



UvA-DARE (Digital Academic Repository)

What happens on the fringes, stays on the fringes?

Information flows in the contemporary media system

Simon, M.

Publication date

2024

[Link to publication](#)

Citation for published version (APA):

Simon, M. (2024). *What happens on the fringes, stays on the fringes? Information flows in the contemporary media system*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 1

Introduction

“From the murky waters of the swamps near a place called Lerna, the hydra would rise up and terrorize the countryside. A monstrous serpent with nine heads, the hydra attacked with poisonous venom. [...] With his club, Hercules attacked the many heads of the hydra, but as soon as he smashed one head, two more would burst forth in its place!”

*(The Second Labour of Hercules,
Perseus digital library)*

Much like the malevolent hydra lurking in the murky swamps of Lerna, the low-moderated dark platforms of our digital age fester on the fringes of the web where political outsiders can operate and extremist narratives can free-flow under the radar. Dark platforms are believed to have the discursive power to influence, not only those who deliberately seek out the shadowy fringes of the web, but also those who are unwittingly exposed to the poisonous narratives that slip through the bounds of the fringes and influence public debate, threatening democracy. Although these concerns are based on predominantly anecdotal evidence, recent scholarly evidence indicated the dark platforms are closely linked to anti-establishment and extremist sentiments and movements (Schulze et al., 2022; Urman & Katz, 2020) that have the potential to stimulate real-world action, including protests and riots (Bakker et al., 2020).

Dark platforms such as Telegram and the notorious image board 4chan/pol have been associated with countless issues and worrying headlines in the recent years with respect to conspiracy thinking (Tuters & Hagen, 2020), extremism (Walther & McCoy, 2021), the spread of misinformation (Zeng & Schäfer, 2021), hate speech (Zannettou, Caulfield, Setzer, et al., 2019), and terrorism (Prucha, 2016). While these dark platforms have been linked to disconcerting issues in the past, it is important to note that they have also played a role in providing safe communication channels for oppressed and marginalized groups in society (Schectman, 2019; Urman et al., 2021).

However, given the predominantly worrying signals about the dangers of dark platforms, several scholars and opinion leaders share the concern that the more secluded corners of the web may function as laboratories of extremist narratives that can seep into and influence the mainstream (Lewandowsky, Smilie, et al., 2020). Such influence is associated with increasing polarization in society that erodes democracy.

Online spaces where political talk occurs (e.g., Reddit, Twitter¹) offer insights about public opinion and provide information that may influence news coverage (Anstead & O’Loughlin,

¹Renamed to X by Elon Musk in 2023. The moderation policies of X are believed to be less strict compared to Twitter, allowing extremist voices to speak up.

2015). But what do we know about the influence of dark platforms from an agenda-setting perspective in the context of socially and politically relevant information in the contemporary media system? Do they truly assert a significant level of control over what we see (i.e., what is being covered by the mainstream media) and what we believe? How does the influence of dark platforms compare to the influence of more mainstream media agendas with respect to the public debate? Although some initial evidence suggest that information from dark platforms can seep into the mainstream and stimulate real-world action (Bakker et al., 2020), little is known about the extent to which information flows from dark platforms shape public debate.

Addressing the need for more “systematic analyses of information flows” (Jungherr, Posegga, & An, 2019, p. 404), the aim of this dissertation is to shed light on the worrying but so far understudied influence of dark platforms. In this dissertation, I take an exploratory, content focused approach that covers information sources and communicative spaces from fringe to mainstream. The guiding question of this dissertation is:

How do dark platforms interact with information flows in the broader media system, and how can we study these dynamics?

Against this backdrop, I argue that in order expand our understanding of influence dynamics within the contemporary media systems we must consider the role of information flows from dark platforms in agenda-setting research. Furthermore, I argue that the complex and dynamic nature of information flows, calls for more fine-grained insights driven by methodological innovation as well as the need to refine existing theories of influence and power dynamics in the contemporary media system. Existing, overly simplistic metaphors such as echo chambers (Sunstein & Vermeule, 2009) or filter bubbles (Pariser, 2011) are already widely contested by empirical evidence (Bruns, 2019). Yet, communication scholars still turn to these to explain phenomena that are much more complex than they appear. One does not simply go down the rabbit-hole of extremism by being continuously exposed to extremist and antagonistic narratives on a single (social media) platform. Many other intertwined aspects of and dynamic information flows are at play that contribute to one’s path to radicalization.

To illustrate, in the contemporary media system a fringe actor, or political outsider (Jungherr, Schroeder, & Stier, 2019) has the ability to introduce potentially dangerous narratives, topics, or issues into discourse on a dark platform such as Telegram or 4chan/pol. This information can then be taken over and amplified by more mainstream actors (e.g., extremist politicians) who post about it on more mainstream platforms (e.g., Facebook). Then, this information can be propelled into the public discourse by also being covered by the mainstream media (Wong & Trilling, 2023).

Alternatively, information first covered by the mainstream media or a politician can be picked up by a fringe actor and end up on dark platforms where users can react to and deconstruct the narrative. Hence, the so far understudied feedback loops across the fringes to the mainstream can provide us with more nuanced and valuable knowledge with respect to the complex interactions across the myriad information sources and communicative spaces that shape the public debate. Understanding the dynamic forces that influence public debate in the contemporary information landscape is crucial to safeguard democratic processes from the influence of extremism.

The contemporary media system and the evolving information landscape

The rise of digital technologies has fundamentally altered the production, consumption, dissemination, as well as the debate of politically and socially relevant information in the contemporary information landscape. The Habermasian public sphere (Habermas, 1962, 2006) – if it ever existed – is now irrevocably divided into numerous interconnected online and offline communicative spaces of varying sizes, accessibility, and publicness, often converging and overlapping in unexpected and serendipitous ways (Bruns, 2023). The interwoven nature of the wide range of communicative spaces within the public sphere form a hybrid media system (Chadwick, 2017) and an “evolving, multi-layered” information ecosystem (Starbird et al., 2018) that fosters rapid and non-uniform information dissemination and opinion formation. On the one hand, it offers a multitude of platforms for information production, consumption, discussion, and debate. On the other hand, it simultaneously creates opportunities “for the continued exploration and transgression of the limits of public debate ”(Bruns, 2023, p. 1) providing convenient avenues for the propagation of false and deceptive information as well as the escalation of conflict, propaganda, manipulation, and harmful speech.

In light of this shift towards an ever-evolving, multi-layered information landscape, scholars argue that we have entered an era of “post-factual relativism” (Van Aelst et al., 2014) where the epistemic status of what is considered to be (truthful) information has become subject for public debate. Others have also referred to this phenomenon as the era of “post-truth” (Lewandowsky et al., 2017) or “poly-truth ” (Harambam et al., 2022). Although several concepts exist to describe the same phenomenon, scholars seem to agree that in the contemporary public sphere facts often become subjects for public debate despite overwhelming scientific evidence, while some opinions are increasingly elevated to the status of facts.

At the same time (political) opportunity structures have shifted in ways that weaken the institutional gatekeepers while enabling so called political “outsiders” to bypass the weakening watchdogs and disseminate their narratives through their own communicative spaces (Jungherr, Schroeder, & Stier, 2019) and “public sphericules” further weakening the position of the

establishment. In this post-truth era outsiders have the “discursive power” (Jungherr, Posegga, & An, 2019) to introduce and maintain anti-establishment appeals in the public debate which helps them to gradually position themselves as influential and relevant actors of the political landscape (Jungherr, Schroeder, & Stier, 2019). Dark platforms play a crucial role in the rise of such outsiders as they enable and facilitate the poisonous narratives that would otherwise not be tolerated by mainstream communicative spaces.

The rise of dark platforms

The rise of “dark platforms” – *“digital platforms that are less regulated and moderated, hence can be used for hosting content and content creators that may not be tolerated by their more mainstream counterparts”* (Zeng & Schäfer, 2021, p. 1) –, can be linked to the advent of the internet and advancements in digital technologies that enable the production and dissemination of content outside of the established media system. The increasing distrust towards mainstream media (Banerjee et al., 2023) fuels the existence of spaces that welcome those who seek “alternative truths”. For instance, while trust in most mainstream news has been rather stable in The Netherlands over the past few years, there appears to be a declining trend regarding trust in mainstream news in other Western European countries such as Germany, Belgium, and Switzerland (N. Newman et al., 2023).

Dark platforms are secluded online spaces that are characterized by *content liberation*, *exile congregation*, and *infrastructure ostracization* (Zeng & Schäfer, 2021). While dark platforms have the same basic affordances as mainstream social media platforms, they attract and host digital counterpublics that disagree with, distrust and defy the mainstream media and established institutions. They offer a safe haven for the propagation of extremist narratives, conspiracy theories, hate speech and disinformation. Unlike the dark web, or dark social media (Al-Rawi, 2019), dark platforms are publicly accessible spaces that may also offer private or semi-private spaces to converse. Despite their easy accessibility, moderation efforts on dark platforms are generally minimal, enabling potentially dangerous narratives to free flow on these “obscure” spaces that virtually anyone could find. Recent research found evidence for an affinity between the contemporary information landscape, platform affordances, and the agendas of political outsiders forming a digital counter public on the dark platform, Telegram, upon the outbreak of the COVID-19 pandemic (Buehling & Heft, 2023).

To illustrate, during the hard lockdown of 2021, a series of “curfew riots” (i.e., “avondklokrellen” in Dutch) took place in the larger cities of the Netherlands. Those unhappy with the government-imposed COVID-19 measures took to the streets to express their disapproval in a violent manner by burning cars, looting shops and businesses, and throwing

rocks at the police who were trying to contain the situation². After a few consecutive days of rioting, the police identified several public Telegram group chats where messages calling to participate in illegal protests were spreading. Some of the implicated Telegram group chats were later banned and the administrators of the groups were prosecuted by the Public Prosecution Service (i.e., Openbaar Ministerie). These groups were not only associated with the “avondklokrellen” but also with the spread of conspiracy narratives and death threats against officials who dealt with the COVID-19 combating measures (Bouma, 2021). Although banning a few notorious public chat groups and prosecuting their administrators implicated in inciting violence may have temporarily stopped the spread of anti-establishment and conspiratorial narratives, thanks to the design of Telegram it is highly likely that malicious actors from the formerly banned groups continued their operations in different, most likely private group chats under the radar.

In response to rising extremism associated with Telegram, some politicians spoke out against dark platforms, pleading for the ban of Telegram as a measure to prevent the escalation of further violence³. However, banning apps like Telegram would be no more effective than severing one of the heads of the mythical hydra. As soon as Telegram gets banned, another similar platform would take its place where criminals could continue their malicious operations without too much trouble, and where ordinary citizens with grievances against the establishment can flock to follow (extremist) deplatformed internet celebrities (Rogers, 2020). In fact, a similar scenario has already taken place. The rise of Telegram can be directly linked to the downfall of Parler, a once-thriving right-wing social media app. Parler’s decline was exacerbated by the stricter content moderation policies enforced by mainstream social media platforms like Facebook and Twitter. As a result, millions of users migrated to darker platforms like Telegram, which welcomed them with open arms (Rogers, 2020). Thus, the end of one dark platform inevitably marks the beginning of another potentially even darker platform.

Another argument against banning more obscure platforms like Telegram is that it would prevent marginalized groups from communicating with each other in a relatively safe environment. Research found that dark platforms are not only hosts for antagonistic behaviour and “dark participation” (Quandt, 2018). They also provide a means of communication to oppressed and marginalized groups of society (Lewandowsky, Smilie, et al., 2020) to discuss sensitive issues they would otherwise be unable to share with others. For instance, Telegram has been a catalyst for empowering and mobilizing marginalized communities in Hong Kong (Urman & Katz, 2020) and Belarus (Walker, 2020). Similarly, although 4chan is more

²<https://nos.nl/artikel/2366099-nederland-zet-zich-schrap-voor-vierde-avondklok-avond>

³<https://nltimes.nl/2023/09/18/amsterdam-pvda-wants-ban-telegram-explosive-attacks>

Chapter 1 • INTRODUCTION

associated with the spread of hate speech, conspiracy thinking, and alt-right meme culture, and “shitposting” (Hagen, 2022; Hagen & Venturini, 2023), the initial goal of this image board was to foster equality and encourage anonymous users to express thoughts they might otherwise hesitate to share (4chan, 2015). Despite 4chan’s bad reputation, some of its users may still find solace in the anonymity and ephemerality of 4chan’s architecture. As such, the banning of more obscure platforms would be detrimental to marginalized voices with no alternative communicative channels.

Finally, Telegram has recently emerged as a prominent platform for the widespread dissemination of news (Lou et al., 2021). I discuss this aspect more in depth in Chapter 2 where I show how following initially “innocent” groups and channels dedicated to discussing news and politics may have exposed users to more radical, extremist, and antagonistic narratives in times of crisis.

In this dissertation I scrutinize two dark platforms: Telegram and 4chan/pol⁴. Although the affordances of these platforms are different in some ways, they both fit the definition of dark platforms as defined by Zeng and Schäfer (2021). Both Telegram and 4chan/pol celebrate content liberation. They welcome digital exiles who were banned from more mainstream platforms. Both platforms possess their own infrastructures for the creation and hosting of content. The darkness of Telegram and 4chan/pol however is difficult to equate. All posts on 4chan/pol are visible to the public on a single image board. However, these posts are ephemeral (Hagen & Venturini, 2023; Tuters & Hagen, 2020) and their expiration date is a function of user engagement, which motivates the generation of more attention grabbing and sensational posts (Bernstein et al., 2011). Moreover, users, “anons” do not need to create an account in order to post which creates the sensation of everybody on the image board being equal, forming a sense of community (Tuters & Hagen, 2020).

Telegram’s architecture is much more complex and fundamentally different from 4chan. On Telegram, dangerous narratives live in countless private and public channels and groups chats that cannot be found as easily as on 4chan/pol. “While public Telegram channels afford collective action — for example, identity and movement-building by organized activists of a movement, public group chats allow for personalized information-sharing and connective action.” (Buehling & Heft, 2023, p. 2). The shelf-life of posts in various Telegram groups and channels can differ to a great extent depending on the level of moderation enacted by channel or group chat administrators. Furthermore, Telegram has a reputation as a platform that enables discourse with enhanced privacy and anonymity, but at the same time it provides fertile ground

⁴4chan’s “/pol/” board, short for “Politically Incorrect,” is a board on the 4chan image board website where users discuss political and controversial topics.

for public identity building, outreach and mobilization (Urman & Katz, 2020). As Chapter 2 will demonstrate, if one wishes to find socially and politically relevant information on Telegram, the right keywords open doors towards an entire ecosystem of public group chats and channels that propagate dangerous, alternative narratives and viewpoints in the context of socially and politically relevant information. Taken together, it is crucial to investigate the role of dark platforms in shaping information flows within the contemporary media system.

Research context

In the course of the three years during which this dissertation was written, several major developments and crises have affected the world and with it the information flows that are studied in this dissertation. Three of these global developments have had a particularly strong influence on this dissertation.

The first and most pivotal one amongst these major developments was a raging global pandemic that marked the inception of the research described in this dissertation. As the virus was claiming thousands of victims across the world for the second year despite strict lockdown measures and the new hope for a vaccine, a simultaneous “infodemic” (Nielsen et al., 2020) poisoned the minds of people across the world. As conspiracy theories and misinformation permeated the public debate, thousands of people began harbouring strong doubts regarding government-imposed measures to combat the pandemic. The widespread uncertainty and increasing apathy surrounding the COVID-19 pandemic drove myriad people to the darker corners of the web where they found solace in collective sensemaking with like-minded others who were equally jaded by seemingly powerless pandemic control measures. While the mainstream media was reporting on daily victims of the pandemic and new government-imposed measures that affected the freedom of millions across the globe, dark platforms enabled people to converse with like-minded others who were both worried about the impact of the pandemic on their loved ones and at the same time felt increasingly negative towards the mainstream media and the establishment. These anti-establishment sentiments fostered connections across various extremists groups, such as the radical right and conspiracy thinkers whose contagious narratives appealed to thousands of people who had enough of the pandemic and looked for alternative views, solutions and leadership in the midst of the COVID-19 crisis.

This is unsurprising, as radical groups to the left and the right of the political spectrum have been capitalising on the lack of confidence and heightened anxiety of people facing critical situations for decades to further their own agendas by offering simple “solutions” to complex issues, blaming the elite and undermining trust in established democratic institutions

Chapter 1 • INTRODUCTION

(Bergmann, 2018). The black and white, simplified thinking of extremists fosters coping with uncertainty and fear via a worldview that is predictable and more understandable (Kruglanski et al., 2006) especially in the wake of crises that are surrounded by a cloud of uncertainty. In the Netherlands, the far right party Forum voor Democratie (FvD) took advantage of people's anti-establishment sentiments amid the pandemic and built an entire election campaign on it. They regularly encouraged their Twitter followers to join their Telegram channel where they could easily expose their followers to misinformation regarding the pandemic that would not have been tolerated by mainstream social media. Encouraging the electorate to not trust information provided by democratic institution and to join dark platforms such as Telegram has been unprecedented before in the history of Dutch politics. FvD's presence on the Dutch Telegramsphere has been very strong as the affordances of this dark platform arguably match the values of far right extremists. In the first empirical chapter of this dissertation (Chapter 2) I delve deeper into the information flows that dominated the Dutch Telegramsphere amid the pandemic. The guiding question of that chapter is:

RQ1: How have information flows influenced network structures and the development of communities within a single dark platform over time?

The second major development that affected the world and this dissertation was the Russian invasion of Ukraine in early 2022. In a chilling televised address on the eve of February 23, Russian president Vladimir Putin declared war on Ukraine, claiming that a “special military operation” will be carried out imminently to “denazify” and “demilitarize” Ukraine and at the same time stop the threatening eastward expansion of NATO. The morning after Putin's shocking speech, Russian forces began the siege of Ukraine that over time morphed into a war with no end in sight. Russia's war in Ukraine displaced millions of people and has claimed thousands of lives on the battle ground. The information warfare surrounding Russia's war in Ukraine however is waged on a digital battleground that enables the spread of contrasting narratives surrounding the invasion⁵.

Thanks to the popularity of social networking sites, news surrounding the invasion have travelled fast, far, and wide, reaching millions of people across the globe. An immense amount of information is being uploaded daily about the invasion to several social networking sites such as Telegram, TikTok, or Facebook both by state and non-state actors in Russia and Ukraine. A large part of this information is false, as the Kremlin has been poisoning the public debate with

⁵<https://foreignpolicy.com/2022/08/22/information-warfare-in-russias-war-in-ukraine/>

disinformation regarding Ukraine since before Putin gave the green light to the invasion. In spite of Western bans of Russian state-backed media outlets such as Russia Today and Sputnik News, Putin's propaganda machine is believed to be thriving as the poisonous narratives, and disinformation propelled by the Kremlin still manage to reach millions of people worldwide through alternative distribution channels such as Telegram, Facebook, TikTok, and WhatsApp (Kahn, 2023). Chapter 3 takes us one step closer to understanding the conundrum of information flows that surround the public debate regarding Russia's war in Ukraine by taking a multi-platform approach. In that chapter I unravel discursive information flows across the traditional media agenda, the social media agenda of politicians, the social media agenda of the public, as most importantly the dark platform agenda. The guiding research question of this chapter was:

RQ2: How are the information flow dynamics between (1) the dark platform agenda, (2) the social media agenda of politicians, (3) the social media agenda of the public, and (4) the traditional media agenda?

Tracing information flows in the interwoven contemporary information landscape represents a challenging task. On the one hand, information flows consist of a static representation of information. These can take the form of topics, named entities, URLs, dictionary and bag-of-words (BoW), or word embeddings generated from textual data. On the other hand, the *flow* aspect of information flows is dynamic in nature. It represents a linkage across representations that are a combination of at least two major indicators: (semantic) similarity across representations and temporal proximity. To date, no consensus exists about a meaningful threshold for either semantic similarity, or temporal proximity. This is where the challenge lies. Establishing a universal threshold for semantic similarity remains a challenge due to the diversity of representation (e.g., topic, named entity, or embeddings), language, as well as similarity measures (e.g., cosine, soft-cosine, or Levenshtein distance between word vectors). Whereas setting a threshold for temporal proximity is equally complex, as the relevance of time can vary across different types of information. What is appropriate for one type of representation (e.g., topic) may not be for another (e.g., news event).

Global developments such as information wars and an infodemic, have further complicated this already difficult endeavour. The last major global development that affected this dissertation was the rise of, or rather the widespread adoption of Large Language Models (LLMs). LLMs show a great promise in conducting research and potentially providing a solution to studying complex phenomena such as information flows by leveraging LLMs' vast knowledge basis. Although it has

Chapter 1 • INTRODUCTION

only been a little over a year since the mainstreaming of this technology, LLMs have already had a significant impact on how research is being conducted. Several scholars of various disciplines have put LLMs to the test to perform tasks that would otherwise be performed by humans. Results of several pioneering studies show that LLMs can perform comparably well or even better than human on various annotation and classification tasks (e.g., Gilardi et al., 2023; Törnberg, 2023a). Albeit, widespread and legitimate criticism surrounds the usage of the so far very much black-boxed and biased LLMs for research purposes (McGee, 2023).

On the one hand, the rise of LLMs further complicate the process of studying information flows as the creation and propagation of (dis)information has become significantly easier faster, and cheaper since their inception. As more and more AI-generated content appears on the internet, scholars and policy makers worry that AI-generated content may exacerbate the already challenging quest of repressing malicious information. Rather than using obscure software to manufacture disinformation, thanks to generative AI malicious actors can use natural language to devise compelling pieces of false information to manipulate the public debate.

Nevertheless, as these models appear to be here to stay, it is important to study them in order to understand their limitations, biases, but also their advantages that may prove to be of great use to solve so far unresolved issues. Even if LLMs in their current state may not be unbiased and or very effective in performing human tasks, in a reliable, and ethical way, we may well be approaching the time where these models will act as AI assistants that can optimize the way research is being conducted.

Circling back to the issue at hand, perhaps the biggest conundrum in studying information flows is the validation of the methods used to scrutinize these phenomena. When thinking about validating automated text analysis approaches we can distinguish between more *static* validation tasks that can be used to benchmark the identification of concepts, and more *dynamic*, document-comparison-based validation tasks that study *flows*. Addressing especially the dynamic aspect of information flows, in Chapter 4, I propose and test to what extent can or should we use LLM-assisted validation for dynamic, flow-based methods such as the detection of information flows. The guiding question of Chapter 4 is:

RQ3: To what extent can or should we use LLMs to optimize the validation of methods tracing information flows?

Taken together, the past three years marked turbulent times that certainly made writing this dissertation a challenging task. Studying the increasingly interwoven but at the same time fragmented public sphere and more specifically the complicated nature of information

dissemination across the multitude of communicative spaces within the contemporary media system represents a highly challenging task. The situation is further complicated by the impact of societal crises such as a global pandemic and a war in Europe, but also by the impact of major technological advances such as the rise of Large language models (LLMs that I outline above). In this sense, studying information flows in the contemporary media system is much like fighting the mythical Hydra. The single research question formulated in early 2020, laying the foundation to this dissertation, has gradually morphed into a hydra-like creature with too many heads to count as addressing one question at the time raised even more questions. Embracing the complexity of the task at hand, this dissertation employed state-of-the-art computational methods to take the discipline one step closer to understanding power dynamics of information flows on the one hand while on the other hand providing scholars with novel approaches to analyze these phenomena as well as to validate their approaches.

Beyond agenda-setting research

After reading the sections above one may wonder why in this dissertation I do not rely on agenda-setting theory to explain the phenomena I am studying. After all, one could argue that agenda-setting could have been a useful theoretical framework to guide the research included in this thesis. Below, I argue why the classical framework of agenda-setting does not entirely fit the scope of the research presented here.

One of the most influential and fundamental theories of communication science is agenda-setting theory (McCombs & Shaw, 1972) which posits that the mass media has the power to influence the public agenda. Since its inception, agenda-setting theory has been extended several times to make it more relevant to studying more aspects of agenda-setting, and to explore more fine-grained phenomena.

The first and fundamental level of agenda-setting research focuses on how the media determines what the public “thinks about”. Second-level agenda-setting scrutinizes the “how”, focusing on dominant characteristics and attributes of the issues and topics that are salient on the public agenda. The first two levels of agenda-setting theory assume “*a logical and hierarchical cognition mechanism*” (Guo & McCombs, 2011, p. 18). Third-level agenda-setting, the most recent extension of the traditional agenda-setting theory, introduces the network agenda-setting model, wherein the media can also shape how people connect various issues and attributes they are exposed to (Guo et al., 2012) through a series of interconnected nodes in complex issue and attribute networks across the media and the public. As such, third-level agenda-setting posits that the media does not only transfer disconnected objects and attributes separately onto the public agenda, but rather an interconnected network of relationships across issues and their

Chapter 1 • INTRODUCTION

attributes, forming a complex picture.

Besides the three levels of agenda-setting theory other connected theories have emerged to explain agenda-setting relationships across politics and the media. Political agenda-setting represents one of these spin-off theories. Political agenda-setting underlines the power of mass media in influencing politics (Van Aelst et al., 2014; Walgrave & Van Aelst, 2006) and the role of politicians influencing the mass media. Some referred to this reciprocal relationship as a “tango” where the “dancers” take turns in leading and following (Van Aelst & Vliegenthart, 2014).

Another relevant extension of agenda-setting theory is intermedia agenda-setting (Danielian & Reese, 1989; McCombs & Shaw, 1993), that specifically focuses on agenda-setting dynamics across media outlets such as newspapers, blogs, television, radio, or social media. Numerous studies have relied on intermedia agenda-setting theory to explain who leads and who follows across a multitude of media systems and contexts (see e.g., Barberá et al., 2019; Gilardi, Gessler, et al., 2022). Figure 1 illustrates the core aspects of agenda-setting theory that are also relevant to this dissertation.

Taken together, agenda-setting theory with all of its extensions has been an extremely useful theoretical framework that guided countless, predominantly media effects scholars in the past few decades when studying the relationship between the public, the media, and politics across a set of predefined issues and topics (see e.g., Gilardi, Gessler, et al., 2022; Van Aelst & Vliegenthart, 2014). As this theoretical framework is more concerned with studying the effect of known issues and topics across a set of “agendas” formed by aggregated news sources, it proves less useful to conducting more exploratory research studying the process of these influence dynamics across a complex, dynamic and multi-layered information ecosystem, especially for lesser known, diverse and so far understudied issues, actors, and platforms.

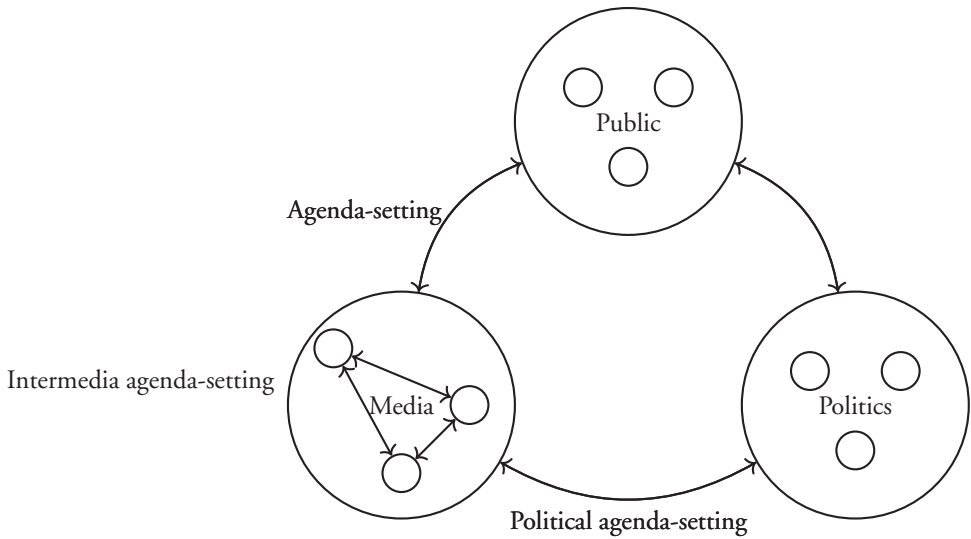


Figure 1: Agenda-setting theory

As the dynamic and multi-layered contemporary information ecosystem (Starbird et al., 2018) offers a multitude of interconnected contributors to the political communication space, scholars argued that there is a need for a theoretical framework that can empirically test the complex interactions and power dynamics across these entities in a systematic manner (Jungherr, Posegga, & An, 2019). This framework should enable the “systematic analyses of information flows” (Jungherr, Posegga, & An, 2019, p. 404). At its core, the concept of information flows delineates how information spreads between entity A and entity B. Figure 2 exemplifies how a basic model for information flows whereby information from one point (A) can spread to another (B), which in turn can be picked up by yet another entity (C) that can further transmit it (D, E).

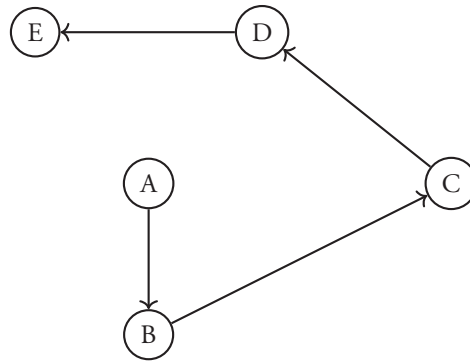
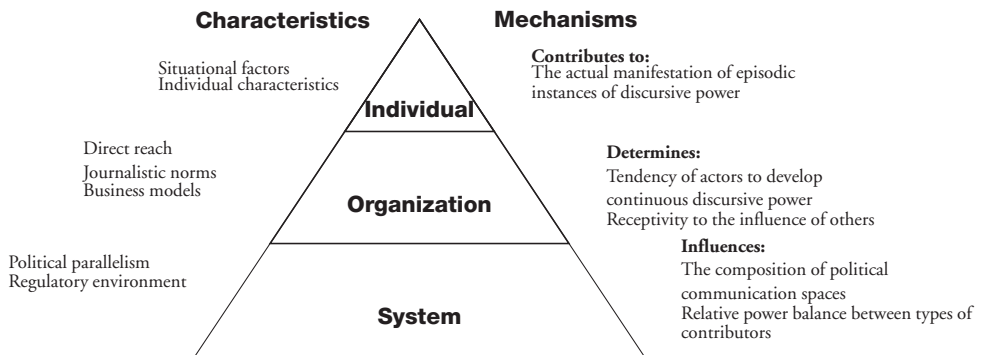


Figure 2: Basic model of information flows

Jungherr, Posegga, and An (2019) proposed the concept of discursive power that “does away with a technologically focused differentiation between “old” or “new” media and instead recognizes that contemporary news organizations share analog and digital operational logics and distribution channels” (p. 419). Furthermore, the concept of discursive power recognizes that the information flows in the hybrid media system are not only shaped by mainstream news sources, but also alternative platforms as well as members of the public who can influence the public debate through a variety of information sources and communication spaces. Figure 3 illustrates the determinants of discursive power based on Jungherr, Posegga, and An (2019).



Note: Model created based on Table 1 in Jungherr, Posegga, and An (2019)

Figure 3: Determinants of Discursive Power

In Chapter 3 of this dissertation I follow the argumentation of Jungherr, Posegga, and An (2019) regarding the limitations of (intermedia) agenda-setting theory in studying information

flows in the contemporary media system. According to Jungherr, Posegga, and An (2019), traditional (intermedia) agenda-setting theory sidesteps the complexities of power dynamics across a variety of contributors to the contemporary communication spaces because aggregating various types of media outlets and other contributors to communication into types or agendas, neglects the differences that exists across these entities. To illustrate, while The New York Times (NYT) and The Kyiv Independent (KI) both are English-language news outlets with an international audience, combining them into a single category without accounting for the differences across these outlets would be a mistake. Not only their size and readership base are fundamentally different, the topics and issues they cover also differ. While the NYT covers local and global topics, KI is more focused on reporting about Russia's war in Ukraine in an accessible format to the rest of the world. Thus, it is crucial to account for these differences when studying information flows in the contemporary media systems.

While I agree with Jungherr, Posegga, and An (2019) that providing a more nuanced picture of influence dynamics across a variety of information sources is crucial, I still recognize the relevance of aggregating certain information sources into aggregate typologies such that would enable scholars to extend their findings to previously studied traditional agenda-setting effects to lesser known issues, actors, and platforms. Therefore, in Chapter 3 besides studying the impact of the mainstream media agenda, I also explore the individual role of KI and the NYT.

As such, this dissertation builds upon both traditional agenda-setting theory as well as on more contemporary theories such as discursive power to understand information flows in the contemporary information ecosystem. In Chapter 3, I introduce the concept of discursive information flows to narrow down the concept of information flows to the domain of studying how the spread of information can shape public discourse. In contrast, Chapter 2 and Chapter 4 focus on information flows more broadly. Chapter 2 relies on a variety of indicators such as user-overlap, URL-overlap, and topic overlap to conceptualize information flows within the Telegramsphere. Chapter 4 provides a new, LLM-assisted approach for scholars to validate any type of flow-based method, including methods that trace discursive information flows. Finally, in Chapter 5 I propose a new model to study information flows in the contemporary media system.

Research design

Data collection

This dissertation is largely exploratory in nature and relies on various content analytic approaches of textual data. The data used in this dissertation has been collected through public APIs and custom scrapers written by either our own research team or the wider network of

Chapter 1 • INTRODUCTION

computational communication scientists in Amsterdam and beyond. These tools were made by researchers, for researchers. Specifically, the data for Chapter 2 was collected via the official Telegram API using the Telethon Python wrapper. This approach allowed us to gather historical chat and channel data from the chats/channels under study. The data collection for Chapter 3 was more complex given that I gathered data from multiple different domains. Part of the dataset for this study was gathered via the 4cat interface (Peeters & Hagen, 2021). I collected posts from 4chan's politically incorrect board, posts and posts from several news-related subreddits. Tweets were collected via the researcher API of Twitter. All tweets of the G7 were collected in the studied time period. Finally, to collect news articles from the New York Times and the Kyiv independent I made use of news-please, a Python package that uses web-URLs to extract news articles from web pages including metadata when available. Finally, for Chapter 4 I made use of a dataset collected by Aaldering et al. (2021), around the 2021 Dutch general elections. These were news articles published in the course of 3 months leading up to and including the election data by 10 different news outlets in the Netherlands.

Data analysis

Neural topic modeling. Topic modeling is a frequently used unsupervised machine learning technique that enables the identification of latent topics in a corpus of documents. The traditional and most widely-used topic modeling approach in the field of communication science is Latent Dirichlet Allocation (LDA) (Blei et al., 2003). LDA is a probabilistic, bag-of-words approach that uses word-concurrences to generate topics across a large corpus of documents. Although LDA can be a suitable approach for topic modeling when applied correctly (Lind et al., 2021). However, advances in machine learning provide better ways to model topical information in texts. Neural topic modeling represents an unsupervised topic modeling technique that relies on a combination of dense vectors generated by transformer models such as BERT and keywords to generate topics based on short documents (Grootendorst, 2022). The neural topic modeling approach this dissertation employed was BERTopic. I used this technique to answer some of the research questions in both Chapter 2 and Chapter 3. Calculating with these dense vectors enables us to map the meaning of words much more clearly by considering the company they keep across hundreds of dimensions. Generating topics while considering their context achieves more easily interpretable topics thereby reducing the bias that may arise from researcher interpretation. Taken together, neural topic modeling enabled me to gain a deeper understanding of the dynamic narratives that dominated various dark platforms (Chapter 2, and Chapter 3) as well as more mainstream information sources (Chapter 3) over time. This method has proven highly lucrative to explore the content of information flows that characterized the contemporary media system.

Network analysis. Network analytic approaches consist of a set of tools and techniques that can be used to model relationships across various entities such as groups, news outlets, or topics. These techniques have proven to be valuable in myriad disciplines (e.g., information science, neuroscience, or sociology), to study relational data (see e.g., Bassett & Sporns, 2017; Borgatti et al., 2009; Chithik Raja et al., 2014). At its core, a network consists of nodes and edges. Nodes are the entities one wishes to study while edges represent the connections between the studied entities. Chapter 2 takes a network approach to understand information flows within the Dutch Telegramsphere. More specifically, in this chapter I adopted network analytic techniques to uncover various levels of connectivity across public Telegram channels and (group) chats. Telegram chats/channels were the nodes while edges were conceptualized as either user overlap, URL overlap, or topic similarity (i.e., cosine). Network analysis helped us understand the overtime development of the ties across chats/channels through the content they propagated and the users they shared. Related, I employed a so called disparity filter, also referred to as a backbone extraction method (see e.g., Gursoy & Badur, 2021; Serrano et al., 2009) to filter out non-meaningful edges (compared to a null model) while maintaining the structural characteristics of the original network, and without eliminating smaller-scale connections that would inevitably fall pray to frequency based edge pruning approaches (e.g., remove all nodes with less than 3 edges). All in all, as demonstrated by Chapter 2, network analytic approaches combined with advanced natural language processing methods (i.e., neural topic modeling) can reveal interesting patterns with respect to community formation around particular narratives over time.

Hawkes processes. Hawkes processes are self-exciting temporal point processes used to model events that occur over time (Hawkes, 1971). Chapter 3 used this statistical framework to study non-linear and dynamic connections across entities by modeling the interdependence between events (i.e., these can be topics, news events, issues) and how they influence one another over time. Figure 4 illustrates these processes. In a nutshell, a Hawkes processes model comprises of k *self-exciting* point processes. Each process is characterized by its *rate of events*, denoting the likelihood of events occurring within that process. These point processes are termed self-exciting because the occurrence of an event triggers a temporary increase in the rate of events within the same or other processes. Specifically, when an event occurs, it can trigger what I refer to as *impulse responses* in the same or other processes, resulting in a short-term elevation in their event rates. The cumulative event rate within a process is determined by two key components: the *background rate* intrinsic to that process and the additional rate increments stemming from impulse responses generated due to the occurrence of events. The background

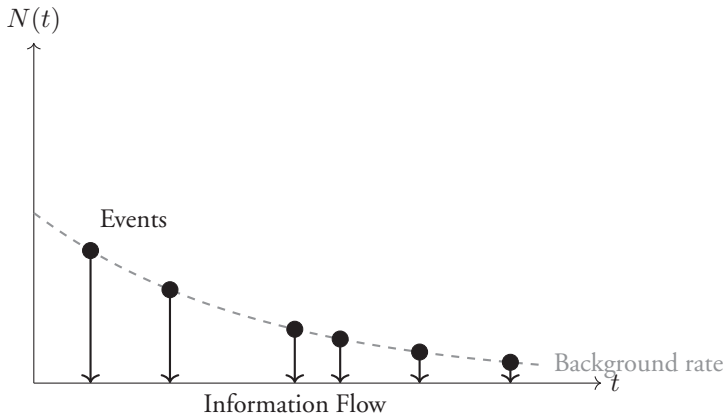


Figure 4: Basic illustration of Hawkes processes

rate represents the base probability of events manifesting within the process.

As mentioned above, we employ this framework in Chapter 3 to study discursive information flows by modeling the influence dynamics across four agendas (i.e., processes) as well as the sixteen individual entities that made up the agendas (i.e., processes), by generating a separate Hawkes models for 548 unique and validated topics (i.e., events) that were shared by the studied entities. Following best practices (Zannettou, 2019) we set up our models in a way that assumed that a given event (in this case a topic) can cause other events within a 12-hour time-frame. To answer our research question we fit the models and calculated two metrics, namely *Influence* (raw influence score) and *Efficiency* (normalized influence score). The raw influence score quantifies the anticipated quantity of events generated in Process B as a result of prior events in Process A. In contrast, the normalized influence score gauges the effectiveness of Process A in instigating new events in Process B. When evaluated together, we can assess the information flows both internally (within agendas) and externally (between agendas) by analyzing these normalized influence scores. Taken together, when combined with neural topic modeling, Hawkes processes can help us trace the dynamics of how the occurrence of certain narratives on one platform can trigger the occurrence of other narratives on other platforms or information sources. Thanks to this technique scholars can obtain a more fine-grained understanding of the dynamics of information flows by modeling the passage of time continuously, rather than across static snapshots.

LLM-assisted human validation scheme. The last methodological contribution of this dissertation is an LLM-assisted human validation scheme devised for automated methods

tracing information flows. The rapid emergence of LLMs has enticed many scholars to test how LLMs compare to human abilities with respect to various annotation and classifying tasks (Pangakis et al., 2023). Although evidence is scarce, and the validity of LLMs is questionable, I argue that these models provide a new opportunity to advance the validation procedure of methods tracing information flows and other comparable flow-based tasks that have been difficult to validate using conventional approaches. In the proposed scheme I involve the help of LLMs as well as an expert annotator in the annotation/validation process. I refer to this as LLM-assisted human validation. Rather than using LLMs to replace human annotators, I propose that the capabilities of LLMs can be leveraged to optimize the validation of various flow-based methods. I argue that if LLMs are able to retain and aggregate information on a large scale, hints provided by these models would lift a significant burden off the shoulders of human annotators who are usually tasked with annotating hundreds or thousands of cases that inevitably leads to mistakes and oversights for even the simplest annotation task. All in all, while the utility of the tested (self-hosted and light-weight) LLMs in the proposed validation scheme has proven to be questionable regarding the performance of flow-based annotation tasks, I provide a series of (error) analyses and recommendations in Chapter 4, and Chapter 5 with respect to the potential, future role of LLMs in complex annotation and validation tasks.

Outline of the dissertation

The dissertation consists of a general introduction, three empirical chapters, and a general conclusion and discussion. The first two empirical chapters delve into case studies of information flows using novel methods and relevant theory, thereby laying the conceptual and methodological foundation of the dissertation. The third empirical chapter ties it all together by testing a Large Language Model-assisted human validation scheme for automated methods tracing information flows.

I begin the journey of studying information flows in the hybrid media system by zooming in on how information spreads within a single but highly controversial dark platform, Telegram (**Chapter 2**). Then I continue with expanding the toolkit as well as the within-platform empirical and theoretical insights by studying discursive information flows across the dark platform agenda, the social media agenda of the public, the social media agenda of politicians, and the traditional media agenda (**Chapter 3**). Finally, in **Chapter 4** I harness the power of LLMs and propose a novel approach to validating methods that aim to trace information flows.

This dissertation begins with an exploratory study. More specifically, **Chapter 2** uses automated content analysis approaches and network analysis techniques to study information flows within the Dutch Telegramsphere of current affairs. I argue that Telegram is more than a

Chapter 1 • INTRODUCTION

tool for political fringe groups or an Instant messaging (IM) app for personal messaging. Instead Telegram has become a source of socially and politically relevant information for many of its users thanks to its affordances. Using sophisticated, state-of-the-art computational methods such as neural topic modeling and social network analysis on a unique dataset of Dutch-language messages ($N = 2,033,661$) sent in news and current affairs-related public Telegram channels and group chats ($N = 174$), I investigate how information flows can reveal often unexpected patterns of connectivity across very different groups. Findings show that over time, conspiracy-themed, and far-right activist groups dominate the Dutch Telegramsphere of current affairs, a pattern that has likely been fuelled by the outbreak of the COVID-19 pandemic on the one hand, and the mass “deplatforming” of extremist and radicalized users from more mainstream social networking sites such as Twitter on the other hand. These results raise concerns about Telegram’s role in fostering polarization and radicalization capabilities in the context of consuming socially and politically relevant information online.

After revealing information flow patterns within a single dark platform, **Chapter 3** continues with expanding the scope of the the dissertation by investigating discursive information flows across the dark platform agenda, multiple mainstream agendas as well as the individual entities that comprised these agendas over time. The aim of Chapter 3 is twofold. First, it investigates how discursive information flow dynamics between (1) the dark platform agenda, (2) the social media agenda of politicians, (3) the social media agenda of the public, and the (4) traditional media agenda. Second, it studies the role of individual entities (i.e., subreddits, Twitter accounts, newspapers) in shaping the discursive information flows across the studied agendas. the case under study in this chapter is the Russian invasion of Ukraine.

Similar to the previous chapter, Chapter 3 employs neural topic modeling, as well as other automated content analysis methods and advanced over-time analyses of fine-grained influence dynamics via Hawkes processes. This study is conducted on a unique, multi-platform, and primarily English-language dataset ($N_{texts} = 1,545,813$) that spans the first six month of the Russian invasion of Ukraine (17.01.2022 - 17.06.2022). Findings reveal negligible discursive information flows from the dark platform agenda to other studied agendas. While the the most influential topics propagated by the dark platform agenda appeared to be harmless, further analysis revealed that the most influential topics on this agenda were significantly more toxic than those on the studied mainstream agendas during the investigated time frame. On the bright side, I find evidence for the continued agenda-setting power of the traditional media agenda which had a substantial influence on the dark platform agenda. This indicates that dark platform may not be as successful in introducing new potentially harmful topics in the hybrid media system, but they may instead take on a primarily reactionary role that reflect on topics

already existing on mainstream agendas. Findings regarding within-agenda discursive information flows show that topics emerging on Zelensky’s Twitter account were by far the most effective in influencing the rest of the individual entities within the studied agendas. Taken together Chapter 3 expands the toolkit of intermedia agenda-setting research by analyzing dark platforms alongside the “usual mainstream suspects” both across and within various agendas.

Chapter 4 is the last empirical chapter of this dissertation. In this chapter, I propose and test a novel LLM-assisted human validation scheme for information flows. The main contribution of this chapter is a series of in-depth insights regarding the usage of LLMs for validating flow-based methods, a novel validation scheme that includes expert verification, and a large validated dataset for information flows on the main topic-level ($N = 990$), and the news events level. The goal of this chapter is to provide social scientists with the necessary evidence about the advantages and disadvantages of using LLMs to optimize the validation of flow-based methods. The main argument behind this chapter is that the validation of complex flow-based methods represent a challenge for human annotators due to the complexity of capturing such concepts. Although leveraging LLMs to provide human annotators with assistance may optimize the validation of more complex tasks, existing local and more condensed models do not seem to perform adequately enough for such a validation scheme to be truly helpful.

In sum, **Chapter 5** brings it all together and discusses theoretical, methodological, and practical implications of this dissertation. The chapter also includes a field guide to computational communication scientists to navigate the complex, and ever-changing challenges of studying information flows in a media system where, quite literally, everything represents a moving target.