can enable incidental exposure.

Once the bundle loosens, the mechanisms of incidental exposure change. If the individual item is shared individually on different platforms, incidental exposure via the original news site's homepage does not occur anymore. Instead, new forms of incidental exposure might emerge, for instance when people are exposed on social media to some content they are not interested in or that originates from a source they do not normally read, but which content has been shared by someone in their circle. As Kümpel, Karnowski, and Keyling (2015) write: »[T]he observation of other people's news sharing activities leads to more (incidental) news exposure and, ideally, to confrontation with other opinions and ideas« (p. 1). News unbundling, thus, is not necessarily bad for incidental exposure, but it fundamentally changes the way how we are exposed to news and which actors and which forces affect this process.

News unbundling as challenge for communication theory

As static news has become liquid and news gets unbundled, classic theories of mass communication have to be revisited. In particular, the theoretical

distinction between mass communication and (mediated) personal communication cannot be upheld (PERLOFF 2015). Accordingly, with regard to social-media use, Schmidt (2014) talks about »personal publics« that »are characterised by the communicative mode of ›conversation‹, where the strict separation of sender and receiver is blurred« (p. 8) and where the primarily intended audience is – among other things, through re-sharing of the message – not identical with the de-facto audience that receives a message (see also the work on egocentric publics by WOJCIESZAK/ROJAS 2011). In such an environment, we can see a lot of shades of gray between, for instance, a mass-media mode in which millions of people subscribe to the feed of a news organization and a personal mode in which a user sends a directed message to a friend. News exposure increasingly happens in this gray area, in which individuals or other actors share and re-distribute news items they found somewhere else. What does this mean for communication theory?

Agendas-setting theory, which explains how media outlets set the agendas of other media outlets, the public, and policy makers, is a prominent example of a theory that needs to be readjusted in a media environment that is characterized by blurring boundaries. Like many other theories, agenda-setting theory assumes that its key concept (in this case, the ›agenda‹) can be meaningfully operationalized as a characteristic of a media outlet. For instance, intermedia agenda-setting explicitly addresses the question how ›the agenda‹ of outlet A influences ›the agenda‹ of outlet B. But what if the boundaries of what constitutes an outlet start to blur? Russell Neuman et al. (2014) even go as far as arguing that the question »who sets the media agenda?« has become »ill structured« (p. 211) because traditional and social media interact and resonate (see also DEUZE 2008). Vice versa, the old two-step flow model, which posits that information spreads via so-called opinion leaders (LAZARSFELD 1944), might become relevant again. Also theories of journalistic decision-making like the news value theory of Galtung and Ruge (1965) might get a new role in explaining how news items get (re-)distributed. For instance, Trilling, Tolochko, and Burscher (2017) departed from the notion of newsworthiness to study the ›shareworthiness‹ of articles, which they used to predict the number of shares news articles receive on Facebook and Twitter. Similar studies have been conducted by Kilgo et al. (2016), García-Perdomo, Salaverría, and Kilgo (2018), and Valenzuela, Piña, and Ramírez (2017).

In short, many classical communication science theories have to be modified and ultimately be integrated in a framework that seeks to understand news dissemination and news exposure. Most importantly, we need conceptual models that allow us to model blurring boundaries. For instance, instead of saying that some article A ›belongs‹ to outlet O, we could rather say that A has a *relation with* (namely being published by) outlet O. Crucially, such a conceptualization also allows A to have *other* relations, such as being shared on social network site S.

A first step towards such a framework has been made by Thorson and Wells (2016), who conceptualize news dissemination as ›curated flows‹ and have applied it in a study in which they combine survey data with tracking data collected via a Facebook app (WELLS/THORSON 2015). Their approach is one example of how to tackle an important methodological issue that arises when we want to analyze the described liquidization of news: How can we measure who has been exposed to which news item?

News unbundling as empirical challenge

The example of curated news flows illustrates how the analysis of unbundled news requires new types of data and new analytical approaches (WELLS/THORSON 2015). Self-reported estimates of news exposure have always suffered reliability problems (PRIOR 2009). Nevertheless, even though people have difficulties in providing accurate frequency estimates, in the traditional media landscape they probably can recall which news sources they never use or which ones they use very frequently. The reason lies in the bundling of news: People used to subscribe to one newspaper rather than buying a different one every day, and there were only a handful of news shows on TV to choose from. In other words, people had relatively stable media repertoires or news diets (e.g., HASEBRINK/DOMEYER 2012; HASEBRINK/POPP 2006; TRILLING/SCHOENBACH 2013, 2015). Again, this has little to do with the question whether news consumption happens online or offline, as the online news site market is very concentrated as well (e.g., HINDMAN 2009; TRILLING 2013). Rather, the challenge has everything to do with the unbundling of news.

In an era of unbundled news, people might very well be exposed to news from very different newspapers, potentially without even realizing it. Visiting newspaper websites directly has become only one out of many different ways to access the news, while links on for instance social network

sites cause a substantial share of the traffic to news articles (e.g., TRILLING et al. 2017). This means that people might well be able to say they got news ›from Facebook,‹ but they might have a difficult time saying from where it originated. When we want to study what news people are exposed to (and not only differentiate whether they got it from social media or a news site), it becomes less and less a viable approach to ask them directly in a survey.

One approach would be qualitative observational studies, which can give us a better understanding of how and where people encounter news. However, such studies do not generalize well, suffer from a low ecological validity (after all, who is having a researcher literally looking over their shoulder every time they sit in front of a computer?), and do not allow for continuous observations over a long period over time.

A big potential lies in the analysis of tracking data or log files (e.g., DVIR-GVIRSMAN/TSFATI/MENCHEN-TREVINO 2014; GENTZKOW/SHAPIRO 2011; MENCHEN-TREVINO/KARR 2012; STRIPPEL/EMMER 2015; TEWKSBURY 2003). Server-side solutions (in which data is collected by the host of a web site, e.g., a news website) do not help much to study unbundled news, as one would only have access to logs from one or few cooperating websites. Suitable technical solutions therefore include client-side solutions (data collection directly on the computer of the participants) and proxy servers (which provide users with access to the internet and can store information about what traffic goes through them). Indeed, comparing such data with self-reported internet use, Scharkow (2016) found that people are rather bad at recalling how they used the Internet.

However, current research usually does not take into account the exact content of the sites people visit: They log the URL but do not store the content behind it. As this approach mainly allows us to tell how often people visit which sites but gives only very limited evidence as to which exact news item they encountered where, it needs to be improved to be suitable for analyzing a media environment of unbundled news. Luckily, this is no unsurmountable challenge. A first approach would be that the researchers themselves simply download the content behind each URL later on (automated with a script). However, this will cause problems when content is dynamically created or when it depends on the user being logged into a service. A second approach therefore would be to not only store the URLs but to intercept the whole traffic. While this is more privacy-invasive and therefore not only technically but also legally and ethically more challenging, it is possible to build such a tool (BODÓ et al. 2017).

Network perspectives on communication

In recent years, communication science has witnessed an increase in studies that analyze social or political networks (EVELAND/HUTCHENS/MOREY 2012). Using a variety of measures of size, centrality, density, and so on, network analysis allows researchers to focus on the relationships between their objects of study. These objects, the nodes in the networks, can be anything: actors, issues, news items, users. The edges between them can also represent a variety of relationships: knowing each other, co-occurring together, reading, retweeting, etc.

For instance, network analysis has extensively been used to analyze social-networking services (mainly Twitter), building networks in which networks represent users or messages and edges retweets, follower- or following relationships, mentions, and so on (for a typical example, see CONOVER et al. 2012). Relatedly, such relationships have been used to study information cascades, i.e., the spread of information on social network sites (e.g., BHATTACHARYA/RAM 2012; CHENG et al. 2014; FRIGGERI et al. 2014). Nevertheless, in other areas, network approaches are not as popular yet, even though the object of study would lend itself excellently to such an approach. Reviewing the literature on news sharing, Kümpel, Karnowski, and Keyling (2015) found that comparatively few studies used network approaches and that »[r]esearch on news sharing networks [as opposed to news sharing research on, e.g., actors or content; DT] is highly focused on technological aspects and thus dominated by scholars from the computer and information sciences« (p. 5).

It is important to note that network approaches can not only be useful for analyzing data structures that have an obvious network characteristic, such as follow- or retweet-networks. More in general, network perspectives can also be used to adapt communication science theories to the demands and characteristics of today's media landscape. For instance, agenda-setting theory has been adapted in such a way. The *network agenda-setting* model explicitly models the similarity of issues as a graph, whereas traditional agenda-setting approaches would treat issues as discrete entities that are simply present or not, without allowing for relationships or overlap (e.g., GUO 2013; GUO/VARGO 2015; VARGO/GUO 2017). Gatekeeping theory has been modified, too: Barzilai-Nahon (2008) proposed a theory of network gatekeeping, which explicitly includes relationships between gatekeepers. Advocating for such an approach in research on news exposure, Thorson

and Wells (2016) write: »[W]hat if we locate the individual as the unit of analysis – as traditional media effects research has – but also build for each individual a network of communication links one step out – as emerging research techniques allow us to do?« (p. 4).

The vast majority of the network approaches in communication science, however, is limited to networks that are limited to *one* type of nodes and *one* type of edges. For instance, in a follower-network, each node represents a user and each edge represents a following-relationship. Or in a topic network, each node is a topic, and each edge might represent their similarity or their frequency of co-occurrence. As I will show in the following section, we can model news exposure as a network in which we allow *multiple* types of nodes and *multiple* type of edges. Rather than analyzing networks of users and networks of items separately, I propose to integrate them into one graph.

Towards a new conceptualization of news exposure

While reviewing related research, I discussed three main themes related to the unbundling of news: theoretical challenges, empirical challenges, and the move towards network perspectives on communication. In this section, I will integrate them and propose a model that conceptualizes news exposure as a network of users and news items. As outlined above, this model offers a methodological improvement, as it allows to operationalize and model news exposure in an ecosystem characterized by unbundled news.

Moving from a tabular data model, in which rows correspond to observations and columns to variables that were measured, to a network model of nodes and edges does not only offer additional flexibility, but it also allows to answer new theoretical questions. Let us again consider the example of agenda-setting. In recent years, studies have addressed the question how social media like Twitter influence the media agenda (e.g., CONWAY/KENSKI/WANG 2015). However, such studies had to treat Twitter as ›just another medium‹, forcing it into the same sort of category as, for example, a newspaper. But this would lump together the Twitter account of a newspaper with the one of a random citizen while not recognizing its relationship to the newspaper's website. If, instead, we see news items as nodes in a network, we can connect them in various ways. This makes it possible to move beyond ›the agenda of newspaper X‹ and lets us answer theoretical questions like: How does issue salience from the media agenda influence the public agenda *via different pathways*?

News exposure as network of users and news items

A first network (graph) could look as follows. The nodes (also known as vertices) represent users and news items. We have thus two different types of nodes: user nodes and item nodes. They are connected by edges (also known as arcs or relationships), for instance to indicate that a user has read a news item. But one could also use edges to indicate relationships between news items, for example them being similar. Having such multiple types of edges, we call our graph a *multigraph*. Additionally, nodes can be tagged with labels, and both nodes and edges can have properties. Therefore, we can say that we employ a *labeled property graph model* (see ROBINSON/WEBBER/EIFREM 2015).

Nodes. The most fundamental nodes in our network are *user nodes* and *news item nodes*. These nodes have certain *properties*: For instance, a user node might have the properties age, gender, education, but also political interest, political orientation, trust in news media, and so on. A news item node represents a single news item, be it a written article, a video clip, or any other format. News items have properties as well: most notably, their full text, but also metadata such as publishing date, where it was published, what its original source was, and so on. Using techniques of automated content analysis (BOUMANS/TRILLING 2016; GRIMMER/STEWART 2013), we could also extract new properties from the full text, for instance, topics or frames. This process is also known as *feature engineering*.

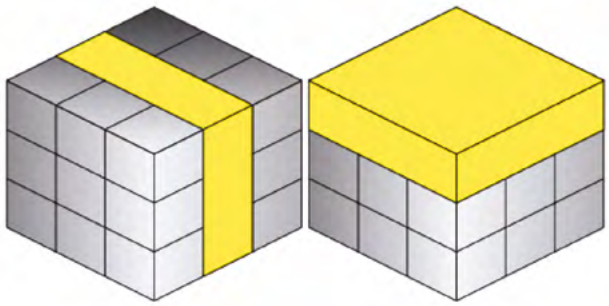
Edges. Because we are interested in news exposure, the most important edges are ›has_read‹ edges. They connect user nodes with item nodes and signify that a user has read a news item. Optionally, such an edge might have properties as well, for instance a timestamp that indicates when the item has been read or how it has been accessed specifically. Using properties, we can also store information on whether the article has been accessed by directly surfing to a news site or via a link on social media. Next to ›has_read‹ edges, we could think of other edge types as well. Most notably, we might want to be able to include information on the relationships between news item nodes. For instance, given that many news items originate from the same press agency releases (BOUMANS 2016; WELBERS et al. 2018), we could connect such items with an ›is_similar‹ edge, which we could determine, for instance, using the cosine similarity. In principle, we could even go further and add edges between persons who know each other, talked to each other, follow each other on social media, and so on. And if we are in-

terested not only in news consumption but also in news sharing, we could introduce edges to that effect in our model, too (see also the suggestions by THORSON/WELLS 2016).

Aggregations and analyses. A main characteristic of the proposed network model is that it avoids premature aggregation of data. Unlike in tabular datasets, we do not have to decide in advance on what constitutes a ›case‹ (i.e., a row) and what constitutes a ›variable‹ (i.e., a column). Instead, we can retrieve any subset of the graph at analysis time. To give an example: We could retrieve all nodes that are connected with a ›has_read‹ edge where the property ›accessed_via‹ equals ›Facebook‹. We could then aggregate the data by the ›topic‹ properties of the item nodes and create a table in which each user is a row and where the columns contain the user properties (like sociodemographics) and the number of articles they read per topic. But we could as well turn it around and create a table in which news items are represented as rows and the aggregated number of times they have been read is one of the columns. If we have connected similar articles with edges, we could also think of collapsing all nearly-identical articles, or aggregating all articles within one outlet, or of many other possibilities, which in the end could lead to a conventional tabular dataset that can be analyzed using familiar techniques like regression analysis. We see that storing our information as a labeled property graph gives us huge flexibility when it comes to answering very different research questions and conducting a huge variety of analyses. We can think of the creation of such datasets as taking a slice out of a much more complex dataset – but rather than doing this in advance, we try to store the information on an as fine-grained level as possible (Figure 1). Of course, we could also refrain from deriving a tabular dataset at all and perform a network analysis on the graph, a subgraph, or any derived graph. For instance, we could derive a graph in which we introduce ›news event‹ nodes, which connect news items that are about the same event and thus allows us to abstract from the individual item.

Summarizing the model. I proposed to conceptualize each news item as an object with a number of properties (like, e.g., news values, origin, topic, etc.). Based on the properties of the item and those of all actors involved (news organizations, anyone sharing an article along the way, the final recipient), it can be predicted if someone is exposed and via which route. Thus, I propose an innovative way of analyzing news exposure with a framework originating in graph theory and used in so-called graph da-

FIGURE 1
Different slices from the same multi-dimensional dataset

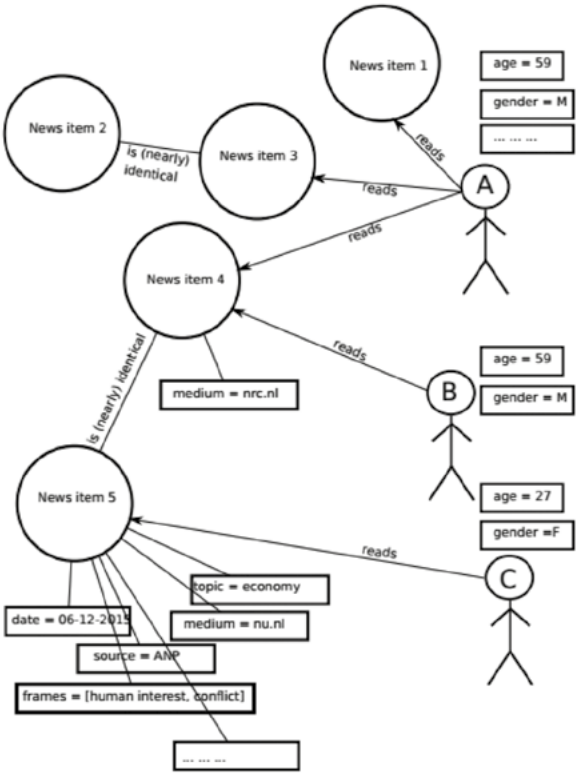


tabases (see ROBINSON/WEBBER/EIFREM 2015). Such a database does not only contain information on cases (which we call nodes) like in common survey or content analysis datasets; it also contains various relationships between these nodes. In our case, the main types of nodes are news items and users; the relationships could be labeled »has_read« and »is_similar«, for instance.

Figure 2 depicts a simplified illustration. News item 4 is read by both A and B, which we could predict with a regression model based on the properties of item 4 (like its topic etc.), A (like age, gender, interests, ...) and B. The crucial improvement of this conceptualization is that we see that C de facto receives the same information: C reads item 5, which – using measures like the cosine similarity – can be identified to contain the same information as item 4. On top, we can estimate which properties predict via which channel the information is received.

Such a conceptualization gives us a huge flexibility in the analyses we can conduct. It allows to predict which item a given person is exposed to – irrespective of the channel. For instance, if a news agency item is published on different sites, the model can account for this and treat them as equal – because their nodes are connected. However, if one is interested in the channel instead, one could also predict the channel. If the news items are time-stamped, one can also investigate how information spreads in the news environment. The possibilities are countless.

FIGURE 2
Graph-based conceptualization of news use



To avoid redundancy, only examples of attached properties are shown, mainly for the bottom node.

Implementation

The model I developed above is a general conceptualization that can be filled in and adapted depending on research interests and the available data. It does not prescribe any specific properties that nodes and edges need to have, and it is open to adding different types of nodes and edges. Let us consider an example: A research team wants to know in how far

the reception of news depends on news values. They have tracking data that allow them to know which news items are read by whom. If they define each news item as a node with properties representing the presence or absence of a news value, and if each reader is another node connected by an edge to the corresponding news item, then they can, for instance, calculate various centrality measures and try to predict them with the associated news values.

The implementation of the model I presented in this chapter does not depend on any specific hard- or software. In this section, I will provide an illustration of how the model can be put into practice. Rather than discussing it along the steps in the research process, starting with data collection and ending with analysis and presentation, I take a different approach and first describe the storage of the data. In the network conceptualization presented here, the question how to store the data is a very fundamental one and, in fact, guides the way data is collected and determines which subsequent analyses are possible (see also GÜNTHER/TRILLING/VAN DE VELDE 2018).

Graph databases. Given that social scientists are often used to working with tabular data, it is worth discussing how to deal with the data structure proposed by the model. Of course, network data can be represented by tables as well (namely by node lists and edge lists), but in our case we want to store much more: We want to be able to distinguish between different types of nodes and edges, and we want to attach properties to them. If one of the properties is the full text of the news item, we quickly reach very large file sizes, and as the number of properties increases, we get very complex data models. Thus, typical programs for network analysis like Pajek or Gephi can help us analyze extracted subsets of the data later on, but their native data formats are not the way to go for storing our data. In fact, what we want to have is some database that is suitable for storing large amounts of news items. But in addition, it should be able to store *relations*. Luckily, such databases exist. Graph databases are specifically designed to store and query labeled property graphs (for an introduction on graph databases in general and the popular Neo4j database in particular, see ROBINSON/WEBBER/EIFREM 2015). In particular, they allow to easily select subgraphs (i.e., subsets of the nodes and edges in a graph) that conform to a specific pattern.

Data for the properties of the nodes. For the user nodes in the network, sociodemographic variables can be measured using traditional surveys. Depending on the research interest, variables as political interest, political

orientation, and so on can be used as additional properties of the nodes. Each participant is thus represented by one node.

Getting news item nodes requires more effort. As discussed above, collecting some type of tracking data that logs the survey participants' online behavior is the most obvious mode of data collection. Each news item that has been read at least once by any of the participants constitutes a node. Getting their properties requires some additional steps. First of all, the full HTML source needs to be acquired, either by downloading the news items, thus based on the URLs the tracking software records, or by having it recorded directly by the tracking software. After parsing the content (i.e., separating the article itself from associated meta data and from irrelevant so-called boilerplate content such as navigation elements; see also GÜNTHER/SCHARKOW 2014), the text can be analyzed using the wide variety of automated content analysis techniques that are available to study digital journalism (BOUMANS/TRILLING 2016; GRIMMER/STEWART 2013). For instance, to each node, one could attach properties like the topic, the author, but also variables of interest like the presence of specific frames. It is important to note that while some of the properties of these nodes have to be stored during data collection (e.g., metadata like URL or publishing date), others can be added afterwards as part of the *feature engineering*, as long as the full text is stored as one of the properties. For instance, based on the full text, features like frames (BURSCHER et al. 2014) or topic (BURSCHER/VLIEGENTHART/DE VREESE 2015; SCHARKOW 2011) can be inferred using Supervised Machine Learning later on.

Connecting the nodes. Connecting the user nodes to the news item nodes is trivial – after all, the tracking software records who has read which article, which enables us to add ›has_read‹ edges. It depends on the exact setup of the data collection software which properties we can add to this edge. A straightforward example would be the so-called ›Referrer-URL‹ that indicates where the user came from. To assess whether two nodes contain essentially the same information (and to connect them with an ›is_similar‹ edge), one can compare them using similarity measures like the cosine distance or the Levenshtein distance. For example, this has been done successfully to identify content overlap between press agency copy and newspaper articles (BOUMANS 2016; WELBERS et al. 2018). If the value of such a similarity measure is above a given threshold, an edge can be drawn between the nodes to indicate their similarity.

3. Conclusion

In this chapter, I have proposed to conceptualize and measure (online) news exposure as a network of users and news items. I argued that news gets unbundled, which makes it less and less useful to ask people how often they use a specific news outlet. To deal with this challenge, it seems necessary to shift the focus towards exposure to an individual news item, recognizing that people can access this item via very different channels. Rather than focusing only on the channel (e.g., asking whether people get their news from Facebook, Twitter, or directly from a news site) or only on the brand (e.g., asking whether they get their news from *nu.nl*, *telegraaf.nl*, or *nrc.nl*), the proposed model allows to integrate both perspectives by seeing channel and brand as properties of edges and nodes, while putting the news item itself in a central role.

In doing so, the proposed approach is in line with the increasing role of network approaches to study communication. While this adds a layer of complexity to news exposure research, its inherent flexibility makes it suitable for studying information flows in complex media landscapes.

What I see as one of the central advantages of conceptualizing and measuring news exposure as a network of users and news items, however, is also one of its disadvantages: By measuring and storing information and relationships on an extremely fine-grained level, it creates complex data structures and huge amounts of data. While this opens up many possibilities for answering innovative research questions based on the data, it also means that researchers have to carefully think about good ways to create meaningful aggregations and to abstract from the individual news item.

Empirically, this conceptualization can be implemented by combining survey data with tracking data in a graph database. In particular, by combining tracking data and survey data to gather information about the nodes in this network, it is possible to create a database of news exposure that allows answering questions that could not be answered before, because it integrates information about users, news items, and exposure to these items, and relationships between the items in one single database. For instance, if we – translating and extending old ideas about a two-step flow of communication into the modern media environment – are interested in the question *via which route* a given piece of information reaches a recipient *and how this can be predicted*, then a network approach and corresponding measures and algorithms (e.g., to find the shortest path between nodes) are necessary.

The proposed approach is one of the first to propose a framework for conceptualizing and measuring news exposure that accounts for the unbundling of news. As such, it still has to prove its usefulness. Future work has to look into possible extensions, for example to include new types of nodes and followers to make it suitable for including active forms (like sharing) next to passive forms (like reading). And, of course, it needs to be filled with empirical data from surveys and tracking tools. Which is what we are working on right now and what hopefully many others will consider doing as well.

References

BARZILAI-NAHON, K.: Toward a theory of network gatekeeping: A framework for exploring information control. In: *Journal of the American Society for Information Science and Technology*, *59*(9), 2008, pp. 1493-1512. doi: 10.1002/asi.20857

BAUMAN, Z.: *Liquid times. Living in an age of uncertainty*. Cambridge, UK [Polity] 2007

BHATTACHARYA, D.; RAM, S.: Sharing news articles using 140 characters: A diffusion analysis on Twitter. In: *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, 2012, pp. 966-971. doi: 10.1109/ASONAM.2012.170

BODÓ, B.; HELBERGER, N.; IRION, K.; BORGESIUS ZUIDERVEEN, F. J.; MOLLER, J.; VAN DER VELDE, B.; DE VREESE, C. H.: Tackling the algorithmic control crisis – the technical, legal, and ethical challenges of research into algorithmic agents. In: *Yale Journal of Law & Technology*, *19*, 2017, p. 133 - 180

BOUMANS, J. W.: *Outsourcing the news? An empirical assessment of the role of sources and news agencies in the contemporary news landscape* (PhD thesis, University of Amsterdam). 2016. http://hdl.handle.net/11245/1.532941

BOUMANS, J. W.; TRILLING, D.: Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. In: *Digital Journalism*, *4*(1), 2016, pp. 8-23. doi: 10.1080/21670811.2015.1096598

BURSCHER, B.; ODIJK, D.; VLIEGENTHART, R.; DE RIJKE, M.; DE VREESE, C. H.: Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis.

In: *Communication Methods and Measures*, *8*(3), 2014, pp. 190 - 206. doi: 10.1080/19312458.2014.937527

BURSCHER, B.; VLIEGENTHART, R.; DE VREESE, C. H.: Using supervised machine learning to code policy issues: Can classifiers generalize across contexts? In: *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 2015, pp. 122 - 131. doi: 10.1177/0002716215569441

CHENG, J.; ADAMIC, L.; DOW, P. A.; KLEINBERG, J. M.; LESKOVEC, J.: Can cascades be predicted? In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY [ACM Press] 2014, pp. 925 - 936. doi: 10.1145/2566486.2567997

CONOVER, M. D.; GONÇALVES, B.; FLAMMINI, A.; MENCZER, F.: Partisan asymmetries in online political activity. In: *EPJ Data Science*, *1*(6), 2012, pp. 1 - 19. doi: 10.1140/epjds6

CONWAY, B. A.; KENSKI, K.; WANG, D.: The Rise of Twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary. In: *Journal of Computer-Mediated Communication*, *20*(4), 2015, pp. 363 - 380. doi: 10.1111/jcc4.12124

DEUZE, M.: The changing context of news work: Liquid journalism and monitorial citizenship. In: *International Journal of Communication*, *2*, 2008, pp. 848 - 865

DE WAAL, E.; SCHOENBACH, K.: Presentation style and beyond: How print newspapers and online news expand awareness of public affairs issues. In: *Mass Communication and Society*, *11*(2), 2008, pp. 161 - 176. doi: 10.1080/15205430701668113

DVIR-GVIRSMAN, S.; TSFATI, Y.; MENCHEN-TREVINO, E.: The extent and nature of ideological selective exposure online: Combining survey responses with actual web log data from the 2013 Israeli Elections. In: *New Media & Society,* (4), 2014. doi: 10.1177/1461444814549041

EVELAND, W. P.; HUTCHENS, M. J.; MOREY, A. C.: Social networks and political knowledge. In: SEMETKO, H. A.; SCAMMELL, M. (Eds.): *The Sage handbook of political communication*. London [Sage] 2012, pp. 241 - 252. doi: 10.4135/9781446201015.n20

FRIGGERI, A.; ADAMIC, L.; ECKLES, D.; CHENG, J.: Rumor cascades. In: *Eighth International AAAI Conference on Weblogs and Social media.* Ann Arbor, MA [AAAI Press] 2014. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8122/8110

GALTUNG, J.; RUGE, M. H.: The structure of foreign news. In: *Journal of Peace Research*, *2*(1), 1965, pp. 64 - 91. http:www.jstor.org/stable/423011

GARCÍA-PERDOMO, V.; SALAVERRÍA, R.; KILGO, D. K.; HARLOW, S.: To Share or Not to Share. In: *Journalism Studies*, *19*(8), 2018, pp. 1180 - 1201. http://doi.org/10.1080/1461670X.2016.1265896

GENTZKOW, M.; SHAPIRO, J. M.: Ideological segregation online and offline. In: *The Quarterly Journal of Economics*, *126*(4), 2011, pp. 1799 - 1839. doi: 10.1093/qje/qjr044

GRIMMER, J.; STEWART, B. M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. In: *Political Analysis*, *21*(3), 2013, pp. 267 - 297. doi: 10.1093/pan/mps028

GÜNTHER, E.; TRILLING, D.; VAN DE VELDE, R.N.: But how do we store it? (Big) data architecture in the social-scientific research process. In: STUETZER, C.M.; WELKER, M.; EGGER, M. (Eds.): *Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications*. Cologne [Herbert von Halem] 2018, pp. 161 - 187

GÜNTHER, E.; SCHARKOW, M.: Automatisierte Datenbereinigung bei Inhalts- und Linkanalysen von Online-Nachrichten. In: SOMMER, K.; WETTSTEIN, M.; WIRTH, W.; MATTHES, J. (Eds.): *Automatisierung in der Inhaltsanalyse*. Cologne [Herbert von Halem] 2014, pp. 111 - 126

GUO, L.: Toward the third level of agenda setting theory: A network agenda setting model. In: JOHNSON, T. (Ed.): *Agenda setting in a 2.0 world: New agendas in communication*. New York, NY [Routledge] 2013, pp. 112 - 133

GUO, L.; VARGO, C.: The power of message networks: A Big-Data analysis of the network agenda setting model and issue ownership. In: *Mass Communication and Society*, *18*(5), 2015, pp. 557 - 576. doi: 10.1080/15205436.2015.1045300

HASEBRINK, U.; DOMEYER, H.: Media repertoires as patterns of behaviour and as meaningful practices: A multimethod approach to media use in converging media environments. In: *Participations*, *9*(2), 2012, pp. 757-779. http://www.participations.org/Volume%209/Issue%202/40%20Hasebrink%20Domeyer.pdf

HASEBRINK, U.; POPP, J.: Media repertoires as a result of selective media use. A conceptual approach to the analysis of patterns of exposure. In: *Communications*, *31*(3), 2006, pp. 369 - 387. doi: 10.1515/COMMUN.2006.023

HINDMAN, M.: *The myth of digital democracy*. Princeton, Oxford [Princeton University Press] 2009

KILGO, D. K.; HARLOW, S.; GARCÍA-PERDOMO, V.; SALAVERRÍA, R.: A new sensation? An international exploration of sensationalism and social media recommendations in online news publications. In: *Journalism*, 2016. doi: 10.1177/1464884916683549

KIM, Y.; CHEN, H. T.; GIL DE ZÚÑIGA, H.: Stumbling upon news on the Internet: Effects of incidental news exposure and relative entertainment use on political engagement. In: *Computers in Human Behavior*, *29*(6), 2013, pp. 2607 - 2614. http://dx.doi.org/10.1016/j.chb.2013.06.005

KÜMPEL, A. S.; KARNOWSKI, V.; KEYLING, T.: News sharing in social media: A review of current research on news sharing users, content, and networks. In: *Social Media + Society*, *1*(2), 2015. doi: 10.1177/2056305115610141

LAZARSFELD, P. F.: The election is over. In: *The Public Opinion Quarterly*, *8*(3), 1944, pp. 317-330. https://www.jstor.org/stable/pdf/2745288.pdf

LECKNER, S.: Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective. In: *Visual Communication*, *11*(2), 2012, pp. 163 - 184. doi: 10.1177/1470357211434029

LEE, J. K.: *Incidental exposure to news: Limiting fragmentation in the new media environment*. PhD dissertation, University of Texas at Austin 2009

MENCHEN-TREVINO, E.; KARR, C.: Researching real-world web use with Roxy: Collecting observational web data with informed consent. In: *Journal of Information Technology & Politics*, *9*(3), 2012, pp. 254 - 268. doi: 10.1080/19331681.2012.664966

PERLOFF, R. M.: Mass communication research at the crossroads: Definitional issues and theoretical directions for mass and political communication scholarship in an age of online media. In: *Mass Communication and Society*, *18*(5), 2015, pp. 531 - 556. doi: 10.1080/15205436.2014.946997

PRIOR, M.: The immensely inflated news audience: Assessing bias in self-reported news exposure. In: *Public Opinion Quarterly*, *73*(1), 2009, pp. 130-143. doi: 10.1093/poq/nfp002

ROBINSON, I.; WEBBER, J.; EIFREM, E.: *Graph databases* (2nd ed.). Sebastopol, CA [O'Reilly] 2015

RUSSELL NEUMAN, W.; GUGGENHEIM, L.; MO JANG, S.; BAE, S. Y.: The dynamics of public attention: Agenda-setting theory meets Big Data. In: *Journal of Communication*, *64*(2), 2014, pp. 193 - 214. doi: 10.1111/jcom.12088

SCHARKOW, M.: Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. In: *Quality & Quantity*, *47*(2), 2011, pp. 761-773. doi: 10.1007/s11135-011-9545-7

SCHARKOW, M.: The accuracy of self-reported Internet use – A validation study using client log data. In: *Communication Methods and Measures*, *10*(1), 2016, pp. 13-27. doi: 10.1080/19312458.2015.1118446

SCHARKOW, M.; BACHL, M.: How measurement error in content analysis and self-reported media use leads to minimal media effect findings in linkage analyses: A simulation study. In: *Political Communication*, *34*, 2017, pp. 323-343. doi: 10.1080/10584609.2016.1235640

SCHMIDT, J.-H.: Twitter and the rise of personal publics. In: ELLER, K.; BRUNS, A.; BURGESS, J.; MAHRT, M.; PUSCHMANN, C. (Eds.): *Twitter and society*. New York, NY [Lang] 2014, pp. 3-14

SCHOENBACH, K.; DE WAAL, E.; LAUF, E.: Research Note: Online and print newspapers: Their impact on the extent of the perceived public agenda. In: *European Journal of Communication*, *20*(2), 2005, pp. 245-258. doi: 10.1177/0267323105052300

SCHOENBACH, K.; LAUF, E.: Another look at the ›trap‹ effect of television – and beyond. In: *International Journal of Public Opinion Research*, *16*(2), 2004, pp. 169-182. doi: 10.1093/ijpor/16.2.169

STRIPPEL, C.; EMMER, M.: Proxy-logfile-Analyse: Möglichkeiten und Grenzen der automatisierten Messung individueller Online-Nutzung. In: HAHN, O.; HOHLFELD, R.; KNIEPER, T. (Eds.): *Digitale Öffentlichkeiten*. Konstanz [UVK] 2015, pp. 85-103

TEWKSBURY, D.: What do Americans really want to know? Tracking the behavior of news readers on the Internet. In: *Journal of Communication*, *53*(4), 2003, pp. 694-710. doi: 10.1093/joc/53.4.694

TEWKSBURY, D.; WEAVER, A. J.; MADDEX, B. D.: Accidentally informed: Incidental news exposure on the World Wide Web. In: *Journalism & Mass Communication Quarterly*, *78*(3), 2001, pp. 533-554. doi: 10.1177/107769900107800309

THORSON, K.; WELLS, C.: Curated flows: A framework for mapping media exposure in the digital age. In: *Communication Theory*, *26*(3), 2016, pp. 309-328. doi: 10.1111/comt.12087

TRILLING, D.: *Following the news: Patterns of online and offline news consumption.* PhD dissertation, University of Amsterdam, 2013. http://hdl.handle.net/11245/1.394551

TRILLING, D.; MÖLLER, J.; HELBERGER, N.; DE VREESE, C. H.: From one-size-fits-all to tailor-made distribution channels: New divides? In: *Etmaal van de communicatiewetenschap.* Tilburg, Netherlands, 2017

TRILLING, D.; SCHOENBACH, K.: Patterns of news consumption in Austria: How fragmented are they? In: *International Journal of Communication*, *7*, 2013, pp. 929-953. http://ijoc.org/index.php/ijoc/article/download/1769/894

TRILLING, D.; SCHOENBACH, K.: Investigating people's news diets: How online news users use offline news. In: *Communications: The European Journal of Communication Research*, *40*(1), 2015, pp. 67 - 91. doi: 10.1515/commun-2014-0028

TRILLING, D.; TOLOCHKO, P.; BURSCHER, B.: From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. In: *Journalism & Mass Communication Quarterly, 94*, 2017, pp. 38-60. doi: 10.1177/1077699016654682

VALENZUELA, S.; PIÑA, M.; RAMÍREZ, J.: Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. In: *Journal of Communication*, *67*(5), 2017, pp. 803 - 826. doi: 10.1111/jcom.12325

VARGO, C. J.; GUO, L.: Networks, Big Data, and intermedia agenda setting: An analysis of traditional, partisan, and emerging online U.S. news. In: *Journalism & Mass Communication Quarterly*, *94*(4), 2017, pp. 1031-1055. Doi: 10.1177/1077699016679976

WELBERS, K.; VAN ATTEVELDT, W.; KLEINNIJENHUIS, J.; RUIGROK, N.: A Gatekeeper among Gatekeepers. In: *Journalism Studies*, *19*(3), 2018, pp. 315 - 333. http://doi.org/10.1080/1461670X.2016.1190663

WELLS, C.; THORSON, K.: Combining Big Data and survey techniques to model effects of political content flows in Facebook. In: *Social Science Computer Review*, 2015, pp. 1-20. doi: 10.1177/0894439315609528

WOJCIESZAK, M.; ROJAS, H.: Correlates of party, ideology and issue based extremity in an era of egocentric publics. In: *The International Journal of Press/Politics*, *16*(4), 2011, pp. 488 - 507. doi: 10.1177/1940161211418226