



Selected Papers of Internet Research 16:  
The 16<sup>th</sup> Annual Meeting of the  
Association of Internet Researchers  
Phoenix, AZ, USA / 21-24 October 2015

## SUBVERTING SOCIAL MEDIA

sava saheli singh  
New York University

### Introduction

Our use of Twitter is shaped by the built-in biases of the people who created it (Winner, 1980; Akrich, 1992; Marx, 2010; etc.). These biases may have come from a well-meaning place, such as the belief in an open space where everyone can reply to and connect with everyone else. But biases like this tend to be quite blind to the needs of vulnerable populations that come to a platform for various, often unpredicted, reasons.

Much of the writing and research around Twitter deals with how we've used the features to perform our identities (e.g.: Papacharissi, 2012; Marwick & boyd, 2010), connect with communities, and incur the wrath of those who feel wronged by us. More recently, there have been numerous mainstream media articles about how much Twitter has changed (Lafrance & Meyer, 2014; Jacobs, 2014; Chimero, 2014; Madrigal, 2014; Stewart, 2014; etc.) - a common nostalgic complaint from those privileged enough to have been early adopters, and privileged further to exercise the choice to reduce engagement or even leave.

As internet researchers, we're familiar with most features of Twitter (Bruns & Moe, 2013; Halavais, 2013) and what they are used for. We're also familiar with how users can be manipulated to change behaviors based on changing features (Liu, Kliman-Silver, & Misllove, 2014). Here, I will discuss how some of these features are repurposed and reinterpreted by users, while highlighting the conflict of features that be used for both useful and harmful actions. Observations here are based on existing literature, reflecting on my own practice of using Twitter, my deep experience with and research of Twitter use (predominantly with scholars and academics), participant observation in my research community, and conversations with other users.

*[Note: This paper specifically addresses features of Twitter's web application on twitter.com – mobile and other applications, while similar, require additional and slightly different considerations and literacies.]*

Suggested Citation (APA): singh, s. (2015, October 21-24) *Subverting Social Media*. Paper presented at Internet Research 16: The 16th Annual Meeting of the Association of Internet Researchers. Phoenix, AZ, USA: AoIR. Retrieved from <http://spir.aoir.org>.

## **Subtweeting**

Indirectly talking about someone is a great way to get opinions and frustrations out in the open without naming them. But, as part of a larger community, there is a very good chance that other users are going to know who is being talked about. That's often the aim of what is known as a subtweet - to say something about someone behind their back, to create a buzz about what's being said or let other people know something about the other person without being obvious. Subtweets allow for "plausible deniability" (Marwick & boyd, 2014) because the subtweeter can always deny a specific target.

There's also another layer of subtweet, through a private account. Private accounts of people who have public twitter accounts are a way for people to create a more private and safe space for them to express themselves, while controlling who gets to see these tweets. Subtweeting can take on an interesting level of passive aggression when someone does so from a private account because only that special private community is privy to this.

## **Block-Unblock Unfollow**

This is the repurposing of a Twitter feature as a way to control and manage followers and appearances on lists. There are some third-party Twitter applications that provide a 'force unfollow' function, but there is no other "official" way to do this. The earliest mention of the tactic I found was from a blogpost (Jorgensen, 2010) in 2010 that provided it as a solution to reduce the number of followers.

The action is simple: if a user is followed by someone they don't want to be followed by, they can block them and then immediately unblock them, which causes an automatic unfollow, without the follower's knowledge. You can also use this tactic to be taken off a list that someone may have added you to.

Forcing an unfollow might seem somewhat counter to the expectation that Twitter is an open space where anyone can "follow" anyone else, but it is a subversion of a Twitter feature that can be used to 'clean up' one's twitter followers and remove bots and people who haven't tweeted in a while. It can also be a form of self-preservation or self-care. People have used this to avoid scrutiny, to discreetly distance themselves from potentially harmful individuals, and reduce the reach of what they might tweet, as a way to control one's own "publicness".

## **Tweet Deletion**

Users might delete social media posts for a few reasons: because there's a typo they want to fix, because they changed their mind, or because they regret something they said (Wang et al, 2011). Some researchers claim that it is possible to predict the kinds of tweets that will be deleted based on factors such as role of the author and certain words (Xu, Burchfiel, Zhu, & Bellmore, 2013; Petrovic, Osborne, & Lavrenko, 2013), but others note that there are no real reasons for tweet deletion that can be surmised through examination of Twitter data because people delete for a variety of reasons

(Almuhimedi, Wilson, Liu, Sadeh, & Acquisti, 2013). The evolving nature of Twitter and how people are learning to protect themselves means it will get harder to predict.

There are also instances in which a user may have had a conversation - maybe an argument or a disagreement - and they want to erase that discussion. It is a way for users to control their content and manage both their external and personal identities, and sense of self. On the receiving end, it can leave other users looking as though they tweeted multiple times at a user and got no response, as though they were trying to force a conversation but were rejected.

Tweet deletion has interesting implications for identity and interaction on Twitter. In 2011, Twitter shut down Undetweetable – a service that let you browse through deleted tweets by entering a username (Hughes, 2011). Ostensibly created in order to provoke conversations about online identity and privacy, it also ran the risk of opening people up to harassment and embarrassment. More recently, the Sunlight Foundation – a non-profit organization that aims to increase the transparency and accountability of public officials in the government – had to shut down Politwoops, a web service that collected deleted tweets from politicians and made them available for people to search through on their website. These deleted tweets from politicians and other government officials visible allowed us to see everything from minor typos to major gaffes, thus exposing the inner workings of political social media.

Both these services were told that they were in breach of Twitter's Terms of Service and were shut down after they had already been in operation for a while. It's hard to say what harm was done to individuals based on these services, but this reaction is indicative of how services like Twitter tend to adopt a just-in-time approach to potentially harmful uses of their services, rather than create a service that addresses these before they can be the conduits of actual harm.

### **Twitter Advertisements and Promoted Tweets**

This may seem out of place on this list, but advertisements and promoted tweets can be effective forms of aggressions and microaggressions. They are unwanted noise. At their most benign, their irrelevance can be an annoyance, but at worst their tone-deaf ad copy can be offensive, and even triggering. As expected, most research and writing on Twitter ads and promoted tweets are in the realm of celebrity, marketing, and business (Wood & Burkhalter, 2014; Dacres, Haddadi, & Purver, 2014; and many more), with not as much consideration to the effects of the features on users beyond getting them to click on something and hopefully buy something.

Promoted tweets allow a user to create content and push this content to specific demographics via a dashboard that lets you target particular audiences based on gender, location, interests, keywords, and other factors ("Increase Twitter Engagement", n.d.). Promoted tweets and sponsored links, either by individuals or companies, manifest much like the pop-up ads of the early internet, forcing users to expend more emotional labor in the effort of refusing them attention. A disturbing case of the use of promoted tweets to spread racist remarks gained attention when a troll used the system to target particular groups of people with white supremacist "ads" (Smith, 2015). As

disturbing as this was, it took a troll to subvert the system for negative purposes in order to reveal the flaws in the system.

## **Open @-ing**

Inserting characters in front of a user's twitter handle makes their identity visible to all of the poster's followers, regardless of whether they follow them or not.

Replying like this to someone can have multiple implications. In its most harmless manifestation, this type of tweet can broadcast information that a tweeter might think the rest of their followers would find relevant or interesting. But as a defense mechanism and a warning to the rest of the community, it can draw attention to a harasser, making public the harassment being enacted with the intention of eliciting support from one's community and also publicizing the harassment. As an offensive or bullying mechanism, it can draw harassment to the recipient, including unwanted comments, abuse, and even shame.

## **Storifies**

Storify is a third party web-based application that allows users to archive conversations and posts from multiple social media applications, often to create a larger narrative outside of the context of the original social media platform. Users can create a record of a series of tweets (or other social media posts) – either from one user or from a conversation. On the surface, this is a useful way to preserve and present particular events or conversations, but these can also be constructed out of context and at the discretion of the creator, injecting that creator's bias into the curation of those tweets. People might ostensibly want to preserve a series of tweets, an argument, or a conversation, but it can also be used as a form of abuse, and longer conversations are rarely completely recorded. Storifies and tweets are also used/embedded in news stories or articles or blog posts and serve as a further magnifier to the tweet or person being reported on. Again, this could work a couple of ways: purely as a record of something, as a way to point would-be harassers at someone, or as a way to point out inequities or harassment.

## **Retweeting and Links to Tweets**

Retweeting is a common activity on Twitter and has gone through several changes over time. In its basic form, it is a way to re-broadcast a tweet by someone else to your followers. People retweet for a variety of reason ranging from amplification to sharing information (boyd, Golder, & Lotan, 2010). One of the earliest forms of retweeting required users to insert 'RT' in front of the text copied from a tweet. Sometimes, users could include a comment before the 'RT' or even change it to 'MT' to indicate that the tweet had been modified, but this meant that both the comment and the text of the retweet had to fit Twitter's 140-character constraint.

Now, you can retweet in a couple of ways: hitting the retweet button transmits the original tweet as a whole and adds a notation that shows it has been retweeted by you, or it gives you the option of adding a comment to an embedded version of the original

tweet. The difference in this type of commenting is that you get the full 140 characters to comment rather than having to share characters with the content of the retweeted tweet.

I describe the historical difference in retweeting format because it is interesting and important to note the implications of this change on user behavior. The first iteration meant you retweeted someone and they'd know you did, but you could tweet a link to their tweet and they wouldn't know – this allowed for more flexibility in how users retweeting. The former approach credited the author with tweeting content, but the latter approach allowed for shaming, drawing attention to a harasser, or making a snarky comment without the author's knowledge. The new format changes how users interact with a retweet and how a retweet can be used to signal different things, and makes it harder to draw negative attention to a tweet/tweeter. You can still copy a tweet and put 'RT' in front of it, but compared to the new retweet, that is time-consuming and would use up the 140 characters limit.

Users sometimes tweet screenshots of a tweet or text – this was not uncommon before the changes to how retweeting works now, but it has gained further traction as a way to subvert the new changes – both to publicly shame and to potentially call out negative twitter behavior without drawing attention to the tweeter.

## **Faving**

Favoriting or faving as a feature was ostensibly meant to be used as a way to bookmark interesting tweets or links in tweets, but it plays many roles now, from, well, bookmarking to signaling different kinds of acknowledgments. Users still use faving as a bookmarking tool, sometimes as a trigger for web bookmarking services like Pinboard using IFTTT (a web application that uses changes in web applications – such as favoriting a tweet – to trigger another web application – such as Pinboard – to perform an action) to send a tweet or a link in a tweet to a bookmarking database which offers an easier way to record these tweets. But there is a vast range of uses outside of bookmarking – as a “like” a la Facebook, to signal agreement, take a side in a Twitter argument between two people you know, or just because something was funny. There might be deeper cultural reasons for favoriting a tweet, such as “participating in a shared context” as Cottom (2015) says, that she prefers to favorite over retweet, to avoid having to explain the context – something she can choose not to do. There are a number of interpretations – real and humorous – that can be inferred from a fav ranging from polite to maybe a little creepy (Smale, 2014).

An interesting phenomenon that brings a new aspect to faving is the “While you were away...” feature. In what seems to be an attempt to emulate Facebook, Twitter introduced this feature to show users tweets they might have missed from people they follow, based on engagement and other metrics (Rosania, 2015). This breaks what used to be the traditional Twitter stream – the ephemerality of Twitter being one of its defining features – and displays tweets from hours ago. Not long ago, if someone favorited someone else's tweets from a while ago, it was more than likely that they were going through the person's timeline looking at their older tweets. But now, if someone faved a tweet from a few hours ago, it could be that they saw it as part of the “While you were away...” feature.

## Conclusion

There are subversive behaviors that I have not listed here, and there are probably new ones being born as I type. None of these are objectively good or bad. As with most systems built with apparent “openness and sharing” at the heart of their design - a fallacy that is used as an excuse for the biases that are built into various platforms - they are not neutral. They are built with very specific biases - often by those at the top of the privilege structure who cannot envision the harm and misuse of the tools that were purportedly meant to allow for the free flow of ideas. Those users who are not part of this demographic must find ways to protect themselves and the most vulnerable, which often involves subverting the uses of the platform to meet their needs.

In the context of researching behavior in and culture of these spaces, these kinds of uses of Twitter features can add messiness – desired and undesired (Tufekci, 2014) – to the data collected. Taking these phenomena into account can provide insight into the ad hoc use of these features and platforms outside of the usual types of data collected, but as features keep changing researchers must keep up with how the community both uses and subverts them. As we continue to conduct research and examine the cultural implications of Twitter, we have to take into account the constantly shifting ground on which we are attempting to base our research.

Twitter will probably go through a lot of changes in response to shareholder expectations and concerns of return on investment. Post-IPO, Twitter is competing with other social media platforms like Facebook, Snapchat, and Instagram, and will feel the pressure to change or introduce features in order to remain relevant. For example, recent changes like the ability to send direct messages to multiple people and lifting the character limit on direct messages have already changed the Twitter experience. Imagine what would happen if Twitter lifted the character limit on tweets – this would be a fundamental change to the experience and there would be a drastic cultural shift.

Our use of Twitter has already changed significantly over the years (Liu, Kliman-Silver, & Mislove, 2014). With every changing feature, we are forced to change our behaviors, find different workarounds, worry about new ways to nurture and protect our communities, and of course, update our research agendas. As Twitter continues to introduce new changes, we will need new literacies to keep up and old cues that we depended on will no longer be valid – how we are online is changing and will continue to change.

## References

- Akrich, M. (1992). The de-scription of technical objects. *Shaping technology/building society*, 205-224.
- Almuhimedi, H., Wilson, S., Liu, B., Sadeh, N., & Acquisti, A. (2013, February). Tweets are forever: a large-scale quantitative analysis of deleted tweets. In *Proceedings of the 2013 conference on Computer supported cooperative work*(pp. 897-908). ACM.
- Bruns, A., & Moe, H. (2013). Structural layers of communication on Twitter. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Pushmann (Eds.), *Twitter and society* (pp. 15–28). New York: Peter Lang.
- Cottom, T. M. (2015, August 19). Faves. *The Message*. Retrieved from <https://medium.com/message/faves-71b8fc71659e>
- Dacres, S., Haddadi, H., & Purver, M. (2013). Topic and Sentiment Analysis on OSNs: a Case Study of Advertising Strategies on Twitter. In *Proceedings of The Second ASE International Conference on Big Data Science and Computing*.
- Halavais, A. (2013). Structure of Twitter: Social and technical. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Pushmann (Eds.), *Twitter and society* (pp. 29-41). New York: Peter Lang.
- Hughes, S. A. (2011, August 4). Undetweetable archives deleted tweets, gets warning from Twitter (Update). *The Washington Post*. Retrieved from [https://www.washingtonpost.com/blogs/worldviews/post/undetweetable-archives-deleted-tweets-gets-warning-from-twitter/2011/08/04/gIQAqlBul\\_blog.html](https://www.washingtonpost.com/blogs/worldviews/post/undetweetable-archives-deleted-tweets-gets-warning-from-twitter/2011/08/04/gIQAqlBul_blog.html)
- “Increase Twitter Engagement”. (n.d.). Retrieved from <https://business.twitter.com/solutions/tweet-engagements>
- Jorgensen, L. (2010, February 20). The Secret to Fewer Followers [guest blog post]. Retrieved from <http://www.twitter-fail.com/2010/02/20/the-secret-to-fewer-followers/>
- Liu, Y., Kliman-Silver, C., & Mislove, A. (2014, May). The tweets they are a-changin': Evolution of twitter users and behavior. In *International AAAI Conference on Weblogs and Social Media (ICWSM)* (Vol. 13, p. 55).
- Marwick, A., & boyd, d. (2010). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience." *New Media and Society*, 13, 96-113.
- Marwick, A. E. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7), 1051-1067.
- Marx, L. (2010). Technology: The emergence of a hazardous concept. *Technology and Culture*, 51(3), 561-577.
- Papacharissi, Z. (2012). Without you, I'm nothing: Performances of the self on Twitter. *International Journal of Communication*, 6, 18.
- Petrovic, S., Osborne, M., & Lavrenko, V. (2013). I wish I didn't say that! analyzing and predicting deleted messages in twitter. *CoRR*, abs/1305.3107, 2013.
- Smale, H. (2014, October 23). Mum asked what a Twitter favourite "meant", so..... GOOD LUCK MUM. [Tweet with image]. Retrieved from <https://twitter.com/HolSmale/status/525370488437350401>
- Smith, S. E. (2015). How a racist troll tricked Twitter into promoting white supremacy. *The Daily Dot*. Retrieved from <http://www.dailydot.com/opinion/weev-supremacist-twitter-ad/>
- Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, 2014.

- Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011). "I regretted the minute I pressed share": a qualitative study of regrets on facebook. In *SOUPS '11: Proceedings of the Seventh Symposium on Usable Privacy and Security*, (pp. 1-16).
- Wood, N. T., & Burkhalter, J. N. (2014). Tweet this, not that: A comparison between brand promotions in microblogging environments using celebrity and company-generated tweets. *Journal of Marketing Communications*, 20(1-2), 129-146.
- Winner, L. (1980). Do artifacts have politics?. *Daedalus*, 121-136.
- Xu, J. M., Burchfiel, B., Zhu, X., & Bellmore, A. (2013). An Examination of Regret in Bullying Tweets. In *HLT-NAACL* (pp. 697-702).