Divergences in Following Patterns between Influential Twitter Users and their Audiences Across Dimensions of Identity

Keywords: Twitter, identity, network, homophily, bridging

Extended Abstract

Identity spans multiple dimensions; however, the relative salience of a dimension of identity can vary markedly from person to person. Furthermore, there is often a difference between one's internal identity (how salient different aspects of one's identity are to oneself) and external identity (how salient different aspects are to the external world). We attempt to capture the internal and external saliences of different dimensions of identity for influential users ("influencers") on Twitter using the follow graph. We consider an influencer's "ego-centric" profile, which is determined by their personal following patterns and is largely in their direct control, and their "audience-centric" profile, which is determined by the following patterns of their audience and is outside of their direct control. Using these following patterns we calculate a corresponding salience metric that quantifies how important a certain dimension of identity is to an individual. We find that relative to their audiences, influencers exhibit more salience in race in their ego-centric profiles and less in religion and politics. One practical application of these findings is to identify "bridging" influencers that can connect their sizeable audiences to people from traditionally underheard communities. This could potentially increase the diversity of views audiences are exposed to through a trusted conduit (i.e. an influencer they already follow) and may lead to a greater voice for influencers from communities of color or women.

For data collection, we modify the "snowball sampling" approach used in Wu et al. to select influencers. We start with a seed set of publicly available celebrity Twitter accounts who are mostly actors/actresses, musicians, or politicians. We consider followees two-degrees away from the seed set, and then filter out to those candidates that have over 10K followers and are American. This yielded a set of 12,593 influencers. Next, we select a small random subset (< 1%) of the followers of the influencers and filter to those that follow at least twenty people in our set of influencers. This yielded a set of 80,288 audience members. We test that our results are robust over different samples of audience members. In order to tag each influencer across the different dimensions of identity, we first find the categories associated with a given dimension of identity. For racial/ethnic, LGBTQIA+, and religious categories, we use lists provided by the NIH, the LGBTQIA+ Health Education Center, and Pew respectively (NIH 2015, LGBTQIA+ 2020, Hackett and Grim 2012). Then, for each user, we intersect these terms with their Wikipedia categories.

Which influencers users choose to follow on Twitter is related to and indicative of the users' interests (Lim and Datta 2012). Therefore, for a given user we construct measures of salience across dimensions of identity based on their followees' identity tags. For race and gender, we use an entropy inspired metric that is maximized when all a user's followees are of the same race/gender (indicating high racial/gender salience). For religion, LGBTQIA+ status, and political orientation, we use a metric that takes into account how many of the user's followees contain a tag corresponding to that dimension of identity (e.g. a user that follows many users with "political" tags, such as "Democrat" or "Republican" would exhibit high political salience).

We find that racial salience is higher for influencers, while religion and politics is more salient for their audiences. While it is difficult to know precisely why we see these divergences, one potential reason that we see more homophily along race for influencers relative to audiences is that influencers often utilize Twitter more professionally (e.g. musicians follow other musicians), and occupation tends to be related to race (Queneau 2009). When investigating the influencers with extremely high gender salience, we find that a significant number of them are football players/coaches that follow other football players/coaches, who are almost exclusively men. Thus, a consequence of occupational homophily for Twitter influencers could lead to the high racial/gender salience we observe. Additionally, since an influencer's followees are public, they may be more careful when following political elites as to not alienate any of their audience. However, we emphasize that a current limitation is that we cannot explain concretely the reason for these differences and would encourage future work in this direction.
One additional application of these metrics is to identify "bridging" influencers that connect their audiences to voices from communities to which they are underexposed. To find these users, we examine influencers that exhibit a large divergence from their audience in salience across gender and race. We find several influencers, such as Dolly Parton and Melinda Gates, who follow significantly more women than their audiences. If these influencers were aware of this fact, they could potentially use their other platform to elevate voices from communities that their audience may be less exposed to.

References

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who says what to whom on Twitter. In Proceedings of the 20th International Conference on World Wide Web (WWW '11) (pp. 705-714). Association for Computing Machinery. doi: 10.1145/1963405.1963504

NIH. (2015). Racial and Ethnic Categories and Definitions for NIH Diversity Programs and for Other Reporting Purposes. Retrieved from https://grants.nih.gov/grants/guide/notice-files/not-od-15-089.html

LGBTQIA+. (2020, February). LGBTQIA+ Glossary of Terms for Health Care Teams. Retrieved December 22, 2021, from https://www.lgbtqiahealtheducation.org/publication/lgbtqia-glossary-of-terms-for-health-care-teams

Hackett, C., & Grim, B. (2012, December). The Global Religious Landscape: A Report on the Size and Distribution of the World's Major Religious Groups as of 2010. Retrieved from Pew Research Center website: https://www.pewforum.org/2012/12/18/global-religious-landscape-exec/

Lim, K. H., & Datta, A. (2012). Finding Twitter Communities with Common Interests Using Following Links of Celebrities. In Proceedings of the 3rd International Workshop on Modeling Social Media (MSM '12) (pp. 25-32). Association for Computing Machinery. doi: 10.1145/2310057.2310064

Queneau, H. (2009). Trends in occupational segregation by race and ethnicity in the USA: Evidence from detailed data. Applied Economics Letters, 16(13), 1347-1350. doi: 10.1080/13504850701367346

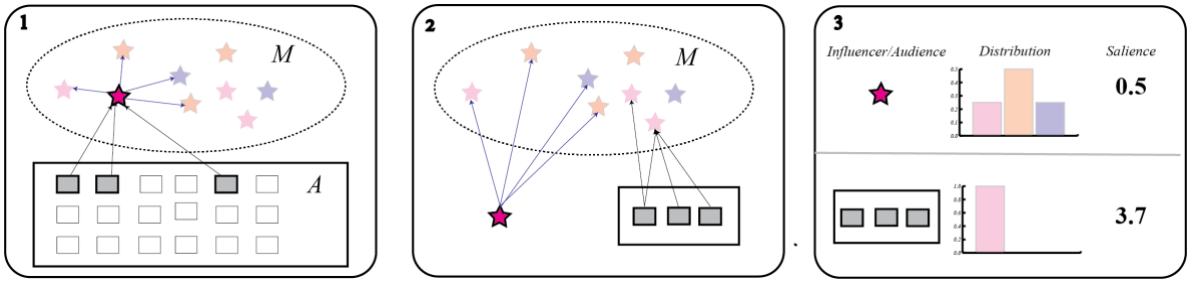

Figure 1. An example of computing identity salience. (1) shows the universe of influencers *M* with the set of audience members *A*. We focus on a single influencer (the dark outlined pink star) and their followers (the dark grey boxes in *A*), with arrows indicating a follow relationship. (2) illustrates how we find the influencer's followees in *M* as well as their audience's followees. Finally, given those follow relationships we find a distribution over identity dimensions and compute a salience metric in (3). Here, since the audience's followees are more homophilous than the influencer's, the audience would have a higher salience score on this dimension of identity.