

Bumper Stickers on the Twitter Highway: Analyzing the Speed and Substance of Profile Changes

Ryan Wesslen,^{1*} Sagar Nandu,¹ Omar Eltayeb,¹ Tiffany Gallicano,²
Sara Levens,³ Min Jiang,² Samira Shaikh¹

¹Department of Computer Science, UNC-Charlotte

²Department of Communications Studies, UNC-Charlotte

³Department of Psychology, UNC-Charlotte

*rwesslen@unc.edu

Abstract

We describe a novel longitudinal study of the frequency and significance of social media users' profile changes. Drawing upon two formative theories from communication and psychology: self-construal and signaling theory, we examine the likelihood that users will change their profiles and what constitutes a significant profile change. Our findings indicate that users are more likely to change their Profile Summaries and Display Names than their Locations and Screen Names (i.e. handles). Further, we used topic modeling to partition users based on their profiles to identify themes and explored how profile changes differ among these thematic groups (e.g., Trump supporters). Last, we identified the most significant word changes by users in their profiles. Our findings provide valuable baseline data for further study of Twitter profiles, including the spread of social contagion through these profiles.

Introduction and Related Work

The motivation to present oneself in an inviting way is rooted in the human need to belong (Goffman 1959). Social media profiles matter because they represent attempts to portray oneself in a way that facilitates relationship cultivation with desired audiences (Schwämmlein and Wodzicki 2012). People define who they are on social media in salient ways through the information they include in their profiles. Consider Figure 1: the Display Name and Screen Name (handle) identify the user, the Profile Summary communicates key elements of identity, and the friends and follower counts indicate social capital and connectivity. In this work, we characterize Twitter profiles as online *bumper stickers*—concise statements about identity that can describe who people are, what they like, and what they value.

Profile changes could signal a significant identity choice at the micro-level (Schwämmlein and Wodzicki 2012) and the rise of a social movement at the macro-level (Raynauld, Richez, and Boudreau Morris 2018). For example, some viewers co-opted the term deplorable while others appropriated the term nasty woman during the 2016 U.S. presidential debates. There is thus significant potential of observing cultural shifts through the lens of profile changes. Although tweets have been extensively studied in the context of identity and social movements (Ray-

nauld, Richez, and Boudreau Morris 2018), the characteristics of *profile changes* are less well understood. Past computational methods investigating social media profile dynamics have focused on attributes such as screen name or follower changes (Jain and Kumaraguru 2016; Mariconti et al. 2017). Previous work has also focused on using Twitter profiles to build predictive models and understand demographics, including psycho-demographic traits, personality, age, and gender (Rao et al. 2010; Burger et al. 2011; Schwartz et al. 2013), using features such as social network, profile images and linguistic content (Liu et al. 2016; Volkova, Bachrach, and Van Durme 2016). However, research is needed to address the *implications* behind profile changes in the emergence and maintenance of social movements. Crucially, how frequently do changes occur and what might it mean when people change their profile information?

To the best of our knowledge, our research is the first to situate the characteristics of profile change in the context of seminal work grounded in communication and psychology (Donath 2007; Goffman 1959). Our research contributes theoretical advances through a novel understanding of micro- and macro-level cultural phenomena, including cultural shifts and social movements, by analyzing content changes in users' online identities.

Data Collection and Method

To address the problem of representativeness and sample selection bias (Tufekci 2014), we designed a longitudinal study of Twitter profile snapshots. First, using Twitter's



Figure 1: Sample Twitter profile and relevant fields.

Streaming API, we captured 3,423,287 tweets on September 28, 2017 (approximately 1% of that day’s tweets). Next, we parsed the tweets to identify 2,290,674 unique profile IDs. Third, we used Twitter’s REST API (Barberá 2016) to systematically ping each profile nine times over a two-week period (October 1 to 14, 2017).¹ Each snapshot took around 36 hours to successfully ping 2.29MM profiles, totaling nine snapshots over a two-week period (called Snapshot 0 to 8). In each snapshot, we retrieved the profile ID, Screen Name (handle), Display Name, Profile Summary, Location, follower count, friend count, statuses count, and creation date. We next describe our research questions and hypothesis.

RQ1: For randomly selected, active Twitter accounts, how often do users change their profile information and to what extent? We calculated the cumulative changes relative to the first observation (Snapshot 0) to identify changes across time horizons and the likelihood of a profile change up to two weeks in the future.

RQ2: How does the rate of Twitter profile updates vary for users with different thematic groups? We ran correlated topic modeling (Blei and Lafferty 2007) on words used in over 654K non-empty English (detected using Ooms and Sites 2017) Profile Summary fields at Snapshot 0 using `stm` R package (Roberts, Stewart, and Tingley 2017). We selected the number of topics ($k=49$) using the anchor-based approach (Lee and Mimno 2014). We labeled the topics using quantitative measures like semantic coherence (Mimno et al. 2011) and FREX (Bischof and Airolidi 2012) and qualitative analysis. Users are partitioned into mutually exclusive groups based on their largest topic proportion.

RQ3: What are the most prevalent types of content changes to Twitter profile summaries? We examined the substance of the profile changes, including the most frequently added, removed, and overall changed words across snapshots. Raw term frequencies might not necessarily indicate the most important terms; thus, we used TF-IDF measure to identify significant terms that were changed across the 49 groups identified in RQ2.

Hypothesis 1: Highly influential users (i.e., via followers) are more likely to update their Twitter profile than less influential users. To test this hypothesis (from Donath 2007), we calculated influence as the number of followers that a user has at Snapshot 0 (Bakshy et al. 2011). To test our hypothesis that there is positive relationship between number of followers (influence) and the likelihood of changing one’s profile, we used logistic regression on the number of followers (log transformed) on a user’s likelihood to change profile fields. We also consider the user’s number of posts and number of friends as additional independent variables.

Results

We describe our results below; we present the implications with respect to two formative theories from communication and psychology in the Discussion section.

RQ1: For randomly selected, active Twitter accounts, how often do users change their profile information and



Figure 2: Population Cumulative Likelihood Change Rates

to what extent? We examined the number of profiles that made any changes (relative to Snapshot 0) to one of four fields: Profile Summary, Display Name, Screen Name, and Location. Figure 2 provides the cumulative change rates for each field relative to Snapshot 0. We observe that users tend to change their Profile Summaries and Display Names the most (16.06% and 13.67% of users), while Location and Screen Names are changed less frequently.

RQ2: How does the rate of Twitter profile updates vary for users with different thematic groups? Figure 3 provides the topic proportions from topic modeling with the top five terms for each topic in descending order by FREX probability. Some topics focus on personal labels (father, husband, activist in Topic 32) whereas others highlight hobbies (video gamer in Topic 22). Figure 4 provides box-plots of the likelihood to change the four profile fields over two weeks. First, organizations and professionals (e.g., Topics 21 and 39) are least likely to change their profiles, perhaps because they want to maintain a consistent brand. Second, users that were categorized into topics relating to teenagers, pop music, and gender identity (Topics 1, 35, and 16) have a higher likelihood of changing their profiles.

RQ3: What are the most prevalent types of content changes to Twitter profile summaries? We examined the most prominent changed words in the users’ profiles for each group using TF-IDF. We find that these words align with the labels assigned in RQ2. For example in Figure 5, the words trump, president, and MAGA were more frequently removed than added for Trump supporters. By observing differences in the frequency with which words are added and removed within a time span, we can identify social shifts.

H1: Highly influential users are more likely to update their Twitter profile than less influential users. Influential users tended to change their profile more frequently than less influential users, supporting Donath’s (2007) hypothesis. Assuming users’ follower count as a measure of influence (Bakshy et al. 2011), we observe that users with more followers are more likely to change three of four profile metrics with 99% significance. Figure 6 provides four regression

¹Code provided on GitHub: <https://github.com/wesslen/twitter-bumper-sticker-icwsm2018>.

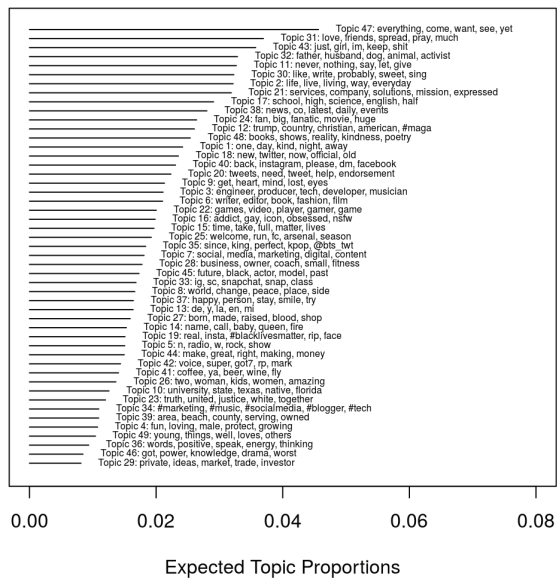


Figure 3: Profile Summary topic proportions with top five words by FREX probability (Bischof and Airoldi 2012).

results, one for each profile metric as the dependent variable. Statuses and friend count have a statistically significant relationship with profile changes. Also, Screen Name and Location models had higher Log Likelihood values, indicating a better model fit.

Discussion

We describe our theoretical foundations and then discuss the implications of our findings in relation to these theories.

Self-construal theory postulates that individuals define themselves in relation to others along three dimensions: 1) independent (e.g. #Gamer), 2) interdependent (e.g. Republican), and 3) relational (e.g. mother) (Cross, Hardin, and Gercek-Swing 2011). **Signaling theory** explains how identity can be consumed and reproduced on social media (Donath 2007). The content of one's profile sends a signal about one's identity—the independent, interdependent, and relational self-construals that a person chooses for self-representation. Social and cultural psychologists view these as three dimensions of the self that virtually all people construct to some degree, but until recently, they have examined self-construal within ethnically or nationally defined cultures (Cross, Hardin, and Gercek-Swing 2011). In the powerful online social media culture that has the potential to catalyze social and political movements (Tufekci 2014), we use self-construal theory to gain insight into how individuals define and monitor themselves within this culture. These theories offer insight into the findings described below.

Finding 1: We see in Figure 2 that users are nearly nine times more likely to change their Display Name than their Screen Name at Snapshot 8. This suggests that Display Names are a prominent field where individuals display high levels of engagement on Twitter to send signals of their collective power. **Finding 2:** From Figure 3, the heterogeneity

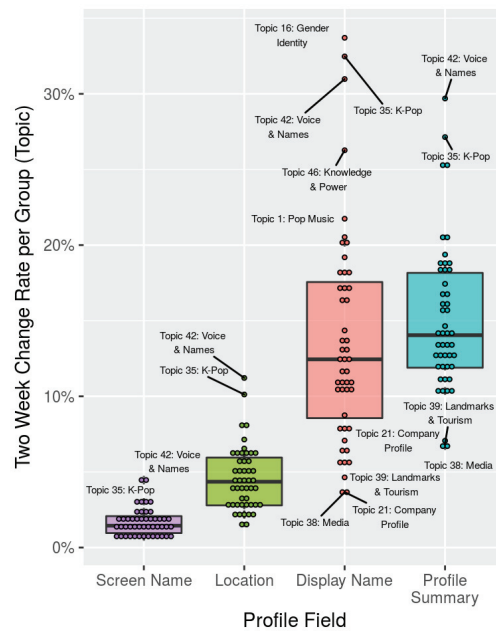


Figure 4: Change rate box-plot by profile field for each group (topic). Point labels include interpreted group labels.

of topics suggests that people are signaling key elements of identity in their Profile Summaries. Users disclose independent (e.g. artist), interdependent (e.g. conservative), and relational (e.g. father or daughter) aspects of self. Users' signaling of personal, dyadic, and group identities represents not only dimensions of identity but also value systems. Using the self-construal categories, we can infer when a user might be more self-minded or more group-minded, which has implications for estimating the potential of a user to be influenced by types of social media content. **Finding 3:** The range in change rates in Figure 4 suggests that certain groups (e.g. consumers of pop music and video games) engage in higher self-monitoring. This suggests that groups can be identified as a function of their need to signal social leadership (trend-setting) in certain domains. Conversely, it is advantageous for other groups (e.g. professions) to maintain a more fixed identity by not updating their display names in this way. **Finding 4:** The removal of keywords from a user's profile may signal shifting user preferences. Researchers can therefore look for the removal of keywords from profiles as evidence of a cultural shift. We highlight these word frequency changes as reflecting a potential social trend revealed through a simple, yet intuitive measure as TF-IDF.

Limitations and Future Work: The limitations of our study point to directions for future work. (1) We could use a longer observation window to identify long-term trends than the two-week period in our current study. (2) Additional profile attributes (profile or background images, followers, tweets before or during observation window and in multiple languages) could be examined to incorporate other ways in which users express themselves (3) Advanced statistical methods such as survival analysis could be used to

Topic Term, its TF-IDF value and (# of times it was added/removed)			
Topic 12: Trump Supporter		Topic 32: Anti-Trump (#Resist)	
military	43.11 (27/63)	nature	16.92 (12/19)
deplorable	40.43 (18/31)	hindu	16.74 (11/19)
potus	39.51 (15/15)	mommy	16.31 (18/17)
dank	35.27 (10/29)	believer	15.00 (10/49)
conservative	31.31 (23/24)	vegan	14.96 (8/23)
trump	29.67 (9/12)	concerned	14.33 (9/8)
maga	26.36 (29/43)	advocate	13.16 (9/27)
nra	24.82 (26/16)	progressive	13.06 (7/18)
president	24.44 (6/20)	groomer	12.71 (6/81)
republican	23.41 (52/33)	blueeyedbaaaby	12.71 (15/20)

Figure 5: The top 10 TF-IDF changed words and their frequency of addition and removal for the two salient topic groups: Trump supporters and Anti-Trump (#Resist) groups.

	Dependent variable:			
	Profile Summary (1)	Display Name (2)	Screen Name (3)	Location (4)
Followers (log)	0.050*** (0.003)	-0.034*** (0.003)	0.094*** (0.007)	0.111*** (0.004)
Statuses (log)	0.051*** (0.002)	0.212*** (0.003)	-0.041*** (0.006)	0.040*** (0.004)
Friends (log)	-0.078*** (0.003)	-0.110*** (0.003)	-0.172*** (0.008)	-0.152*** (0.005)
Constant	-2.041*** (0.020)	-2.926*** (0.022)	-3.289*** (0.052)	-3.225*** (0.035)
Observations	642,738	642,738	642,738	642,738
Log Likelihood	-269,822	-250,490	-53,245	-115,340
Akaike Inf. Crit.	539,651	500,987	106,499	230,688

* p<0.1; ** p<0.05; *** p<0.01

Figure 6: Logistic regressions to estimate the effect of followers, statuses, and friends on changing four profile fields.

measure duration (time to profile change), or mixed effects regression analysis could be used to analyze group-level effects over simple statistical techniques like TF-IDF. (4) Importantly, we can analyze to what extent the bumper sticker phenomenon is manifested on different social media platforms, given the different affordances these platforms offer to construct and change one’s online identities.

References

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone’s an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 65–74. ACM.

Barberá, P. 2016. tweetscores: Tools for the estimation of ideology scores with twitter data. R package version 0.1.2.

Bischof, J., and Airoidi, E. M. 2012. Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, 201–208.

Blei, D. M., and Lafferty, J. D. 2007. A correlated topic model of science. *The Annals of Applied Statistics* 17–35.

Burger, J. D.; Henderson, J.; Kim, G.; and Zarrella, G. 2011. Discriminating gender on twitter. In *Proceedings of the Con-*

ference on Empirical Methods in Natural Language Processing, 1301–1309. Association for Computational Linguistics.

Cross, S. E.; Hardin, E. E.; and Gercek-Swing, B. 2011. The what, how, why, and where of self-construal. *Personality and Social Psychology Review* 15(2):142–179.

Donath, J. 2007. Signals in social supernets. *Journal of Computer-Mediated Communication* 13(1):231–251.

Goffman, E. 1959. The presentation of self in everyday life. Garden City, NY.

Jain, P., and Kumaraguru, P. 2016. On the dynamics of username changing behavior on twitter. In *Proceedings of the 3rd IKDD Conference on Data Science, 2016*, 6. ACM.

Lee, M., and Mimno, D. 2014. Low-dimensional embeddings for interpretable anchor-based topic inference. In *Proceedings of Empirical Methods in Natural Language Processing*.

Liu, L.; Preotiuc-Pietro, D.; Samani, Z. R.; Moghaddam, M. E.; and Ungar, L. H. 2016. Analyzing personality through social media profile picture choice. In *ICWSM*, 211–220.

Mariconti, E.; Onaolapo, J.; Ahmad, S. S.; Nikiforou, N.; Egele, M.; Nikiforakis, N.; and Stringhini, G. 2017. What’s in a name?: Understanding profile name reuse on twitter. In *Proceedings of the 26th International Conference on World Wide Web*, 1161–1170. International World Wide Web Conferences Steering Committee.

Mimno, D.; Wallach, H. M.; Talley, E.; Leenders, M.; and McCallum, A. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the conference on empirical methods in natural language processing*, 262–272. Association for Computational Linguistics.

Ooms, J., and Sites, D. 2017. cld2: Google’s Compact Language Detector 2. R package version 1.1.

Rao, D.; Yarowsky, D.; Shreevats, A.; and Gupta, M. 2010. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 37–44. ACM.

Raynauld, V.; Richez, E.; and Boudreau Morris, K. 2018. Canada is# idlenomore: exploring dynamics of indigenous political and civic protest in the twitterverse. *Information, Communication & Society* 21(4):626–642.

Roberts, M. E.; Stewart, B. M.; and Tingley, D. 2017. Stm: R package for structural topic models. R package version 1.3.0.

Schwämmlein, E., and Wodzicki, K. 2012. What to tell about me? self-presentation in online communities. *Journal of Computer-Mediated Communication* 17(4):387–407.

Schwartz, H. A.; Eichstaedt, J. C.; Kern, M. L.; Dziurzynski, L.; Ramones, S. M.; Agrawal, M.; Shah, A.; Kosinski, M.; Stillwell, D.; Seligman, M. E.; et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS one* 8(9):e73791.

Tufekci, Z. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *ICWSM* 14:505–514.

Volkova, S.; Bachrach, Y.; and Van Durme, B. 2016. Mining user interests to predict perceived psycho-demographic traits on twitter. In *Big Data Computing Service and Applications (BigDataService), 2016 IEEE Second International Conference on*, 36–43. IEEE.