ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/271707566

A methodology for mapping Instagram hashtags

Article *in* First Monday · January 2015

DOI: 10.5210/fm.v20i1.5563

CITATIONS
245

READS
8,194

2 authors, including:

Tama Leaver
Curtin University

93 PUBLICATIONS **1,418** CITATIONS

SEE PROFILE

All content following this page was uploaded by Tama Leaver on 06 February 2015.

The user has requested enhancement of the downloaded file.

First Monday, Volume 20, Number 1 - 5 January 2015

PEER-REVIEWED JOURNAL ON THE INTERNET

A methodology for mapping Instagram hashtags

by Tim Highfield and Tama Leaver

Abstract

While social media research has provided detailed cumulative analyses of selected social media platforms and content, especially Twitter, newer platforms, apps, and visual content have been less extensively studied so far. This paper proposes a methodology for studying Instagram activity, building on established methods for Twitter research by initially examining hashtags, as common structural features to both platforms. In doing so, we outline methodological challenges to studying Instagram, especially in comparison to Twitter. Finally, we address critical questions around ethics and privacy for social media users and researchers alike, setting out key considerations for future social media research.

Contents

Introduction
Social media research context
Twitter studies
Hashtags
Instagram: Challenges and questions
Ethics and privacy beyond the binary
Conclusion

Introduction

Social media, as platforms for content sharing, information diffusion, and publishing thoughts and opinions, have been the subject of a wide range of studies, examining online activity within fields from politics and media to health and crisis communication. However, for various reasons, some platforms are more widely represented in research to date than others, particularly when examining large-scale activity captured through automated processes, or datasets reflecting the wider trend towards 'big data'. Facebook, for instance, as a closed platform with different privacy settings available for its users, has not been subject to the same extensive quantitative and mixed methods studies as other social media, such as Twitter. Indeed, the accessibility of Twitter has led to it being one of the most studied social media platforms across myriad contexts: the strict character limit for tweets and the common functions of hashtags, replies, and retweets, as well as the more public nature of posting on Twitter, mean that the same processes can be used to track and analyse data collected through the Twitter API, despite presenting very different subjects, languages, and contexts (Bruns, *et al.*, 2012; Moe and Larsson, 2013; Papacharissi and de Fatima Oliveira, 2012).

Building on the research carried out into Twitter, this paper outlines emerging methods to study uses and activity on the image-sharing app and social media platform Instagram. While the content of the two social media platforms is dissimilar — mainly short textual comments versus images and video — there are some architectural parallels encouraging the extension of methods from one platform to another. The importance of tagging on Instagram, for instance, has conceptual and practical links to the hashtags employed on Twitter (and other social media and 'Web 2.0' platforms), with tags serving as markers for the main subjects, ideas, events, locations, or emotions featured in tweets and images alike. The Instagram Application Programming Interface (API) allows queries around user-specified tags, providing extensive information about relevant images and videos, similar to the results provided by the Twitter API for searches around particular hashtags or keywords. The information provided allows for the analysis of collected data to incorporate several different dimensions; for example, the information about the tagged images returned through the Instagram API allows us to examine patterns of use around publishing activity (time of day, day of the week), types of content (image or video), and locations specified around these particular terms.

This is an exploratory study, developing and introducing methods to track and analyse Instagram data; it builds upon the methods, tools, and scripts used by Bruns and Burgess (2011a) in their large-scale analysis of Twitter datasets. These processes allow for the filtering of the collected data based on time and keywords, and for additional metrics around time intervals and overall user contributions. Such tools allow us to identify quantitative patterns within the captured, large-scale datasets, which are then supported by qualitative examinations of the resulting filtered datasets.

Social media research context

Our exploration of Instagram occurs in the wider context of extended research into social media, from both qualitative and quantitative foundations. These take into account a wide range of social media platforms, including apps as well as Web sites, and studying a wealth of practices and functions, concepts and affordances. Research into social media has been carried out within and across numerous disciplines, further demonstrating the use of social media for multiple purposes and contexts (see, for instance, the perspectives represented in Weller, *et al.*, 2014). There has been a general trend, though, within Internet Studies and other, related fields, towards quantitative-driven, large-scale projects using automated processes to capture and analyse activity on social media platforms. Using APIs provided by the platforms, publicly available data is collected and processed in large numbers, with rich datasets gathered. Such projects provide the means for researchers to study extended online activity, uses of social media platforms and their evolution over time and across topics, and populations.

The prevalence of big data, aided by the availability of tools for capturing and processing large datasets, has been noted across disciplines, as have its limitations and pitfalls (see boyd and Crawford, 2012). Simply reporting metrics around activity per day or measuring the most active users or most-mentioned users and content, while an important output of the dataset, does not necessarily account for why this has happened, and may overlook reasons and behaviours influencing social media use. As boyd and Crawford (2012) note: "Too often, Big Data enables the practice of apophenia: seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions." [1] Of course, big data is not synonymous with quantitative analysis (nor small data with qualitative), and there are important insights to be garnered from large-scale analysis and tracking; we advocate for mixed methods, combining quantitative and qualitative tools and insights, to take full advantage and explore more thoroughly the rich datasets collected from social media.

Spending time studying data through qualitative means is also critical to understanding how social media are used in non-standard ways, identifying practices that might easily be missed through automated analyses. For Twitter, for instance, while retweets and replies can be measured and connections between users identified through following relationships and @mentions, there are several practices that imply connections without necessarily creating a formal, structural link. Tufekci (2014) outlines some of these practices, including sub-tweeting (discussing others without explicit mention of their name or Twitter handle) and screencapping (including screenshots of comments in tweets). Such practices might be overlooked in large-scale, automated analyses of Twitter data, however, and in the case of sub-tweeting the lack of context for the tweets might further mask their meaning. In such cases, the quantitative patterns might suggest one thing, while qualitative analysis finds alternative interpretations or phenomena at play — a connection between users does not automatically imply endorsement of their comments. For visual platforms such as Instagram, this is even more critical, for the textual and graphic components of a post each offer key information, and analysis needs to take into account both aspects — the text-driven focus of Twitter studies, for instance, can risk overlooking visual and mixed media within tweets, and developing methods for analysing images as well as text is an important direction for social media research.

So far, textual content has dominated social media research, particularly for large-scale analyses; this is in part due to the ease of collecting and processing text in comparison to images, especially in the case of Twitter where tweets conform to a strict maximum character limit. The promotion of tools for collecting data from Twitter, and the standardisation of methods for comparable research projects, has also made studying this platform more straightforward than others. Visual content, though, is an important aspect of social media activity, including on Twitter (see Hjorth and Burgess, 2014, for instance). There is a need for expanded social media research methods which more directly and consistently incorporate visual content, from collecting data from image sharing platforms to approaches to processing and analysing photographs, images, or videos. Preliminary research into Instagram has offered large-scale analyses of images from specific locations (Hochman and Manovich, 2013), but there are many user practices, shared experiences, and approaches to content that have not yet been researched in any significant detail.

Our approach here offers a transition between established methods for Twitter research and ways for studying newer platforms such as Instagram. As with other social media platforms, the Instagram API gives researchers access to metadata around public content, including photographs, videos, and comments. Given the development of extensive methods for analysing Twitter data (see Bruns and Burgess, 2011a, for instance), and the presence of shared content elements such as hashtags on both Twitter and Instagram, we use existing Twitter research as a starting point for preliminary Instagram research. This is designed for comparability of activity across the two platforms, as well as offering an initial means of examining Instagram data. There are other options for both querying the Instagram API, such as using location-specific queries, and for studying content; at this preliminary stage, we mostly focus on the metadata of Instagram content, rather than the images or videos themselves. However, this method outlines how hashtags may be used to define the scope of research into Instagram and the types of information available for such projects, which can then be supplemented by visual and content analysis of relevant media.

Twitter studies

For our mapping and tracking of social media use, focusing on Instagram, starting with established Twitter methods allows us to draw on a rich research literature exploring Twitter from many contexts and paradigms (for example, Weller, *et al.*, 2014). Research has focused in part on topical datasets, using similar methods to gather data around varied subjects, from breaking news (Vis, 2013), politics (Ausserhofer and Maireder, 2013), and crises (Bruns, *et al.*, 2012), to television broadcasts and popular culture (Giglietto and Selva, 2014; Highfield, *et al.*, 2013), and sports (Bruns, *et al.*, 2013). In these cases, despite the widely different topics in question and their contexts (including different countries, languages, and time periods), common approaches are used in order to identify relevant content.

The wealth of research into Twitter has allowed a standardisation of approaches for both collecting and analysing activity on this platform. There are numerous tools and services, including commercial on-sellers and open source, freely available applications, which researchers can use to obtain data, from large-scale, ongoing captures to snapshots and user-specific information. Although accessing the entirety of Twitter activity requires finances and infrastructure that are likely unavailable for most researchers, the results gathered from the Streaming and Search APIs, while only representing a percentage of the total activity (Morstatter, *et al.*, 2013), are still sufficiently rich data sources for analysis.

What has helped to foster standardisation for Twitter research is the consistency of Twitter data. In particular, the 140 character limit for tweets makes their capture — as small strings of text — and processing a more straightforward proposition than messages of more variable length, mixed media content, or images, videos, or audio content. Twitter's approach to public and private accounts also sets it apart from other social media. Unlike Facebook, for instance, where users might have several different settings for account privacy, Twitter users are either public or private: content is either visible to everyone (including people not signed in to Twitter) or only to approved users. Since tweets from private accounts are not retrieved by the Twitter API, and cannot be automatically retweeted by other users, tweets captured from the API for Twitter research are known to be publicly available comments (although they may of course be deleted after publication, but still be captured before deletion).

The structure of tweets also provides some consistency for analysis, beyond being short textual content. Practices have developed on Twitter, both created by the platform and user-led, which have become commonplace, from the use of hashtags to denote key subjects, contexts, or emotions, to conventions of mentioning other users. Again, the consistent form of hashtags and @mentions — and common, if not universal, approaches to retweeting — enable the use of automated processes to identify these elements within tweets. Figure 1 demonstrates the common elements of an individual tweet. Furthermore, though, a tweet is its own datapoint: once created, it exists in isolation. A response to the tweet in the form of a reply or retweet creates a new tweet on the part of the responding user, rather than appending a comment to the original tweet. While Twitter timelines

might show replies as more threaded conversations, the creation of unique tweets for each comment means that capturing Twitter activity has fewer concerns around dynamic actions: a tweet's follow-ups are independent of the first post. There are some metrics which are more dynamic, and which continue to change following data capture, such as the numbers of automated retweets and favourites a tweet receives; perhaps for this reason, these metrics are less extensively analysed than hashtag use and networks of @mentions and @replies within unfolding topical discussions.

Figure 1: Example tweet and selected elements.

While the advantages of consistent, public data, and established capture and analytics methods mean that Twitter is an appealing social media platform for research, it is important to remember that it is far from the only example of popular social media. The comparative ease of access to data and less blurry questions around public/private comments than for platforms such as Facebook may mean that, while Twitter is eminently examinable, it is over-represented in research — or indeed that certain practices are over-emphasised to the detriment of others (including approaches to communication on Twitter which ignore or circumvent common practice, as noted by Tufekci, 2014). Yet these same Twitter approaches give us a strong foundation for developing research methods for studying other social media, moving beyond short snippets of text to more dynamic, visual social media content. Other platforms create their own methodological challenges, but building on the wealth of research into Twitter allows us to start with comparative work that can evolve into standalone, more detailed platform-specific methods; rather than moving back to single-platform studies, though, the development of methods for gathering and analysing rich data for different platforms will afford the opportunity to carry out more comparative work, especially in the absence of overlapping conventions or common elements to provide an initial hook for comparisons.

For our initial work, though, common structural elements are crucial for setting up the application of Twitter methods to other platforms. Hashtags act as a point of departure for our research, due to their adoption on additional social media to Twitter, including Instagram, Facebook, and Pinterest (and as tags more generally on platforms predating Twitter, such as Flickr). On Twitter, the use of keywords and, especially, hashtags

(such as #ir15) is a primary means for finding topically relevant tweets. Tools such as yourTwapperKeeper use defined terms as search queries, accessing the Twitter API to find relevant content; searching for a specific hashtag, for instance, locates public tweets containing that marker. By including a hashtag in a tweet, a Twitter user is making an active decision, and this choice connects their post to other comments around a common topic, event, or theme. Although hashtags may represent different intentions on the part of the individual user and the platform chosen, their use still reflects particular social media affordances and practices, and are thus useful filters for defining the scope of research projects.

Hashtags

User tagging of entries, media and other information has been a central feature of many 'Web 2.0' and social media platforms for more than a decade (Bruns, 2008). As Adam Mathes (2004) explained in relation to the social bookmarking tool Delicious and photo-sharing platform Flickr, social tagging leads to a folksonomy — the combination of the words folk and taxonomy — indicating the collective social organisation and description of information at the metadata level. A folksonomy is emergent in the sense that it is always in the process of being generated by users and updated with new contributions. Importantly, and in contrast with a traditional taxonomy, a folksonomy does not follow a pre-determined organisation or indexical structure, but is, instead, a socially and collectively produced alternative. Tagging and folksonomies can be harnessed not just for the basic organisation of media and information, but via the tools and access provided by many platforms can lead of new social and technical ways of manipulating and harnessing data, such as detailed explorations of specific places enabled by the locative and tagging data created by Flickr users (Kennedy, *et al.*, 2007).

When Twitter was initially launched in 2006, there was no technical or social convention for replying to another user, for organising tweets together, or for indicating that a tweet was part of a broader topic. All of these affordances emerged from the social uses of Twitter's early users, with the '@' reply convention, retweeting, and, hashtags all becoming technically formalised in Twitter's architecture (Halavais, 2013). The use of the hash (or sometimes called the pound or number sign) character before certain terms indicated a desire to group tweets socially, or as Halavais explains, these are "... a way of indicating textually keywords or phrases especially worth indexing ... by using the # character to mark particular keywords, Twitter users communicate a desire to share particular keywords folksonomically" [2]. Examples of hashtags can include political conversations, including #auspol for Australian politics (Bruns and Highfield, 2013); discussions centred on television events such as #Eurovision for the live broadcast of the Eurovision song contest (Highfield, *et al.*, 2013); information, discussion and relief coordination during natural emergencies (Bruns, *et al.*, 2012); or even external discussions about a particular unit of study at university (Leaver, 2012).

As Bruns and Burgess (2011b) have argued, hashtags can allow certain types of communities to emerge and form, including ad hoc publics, forming and responding very quickly in relation to a particular event or topical issue. These publics or communities may not persist for long periods, but can be extremely efficient and significant even if only existing for a short time. Yet even on a single platform, since hashtags emerge socially, it can take time before a predominant agreed upon tag emerges for a given theme or event. This can be exacerbated by particular pressures, including the length of time individuals have been using a platform, an issue of significant note in relation to Twitter use during natural disasters (Potts, *et al.*, 2011).

It is also important to note that while some hashtags on Twitter may lead to the formation of publics or communities, many hashtags do not, and were not used with that intention in mind (Bruns and Burgess, 2011b). Thus ascribing intentionality to hashtags — presuming that using a tag is an indication that the user intended this tweet, or photo, to be grouped together meaningfully with other tweets or photos using that tag — cannot be taken as a given. Tags and hashtag use is not only platform-specific at times, but also contextually specific to the individual user. A photograph or tweet with the hashtags #longday #tired #cantwaitforbed may be used to enhance (or simply be) the caption or description of that photograph; it is unlikely the user purposefully intends that image or tweet to be grouped together with all other #tired or #cantwaitforbed instances. As we move our attention to the specifics of Instagram hashtags as a way of accessing and organising certain ideas, discussions or visualisations, we must keep in mind that while we can, and do, access Instagram photos using hashtags as a means of organisation, in many cases this affordance is likely to be more the focus of researchers, not individual users.

Instagram: Challenges and questions

As on September 2014, Instagram has 200 million monthly users, with 65 percent of those users outside of the United States, with over 20 billion photos shared, an average of 1.6 billion likes each day, and an average of 60 million photos posted per day (Instagram, 2014b). Like Twitter, Instagram offers researchers the opportunity to study how users document elements of their everyday lives, in this case in a predominately visual context, and how these are presented online. As a smartphone app, Instagram also represents the confluence of different practices and interests: the ubiquity of smartdevices, internet and mobile data connections, and cameras within these devices; the shared experiences promoted through social media, including the instant publication of statuses and images from the scene of the experience, tagging friends present, and liking and commenting on others' content; the performative aspects of social media activity, including online visual culture, from selfies (self portrait photographs) and images specifically indicating that they have not had additional filters added (see Rettberg, 2014); and the promotion of small, standardised bits of information — photos, 15 second videos, tweets — rather than extended commentary. While these practices are not limited to Instagram, for users share content on multiple platforms simultaneously, they highlight key factors impacting upon how Instagram is used and how to study user activity on the platform.

As noted previously, a common approach to Twitter research is to track specific hashtags or keywords over time, using the Twitter Search and Streaming APIs. The Instagram API provides a search hook dedicated solely to tags, and this offers an immediate comparative opportunity. Querying the Instagram API for selected tags then provides results akin to similar Twitter projects, although with different metadata, resulting in further methodological questions. These queries also do not archive the image or video content retrieved using the searches (which would be against the Instagram terms of service), but provide links to the media in question, which can then be verified by the researcher.

Figure 2 provides an example of a image shared on Instagram and its common elements, from user information to comments and tags. An Instagram API query provides a wealth of metadata for relevant media shared on the platform, extending beyond what might be immediately visible in Figure 2. For each media object matching the tag query, the API also returns not just its unique identifier (id) and links to the low and standard-resolution versions of the content (whether image or video), but also metadata including usernames, time and date of creation, caption, comments (and user and time information for comments), tags, likes, and location information when a user has geotagged their media. Such data allows for quantitative and qualitative analyses, whether counting the amount of content over time, users, or tags, mapping media based on location data, or looking at the content of the media and their captions, for example.

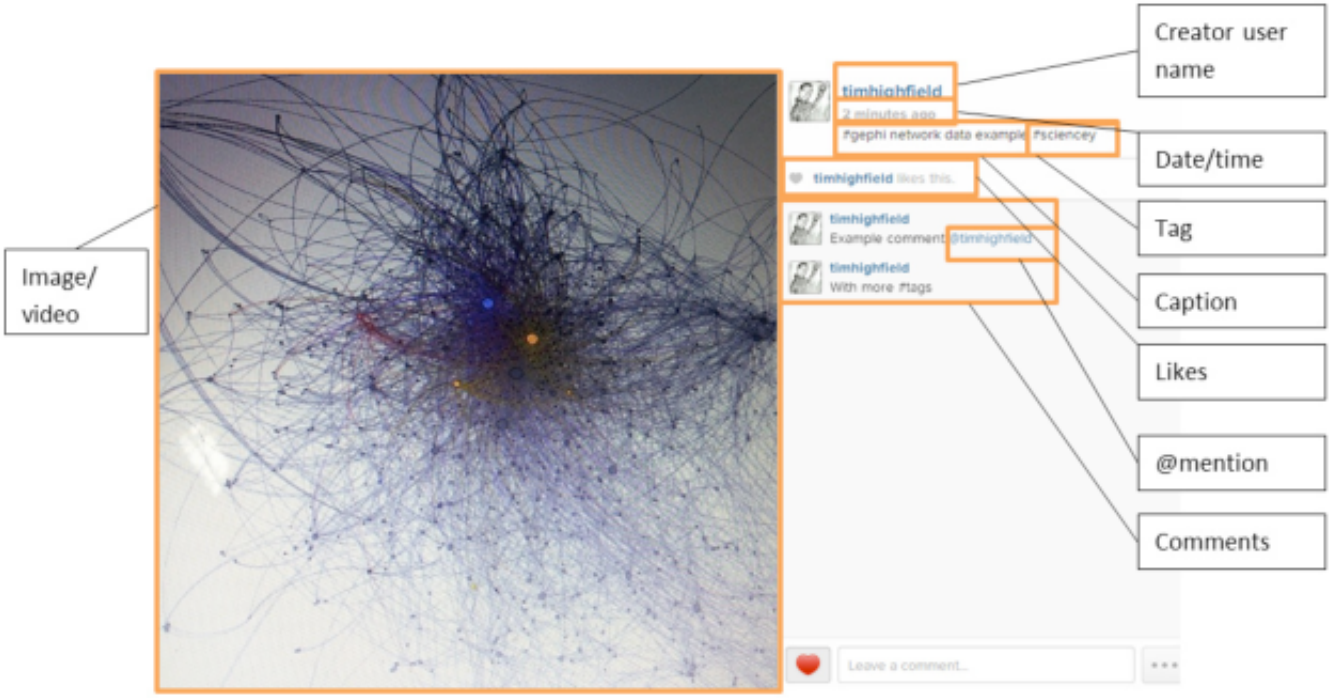

Figure 2: Example Instagram post and selected elements.

However, the nature of Instagram raises methodological questions without clear precedents from Twitter research. A tweet is a standalone data point: once published, it can only be deleted, not edited. Other users may reply to it, favourite it, or share it (whether by using the retweet button or creating a new tweet quoting the original), but for the most part these actions create new data points rather than altering the original — a reply or a retweet is a new tweet, not an extension of the original. Favouriting is the exception here, and the challenges capturing this action raises for automated capture of more dynamic data may account for its relative absence from much large-scale Twitter analysis (for methodological reasons as well as analytical — favouriting can be a less widely visible act in response to a tweet than either retweeting or replying).

Instagram content are more dynamic data points than tweets, though. Again, each image or video is its own data point. However, if a user responds to an image by leaving a comment, that becomes an addition to the original data point — it is part of the comments thread for that media, rather than being a distinct entity. This has two major implications for our research methodology: first, the challenge of tracking comments as part of user practices and responses to chosen topics, and second the impact comments have on querying the Instagram API.

The first concern has numerous possible solutions: filtering out comments and studying just the original media posted on Instagram, or selecting a

particular date range and only studying comments up to that point (as users may leave comments on media days, weeks, or years after it was originally posted), for example. Since media might attract many or no comments, there is no consistency between data points: while the results can be stored in a database, analysing variable comment threads which may change over the course of the data capture is a new methodological concern which does not affect Twitter research (although the tracking on conversations on Facebook pages, which have no technically enforced end point to conversation threads, provokes similar issues). Approaches for tracking Instagram activity can take Twitter methods and common elements as an inspiration, but addressing aspects such as comments is one point where these two platforms and their analysis will diverge.

The second concern relates to how results are returned from the Instagram API. Querying for a specific tag will retrieve information about media published with the relevant tag in the caption for that media — that is, an image or a video that the original user has explicitly chosen to accompany with the tag in question. However, the same search will also return media which includes the tag in comments — even if the original caption does not feature the tag, and if the comment and tag were not posted by the original user. This then raises questions about authorship and intentionality when it comes to tags and Instagram content: the inclusion of a tag in a comment posted by another user might reflect different intentions than the original user in posting the content (and be irrelevant or contradictory to the media content, including spam comments). Including the tag might also give the media a public prominence that was not previously sought, adding to the distinctions required for an individual's possible uses and intentions around tagging.

For this reason, in our pilot study of tagged content on Instagram, we have filtered our dataset to include only the media featuring the specified tags in the original caption. These media then demonstrate the explicit decision by the person posting the media to include the tag at the time of publication, making their intentions clearer. While this filtering process then excludes any content where the original user has included the tag in a later comment rather than the caption, we have chosen to focus solely on captions for methodological ease at this initial stage (since our pilot study's data collection, Instagram has introduced editing capability for captions, creating further methodological questions to be addressed in future iterations of this research).

Adapting methods used for analysing large-scale Twitter data, we are also able to filter and process the Instagram metadata based on variables such as time and date, keywords in captions, additional tags, and user names. Our interest here, though, extends beyond this quantifiable information, with our analyses also looking at the other types of information users provide (whether intentionally or not) — this includes details provided in comments and in the images and videos themselves. Such information again raises further methodological and also ethical concerns: while all of these details are being made publicly available, the responsibilities of researchers in how to treat and represent this in their studies has been inconsistent for social media research (see Zimmer and Proferes, 2013, for example). Addressing anonymity and pseudonymity of social media users, for instance, has not had a clear direction and differs on a case-by-case basis. However, this is an obvious and real concern, especially when users offer highly personal and identifiable information (such as real names or addresses) in their media content when such information is not provided in their profiles (and may, at times, be provided accidentally).

Ethics and privacy beyond the binary

The question of privacy in relation to social media platforms of all types remains an ongoing issue. Privacy at a technical level, the individual and collective experiences of privacy, and what private actually means, are often conflated but can be imagined and understood in quite different ways by different people. As boyd [3] argues, software engineers tend to view privacy in binary terms, at the level of code, not experience:

> The tech world has a tendency to view the concept of 'private' as a single bit that is either 0 or 1. Data are either exposed or not. When companies make a decision to make data visible in a more 'efficient' manner, it is often startling, prompting users to speak of a disruption to 'privacy'.

Along similar lines, the 'Ethical Decision-Making and Internet Research Recommendations from the AoIR Ethics Working Committee' by Markham and Buchanan [4], and their collaborators, warns against a technical or any singular understanding of privacy online:

> Individual and cultural definitions and expectations of privacy are ambiguous, contested, and changing. People may operate in public spaces but maintain strong perceptions or expectations of privacy. Or, they may acknowledge that the substance of their communication is public, but that the specific context

> in which it appears implies restrictions on how that information is — or ought to be — used by other parties. Data aggregators or search tools make information accessible to a wider public than what might have been originally intended.

Researchers who collate datasets may also, inadvertently, reduce the experience of privacy by removing the original context of social media posts, including Instagram images and videos and metadata about them. By highlighting examples, or surfacing particular notable users from data not necessarily visible to everyday users, there is potential to alter the experience of privacy for an Instagram user (while not altering privacy at a technical level).

Given the seeming simplicity of Instagram's privacy settings — a binary choice between a public or private account, similar to Twitter's options (Zimmer and Proferes, 2013) — privacy may, at least ostensibly, appear relatively straight forward to understand, along the same lines as Twitter. However, an example demonstrating the opposite, in answer to the question "What happens if I share my photo to another social network?", Instagram (2014a) provides the following explanation:

> If someone with a private profile shares a photo or video to a social network (like Twitter, Facebook, Foursquare and so on) using Instagram, the image will be visible on that network and the permalink will be active. In other words, the photo will be publicly accessible by anyone who has access to its direct link/URL. Keep in mind that sharing a photo or video to a social network doesn't mean that the image will be visible in Instagram.

So the fact that Instagram is part of a complex ecosystem of social media platforms actually means that it has more nuanced privacy options than initially seems the case. Given the seeming simplicity of Instagram's privacy settings, the full implications of sharing an Instagram image via Twitter, Facebook, or another social networking service may not be immediately obvious to users. While this complexity is clearly important to individuals managing their Instagram use, it also begs some interesting questions for researchers collating Instagram datasets. A private Instagram image or video purposefully shared on other social media would not be returned when querying the API, but if it was shared with Twitter, refined searching of the Twitter API for Instagram posts would capture this media (or, at least, the link to it). Thus, comparative work looking at, for example, Twitter and Instagram use side by side may come up with anomalies due to this sort of complex sharing and privacy landscape, which provokes equally important questions about whether research should find ostensibly private Instagram posts, or whether searching via another platform such as Twitter could differentiate between public and private-but-shared-via-another-platform Instagram links (there is no obvious way to sort between these two types at face value using the Twitter API).

It is also important to remember that the experience or perception of privacy, or relative privacy, for Instagram users may shift over time as the platform and apps have become more accessible and widespread. As the timeline in Table 1 indicates, the Instagram app was originally launched in October 2010 and was only available as an iOS app on Apple devices for the first 18 months. While the early app had the same choice of having a public or private account, for the vast majority of users, it was only other iPhone or iPad users who could see their public Instagram stream, so there may have been a relative sense of privacy in that users on other platforms, or the Web, could not access their Instagram media. Over time, this changed, with the April 2012 release of a version for Android phones, the purchase of Instagram in the same month by Facebook for $US1 billion, the subsequent establishment of Instagram profiles on the Web, and finally the ability to embed Instagram media directly into other Web pages using an official embed code. Over that period the number of users has grown from just a few to 200 million active users today. Thus, for early users whose Instagram photos could only be seen by the small number of other Instagram app users, only on Apple devices, and only on mobile devices, not the Web, there has been a steady shift to their Instagram material becoming accessible to more and more people. At a technical level, their material may have always been public, but at an experiential level, relative privacy (or limited exposure) has been lost. Moreover, many users are probably unaware that using the API, many third party tools, apps and Web-based services can also access and share Instagram media. As researchers, it is important to consider this shifting experience of visibility and privacy on Instagram, even if nothing has changed in terms of the technical or code-level structure of having a public versus a private Instagram account. It is also worth noting that in December 2013 Instagram released an internal direct messaging tool, and while there are no official statistics about the number of people using this tool, it can be inferred that at least a proportion of users now probably share their more personal photos and videos with more limited audiences using this tool.

Table 1: Instagram timeline.

16 October 2010	Instagram app launched via Apple's App Store
12 December 2010	1 million registered users
3 August 2011	150 million photos uploaded
September 2011	10 million registered users
3 April 2012	Instagram releases Android version
9 April 2012	Facebook purchases Instagram for $US1 billion
26 July 2012	80 million registered users
16 August 2012	Instagram Photo Maps launched
5 November 2012	Instagram Profiles for the Web launched
5 December 2012	Instagram removes ability for photos to appear as 'cards' on Twitter
17 December 2012	Instagram alters Terms of Use
18 December 2012	Instagram reverts to previous Terms of Use after public backlash
26 February 2013	100 million active monthly users
20 June 2013	Instagram adds video (15 seconds maximum)
10 July 2013	Instagram adds native Web embedding for photos and videos
6 September 2013	150 million users
12 December 2013	Instagram Direct messaging service added
24 March 2014	200 million users
26 August 2014	Instagram/Facebook release Hyperlapse app via Apple App Store

Instagram, similar to many other apps and platforms, is in the grip of a social media contradiction (Leaver and Lloyd, 2015): while users often focus on platforms as a means of communication, often considered ephemeral akin to a telephone call in its permanence, Instagram themselves (and the parent company Facebook) are more focused on the media side, on the longevity of media both in terms of photos and videos, but also the associated metadata and further metadata generated by user interactions, which can be analysed, aggregated and used for commercial ends, including focusing targeted advertising. As researchers, we need to balance this contradiction with the relative ease of extracting Instagram metadata, and the value of the research and insights produced by analysing it. As it stands, Instagram's Terms of Use prohibit the large scale saving of media files from the platform, so qualitative researchers might view examples on the Web or use a bespoke tool or script to automate the viewing of photos or videos from a dataset without permanently storing a copy. To minimise the risk to user privacy, or the experience of privacy, where practical, the results from analysing hashtag-based Instagram datasets should be reported at the aggregate level, such as the number of Instagram images and videos posted over a set period of time, the most frequently associated hashtags, the most prominent geographic locations of this media, and so forth. To minimise the risk of doing any harm by increasing the exposure or visibility or particular Instagram users, we have chosen not to share the large datasets gathered. While this protects Instagram users, it does have the obvious downside that other researchers cannot immediately replicate the same queries and verify our results. In some cases the balance between respecting users' privacy and the thoroughness of research may best be served by sharing datasets but this would have to be carefully weighed on a case-by-case basis. Where illustrative examples of Instagram images or videos are needed for the purposes of demonstration, these media should be as de-identified as is practically possible.

While this paper has focused on the Instagram API and the official Instagram app, it is important to remember that literally hundreds of different apps work using the Instagram API and can bring affordances and functionality beyond that of the official app. For example, while there is no Instagram equivalent of the Twitter retweet, the familiarity of this functionality from Twitter and demand for it has led to the release of dozens of 'regramming' apps which provide this affordance. Regramming, or replicating the image content — usually with an acknowledgement of the source Instagram account, at least by default — does not necessarily leave any trace at the metadata level. While some regramming apps either replicate or acknowledge the caption that went with the original image or video, and may use bespoke hashtags such as #regram to explicitly flag that this is a replicated piece of media, users can easily choose not to post this in their new caption; and other regramming apps do not replicate any caption data at all. This is also an instance which highlights the importance of mixed method approaches to studying Instagram datasets; while regrams may be invisible at the metadata level, even a cursory qualitative eye cast over a series of images would easily notice replicated images, especially if these are widespread in a dataset. Moreover, regramming reminds us that understanding and studying Instagram is not just about understanding the app, or the API, or the platform as a whole, but also about understanding the app ecosystem which utilises the Instagram API and, more generally and probably more importantly, the broader technical and social contexts in which individuals and groups utilise, share, manipulate and regram Instagram images and videos.

Conclusion

This paper has set out our rationale for hashtags as an initial point of departure for studying activity on Instagram, adopting established methods from Twitter research as a means of enabling long-term comparative studies as well as single-platform analysis. A pilot study testing these new methods is currently underway, with the outcomes of this project highlighting additional questions to address, challenges to overcome, and affordances of Instagram to explore further.

While our analysis is still in the exploratory stage, what is clear from research in this field is that studies of Instagram — and social media in general — need to reconcile several key, and not insignificant, considerations influencing how and why people use different platforms, and how to best analyse this. The question of public and public-ness, from an ethical perspective but also regarding how public and private are realised and performed on social media, is a critical element of social media research, especially around personal information revealed by users (whether deliberately or inadvertently). There is a clear need for richer quantitative and qualitative methods for studying social media activity, including new methods for emerging platforms, and an understanding of the different strengths these approaches bring to analysis and how they can complement one another to provide extended, detailed examinations of social media data.

Researchers also need to address the types of data and content made available by users and social media APIs alike: this includes text and image content, as well as their contexts and associated metadata, and the need for appropriate analytical methods for each type. Similarly, an awareness of the evolving practices and their contexts on social media platforms is essential, since uses and affordances change over time; these include changes enforced by the platforms themselves, and new practices developed informally but spread widely by users. Finally, but most importantly, research needs to reconcile platform-specific studies with social media use overall. Social media platforms are not used in isolation, and activity is influenced by how individual users choose to employ different platforms. Content is not limited to one single platform, but is shared through diverse networks, critiqued, transformed, and remixed. While our methodology here focuses initially on Instagram, it is designed to also enable comparative analysis of Twitter and Instagram activity around specific hashtags. The next step for research in this field is to further these approaches, providing extended comparative approaches for qualitative and quantitative explorations of multiple social media platforms.

About the authors

Tim Highfield is a Postdoctoral Research Fellow in Media and Communication at Queensland University of Technology.
Web: http://timhighfield.net
E-mail: t [dot] highfield [at] qut [dot] edu [dot] au

Tama Leaver is Senior Lecturer in Internet Studies at Curtin University, Perth, Australia. He is the author of *Artificial culture: Identity, technology, and bodies* (Routledge, 2012) and co-editor of *An education in Facebook? Higher education and the world's largest social network* (with Mike Kent,

Routledge, 2014).
Web: http://tamaleaver.net
E-mail: t [dot] leaver [at] curtin [dot] edu [dot] au

Notes

1. boyd and Crawford, 2012, p. 668.

2. Halavais, 2013, p. 36.

3. boyd, 2008, p. 14.

4. Markham and Buchanan, 2012, p. 6.

References

Julian Ausserhofer and Axel Maireder, 2013. "National politics on Twitter: Structures and topics of a networked public sphere," *Information, Communication & Society*, volume 16, number 3, pp. 291–314.
doi: http://dx.doi.org/10.1080/1369118X.2012.756050, accessed 18 December 2014.

danah boyd, 2008. "Facebook's privacy trainwreck: Exposure, invasion, and social convergence," *Convergence*, volume 14, number 1, pp. 13–20.
doi: http://dx.doi.org/10.1177/1354856507084416, accessed 18 December 2014.

danah boyd and Kate Crawford, 2012. "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Communication & Society*, volume 15, number 5, pp. 662–679.
doi: http://dx.doi.org/10.1080/1369118X.2012.678878, accessed 18 December 2014.

Axel Bruns, 2008. *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York: Peter Lang.

Axel Bruns and Tim Highfield, 2013. "Political networks on Twitter: Tweeting the Queensland state election," *Information, Communication & Society*, volume 16, number 5, pp. 667–691.
doi: http://dx.doi.org/10.1080/1369118X.2013.782328, accessed 18 December 2014.

Axel Bruns and Jean Burgess, 2011a. "Tools," *Mapping Online Publics*, at http://mappingonlinepublics.net/resources/, accessed 14 October 2014.

Axel Bruns and Jean Burgess, 2011b. "The use of Twitter hashtags in the formation of ad hoc publics," paper presented at the European Consortium for Political Research conference, Reykjavik (25–27 August), at http://eprints.qut.edu.au/46515/, accessed 14 October 2014.

Axel Bruns, Katrin Weller and Stephen Harrington, 2013. "Twitter and sports: Football fandom in emerging and established markets," In: Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 263–280.

Axel Bruns, Jean Burgess, Kate Crawford and Frances Shaw, 2012. "#qldfloods and @ QPSMedia: Crisis communication on Twitter in the 2011 South East Queensland floods," Brisbane: ARC Centre of Excellence for Creative Industries and Innovation, at http://eprints.qut.edu.au/48241/, accessed 14 October 2014.

Fabio Giglietto and Donatella Selva, 2014. "Second screen and participation: A content analysis on a full season dataset of tweets," *Journal of Communication*, volume 64, number 2, pp. 260–277.
doi: http://dx.doi.org/10.1111/jcom.12085, accessed 18 December 2014.

Alexander Halavais, 2013. "Structure of Twitter: Social and technical," In: Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius

Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 29–41.

Tim Highfield, Stephen Harrington and Axel Bruns, 2013. "Twitter as a technology for audiencing and fandom: The #Eurovision phenomenon," *Information, Communication & Society*, volume 16, number 3, pp. 315–339.
doi: http://dx.doi.org/10.1080/1369118X.2012.756053, accessed 18 December 2014.

Larissa Hjorth and Jean Burgess, 2014. "Intimate banalities: The emotional currency of shared camera phone images during the Queensland flood disaster," In: Gerard Goggin and Larissa Hjorth (editors). *The Routledge companion to mobile media*. London: Routledge, pp. 499–513.

Nadav Hochman and Lev Manovich, 2013. "Zooming into an Instagram city: Reading the local through social media," *First Monday*, volume 18, number 7, at >http://firstmonday.org/article/view/4711/3698, accessed 14 October 2014.
doi: http://dx.doi.org/10.5210/fm.v18i7.4711, accessed 18 December 2014.

Instagram, 2014a. "Privacy and safety center," at https://help.instagram.com/442837725762581, retrieved 3 September 2014.

Instagram, 2014b. "Stats," at http://instagram.com/press/ (3 September), retrieved 3 September 2014.

Lyndon Kennedy, Mor Naaman, Shane Ahern, Rahul Nair and Tye Rattenbury, 2007. "How Flickr helps us make sense of the world: Context and content in community-contributed media collections," *MULTIMEDIA '07: Proceedings of the 15th International Conference on Multimedia*, pp. 631–640.
doi: http://dx.doi.org/10.1145/1291233.1291384, accessed 18 December 2014.

Tama Leaver, 2012. "Twittering informal learning and student engagement in first-year units," In: Anthony Herrington, Judy Schrape and Kuki Singh (editors). *Engaging students with learning technologies*. Perth, Western Australia: Curtin University, pp. 97–110.

Tama Leaver and Clare Lloyd, 2015. "Seeking transparency in locative media," In: Rowan Wilken and Gerard Goggin (editors). *Locative media*. London: Routledge, pp. 162–174.

Annette Markham and Elizabeth Buchanan, 2012. "Ethical decision-making and Internet research recommendations from the AoIR Ethics Working Committee (version 2.0)," at http://aoir.org/reports/ethics2.pdf, retrieved 14 October 2014.

Adam Mathes, 2004. "Folksonomies — Cooperative classification and communication through shared metadata," at http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf, retrieved 14 October 2014.

Hallvard Moe and Anders Olof Larsson, 2013. "Untangling a complex media system: A comparative study of Twitter-linking practices during three Scandinavian election campaigns," *Information, Communication & Society*, volume 16, number 5, pp. 775–794.
doi: http://dx.doi.org/10.1080/1369118X.2013.783607, accessed 18 December 2014.

Fred Morstatter, Jürgen Pfeffer, Huan Liu and Kathleen M. Carley, 2013. "Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose," *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, at http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6071, accessed 18 December 2014.

Zizi Papacharissi and Maria de Fatima Oliveira, 2012. "Affective news and networked publics: The rhythms of news storytelling on #Egypt," *Journal of Communication*, volume 62, number 2, pp. 266–282.
doi: http://dx.doi.org/10.1111/j.1460-2466.2012.01630.x, accessed 18 December 2014.

Liza Potts, Joyce Seitzinger, Dave Jones and Angela Harrison, 2011. "Tweeting disaster: Hashtag constructions and collisions," *SIGDOC '11: Proceedings of the 29th ACM International Conference on Design of Communication*, pp. 235–240.
doi: http://dx.doi.org/10.1145/2038476.2038522, accessed 18 December 2014.

Jill Walker Rettberg, 2014. *Seeing ourselves through technology: How we use selfies, blogs and wearable devices to see and shape ourselves*. Basingstoke: Palgrave Macmillan.

Zeynep 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, at http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062, accessed 18 December 2014.

Farida Vis, 2013. "Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots," *Digital Journalism*, volume 1, number 1, pp. 27–47.
doi: http://dx.doi.org/10.1080/21670811.2012.741316, accessed 18 December 2014.

Michael Zimmer and Nicholas Proferes, 2013. "Privacy on Twitter, Twitter on privacy," In: Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt and Cornelius Puschmann (editors). *Twitter and society*. New York: Peter Lang, pp. 169–181.

Editorial history

Received 16 October 2014; accepted 20 December 2014.

This paper is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

A methodology for mapping Instagram hashtags
by Tim Highfield and Tama Leaver.
First Monday, Volume 20, Number 1 - 5 January 2015
http://firstmonday.org/ojs/index.php/fm/rt/printerFriendly/5563/4195
doi: http://dx.doi.org/10.5210/fm.v20i1.