Identity Multipliers and the Mistaken Twittering of 'Birds of a Feather'

David M Cook, Edith Cowan University

Available at: https://works.bepress.com/david_cook/16/
Identity multipliers and the mistaken twittering of ‘birds of a feather’.
David M. Cook
Edith Cowan University Security Research Institute, Joondalup, Australia
d.cook@ecu.edu.au

Abstract: New Media usage has expanded in unexpected directions. This qualitative study pulls together two new media phenomena that demonstrate widespread engineered influence through Twitter dispersion. The first aspect is the practice of sock-puppetry deception that occurs when social media identities are used without proof of identity. The relative ease in creating a single invented online identity is overshadowed by the more recent practise of harnessing thousands of created identities to accentuate the standing of a single identity. The second aspect is the use of metadata to fetch and find increased nodes of commonality. This promulgates the perception of connectivity and homogeneity in terms of information transfer. Identity masking in Twitter, and the use of homophily to manipulate authority, public sway, and social media deception are a formidable combination. This study considers the increasing use of sock-puppetry in conjunction with techniques that exploit homophily to sway the perceptions of others. The ethical and socio-political issues of identity deception and their subsequent impact are discussed. Using the publicly available Twitter accounts of six high profile political figures, the study compares retweets that retain metadata from original tweets. Using both real and fake accounts, the paper tests the hypothesis that non-genuine Sybil Twitter accounts are deployed in the followings of popular figures by using metadata in retweets to establish credibility amongst fake entities.

New Media experiences augmented manipulation in multiple ways. The use of metatagging to expose greater commonality increases both the frequency and the amplitude of information diffusion when delivered through social networking and microblogging. A comparison of the associated tweets, followers, and retweets of high profile leaders indicate pervasive sock-puppetry and Sybil deception. With homophily embedded into Tweets and Retweets, sock puppetry in social media presents a challenge. Where homophily fits traditional notions of “similarity between two nodes”, increasing the likelihood of likeness to occur is problematic. If homophily and false identity coalesce, then the use of twitter as a reliable social medium is contingent on discrete user identity in social media.

Keywords: Homophily, Twitter, Deception, Sock-puppetry, Metatagging, Sybil Accounts

1. Introduction: The Deceptive Progression of New Media

New Media usage has expanded beyond the expectations of society in general. It has replaced print media, postal mail, and a variety of community-based communications platforms. Twitter in particular, with its high speed delivery and narrative brevity, has bridged the gap between formally reported journalism and the public and civic reporting that has been informally recounted through the community (Jewitt, 2009). Now there is simply news. Individual tweets are now more readily accepted as factual, backed up by smartphone photography, crowd sourced verification, and Twitter’s lightning fast diffusion of information (Wilson, 2011). However, as New Media continues its rapid expansion, the volume and frequency of digital telecommunication has accelerated the proliferation of exploits that deceive the world by means of technology force multipliers (Christensen, 2011).

The business community has embraced the marketing value of new media, using the mechanisms of mass connectivity for profit and market share (Wilson, 2009). Users are best described as ‘prosumers’, combining content production and content consumption in a single descriptor (Leavitt, Burchard, Fisher and Gilbert, 2009). Twitter holds prominence within marketing circles because of its ability to connect people with similar thinking, ‘like’ desires, and, most importantly, analogous decision-making. In business terms this equates to purchases, commitments, profits and revenue (Nielsen, 2009). Such commonalities are tremendously sought after. The business community is repeatedly contemplating whether to use false identities as microblogging entities in order to...
exaggerate and to substantiate the power and influence of any given message, narrative, or person (Miller, 2009; Stringhini, Wang, Egele, Kruegel, Vigna Zheng, and Zhao, 2013). One of the fundamental golden pillars upon which Twitter is built is the speed and ease of information diffusion. The more people in a social network, and the more connected they are in closeness, the faster and stronger any given message is circulated (Suh, Hong, Pirolli, and Chi, 2010).

1.1 What damage can fake identities do?

Fake identities manipulate, deceive, and grow the uncertainty and mistrust of online communities (Ratkiewicz, Conover, Meiss, Goncalves, Flammini, and Menczer, 2011; Romero, Meeder and Kleinberg, 2011). A fake identity on Facebook or Twitter is most commonly not operating in isolation (Thomas et al., 2012; Ratkiewicz et al, 2011). Whilst sock-puppetry does occur on a personal level expressed in the form of one or two additional extensions of an individual's own discrete identity, it is the practice of botnet-enabled Twitter puppets (Sybils) that can be deployed by the thousand, that is more influential in propagating Twitter identity deception (Krebs, 2011; Wheatley, 2013; Parmelee & Bichard, 2013).

The ethical and legal question about whether the wider netizenry should accept the use of multiple invented personas by individuals is most explicitly identifiable in an examination of Twitter entities (Parmelee & Bichard, 2013). The question arises as to whether the public have the right to know when one person's message is deceptively repeated through multiple fake entities to give the impression of widespread support for a person, their theme, or their narrative. Even though the practice of sock-puppetry far outdates the invention of web-based information transfer (Rollins, 1993), the use of technology-enabled puppetry presents a much more powerful and highly augmented set of possibilities in terms of multiple automated microblogging (Chu, Gianvecchio, Wang, & Jajodia, 2010).

The use of such deception is visible and prominent across a wide range of Twitter disciplines, whether it be in business (Streitfield, 2012), violence and extremism (Conway, 2012), or government and politics (Cogburn & Espinoze-Vasquez, 2011). There are already established and institutionalised practices that deliberately augment followings for their own means and ends, whilst at the same time the presence of widespread acknowledgement and recognition of the practices seems largely overshadowed by the more prominent behaviour of ignorance, passive-agnosticism, and tolerance, best generally described under the terminology of slacktivism. Fake identities in large enough numbers can dominate or at least influence global outcomes in terms of purchasing behaviours, radicalisation, governance and democracy (Waugh, Abdipanah, Hashemi, Abdul Rahman, and Cook, 2013).

2. New Media platforms encourage Multiple Identities

At a time when individuals are beginning to understand that their data can be used by corporations to great commercial benefit, the notion that social media platforms knowingly continue to promote practices that encourage multiple identities is divisive. Profit, power, influence and authority are drivers of the practice of using botnets and automation to harness the advantages of vastly enlarged support from multiple followings (Stringhini et al, 2013). The Russian parliamentary elections of 2011 saw pro-Kremlin and anti-Kremlin parties endure the deployment of 25860 fake Sybil Twitter accounts used to leverage over 400,000 tweets in a swarm attack that disrupted and confused online narrative homophily beyond expectations (Thomas, Grier and Paxson, 2012). The tactic of injecting creative narrative into social networks at key moments reveals one of the major drawbacks of a social media platform that does not require identity authentication beyond the originating CAPTCHA authentication (Twitter, 2013). This would suggest that the organisational side of Twitter is comfortable with the knowledge and expectation that there are many fake Twitter entities, and that large scale sock-puppetry is alive and well in the Twitter community (D’Yonfro, 2013). Twitter does not offer or promote
any serious effort towards the filtering of spam (Grier, Thomas, Paxson and Zhang, 2010). Thus, for those seeking to exact influence and authority, Twitter has sufficient leeway and latitude to suggest that multiple identities are accepted as an applicable component of new media.

3. The effect of Homophily upon Twitter-based Social Network influence

The ‘Birds of a Feather’ understanding of homophily describes the phenomena where communication amongst “similar people occurs at a higher rate than among dissimilar people” (McPherson, Smith-Lovin, and Cook, 2001). Indeed, in most social networks the key pointer to their homophily is the reciprocity which denotes their sharing of one or more interests. The reciprocity of social connections shows the exchange and interaction between people and although their reciprocity need not be equally mutual, there is an expectation that the interaction is not all one-way traffic (Weng, Lim, Jiang and He, 2010). However, the same is not true for popular Twitter accounts where a person of celebrity and reputation will typically have many followers but have very few followings (Kwak, Park and Moon, 2010).

Twitter uses homophily to draw attention to both the trending themes and ideas of the day as well as those Twitter accounts that command the greatest numbers of followers, and the greatest number of accounts that retweet across those themes and ideas. These trends are readily identified through a range of metatagged expressions that show connection with (although not necessarily support for) a given idea, topic, or narrative expression (Blanquart and Cook, 2013; Waugh et al, 2013). We use the term metatagging here to include the use of identifiable search physiognomies such as hashtags, URLs, attached pics, @ symbols and the acronym RT. These characteristics form the metadata of choice within the Twitter community (Romero, Meeder, and Kleinberg, 2011). In combination with idioms, sarcastic phrasing, and emotive expressions, they shape much of the trend-creating behaviour that is magnetic to twitterbots, cyborgs, and other Twitter manipulators (Chang, 2010; Romero et al, 2011; Wang, Wei, Liu, Zhou, and Zhang, 2011).

The interdependency of entities / actors, and the manner in which their interactions overlap, forms the cornerstone upon which social network influence can be understood (Kwak, Lee, Park, and Moon, 2010). In its most basic sense, the very use of Twitter marks a form of technology acceptance and an approval of the system of microblogging as an appropriate method of dispersing information. Rogers’ (2003) work on diffusion is very clearly visible and accurately demonstrated when examining the manner in which an original tweet becomes transferred and interrelated through a network of other retweeted narratives (Chang, 2010). For those in the business of seeking influence and profit, homophily is the agent that augments that influence. At the same time, homophily should still be understood as the characteristic of individuals who are serious about finding other people (or networks of other people) who share some of their own “sociodemographic, behavioural, and intrapersonal characteristics” (McPherson, Smith-Lovin, and Cook, 2001).

There is evidence to suggest that homophily is not quite as innocent a characteristic as often portrayed. In particular, when combined with the inherent multipliers of its metadata, homophily and sock-puppetry represent a dangerous combination, in the sway and influence of social media activity (Ratkiewicz et al., 2011; Romero, et al., 2011, and Suh et al., 2010). This combination is an obvious characteristic of our communications such that in normal, everyday terms its effect goes largely unnoticed. Twitter users gravitate towards friends and colleagues assisted by a range of communications that hold them to agree with or share interest in a given person, message, story or outcome. People do not have to like the other person’s view, they simply have to be drawn towards the general theme that is of appeal to them. (McPherson et al, 2001; Kwak et al 2010). Metadata has become the catalytic agent of choice for the discerning Twitter entity looking to exert greater influence and authority Ratkiewicz et al., 2011).

3.1 Homophily and Influence from Actionable Content
There are two classifications of tweets on the Twitter spectrum. If a Tweet is ‘conversational’ then the original tweet and retweet (RT) may respond back and forth as a conversation of sorts ensues. If a tweet (especially from a celebrity or person of prominence) is ‘content based’ then the direction of the original tweet and any subsequent tweets is more one directional. Subsequent retweets continue to diffuse the content through a wider and wider network. In these situations the original tweet is less likely to give direct responses to replying retweets, and will instead post further ‘content-based’ tweets as the wider community respond and react. Thus the one directional actionable content is of specific interest in determining the true measure of influence on the Twitter platform.

Leavitt et al, (2009) suggest that there are two extrinsic classifiers that relate to this form of actionable content: the mention and the attribution.

In the first instance a user may express something like:

*Mention: {content} @username ((content))
Watching @BarackObama speak in Colorado on @CNN
RareAir24 (on 2009-08-15 at 19:08:51)*

In the second instance the user may narrate as follows:

*Attribution: {content} via @username ((content))
Fire at Kuwaiti wedding kills dozens, official media says http://bit.ly/wn95A (via@cnnbrk)
ChilliGaz (on 2009-08-15 at 19:40:18)*

Both mentions and attributions deliver content that when augmented by metadata, (for example hashtag, @CNN, URL, pic) can achieve much higher rates of diffusion through perceived homophily. These metadata not only work as organisational markers that assist Twitter in determining its ‘trending topics’ (Chang, 2010), they also build trust and credibility into narrative attempts at projecting social norms, opinion, and innovation and change (Rogers, 2003). There is widespread evidence to indicate that Twitter homophily is augmented by the addition of metadata (Chang, 2010; Blanquart & Cook, 2013; Waugh et al, 2013). This is especially clear where metadata takes the broader meaning to include hashtags, URLs, pics, idioms, RTs, and sarcastic or hateful phrasing (Suh et al, 2010; Kwak et al, 2010; Ratkiewicz et al, 2011; Davidov, Tsur and Rappoport, 2010).

4. Research Method

In order to test the followership of high profile political figures for the combinational occurrence of both fake personas and retweets with metadata-augmented homophily, the following sample was taken. The Twitter postings and subsequent retweets were captured for six discrete high profile (trending) Twitter accounts. The data was obtained using Twitter’s publicly available API and data was gathered by crawling Twitter in sixty day periods between March 2013 and December 2013. All six accounts
were political figures who had posted more than 500 tweets and who had over 50,000 followers. The six accounts were selected at random from a predetermined selection of 24 political figures who met the tweet and follower criteria (Table 1). Their subsequent retweeting entities (active followers) were then examined using the Chu et al. (2010) four way test for fake entities. From these followers a random sample of 10 suspected Sybil entities and 10 genuine entities were selected for each of the six original accounts (Table 2).

<table>
<thead>
<tr>
<th>Political Figures</th>
<th>Username</th>
<th>Tweets</th>
<th>Followers</th>
<th>Following</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Cameron - UK Prime Minister</td>
<td>@David_Cameron</td>
<td>674</td>
<td>579300</td>
<td>331</td>
</tr>
<tr>
<td>Geert Wilders - Dutch Parliament Chair Party for Freedom</td>
<td>@geertwilderspvv</td>
<td>1006</td>
<td>311000</td>
<td>0</td>
</tr>
<tr>
<td>Tony Abbott - Prime Minister of Australia</td>
<td>@TonyAbbottMHR</td>
<td>1570</td>
<td>278000</td>
<td>31500</td>
</tr>
<tr>
<td>Ed Miliband - UK Leader of Labour Party</td>
<td>@Ed_Miliband</td>
<td>3000</td>
<td>287600</td>
<td>1446</td>
</tr>
<tr>
<td>Bill Shorten - Leader of the Australian Labor Party</td>
<td>@billshortenmp</td>
<td>1162</td>
<td>51340</td>
<td>10600</td>
</tr>
<tr>
<td>Dr Manmohan Singh - Prime Minister of India</td>
<td>@PMOIndia</td>
<td>4206</td>
<td>1068317</td>
<td>48</td>
</tr>
</tbody>
</table>

Table 1. Political figures from which sample retweets were captured.

A total of 120 retweeting entities (60 genuine and 60 Sybil) were examined for signs of metadata used as actionable content. A total of 6,895 retweets drawn from all 120 entities were reviewed for their use of like-metadata. In cases where a retweet used metadata that was different from the metadata in the original tweets, the results did not record the retweet as showing homophily-type metadata. For example if a different hashtag was used in a retweet, then the metadata component was not recorded even though the retweet was included in the sample. Only retweets derived from one of the six chosen figures were considered in the sample data. Retweets were considered on the basis of restated metadata (notably the use of URL, hashtags, Pics, Youtube links, @username, and phrases constituting sarcasm, hateful commentary and idioms). A comparison was then drawn between the genuine and Sybil accounts in order to see whether non-genuine accounts showed greater levels of homophily-generating metadata.

<table>
<thead>
<tr>
<th>Four Way Test for non-genuine Twitter entities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Entropy Test - Measure Retweet intervals</td>
</tr>
<tr>
<td>2. Spam and Miscreant Test - Check for Benign or Malicious content</td>
</tr>
<tr>
<td>3. Account Properties – Does the Account have subscriber details or does it look hollow</td>
</tr>
<tr>
<td>4. Discrimination Analysis – combining Entropy, Spam, and Account Properties to evaluate all three indicators</td>
</tr>
</tbody>
</table>

Table 2. Chu et al. (2010) 4 way test to distinguish genuine from Sybil accounts

5. Results

A comparison between the sixty genuine followers and the sixty non-genuine (Sybil) accounts showed a number of areas where retweeted narrative shared similar behaviour. Since the behaviour of the non-genuine accounts was assumed to use some form of bot-like automation, it was postulated that Twitter-bots, cyborgs and automated Sybil accounts would show equal or greater numbers of discrete retweets that also incorporated one or more of the metadata markers as described above. The results were highly consistent with this hypothesis in terms of URLs, hashtags, and pics. In these three areas the Sybil accounts showed significantly greater numbers of discrete retweets where the metadata was also passed on as part of the retweeted message. The more often a retweet had one or more of these
metadata, the more likely it was to be retweeted forwards. Across both genuine and Sybil entities, the strongest indicator of homophily metadata was shown as links to Pics and Youtube clips. Pics and Youtube links in particular showed the highest levels of homophily in fake (Sybil) accounts. Since these characteristics are ascribed to fake accounts, it is the perceived commonalities that we describe as homophily, rather than actual like-mindedness, since the retweets in question represent contrived puppetry. Nevertheless, these characteristics are indeed consistent with homophily as expressed through higher numbers of retweeted actionable content, regardless of whether they are mentions or attributions.

In contrast the @username metadata showed no discernable differences between fake and real retweets. This would suggest that for the most part, botnets are not relying on the @username metadata to extend or augment the influence and authority of any given Twitter narrative. This is consistent with marketing and business research (Suh et al, 2010) that places greater emphasis upon links to websites, pictures and youtube clips.

6. Discussion

Whilst the retweeting of @username was present in both genuine and Sybil accounts, its presence was lower than that of the URL based objects such as links to Pics, Youtube and an assortment of extended blogs and web pages. This presents an intriguing comparison. The notion of homophily is defined as homogeneity to intrapersonal characteristics (McPherson et al. 2001; Weng, et al. 2010). Therefore the @username might be expected to retain a greater overall prominence in retweets. However it is the subject-based URLs rather than the people-based @username that carries the lion’s share of the homogenous linkages.

This study also examined descriptive phrasing and narrative in terms of its metadata potential. Phrases, the use of sarcasm, and idioms were all included in the analysis of retweets. These characteristics had a much lower percentage of homophily with fake retweet accounts. We suggest that this is because phrases and idioms require individual and personal scrutiny and do not present the same ease of identification as the more obvious and object-oriented metadata such as URLs, Pic links and hashtags. In fact, sarcastic idioms and hateful phrases can become singleton narratives, since they often hold a level of intensity that does not maintain the same fervor through subsequent retweeting entities. Some phrases, particularly where explicit or passionate phrases are narrated, are highly personal and therefore may encourage restraint rather than connection. Those retweets that showed idiom and sarcastic phrasing were usually re-messaged because they also contained hashtags or other metadata as well. In the sample that was collected, there were no significant numbers (less than 0.1%) of tweets where idiom of sarcasm was retweeted unless other metadata was also present in the original tweet.

The identification of non-genuine Twitter entities was determined using the Chu et al (2010) four way test (Table 2). The test identifies automation in bot crawlers through entropy, the timing of retweets and their relevant intervals. It also looks at the spam and miscreancy behaviour of entities, as well as the account details, its stature and potential hollowness. In combination, these patterns show clearly visible differences between Twitter-bots and genuine human Twitter accounts. Since the presence of Twitter-bots and Sybil accounts was predicted as an underlying assumption of the study, the relative ease with which they were identified suggests their widespread integration into political audiences of the Twitter community.
<table>
<thead>
<tr>
<th>Retweeting Entity Classification</th>
<th>Real or Sybil</th>
<th>URL %</th>
<th>Hashtags %</th>
<th>Pics &amp; Youtube %</th>
<th>@username %</th>
<th>Sarcasm / Hateful / Idiom %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real RTs for Cameron</td>
<td></td>
<td>22.4%</td>
<td>21.2%</td>
<td>39.0%</td>
<td>17.9%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Sybil RTs for Cameron</td>
<td></td>
<td>47.4%</td>
<td>34.3%</td>
<td>68.9%</td>
<td>18.5%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Real RTs for Wilders</td>
<td></td>
<td>22.3%</td>
<td>17.0%</td>
<td>42.8%</td>
<td>7.6%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Sybil RTs for Wilders</td>
<td></td>
<td>37.8%</td>
<td>25.1%</td>
<td>58.4%</td>
<td>7.6%</td>
<td>5.8%</td>
</tr>
<tr>
<td>Real RTs for Abbott</td>
<td></td>
<td>23.3%</td>
<td>22.0%</td>
<td>39.3%</td>
<td>15.6%</td>
<td>4.9%</td>
</tr>
<tr>
<td>Sybil RTs for Abbott</td>
<td></td>
<td>36.5%</td>
<td>36.4%</td>
<td>74.2%</td>
<td>15.8%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Real RTs for Miliband</td>
<td></td>
<td>19.5%</td>
<td>16.2%</td>
<td>34.1%</td>
<td>11.1%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Sybil RTs for Miliband</td>
<td></td>
<td>35.2%</td>
<td>33.0%</td>
<td>53.3%</td>
<td>10.4%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Real RTs for Shorten</td>
<td></td>
<td>9.3%</td>
<td>2.3%</td>
<td>42.9%</td>
<td>13.9%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Sybil RTs for Shorten</td>
<td></td>
<td>9.1%</td>
<td>3.3%</td>
<td>43.2%</td>
<td>14.3%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Real RTs for Singh</td>
<td></td>
<td>3.1%</td>
<td>4.9%</td>
<td>46.3%</td>
<td>2.1%</td>
<td>5.6%</td>
</tr>
<tr>
<td>Sybil RTs for Singh</td>
<td></td>
<td>3.3%</td>
<td>3.7%</td>
<td>63.9%</td>
<td>1.9%</td>
<td>4.4%</td>
</tr>
</tbody>
</table>

Table 3. percentage of retweets from sample that deployed one or more pieces of the same metadata expressed by an original tweets as nominated in Table 1.

Note: This table shows the percentage of retweets that included a meta-marker from an original tweet. It does not include those instances where a retweet has included a fresh item of metadata. Thus it is possible that the figures for both real and Sybil retweets may have a slightly higher total number of tweets that include some form or otherwise homophily-like metadata.

6.1 Hypotheses Results

This study’s hypothesis asked whether large numbers of fake, automated retweets were establishing a false credibility by including specific objects of metadata in retweeted messages. Are Twitter force multipliers in the form of metadata a means to promulgate influence? If sock-puppetry and homophily could be combined, it would be possible for politically driven groups and associations to exercise significantly greater influence upon mass audiences. Thus by deploying large numbers of Sybil accounts that carry greater authority and trustworthiness by establishing desirable homogeneity between the retweeting entities and their original high-trending counterparts, a narrative of sway can be significantly amplified. It is clear from the results that URLs that link to pictures, youtube clips, and other extended blogs form the type of homophily that can both augment the credibility of a fake twitter account, and at the same time mask its hollowness.

7. Conclusion

Many researchers have attempted to understand the exact metrics that correctly indicate the size and speed of the influence that can be attributed to each twitter narrative, and subsequently to broader themes and Twitter figures of prominence. To date there are many algorithms but none that accurately and predictably measure genuine influence (Weng et al, 2010). However the presence of Sybil accounts dilutes the value of most metrics which include the growing number of puppet-controlled Twitter accounts. The combination of sock-puppetry and homophily-attributed metadata challenges the notion that “birds of a feather flock together”. Twitter influence is at risk of great
deception through the disguised and manipulated persuasion using force-multiplied objects of metadata. The descriptor *homophily* has its historical origins in the homogeneity of people and their likes, yet in the Twitter network its power lies more closely with the eye-catching gimmickry of shared pictures and videos.

References


Twitter. (2013). Twitter; An Information Network, retrieved 13th October 2013 from: https://twitter.com/about


