

When URLs on social networks become invisible: Bias and social media logics in a cross-platform hyperlink study

by Suay Melisa Özkula, Maria Lompe, Mariangela Vespa, Estrid Sørensen, and Tianshi Zhao

Abstract

Extant research has addressed various concerns of representativeness in digital social research including: bias in researchers' selection of online spaces, foci on single-platform approaches, and limited or skewed samples due to API (application programming interface) restrictions. This paper adds to that work through an illustration of tool bias towards specific social media logics (*e.g.*, Twitter logics) in a URL-based network across/within social media sites (illustrative case study = greenwashing). These “biases” are implicit in design, mirror extant societal trends, and are reinforced through platform biases. As such, researchers using such tools (above all, non-computational scholars) may have little awareness of these subliminal influences. The paper consequently argues that (a) tool choices often fall prey to issues in representation, reinforcing existing biases on a subliminal level; and, that (b) non-platform-specific creative situational approaches (like cross-platform URL explorations) provide a much-needed understanding of wider platform dynamics that highlight such biases.

Contents

[Introduction](#)

[Research context](#)

[Summary of research protocols](#)

[Methodological reflections](#)

[Discussion and conclusion](#)

Introduction

Digital methods constitute a popular approach in contemporary research on online dynamics and phenomena, above all on social networking sites (SNS) such as Twitter and Facebook (in the Western

hemisphere) [1]. Compared to the wider field of digital social research, ‘digital methods’ are characterised by the use of software and automated processes for collecting and interpreting digital data (as per Rogers, 2019). Digital methods approaches rely on reusing trace data and methods from digital objects (Rogers, 2018; Venturini, *et al.*, 2018). Despite their growing popularity, critical scholarship has at times raised questions about the representativeness of the data generated through these approaches. Concerns about these issues often stem from choices made by scholars in the research process, limitations set by social media platform providers, and difficulties in navigating complex multi-user terrains that are shaped by commercial and institutional entities (see Burgess and Bruns, 2015; Hargittai, 2020; Hino and Fahey, 2019; Tromble, 2021).

One common concern in representativeness has been the dominance of single-platform approaches, in particular Twitter-focused research (*e.g.*, Blank, 2017; Bruns and Burgess, 2015; Gerlitz and Rieder, 2013; Özkula, *et al.*, 2022; Venturini, *et al.*, 2018). While single-platform studies have contributed significant knowledge for understanding networked publics, a range of scholars have also called for more diversity in digital methods research, above all through cross-platform approaches and the study of alternative platforms such as Reddit, Tumblr, and non-Western platforms (*e.g.*, Tufekci, 2014; Pearce, *et al.*, 2020). Cross-platform research is richer as it considers variation and therefore allows for triangulation, an asset as most phenomena extend beyond the confines of a single platform [2]. Some work has already engaged with cross-platform approaches, for example Rogers’ (2018) and Venturini, *et al.*’s (2018) discussions of what cross-platform research encompasses (in comparison to merely multi-sited approaches), Pearce, *et al.*’s (2020) work on visual cross-platform analysis, and d’Andrea and Mintz’s (2019) work on tracing image circulation across platforms through the Google vision API.

Nevertheless, cross-platform approaches remain comparatively rare (see, *e.g.*, systematic review on digital activism research, Özkula, *et al.*, 2022). They often take user-rich Western platforms such as Twitter as starting points (examples: d’Andrea and Mintz, 2019; Horawalavithana, *et al.*, 2020). Thus, even where cross-platform approaches are applied, these typically rely on the social media logics of more commonly researched platforms. Social media logics describe ‘processes, principles, and practices through which these platforms process information, news, and communication, and more generally, how they channel social traffic’ [3]. Although these platforms may constitute justified starting points given a particular focus or case, the majority of comparative research remains multi-platform rather than cross-platform.

This paper addresses this prevalence of social media logics through a discussion of logics *assumed* or *built into* digital methods tools. The starting point of this exploratory work was a foundational material element used across platforms: hyperlinks to Web sites used within social media sites. Hyperlinks with high engagement (*i.e.*, frequent sharing) on Facebook, Twitter, and Reddit were explored in terms of how they circulated in a wider network, and in terms of the discursive features of their Web site content. This paper aims to reflect on a range of difficulties in attaining the pursued data, an issue tied to tool settings and logics. It argues that (newer) tools developed for digital research often prioritise mechanisms of popular SNS, reinforcing existing digital biases, above all in research that uses individual platforms as starting points. This argument is constructed as follows. The next section will outline common issues of representativeness in digital social research towards making a case for cross-platform and situational analytics approaches. Then, the project’s research protocols will be outlined and subsequently discussed as part of a reflexive account of tool- and platform-based limitations in hyperlink work. To illustrate this, an illustrative case of “greenwashing” will be used. This article will argue in favour of non-platform-specific creative situational approaches (that do not take individual social media platforms as starting points; *e.g.*, hyperlink-focused approaches) and more critical engagement with implicit software features.

Research context

URLs, situational analytics, & cross-platform research

Hyperlinks are significant conversation markers or digital tagging mechanisms [4]. As one of the earliest tags on the commercial Web, they inform user commentary through tagged imagery and headlines, as factual evidence, news, and as secondary voices that act as experts, gatekeepers, or second opinions. As such, they are a significant part of the fabric through which meaning is created. Social media comments heavily relying on hyperlinks are therefore difficult (if not incomplete) for interpretation if the linked Web site contents and authors/sources are excluded. In part, research disregarding linked materials may well be founded on user behaviour, as studies have shown that user engagement with link content varies significantly across platforms and link type (e.g., Pew Research Center [U.S.] data, Shearer and Mitchell, 2021). Even so, social media publics often accumulate around specific events and issues, for which hyperlinks constitute the starting point. For example, political discussion often develops in response to shared hyperlinks to news Web sites. As such, hyperlinks form a foundational component of social media debates.

Even so, due to trends in *Web* platformisation (see Helmond, 2015) URLs have more rarely become the primary or direct entry points to *Web* content. At times described as a shift from *Web* 1.0 (= the broadcasting Web; the *Web* as an information source) to *Web* 2.0 (= the social Web) or the participatory turn (Scholz, 2010), platformisation describes platforms' increasing dominance as an infrastructural and commercial model of the *Web* (see Helmond, 2015). Using the example of Facebook, Gerlitz and Helmond (2013) describe these infrastructures as a new form of connectivity that relies on engagement buttons (e.g., shares and likes) and their metrics (in their words — a 'like economy'). In doing so, platforms emerging above all in the mid-2000s have de- and re-centralised the Web (Gerlitz and Helmond, 2013). While hyperlink network analyses have been a popular method for exploring Internet infrastructure in early Internet research (Park and Thelwall, 2003), some research has also shown that users continue to use Web sites as significant meeting points (Beaulieu, 2005; Garrido and Halavais, 2003; Howe and Bisel, 2020). While this has meant that URLs have fallen somewhat (though not entirely) out of favour as primary entry points to participation, social media platforms commonly still connect to Web sites through hyperlinks.

Unlike many other markers, URLs and hyperlinks are not platform-specific or even necessarily platform-internal. They are often cross-platform tagging markers. As such, they lend themselves particularly well to extended cross-platform analyses and what Marres (2020) calls 'situational analytics', e.g., in wider (social) media environments. The latter is based on a perspective that allows for the consideration of multiple entities in a given situation, such as social, contextual, and technological factors in how situations evolve (Marres, 2020, based on Clarke, 2003). When applied to complex computational situations (e.g., multi-actor, -sited, -cultural), situational analysis allows for 'situational analytics' (Marres, 2020). These propose a way for automating and scaling up the interpretative study of such situations by combining computational with interpretive data [5]. Such approaches render visible otherwise invisible/latent elements in digital infrastructure (Marres, 2020), also an endeavour in webometrics (see Thelwall, 2012) and wider media ecosystems research (see Zuckerman, 2021), and are particularly beneficial where little is known about cross-platform movements of digital publics (example: Fletcher, *et al.*, 2021, who used pc trackers for understanding user behaviour across the Web).

Beyond tracing user behaviours, situational analytics and extended cross-platform analysis allow for an understanding of media differences. Media are not neutral and play an active role in shaping social relations, a material connection already identified in actor-network-theory [6]. They are part of the processes of meaning-making (Venturini, *et al.*, 2018, drawing on James Carey). While this notion has commonly been applied to traditional media forms, material differences have also been acknowledged between social media platforms. Extant research demonstrates that individual platforms differ in more than features and aesthetics; they are characterised by different social media logics, affordances, logics, grammars, and vernaculars (Venturini, *et al.*, 2018; Gibbs, *et al.*, 2015; Pearce, *et al.*, 2020; Rogers, 2018; Van Dijck and Poell, 2013) — essentially *platform-cultural* differences. A range of studies has already shown that platform choice may impact how a given topic is articulated, framed, and distributed (e.g., Mosca and Quaranta, 2021; Theocharis, *et al.*, 2021). These differences make platform data difficult to compare, particularly in digital methods approaches relying on queries.

Extended cross-platform analyses and situational approaches then allow for capturing these differences since they link and distinguish platforms as part of “what makes a difference” in a given situation (*i.e.*, the variable that matters, *e.g.*, actor, platform, content). In doing so, situational approaches also help circumvent issues of “source adequacy”. The term describes/questions to what extent a studied medium adequately reflects the phenomenon or practice of interest, *i.e.*, the match between platform and phenomenon [7]. Flanagin (2020) therefore proposes for researchers to highlight what a given technology or tool represents (*i.e.*, manifestations of underlying dynamics). In part, situational analytics addresses this problem, as a significant part of its purpose is to understand the role of the medium in the phenomenon under study. It contextualises a given media space as well as reflects on what it represents. Thus, situational approaches should circumvent some common issues in representativeness in digital social research.

To some extent, these issues also complicate contextualised digital social research as situations become difficult to trace in diffused, multi-access/-perspective, and “synthetic” environments (Knorr Cetina, 2009), an issue tied to media diversity and architecture. Venturini, *et al.* [8] explain that these complex situations create a risk for researchers to mistake medium characteristics for the phenomena under observation. For example, a study on a hashtag Twitter movement reflects, to an extent, architectural properties of Twitter that facilitate (if not produce) this kind of activism. Multi-platform research may therefore constitute merely a reflection of platform differences. While this point is also noted by Marres (2020; 2015), she suggests that bias is not the enemy of social research. Rather, she observes that any media — digital or analogue — have effects on how phenomena are represented. She suggests [9] an affirmative approach in which platform bias is understood as an aspect of the research object rather than something external to it. As such, this complex enmeshing of medium, content, architecture, and dynamics creates situations that are in themselves worthy of analysis (also argued by Knorr Cetina, 2009; Marres, 2020). Above all, situational analytics should be able to account for how platform architecture and culture create or affect biases and subsequent decision-making by researchers. It may also, as argued here, show how the development of methods and tools conversely affect researcher choices, practices, and knowledge production.

Bias in digital social research

Issues of partiality in digital social research have been described as digital bias (Marres, 2017), selection bias (Blank, 2017), or sampling/representation bias (Hargittai, 2020). They describe issues in representativeness whereby methodological choices allow for only partial views of (digital/online) social phenomena that remain difficult to generalise or contextualise beyond single spaces or demographics. Marres distinguishes here between bias in content selection (*i.e.*, research-induced), research instruments (*i.e.*, built-in, *e.g.*, algorithmic or API-driven), and wider methodological arenas. The former relates to choices made in the research process, including the chosen methodological approaches, digital platforms, or spaces (*e.g.*, source adequacy or platform preferences), query and/or search formation, as well as efforts made in contextualising and justifying these choices. In comparison, built-in or infrastructural issues relate to restrictions in site access and tool versatility. They include skewed representations arising from missing data due to API restrictions (*e.g.*, timeline or data limits on Twitter), deletions (above all an issue in retrospective scraping), differences in user behaviour (*e.g.*, cultural uses of tagging markers such as hashtags or mentions), non-random API sampling, and differing platform demographics (see Blank, 2017; Flanagin, 2020; Gerlitz and Rieder, 2013; Hargittai, 2020; Hino and Fahey, 2019; Lorentzen and Nolin, 2017; Tromble, 2021; Tufekci, 2014).

Some of these choices are subject to little control as they are underpinned by socio-economic and socio-political conditions that govern methodological decision-making. Researchers rely on access provided by platform providers, particularly as prescribed by API limitations and platform terms & conditions. Changes in API access (particularly following data scandals) have made certain platforms easier to research than others, above all Twitter (Burgess and Bruns, 2015; Rogers, 2018; Tromble, 2021). Scholars have therefore distinguished between a ‘Data Golden Age’ with greater access to data and a subsequent ‘post-API era’/ ‘API-calyse’ (*e.g.*, Bruns, 2019; Tromble, 2021). The history of computational research is as such closely linked to histories of API access. Newer terms may yet again emerge in response to recent developments in opening up (and potentially closing of) researcher access to data through academic Twitter data access and

Facebook-owned monitoring tool CrowdTangle, avenues that allow researchers to directly apply to the platforms for data access. Hence, despite researchers' efforts at cross-platform and situational analyses, these infrastructural and socio-economic conditions stymie situational research to a certain extent, making it difficult for researchers to compare complete like-for-like data.

These research choices or latent research influences (a difference which is often difficult to determine) do not, in themselves, render data non-viable or necessarily niche. While extant research has disproportionately produced single-platform Twitter findings (*e.g.*, argued by Burgess and Bruns, 2015; Hargittai, 2020; Özkula, *et al.*, 2022), this type of research has made significant contributions to scientific knowledge. Even so, it often focuses on specific/narrow sets of social acts and dynamics. Those include, above all, (hashtag-based) network dynamics of elite user bases (*i.e.*, high-earning, educated, and Western) in weaker-tie SNS (Blank, 2017; Hargittai, 2020). Choices in content selection therefore require researchers to critically examine what their data may realistically represent (= a situational question), *e.g.*, with regards to demographics, spaces, or dynamics. Where possible, they need to circumvent opaqueness stemming from omissions in research reporting, missing insights through what has become known as the black box of data flows (Driscoll and Walker, 2014), and comparatively little cross-platform, contextual, or situational analysis (see Lorentzen and Nolin, 2017; Marres, 2020; Özkula, *et al.*, 2022; Venturini, *et al.*, 2018; Vicari and Kirby, 2022). Research omitting such reflections otherwise remains difficult to evaluate in terms of its application and wider contextual dynamics.

As yet, existing concerns in digital bias revolve predominantly around researcher choices, platform characteristics, and API access (or similar technical) restrictions. A sub-area that has received less attention are limitations set by developments in research tools. This paper turns the focus to such tool and platform biases, specifically restrictions stemming from expectations of how such tools may be utilised. Some research, above all Borra and Rieder (2013; as well as Borra, Peeters, and Rieder's follow-on work on Capture and Analysis Tools for Social Media Research (CAT4SMR), see <https://cat4smr.humanities.uva.nl/>) has connected tool design to epistemological questions on methodology. In line with their work, we propose that not only are digital platforms not neutral, but neither are the tools that scrape their data, as they are a consequence of pre-existing biases, and are, as such, reproducing them. As a result, we argue that (a) these mediations must necessarily be taken into account by researchers, and that (b) situated and normative logics are inscribed into choices for digital methods software design (and consequently research tools) in the process of their development. As such, they are limited to popular platforms and logics at a given time.

These insights are reached through reflections on biases involved in a hyperlink project on greenwashing, conducted at the Digital Methods Initiative (DMI; in Amsterdam) in 2021. This project serves as an illustrative case for our methods discussions. Before turning to the methodological reflections on bias and representativeness (= the core interest of this paper), we will present the research protocols on the illustrative case.

Summary of research protocols

Illustrative case: Greenwashing

We operationalised our exploratory interest in hyperlinks in social media to Web contents by choosing a subject of analysis that would be well suited to combining media and content (as per Venturini, *et al.*, 2018) or infrastructure and issue (as per Marres, 2015) [10]: “greenwashing”. The term has emerged in response to corporate entities and organisations promoting green values without adhering to their claims on environment-conscious behaviour (de Freitas Netto, *et al.*, 2020; Gatti, *et al.*, 2019). It describes this ‘divergence between socially responsible communication and practices’ [11], or, in more critical terms, organisations’ positive communication on their environmental impact when it is, in fact, comparatively

poor in performance [12].

Due to its positioning between environmental and commercial concerns, the term is relevant in diverse disciplinary fora. While this arguably makes greenwashing difficult to trace online, it also made it a valuable case for situational analysis that considers both content and infrastructure. It allows for a consideration of how platform choice affects contents. Research designed this way offers a chance for identifying how and to what extent technological infrastructures shape controversies and vice versa (as per Marres, 2015; Venturini, *et al.*, 2018). Greenwashing lent itself particularly well to such an enquiry due to the potentially strongly divergent sharing and framing of the issue across media contexts, as well as its comparatively easy application as a query. It was explored in terms of if and how the term was associated with pro-environmental movements on Web sites shared on social media platforms (for full research questions, see Digital Methods Initiative (DMI) Wiki Greenwashing, 2021).

Research design

The research design combined an examination of (1) issue language in Web site contents (*i.e.*, discursive features); and, (2) a mapping of the wider URL network dynamics of the shared Web sites. In doing so, this project constitutes what has been described as a combination of ‘computational heuristics and close reading’ [13] or automated processes and contextual/situational readings (Marres, 2020). It builds on early webometrics research that used link analysis such as URL networks towards extracting patterns of interconnection (Thelwall, 2012). However, it additionally provides opportunities for what Marres (2020) describes as identifying ‘what makes a difference’ by considering the relationship between URLs and individual platforms.

This orientation allowed for the consideration of the wider contextual and ecological relationships around greenwashing, beyond a single platform, an individual hyperlink network, or a specific social media public. A significant part of this was the project’s cross-platform orientation, for which highly-engaged-with URLs on the platforms Facebook, Reddit, and Twitter were compared. This procedure focused not on user behaviour and user-generated-content across platforms, and thus differs from what is more commonly imagined as a cross-platform approach. Instead, its focus lies on differences in hyperlink engagement across platforms (*i.e.*, situational analytics). Here, the starting point for most of the enquiry were hyperlinks within SNS to Web sites and not individual movements, keywords, or hashtags for collecting social media commentary. As such, this process can be described as an *extended cross-platform* analysis.

This choice created a unique set of methodological challenges. The focus was not merely on understanding a single URL network and the issue language surrounding greenwashing on their Web contents, but on comparing these across platforms. As such, the platforms’ specific technological mechanisms, tagging markers, and affordances needed to be considered. In part, these difficulties were tied to common concerns in cross-platform research as it carries the necessity to tune ‘operational definitions’ to the individual platforms under study, but also to achieve a degree of consistency in methodological approaches across platforms [14]. Even so, the study was based on a methodological approach that differed both from single-platform studies and more conventional cross-platform analysis. This innovative approach was deliberately chosen towards generating an alternative/creative pathway for exploring platform differences and cross-platform hyperlink dynamics across platforms towards understanding “what matters” in greenwashing link sharing online. As such, it is particularly relevant for non-medium-centric bias research as it questions the existence of “pure” platform data and research. Platform studies are, in that regard, necessarily mediated by Web site contents, a complex situation, which is in itself analysable as tool combinations mediate these pathways in different ways.

Initial dataset: Building a link list

The initial data corpus for this enquiry consisted of a URL list created through Buzzsumo (<https://buzzsumo.com/>), a content-marketing tool that allows for scraping the most frequently shared URLs in relation with queried phrases. Buzzsumo was queried on posts relating to greenwashing to return the

most frequently shared (*i.e.*, most engaged with) URLs on Facebook, Twitter, and Reddit over the past 12 months. Toward creating a consistent dataset in which terms were comparable, only English texts were selected. The resulting URL list consisted of 2,227 links, which was then split into four datasets based on sorting:

Ranked by top engagement:

- on Facebook
- on Twitter
- on Reddit
- across platforms

Additionally, capped lists of the top 100 (for text-mining) and top 20 links (for CrowdTangle networks) for each platform were created. The choice to cap the lists emerged from disparities in engagement. For example, Reddit showed considerably less URL engagement on greenwashing than Facebook. This meant that some of the URLs appearing in the Reddit sorting showed little to no engagement. The capped lists were therefore more reflective of platform-specific engagement than the original corpus. Text-mining and social networks were conducted on this corpus.

Text mining

The textual contents of the Web sites to which the URLs linked were scraped through a custom extraction extension (created by Stijn Peeters) and entered into Voyant (<https://voyant-tools.org/>) for a semantic overview, and, for a more contextualised visualisation, on WordTree (<https://www.jasondavies.com/wordtree/>). While content analysis also lends itself to textual interpretation, the texts explored here were often long-form (*i.e.*, full-length articles) with sometimes few mentions of “greenwashing” (relative to article length). Contextual uses such as provided by word trees were therefore of higher relevance.

URL social media networks

To complement the text-mining, network dynamics were captured through issue-crawling the wider networks of URLs linked to from the social media platforms (= capped hyperlink list of top 100 URLs engaged with on Twitter, Reddit, Facebook; <https://www.issuecrawler.net/>). A snowball analysis was conducted with the aim of determining where and how shared URLs related to greenwashing spread across Web sites. This provided a sense of the wider networks surrounding the debate.

Network dynamics were additionally explored through CrowdTangle (<https://www.crowdtangle.com/>), a public insights tool from Facebook that allows for scraping data (content & metadata) shared by pages or groups. The aim was to create a database that listed the Facebook groups and Reddit sub-groups sharing these URLs including the surrounding social media commentary. While it was also possible to generate such a database for Twitter, CrowdTangle limited the data extraction to seven days and the data contained individual rather than group or public accounts. Due to the appearing lack of comparability, Twitter data was not considered in the analysis. Capped lists of 20 URLs per platform (Buzzsumo lists for Facebook, Twitter, Reddit) were entered into the search bar, separated by Boolean operator “OR”. The resulting CSV files included the Facebook pages and Reddit sub-groups, type of groups (*e.g.*, NGO, political, news), number of reactions to links, shared links, and post contents where links were published (*cf.*, [Figure 1](#)).

Data refinement & visualisation

The capped corpus included 400 URLs (= 100 each for Facebook, Twitter, Reddit, and one combining these), compared to an initial dataset of 2,227 URLs. The Web contents scraped from these URLs produced 8,000 A4 pages of textual content, plus 128 A4 pages of Facebook posts (obtained through CrowdTangle) in a standard font. Data refinement took place alongside data collection as part of a reiterative process, using WordTree, IssueCrawler, and CrowdTangle. Data were then visualised through both relational

visualisations (alluvial diagrams, snowball networks) and content-based visualisations (= word trees). For a more detailed account of the research protocols and tools, see [Table 1](#) and the Digital Methods Initiative (DMI) Wiki Greenwashing (2021).

Table 1: Research protocols.	
	Protocols
Case study & query	Greenwashing
Dataset	Initial dataset: 2,227 URLs Capped datasets: 400 URLs (=100 each for Facebook, Twitter, & Reddit) and 80 URLs (=20 each), equivalent to 8,000 pages of textual Web site content data plus 128 pages of comments
Digital methods used	Text-mining [tools: WordTree, Voyant, custom program] Issue crawling [tool: IssueCrawler] Network analysis [tools: IssueCrawler, CrowdTangle]
Tools	Buzzsumo WordTree Voyant IssueCrawler CrowdTangle Custom program for scraping URL contents
Processed datasets & visualisation	Word trees (from Word Tree) Issue networks (from IssueCrawler) Network alluvials (CrowdTangle)

Limitations

Tools were chosen based on their benefits in terms of access, availability, and suitability for the research purpose. Where applicable, preference was given to full versions (*e.g.*, Buzzsumo & CrowdTangle). In some cases, no alternatives existed. For example, CrowdTangle was problematic in that it is a Facebook-owned product and imposes certain data restrictions (*e.g.*, limited to public Facebook groups). As such, it is subject to the black box of mechanics with difficult-to-define operating parameters and only partial access to data. Even so, CrowdTangle was the only tool that allowed for access to Facebook data. Others limitations remained, *e.g.*, with regards to other black boxed issues in commercial and platform-provided tools, comparability of data collected across tools, and the use of hyperlinks as queries (for detailed discussion *cf.*, methodological reflections).



Methodological reflections

The situated analytics [or extended cross-platform analysis] of greenwashing gave rise to a number of considerations concerning representativeness that were split into (1) tool biases, and (2) platform biases.

Tool biases

Beyond the limitations of the specific methodological approach and the selected tools, the project revealed a range of implicit biases, starting with the query. Queries have been controversial in computational methods due to the limitations posed by ambiguity and interpretative differences in term selection (see Rogers, 2019). While this is in itself well-known, the query limitations here were of a different nature. It was not an issue of which precise description best fit the issue of greenwashing, a case where researchers might need to use synonyms or similar descriptors (*e.g.*, “climate change” versus “global warming”). Instead, the issue was rooted in the choice of entering URLs rather than a keyword (here “greenwashing”) and the related word count restrictions. Although using a longer link list was preferred in the CrowdTangle analysis, the tool was not designed for such long queries. It is by default used to search for specific Facebook groups (in part a response to ethical concerns), not groups sharing links (an issue not explicitly mentioned or acknowledged in its site settings). Although link lists could be searched in several sets, combining these datasets would have created additional complications in comparability. As such, the assumed tool usage related to social media commentary, as identified through keywords or hashtags, and not infrastructural or material elements such as hyperlinks and their contents.

Underlying logics also played a key role in data cleaning where stop lists or sanitised versions were based on common wording and social media tagging markers such as @ or #. In comparison, they rarely considered common hyperlink parts. Tools typically did not deploy mechanisms for deleting terms such as www, http://, or domains. This created a considerable obstacle in the word trees where Web site content constituted more common word stems (particularly in the Facebook comments). The frequent embedding of the URLs subsequently needed to be mitigated through further manual cleaning processes. Even so, it did not entirely remove issues encountered in long-form data as Voyant would not load larger corpuses, a deficit only discovered through reiterative loading of differently sized corpuses for testing. These restrictions were not acknowledged as part of the otherwise very extensive tool guidance — a black box issue. Thus, while some tools (*e.g.*, Voyant) offered sanitised text options for URLs, these tend to be fairly new, experimental, and/or require additional technical knowledge.

While word count restrictions may be purely pragmatic (rather than a deliberate preference towards social media logics), they still cater more to short-form media, privileging, above all, data from micro-blogging sites over long-form data (*e.g.*, Web page contents, articles, long-form blogs), even when these are shared and used within such sites. This issue is exacerbated by limitations in free software versions. A wide range of free versions of monitoring tools (including Buzzsumo and CrowdTangle) have word count and timeline restrictions, issues not always visible due to the black box of tool mechanics. As such, researchers lacking the necessary funding are by default limited to platforms that lend themselves to short queries, short-form data, and shorter timeframes, such as is famously the case with Twitter. In digital methods tools, keywords and hashtags (= default tools for organising data in social media logics) have long been the assumed default entry points. Although hyperlinks have been widely used in issue crawling, newer tools have shifted the focus to popular platforms such as Facebook and Twitter (*e.g.*, CrowdTangle) and consequently turned analytical attention to platform mechanics, such as likes and hashtags (*i.e.*, logics of prominent SNS) within these.

Similar issues appeared in visualisations where hyperlinks constituted a visually undesirable option due to their length and lack of informational insights. Although options existed for sanitising social media tagging markers in text-mining visualisations (see Figures 2, 3, and 4), technical options for extracting Web site titles, word stems, or shortening them for visualisation were largely absent. Readable and meaningful visualisations required contextual knowledge of the individual platforms and sub-pages, as well as the URL authors and/or source pages (*e.g.*, investment, news, or movement pages addressing greenwashing). While visualisations may generally fail to convey complexity in networks, this was a particularly inhibiting factor for hyperlinks as they are not as easily identifiable or presentable as hashtags (*e.g.*, #greenwashing),

keywords, or @mentions (*e.g.*, targets of greenwashing such as Shell or Nutella).

Although biases may be expected in computational research, these processes highlighted the comparatively difficult, limiting, and time-consuming nature of *hyperlink-based (i.e., non-platform-centric) work within and across SNS* (rather than standalone hyperlink networks of Web sites of the Web 1.0 generation), where query formation, corpus loading, and cleaning were opaque time- and resource-intensive processes. This work required a high degree of technical creativity and manual adaptations for circumventing settings based on underlying popular social media logics. These restrictions by design (based on underlying social media logics and platform interests) consequently reinforce existing preferences and biases towards single-platform studies and either user networks or comments.

Platform biases

The greenwashing case additionally highlighted that platforms mattered in myriad ways, suggesting that research using social media platforms as entry points would produce significantly different results on a given case. Web site materials did not necessarily dictate or influence the frames in social media posts, *i.e.*, Web site contents were surprisingly not what made a difference in SNS (*i.e.*, not a significant variable in these situations). Instead, variations could be attributed to the distinct platform vernaculars, affordances, or user associations of these platforms that determine issue framing in social media commentary. Platform choice “mattered” (situationally) in actor and framing diversity (see [Figure 1](#)) and potential user journeys suggested by out-links (issue networks, see [Figures 2, 3, and 4](#)). This suggests that platform structures, affordances, and vernaculars carry significant weight in issue framing, even where the materials used to substantiate these claims are largely consistent. To some extent, this effect relates to the general purpose of hyperlinks, *e.g.*, as evidence, context, or discussion points. Even so, these dynamics suggest “platform identity” may potentially carry more weight in social media commentary than supporting materials, a confluence only made visible through the extended cross-platform analysis applied here. As such, platform-centric research carries the potential to reinforce tool biases.

Platform choice mattered across all datasets, particularly in terms of platform structure and logic. For example, the CrowdTangle analysis showed more source diversity on Facebook compared to Reddit (see diversity of content streams in Facebook graphs compared to Reddit graphs, based on number of colour-coded waves, [Figure 1](#)), which resulted in more content diversity. Reddit showed broader thematic discussions involving actors with different views such as links from “liberal” media being captured and published with different commentary on “conservative” or sceptical groups. While links from *The Guardian* were popular, they appeared in fewer forums (here subreddits), for example in the “sceptic”, “British Labour Party”, and “the collapse of civilization” theme groups. In comparison, *The Guardian* and other news pages were more widely shared on Facebook (on pages owned by these news agencies or aligning with their views), an issue in part tied to infrastructure. Facebook’s infrastructure is built on groups and pages, meaning news outlets have their own spaces in which to distribute/post their URLs, compared to Reddit which uses thematic streams called subreddits. As such, it constituted a more alternative and independent forum for sharing news, where site owners did not necessarily suggest or dictate content. While these infrastructural differences are in themselves well-known in the academic community, this contextualised approach provided a view into how these spatial differences affected hyperlink diversity and prominence, and consequently results obtained through non-platform-centric approaches.

These platform logics also became a hindrance for obtaining cross-platform and contextual data. CrowdTangle provided access to social media commentary on Facebook, but Twitter data was limited to one week (an API restriction). For Twitter, other options would have been needed — either commercial, academic, or platform-provided. Although this was considered, issues existed in the availability, access, and comparability of these data. Although Twitter access for academics has recently been granted by the platform, the different ownership meant that different tool skills were needed to acquire the data. Even with skills present, the black box issue persisted. Given that two separate black boxes would have existed (for CrowdTangle and the Twitter Developer Platform), it would not have been possible to evaluate the similarity or comparability of the data collected across tools. As a result, more contextualised data on the

use of the greenwashing hyperlinks could not be obtained or compared across platforms. While this may seem natural given the platform provider's interest, it also meant that platform provider preferences shaped research design, limiting creative and cross-platform approaches.

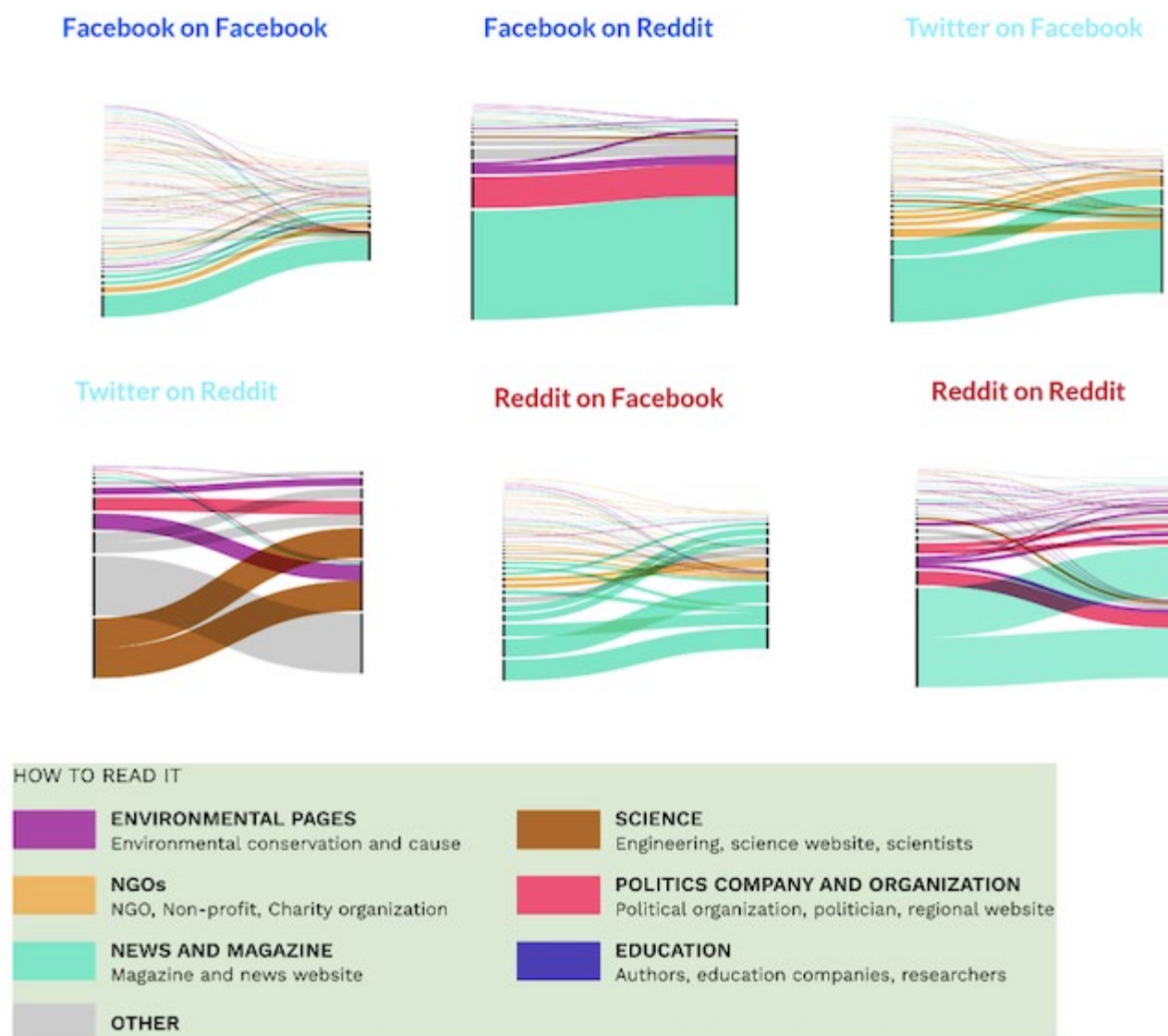


Figure 1: Diverging Facebook & Reddit commentary by engagement on Facebook, Twitter, & Reddit (Digital Methods Initiative (DMI) Wiki Greenwashing, 2021). Design by Matteo Bettini, Valentina Pallacci, Fabiola Papini, and Mara Paula Vargas Triana.

In the IssueCrawler network graphs, platform logics also mattered as a variable for the wider URL networks. The network visualisations (Figures 2, 3, and 4) show the interlinkages between URLs that are linked to from pages that the seed pages link to (*i.e.*, two clicks away from the seed). These outlinks shape user journeys and the contents they potentially consume. A comparison by hyperlink list (= lists ranked by engagement on Twitter, Facebook, and Reddit) showed differences in the types of content of the outlinked Web pages. The URL networks emerging from the three platform lists all linked to finance and investment

Web pages, but differences appeared in the smaller clusters. The URLs following from Reddit showed a cluster on left-leaning news (e.g., *The Nation*), whereas the URLs engaged with on Facebook showed a cluster of Republican news contradicting greenwashing accusations. Another difference was that URLs engaged with on Twitter included a large cluster of Chinese platforms (e.g., *Weibo*, *China Daily*, *FT Chinese*), while Facebook URLs primarily linked to U.S. pages, and Reddit to U.S. and Korean pages (e.g., *Korean Times*). As such, despite initially somewhat familiar hyperlink lists, larger differences emerged when following the URLs further away from the seeds into potential user journeys depending on the platform these URLs were mostly shared.

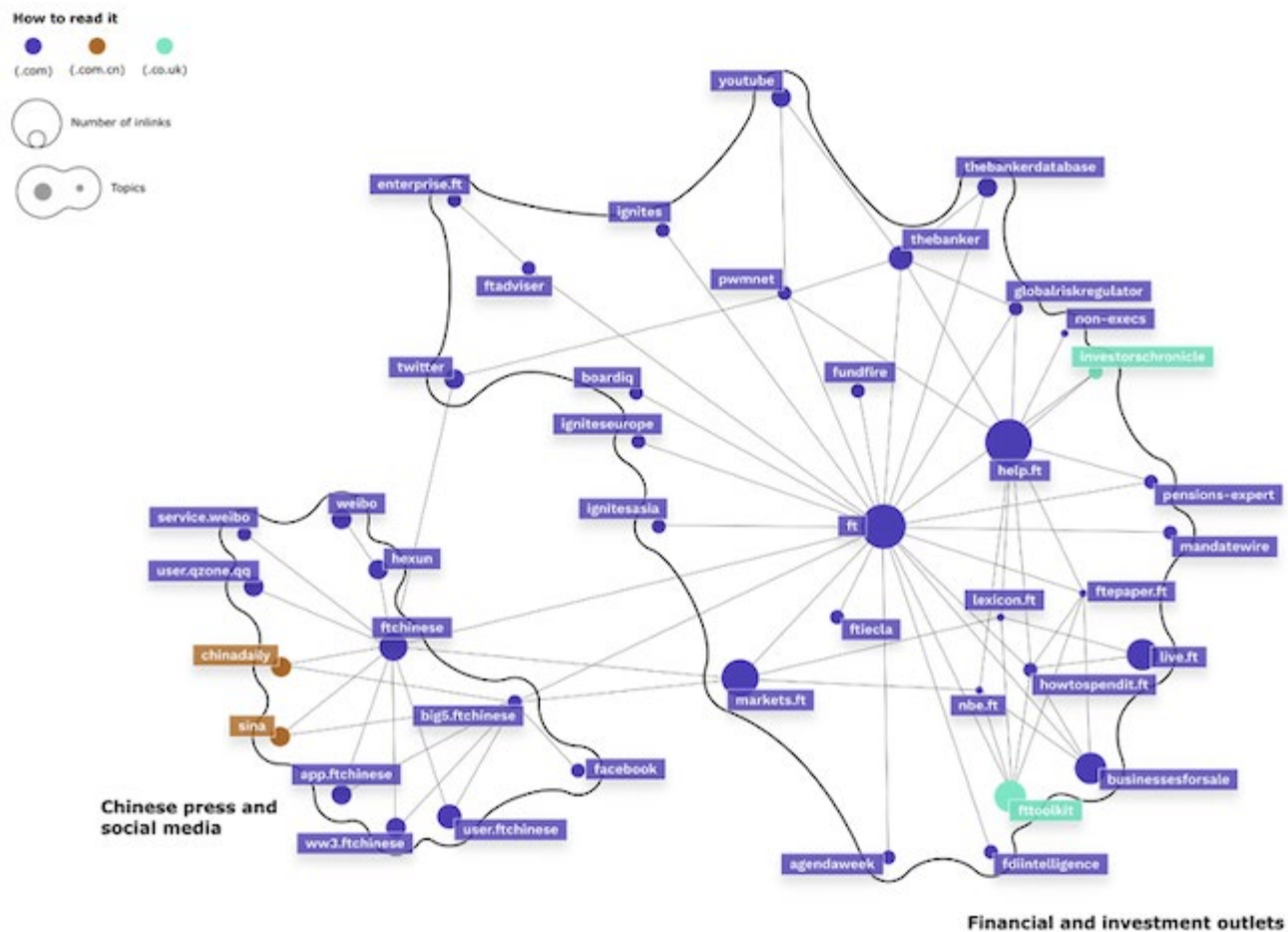


Figure 2: IssueCrawler Twitter Network.

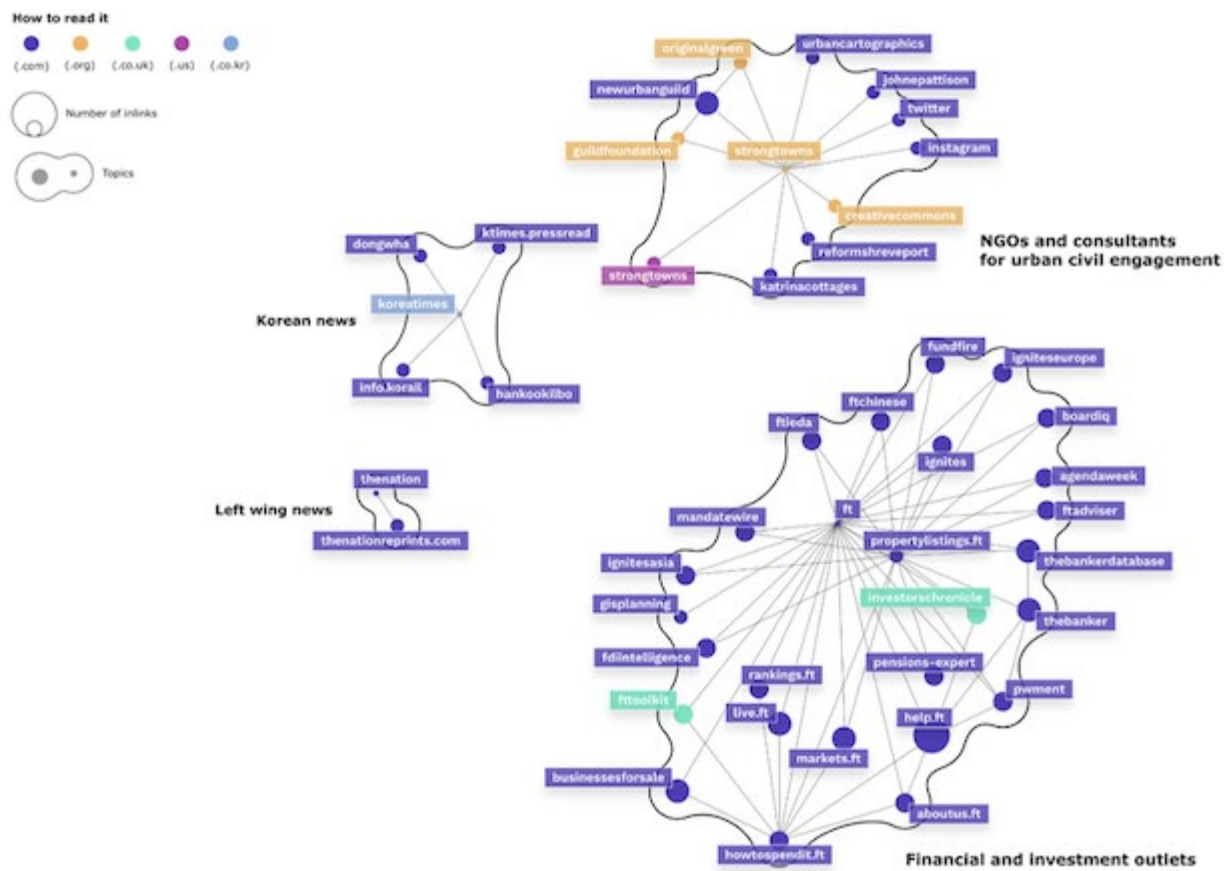


Figure 3: IssueCrawler Reddit Network.

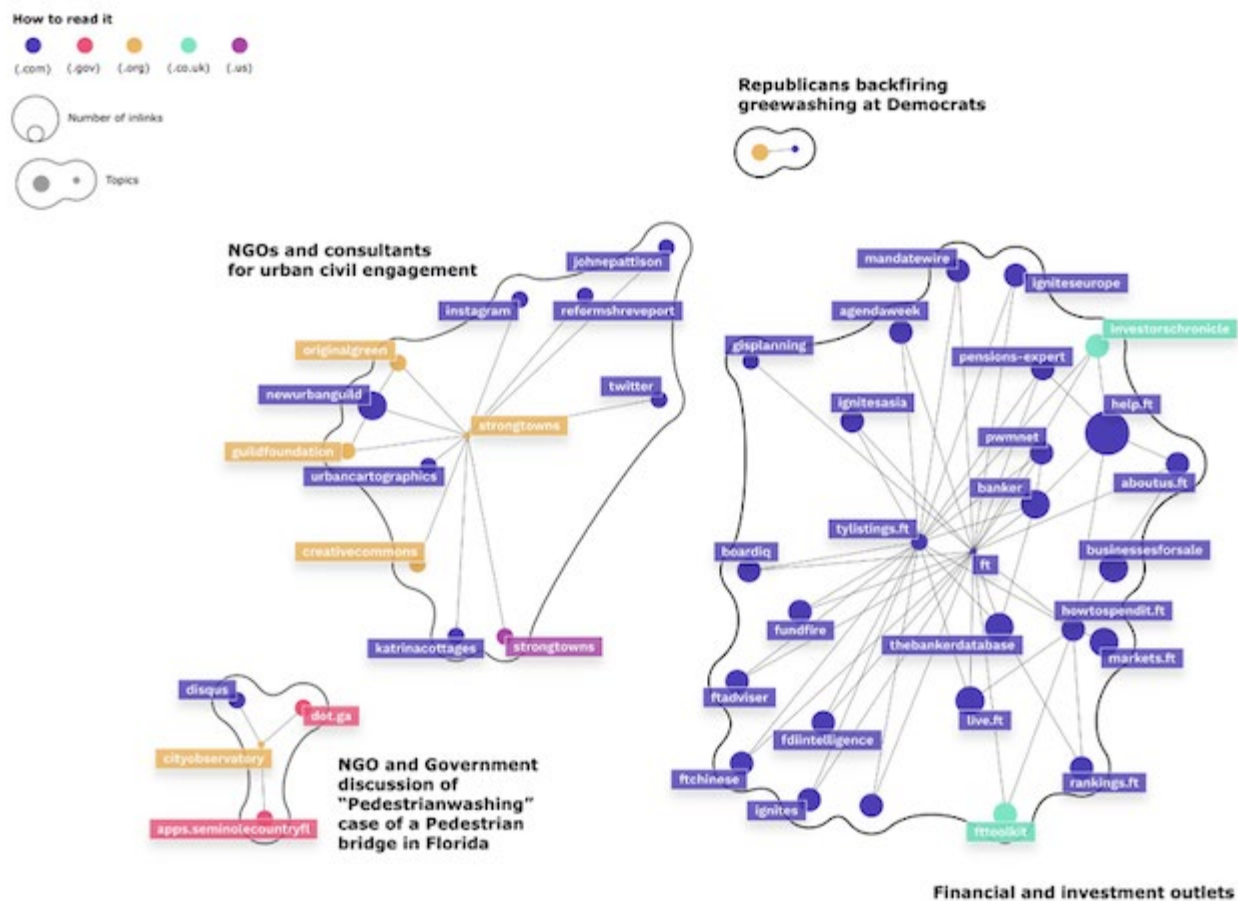


Figure 3: IssueCrawler Facebook Network.

Similar effects were observed in the word trees, where discursive differences appeared in Facebook commentary surrounding the hyperlinks. While differences in commentary are generally expected to be found in comparative or cross-platform research (indeed, they constitute the very reason for it), the differences here emerged more so from the contextual framing (*i.e.*, comments) across different Facebook spaces as the top links were similar across the lists, but the Facebook commentary around them showed more discrepancies (by platform: Facebook, Twitter, and Reddit). Thus, while the top URLs within the SNS (and their respective contents) were fairly similar, differences existed in the wider URL networks (out-links), the precise spaces where links were shared, and the surrounding social media commentary. While the Facebook commentary itself did not allow for capturing differences in commentary across platforms, the platform’s versatility in terms of actors, pages, and groups (see [Figure 1](#)) meant that significantly greater diversity existed in how the hyperlinks were framed than in their contents.

Thus, URLs’ *Web* site contents did not dictate issue framing in social media posts, but the choice of platform where these URLs were shared *did*, *i.e.*, platform choice “mattered” as a variable. In terms of digital bias, this means that platform-centric research is not only subject to differing platform affordances, dynamics, and vernaculars, but these also produce qualitatively different datasets — even when the subject of analysis (here: URL lists and their Web contents) are highly similar. As such, a focus on specific platforms and their data reinforces social media logics already inscribed into the tools (*i.e.*, tool bias) and therefore *pre-existing biases* as platforms themselves constitute significant variables.

Discussion and conclusion


This paper provided methodological reflections on an extended cross-platform analysis using URLs shared within SNS. Limitations were tied to settings that did not work well for hyperlink-based research in and across SNS. URLs could either not be entered as queries, were not considered to be an intended starting point for social media research, or site settings were not conducive to URL research. An account of these limitations and platform dynamics was used for a dual purpose: the paper (1) emphasised that tool design often falls prey to issues in representation (here: a focus on specific & interest-driven social media logics), *reinforcing* existing biases; and, (2) argued that non-platform-specific creative situational approaches such as this cross-platform URL project provide a much-needed understanding of wider platform dynamics and their implicit biases. In doing so, this paper adds to the repertoires of literature identifying issues of digital bias (e.g., Marres, 2017; Hargittai, 2020) and methodological work exploring cross-platform, ecological, and situational approaches (e.g., Marres, 2020; Pearce, *et al.*, 2020; Zuckerman, 2021; Rogers, 2018; Venturini, *et al.*, 2018), above all in relation to research design.

Whether these findings apply across the spectrum of digital methods tools remains to be seen. While a range of tools rely on URL input for pragmatic tasks such as shortening URLs, completing keyword strings, visualising data, or as a social media starting page, URLs often do not constitute the target of analysis beyond early Web site network analyses (even though they constitute a significant part in how social media contents are created, discussed, and linked). Instead, a wide array of popular tools from diverse sources appear to focus on social media commentary and/or networks, above all Twitter (e.g., Node XL, Tweetdeck, Mozdeh, Web Data Research Assistant, DiscoverText, and as identified through keywords and hashtags). Despite the popularity of tools that use hyperlinks for analysis, they are typically treated not as objects that form a foundation or network within SNS (except for link analysis), but rather as starting points for keyword-based analysis. To an extent, these trends reflect technological changes, as (static) Web sites represent early *Internet* environments (often dubbed *Web 1.0*) and social media more so later dynamic media systems (*Web 2.0*), phases that are not mutually exclusive, but may be treated as separate. Thus, while tool usage is difficult to trace, the prevalence of social media scraping tools and the (often criticised) focus on specific platforms suggest that digital social research is increasingly turning towards the application of popular social media logics, turning invisible some of their foundational elements: URLs.

Arguably, the preference of SNS and their underlying logics over alternatives reflects current trends. Tool design may merely be reflecting the popularity, extent of usage, and/or increasing social significance of specific social media platforms, such as, for example, Twitter and Facebook (e.g., around surveillance advertising). In comparison, long-form Web sites are often perceived as an earlier form of digital technologies (e.g., webometrics). Difficulties in attaining access and the underlying power dynamics potentially also make platforms more significant as research spaces as they address current societal issues such as platform power. As such, the dominance of certain media spaces in research may be founded on other criteria that justify these choices. Even so, such research potentially reinforces existing methodological biases in digital research, for example through the form of a self-fulfilling prophecy, in which more platform studies and platform-centric tools create a sense that this area is more relevant compared to other fields, and potentially creating the demand for similar tool design (creating a circular economy in research). These trends may also lead to certain elements of Internet infrastructures such as URLs becoming overlooked (within or outside of SNS), as they are considered an early and therefore potentially less current form of digital participation, despite their continued wide application within SNS.

This may mean that certain types of studies may fade over time, particularly if few viable alternatives exist. A focus on social media logics and the difficulty in pursuing creative approaches within tight budgets and time limits will likely lead researchers to predominantly focus on user commentary and dynamics. In comparison, it may limit research providing wider structural or cross-platform overviews, critically

analysing technological infrastructures, and reflecting more broadly on representativeness, such as the URL-based greenwashing project. In combination with other shaping factors such as API restrictions and provider-led tool design, this means that researchers may have little wriggle room for alternative approaches (at least in computational research). Given that API restrictions and platform-centric design favour social media logics and single-platform research, it appears unlikely that researchers would be able to escape tool and platform biases without intervention.

Going forward, we therefore hope (like Marres, 2020) to see the application of more situational and ecological approaches in digital social research, particularly compared to single-platform research and movement-/ space-/ or actor-centred approaches (see Moats, 2019). Such approaches would be particularly useful in revealing how new opportunities in digital social research such as platform-provided or -driven tool design may be affecting current research. Some of these concerns may be addressed by scholarly communities contributing to tool repertoires, for which we are immensely thankful. The larger issue, however, is tied to researchers' reliance on commercial organisations and tools provided by these. While the provision of new commercial tools and access options has been seen as a significant turn in research access, it remains questionable how much freedom these changes actually provide. This becomes particularly an issue where these tools are not only developed by commercial entities, but by the platform's providers themselves, such as was the case with CrowdTangle. These seemingly innocuous exclusive opportunities shape research in providers' interests, limiting comparable cross-platform research, and, when no other viable access options exist, make researchers reliant on them. Thus, beyond issues in software capacity or flexibility, tool settings and rules for certain platforms are determined by the tool providers. As such, beyond issues of API access, they may well govern the type of research we produce in the future. 

About the authors

Suay Melisa Özkula is a Marie Skłodowska-Curie Research Fellow in the School of International Studies at the Università degli Studi di Trento, Trento, Italy.

Direct comments to: suaymelisa [dot] ozkula [at] unitn [dot] it

Maria Lompe is a Ph.D. candidate in the Doctoral School of Humanities, Theology and Art, Nicolaus Copernicus University, Poland and a Research Fellow at the Collaborative Research Center "Media of Cooperation" at Siegen University, Germany.

E-mail: maria [dot] lompe [at] doktorant [dot] umk [dot] pl

Mariangela Vespa is a Marie Curie Early Stage Researcher in the Department of Environmental Psychology, Institute for Future Energy and Material Flow Systems (IZES gGmbH), Saarbrücken, Germany and a Ph.D. student at Universität des Saarlandes, Germany.

E-mail: vespa [at] izes [dot] de

Estrid Sørensen is a Professor of Cultural Psychology and Anthropology of Knowledge in the Department of Social Science at Ruhr-University Bochum, Germany.

E-mail: estrid [dot] sorensen [at] rub [dot] de

Tianshi Zhao is a graduate student at the University of Pennsylvania's Graduate School of Education.

E-mail: ztianshi [at] upenn [dot] edu

Acknowledgements

S.M.Ö. is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 101027274. Maria Lompe (ML) is supported by Doctoral School of Humanities, Theology and Art of Nicolaus Copernicus University and by

Deutsche Bundesstiftung Umwelt (DBU) Fellowship for Central and Eastern Europe (MOE), grant number 30021/933-46. Mariangela Vespa (MV) is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie action grant agreement number 813837. The article is based on a data sprint at the Digital Methods Initiative Summer Certificate Program, University of Amsterdam. We would like to offer our sincere thanks to the co-ordinators Prof. Richard Rogers, Prof. Sabine Niederer, Dr. Esther Weltevrede, and Guillén Torres Sepulveda; to Prof. Richard Rogers for his guidance in research design; to Dr. Stijn Peeters for instrumental technical support; to Matteo Bettini, Valentina Pallacci, Fabiola Papini, and Maria Paula Vargas Triana for their invaluable design assistance; to Dr. Stefan Laser as creator/lead of the wider project. The authors are grateful to readers and reviewers for their helpful comments, especially Dr. Paul Reilly and Dr. Stefania Vicari.

This article is the result of joint work by all authors based on a joint project during a data sprint. All authors engaged in the project development, data collection, as well as the critical discussions presented here. The research protocols were jointly written; all authors engaged in the overall editing and theoretical development of the paper; S.M.Ö. wrote the majority of introduction, research context, methodological reflections, and conclusion; M.L. wrote sections on CrowdTangle and parts of the methodological reflections; M.V. wrote parts of the methodological reflections; E.S. wrote the sections on IssueCrawler; T.Z. wrote the sections on WordTree.

Notes

- [1.](#) SNS here describe a type of social media focusing on social networks/-ing (*e.g.*, Twitter), compared to, for example, bookmarking services (*e.g.*, Pinterest) or mobile instant messaging services (*e.g.*, WhatsApp, Telegram; see Mosca and Quaranta, 2021), although these distinctions have in recent years been waning.
- [2.](#) Venturini, *et al.*, 2018, p. 4,207.
- [3.](#) Van Dijck and Poell, 2013, p. 5.
- [4.](#) See Marres, 2015, p. 667; Rogers, 2019.
- [5.](#) Marres, 2020, p. 14.
- [6.](#) Venturini, *et al.*, 2018, p. 4,199.
- [7.](#) Venturini, *et al.*, 2018, p. 4,200.
- [8.](#) Venturini, *et al.*, 2018, p. 4,198.
- [9.](#) Marres, 2015, pp. 664–665.
- [10.](#) This choice also reflects the nature of the chosen methods forums, which guided research design.
- [11.](#) Gatti, *et al.*, 2019, p. 1.
- [12.](#) de Freitas Netto, *et al.*, 2020, p. 2, based on Delmas.
- [13.](#) Rogers, 2019, p. 43.
- [14.](#) Venturini, *et al.*, 2018, p. 4,208.

References

- A. Beaulieu, 2005. "Sociable hyperlinks: An ethnographic approach to connectivity," In: C. Hine (editor). *Virtual methods: Issue in social research on the Internet* eds. Oxford: Berg, pp. 183–198.
doi: <https://doi.org/10.5040/9781474215930.ch-013>, accessed 14 May 2022.
- G. Blank, 2017. "The digital divide among Twitter users and its implications for social research," *Social Science Computer Review*, volume 35, number 6, pp. 679–697.
doi: <https://doi.org/10.1177/0894439316671698>, accessed 14 May 2022.
- E. Borra and B. Rieder, 2014. "Programmed method: Developing a toolset for capturing and analyzing tweets," *Aslib Journal of Information Management*, volume 66, number 3, pp. 262–278.
doi: <https://doi.org/10.1108/AJIM-09-2013-0094>, accessed 14 May 2022.
- A. Bruns, 2019. "After the 'APicalypse': Social media platforms and their fight against critical scholarly research," *Information, Communication & Society*, volume 22, number 11, 1,544–1,566.
doi: <https://doi.org/10.1080/1369118X.2019.1637447>, accessed 14 May 2022.
- A. Bruns and J. Burgess, 2015. "Twitter hashtags from ad hoc to calculated publics," In: N. Rambukkana (editor). *Hashtag publics: The power and politics of discursive networks*. Bern: Peter Lang, pp. 13–27.
- J. Burgess and A. Bruns, 2015. "Easy data, hard data: The politics and pragmatics of Twitter research after the computational turn," In: G. Langlois, J. Redden, and G. Elmer (editors). *Compromised data: From social media to big data*. London: Bloomsbury Academic, pp. 93–111.
doi: <https://doi.org/10.5040/9781501306549.0010>, accessed 14 May 2022.
- A.E. Clarke, 2003. "Situational analyses: Grounded theory mapping after the postmodern turn," *Symbolic Interaction*, volume 26, number 4, pp. 553–576.
doi: <https://doi.org/10.1525/si.2003.26.4.553>, accessed 14 May 2022.
- C. d'Andrea and A. Mintz, 2019. "Studying the live cross-platform circulation of images with computer vision API: An experiment based on a sports media event," *International Journal of Communication*, volume 13, at <https://ijoc.org/index.php/ijoc/article/view/10423>, accessed 14 May 2022.
- S.V. de Freitas Netto, M.F.F. Sobral, A.R.B. Ribeiro, and G.R. da Luz Soares, 2020. "Concepts and forms of greenwashing: A systematic review," *Environmental Sciences Europe*, volume 32, number 1, article number 19.
doi: <https://doi.org/10.1186/s12302-020-0300-3>, accessed 14 May 2022.
- Digital Methods Initiative (DMI) Wiki Greenwashing, 2021. "Greenwashing, in authenticity & protest. Following the dynamics and relations of an ambiguous term on Twitter and the Web (Group 2)," at <https://wiki.digitalmethods.net/Dmi/DmiSummerSchoolGreenwashing>, accessed 14 May 2022.
- K. Driscoll and S. Walker, 2014. "Working within a black box: Transparency in the collection and production of big Twitter data," *International Journal of Communication*, volume 8, pp. 1,745–1,764, and at <https://ijoc.org/index.php/ijoc/article/view/2171/1159>, accessed 14 May 2022.
- A.J. Flanagan, 2020. "The conduct and consequence of research on digital communication," *Journal of Computer-Mediated Communication*, volume 25, number 1, pp. 23–31.
doi: <https://doi.org/10.1093/jcmc/zmz019>, accessed 14 May 2022.
- M. Garrido and A. Halavais, 2003. "Mapping networks of support for the Zapatista Movement: Applying social network analysis to study contemporary social movements," In: M. McCaughey and M. Ayers (editors). *Cyberactivism: Online activism in theory and practice*. New York: Routledge, pp. 165–184.
doi: <https://doi.org/10.4324/9780203954317>, accessed 14 May 2022.

- L. Gatti, P. Seele, and L. Rademacher, 2019. “Grey zone in — greenwash out: A review of greenwashing research and implications for the voluntary-mandatory transition of CSR,” *International Journal of Corporate Social Responsibility*, volume 41, number 1, article number 6.
doi: <https://doi.org/10.1186/s40991-019-0044-9>, accessed 14 May 2022.
- C. Gerlitz and B. Rieder, 2013. “Mining one percent of Twitter: Collections, baselines, sampling,” *M/C Journal*, volume 16, number 2.
doi: <https://doi.org/10.5204/mcj.620>, accessed 14 May 2022.
- M. Gibbs, J. Meese, M. Arnold, B. Nansen, and M. Carter, 2015. “#Funeral and Instagram: Death, social media, and platform vernacular,” *Information, Communication & Society*, volume 18, number 3, pp. 255–268.
doi: <https://doi.org/10.1080/1369118X.2014.987152>, accessed 14 May 2022.
- E. Hargittai, 2020. “Potential biases in big data: Omitted voices on social media,” *Social Science Computer Review*, volume 38, number 1, pp. 10–24.
doi: <https://doi.org/10.1177/0894439318788322>, accessed 14 May 2022.
- A. Helmond, 2015. “The platformization of the Web: Making Web data platform ready,” *Social Media + Society* (30 September).
doi: <https://doi.org/10.1177/2056305115603080>, accessed 14 May 2022.
- A. Hino and R.A. Fahey, 2019. “Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints,” *International Journal of Information Management*, volume 48, pp. 175–184.
doi: <https://doi.org/10.1016/j.ijinfomgt.2019.01.019>, accessed 14 May 2022.
- S. Horawalavithana, K.W. Ng, and A. Iamnitshi (editors), 2020. “Twitter is the megaphone of cross-platform messaging on the white helmets,” In: R. Thomson, H. Bisgin, C. Dancy, A. Hyder, and M. Hussain (editors). *Social, cultural, and behavioral modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings. Lecture Notes in Computer Science*, volume 12268. Cham, Switzerland: Springer, pp. 235–244.
doi: https://doi.org/10.1007/978-3-030-61255-9_23, accessed 14 May 2022.
- W.T. Howe, Jr. and R.S. Bisel, 2020. “Third-party online organizational reviews: Explaining review-and-rating patterns of the United States military and large corporate organizations,” *Computers in Human Behavior Reports*, volume 1, 100006.
doi: <https://doi.org/10.1016/j.chbr.2020.100006>, accessed 14 May 2022.
- K. Knorr Cetina, 2009. “The synthetic situation: Interactionism for a global world,” *Symbolic Interaction*, volume 32, number 1, pp. 61–87.
doi: <https://doi.org/10.1525/si.2009.32.1.61>, accessed 14 May 2022.
- D.G. Lorentzen and J. Nolin, 2017. “Approaching completeness: Capturing a hashtagged Twitter conversation and its follow-on conversation,” *Social Science Computer Review*, volume 35, number 2, pp. 277–286.
doi: <https://doi.org/10.1177/0894439315607018>, accessed 14 May 2022.
- N. Marres, 2020. “For a situational analytics: An interpretative methodology for the study of situations in computational settings,” *Big Data & Society* (2 December).
doi: <https://doi.org/10.1177/2053951720949571>, accessed 14 May 2022.
- N. Marres, 2017. *Digital sociology: The reinvention of social research*. Cambridge: Polity.
- N. Marres, 2015. “Why map issues? On controversy analysis as a digital method,” *Science, Technology, &*

Human Values, volume 40, number 5, pp. 655–686.

doi: <https://doi.org/10.1177/0162243915574602>, accessed 14 May 2022.

D. Moats, 2019. “From media technologies to mediated events: A different settlement between media studies and science and technology studies,” *Information, Communication & Society*, volume 22, number 8, pp. 1,165–1,180.

doi: <https://doi.org/10.1080/1369118X.2017.1410205>, accessed 14 May 2022.

L. Mosca and M. Quaranta, 2021. “Are digital platforms potential drivers of the populist vote? A comparative analysis of France, Germany and Italy,” *Information, Communication & Society*, volume 24, number 10, pp. 1,441–1,459.

doi: <https://doi.org/10.1080/1369118X.2021.1894211>, accessed 14 May 2022.

S.M. Özkula, P.J. Reilly, and J. Hayes, 2022. “Easy data, same old platforms? A systematic review of digital activism methodologies,” *Information, Communication & Society* (9 January).

doi: <https://doi.org/10.1080/1369118X.2021.2013918>, accessed 14 May 2022.

H.W. Park and M. Thelwall, 2003. “Hyperlink analyses of the World Wide Web: A review,” *Journal of Computer-Mediated Communication*, volume 8, number 4.

doi: <https://doi.org/10.1111/j.1083-6101.2003.tb00223.x>, accessed 14 May 2022.

W. Pearce, S.M. Özkula, A.K. Greene, L. Teeling, J.S. Bansard, J.J. Omena, and E.T. Rabello, 2020. “Visual cross-platform analysis: Digital methods to research social media images,” *Information, Communication & Society*, volume 23, number 2, pp. 161–180.

doi: <https://doi.org/10.1080/1369118X.2018.1486871>, accessed 14 May 2022.

R. Rogers, 2019. *Doing digital methods*. London: Sage.

R. Rogers, 2018. “Otherwise engaged: Social media from vanity metrics to critical analytics,” *International Journal of Communication*, volume 12, pp. 450–472, and at

<https://ijoc.org/index.php/ijoc/article/view/6407/2248>, accessed 14 May 2022.

T. Scholz, 2010. “Infrastructure: Its transformations and effect on digital activism,” In: M. Joyce (editor). *Digital activism decoded: The new mechanics of change*. New York: International Debate Education Association, pp. 17–32, and at <https://www.opensocietyfoundations.org/publications/digital-activism-decoded-new-mechanics-change>, accessed 14 May 2022.

E. Shearer and A. Mitchell, 2021. “News use across social media platforms in 2020,” *Pew Research Center* (12 January), at <https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/>, accessed 14 May 2022.

M. Thelwall, 2012. “A history of webometrics,” *Bulletin of the American Society for Information Science and Technology*, volume 38, number 6, pp. 18–123.

doi: <https://doi.org/10.1002/BULT.2012.1720380606>, accessed 14 May 2022.

Y. Theocharis, A. Cardenal, S. Jin, T. Aalberg, D.N. Hopmann, J. Strömbäck, L. Castro, F. Esser, P. Van Aelst, C. de Vreese, N. Corbu, K. Koc-Michalska, J. Matthes, C. Scherer, T. Sheaffer, S. Splendore, J. Stanyer, A. Stpińska, and V. Štětka, 2021. “Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries,” *New Media & Society* (9 October).

doi: <https://doi.org/10.1177/14614448211045666>, accessed 14 May 2022.

R. Tromble, 2021. “Where have all the data gone? A critical reflection on academic digital research in the post-API age,” *Social Media + Society* (19 January).

doi: <https://doi.org/10.1177/2056305121988929>, accessed 14 May 2022.

Z. Tufekci, 2014. “Big questions for social media big data: Representativeness, validity and other methodological pitfalls,” *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, pp. 505–514, and at <https://ojs.aaai.org/index.php/ICWSM/article/view/14517/14366>, accessed 14 May 2022.

J. van Dijck and T. Poell, 2013. “Understanding social media logic,” *Media and Communication*, volume 1, number 1.
doi: <https://doi.org/10.17645/mac.v1i1.70>, accessed 14 May 2022.

T. Venturini, L. Bounegru, J. Gray, and R. Rogers, 2018. “A reality check (list) for digital methods,” *New Media & Society*, volume 20, number 11, pp. 4,195–4,217.
doi: <https://doi.org/10.1177/1461444818769236>, accessed 14 May 2022.

S. Vicari and D. Kirby, 2022. “Digital platforms as socio-cultural artifacts: Developing digital methods for cultural research,” *Information, Communication & Society* (30 January).
doi: <https://doi.org/10.1080/1369118X.2022.2027498>, accessed 14 May 2022.

E. Zuckerman, 2021. “Why study media ecosystems?” *Information, Communication & Society*, volume 24, number 10, pp. 1,495–1,513.
doi: <https://doi.org/10.1080/1369118X.2021.1942513>, accessed 14 May 2022.

Editorial history

Received 25 March 2022; revised 28 April 2022; accepted 4 May 2022.



This paper is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

When URLs on social networks become invisible: Bias and social media logics in a cross-platform hyperlink study

by Suay Melisa Özkula, Maria Lompe, Mariangela Vespa, Estrid Sørensen, and Tianshi Zhao.

First Monday, volume 27, number 6 (June 2022).

doi: <https://dx.doi.org/10.5210/fm.v27i6.12568>