Emoji and Group Identity on Twitter

CARL COLGLAZIER, North Carolina State University, USA
ZACKARY ALLEN, North Carolina State University, USA

We propose a measure for group identity on Twitter using emoji displayed by users in their names or biographies. Viewing the use of emoji as social phenomenon, we introduce methods for measuring how often emoji users follow/friend other emoji users compared to non-emoji users. Furthermore, we introduce methods to measure quantify the frequency users friend/follow users with similar emojis, concluding with visual analysis of what ties between certain emoji look like.

1 BACKGROUND

Many Twitter users engage with the platform to present a formed identity to an imagined audience [1]. This activity often flattens identity as individuals try to fit in to larger groups [2]. At the same time, Twitter limits the metadata associated with each profile by design.

Twitter has long used "hovercards" which show a small set of information about a profile in the timeline [3]. Thus when someone encounters an unfamiliar profile, they can quickly view metadata including the user's display name, handle, and a short biography. By elevating these short pieces of metadata, Twitter creates an incentive for users to succinctly spotlight pieces of their identity to potential audience members through these attributes. It is reasonable to assume many users use emoji here in large part because the limit for a display name is 50 characters and the limit for a bio is 160 characters.

Twitter's main feature is the ability to post and view Tweets—short messages limited to 280 characters. Tweets are typically intended for the author's followers, though they often get spread beyond just the user's followers thereby encouraging a social community. Tweets can spread in a variety of ways: they can be retweeted directly, quote retweeted, or shown in another user's feed when someone they follow likes the post. Each of these methods let a user see another user's Tweet organically even without following them. Thus Twitter users may be exposed to people outside of their immediate social circles and make judgements about these accounts on the basis of presented metadata rather than social association.

Symbols often represent groups and can be political in nature. Following Benedict Anderson's concept of "imagined communities" to describe the concept of nationalism [4], for instance, we can think of national symbols such as flags or seals as materialized representations of group identity. Indeed, national flags make up a good part of the set of standardized emoji [5]. These symbols–even within the context of an international, standardized system like emoji–can prove controversial. The Information and Communication Ministry of Indonesia, for instance, sought to remove emoji depicting same-sex couples [6]. The standardization of emoji has also raised concerns over the extent of their representation and diversity.

Twitter has supported emoji on their platform since 2014. In fact, Twitter even developed an open-source emoji library. The site's users have likewise embraced emoji characters and use them frequently. Emojitracker, a service which logs how Twitter accounts use emoji, has logged over a two billion instances of the most popular emoji on the list, the "Face with Tears of Joy" emoji [7]. This emoji was declared the 2015 word of the year by Oxford Dictionaries.

Authors' addresses: Carl Colglazier, North Carolina State University, Raleigh, NC, 27695, USA; Zackary Allen, North Carolina State University, Raleigh, NC, 27695, USA.

Based on the hypothesis that many Twitter users include emoji in their profile metadata as a signal for group identity, we predict that users with certain emojis in their name or biography will connect at a higher rate with users that have similar emojis. We further expect some emoji to be associated with this behavior more than others. For instance, the rose emoji, which is used by members of the democratic socialist movement, would more likely show homophilic behavior compared to more generic emoji such as the red heart emoji.

Understanding the social intricacies of emoji use on Twitter can be valuable for marketing and public relations. For organizations that use Twitter, emojis can be used to make their organization seem more approachable. If done correctly, those that use these symbols correctly can pull on the sense of identity that it brings to blend into the group. By taking advantage of the way emojis can create a sense of belonging, organizations could strengthen the effectiveness of their social media appearance.

1.1 Research Questions

- R1: Do users with emojis make more connections?
- R2: Do users tend to follow/friend more users with similar emoji?
- R3: What does the overall emoji network look like?

2 METHOD

From the Twitter API, we amassed a significant number of Twitter user-ids. With these ids, we created two datasets: a control dataset of users without emojis in their screen names or biographies and a sample dataset of users with emojis in their screen names or biographies. From there, we constructed a third dataset of secondary user information on their connections (accounts they follow and accounts that follow them).

The sample and control accounts were originally sampled from a set of 400,000. The secondary accounts were pulled for 2,500 users in the sample and 2,500 users in the control group. As some of the users in both groups had updated their metadata in the time since the set of 400,000 users was collected, a subset of 1,000 users from the sample and control were randomly chosen. Those in the sample group had an emoji in their name or biography the first time their profile was viewed and continued to use emoji at the time of the study. Those in the control group did not use emoji during either period of time.

2.1 Challenges

- How can we scrape useful amounts of data using only a free Twitter API key?
- How can we simultaneously draw conclusions about users and emojis?
- How can we graph a network of secondary users and their emojis?

3 RESULTS

3.1 Sample Analysis

Before any analysis can occur, we must first determine if parametric statistics apply. We can test for the normalcy of the distribution using D'Agostino and Pearson's test for departure from normalcy.

As Table 1 shows, the null hypothesis that each sample comes from a normal distribution can be rejected. The Mann-Whitney rank test is used instead to test the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.

Table 1. D'Agostino and Pearson's test on variables.

Group	p-value	statistic
sample $\text{listed}_{\text{count}}$	3.66×10^{-295}	1355.9307
sample $\text{followers}_{\text{count}}$	1.31×10^{-291}	1339.5606
sample $\text{friends}_{\text{count}}$	$0.00\times10^{+00}$	1483.7946
control $\text{listed}_{\text{count}}$	$0.00\times10^{+00}$	2114.8148
control $\text{followers}_{\text{count}}$	$0.00\times10^{+00}$	1779.9979
control $\text{friends}_{\text{count}}$	1.10×10^{-248}	1141.8983

Table 2. Mann-Whitney rank tests.

Variable	p-value	statistic
$\text{listed}_{\text{count}}$	4.23×10^{-02}	478001.5
$\text{followers}_{\text{count}}$	5.39×10^{-23}	373393.5
$\text{friends}_{\text{count}}$	7.99×10^{-16}	397093.0

Table 2 suggests a statistically significant difference in the distributions between the sample and control groups for the number of lists in which users appear, the number of accounts users follow, and the number of followers for each user.

Figure 1 shows the distribution of the ranks for the variables. Matching the results from 2, the number of lists in which users appear is the least significant of the three variables. The number of followers and accounts followed (friends) have differences in distributions and rank means which are visually apparent.

3.2 Network Analysis

The network analysis looks at the accounts that follow users in the sample and control group and the accounts that users in the sample and control group follow. The total size of this population was $N = 4192099$.

Table 3. A comparison of the number of followers of users in the sample and control groups for the most-used emoji.

Emoji	Total Count	Sample Followers	Control Followers
✨	26895	7743	2210
🇺🇸	24072	15350	8901
👑	12985	3824	1163
❌	11830	8596	4874
🌹	11396	3897	1529
🌊	10740	6892	4896
❤️	9584	3444	1422
🌻	7643	2250	845
🎄	7417	1955	923
💜	7371	2189	484

Table 3 and Table 4 show that for each of the top ten most-used emoji, users in the sample group had more followers using each emoji in their profiles and followed more users using each emoji in their profiles.

Table 5 and Figure 2 demonstrate that for almost all of the most popular emoji, users were followed most commonly by profiles that shared the same emoji.

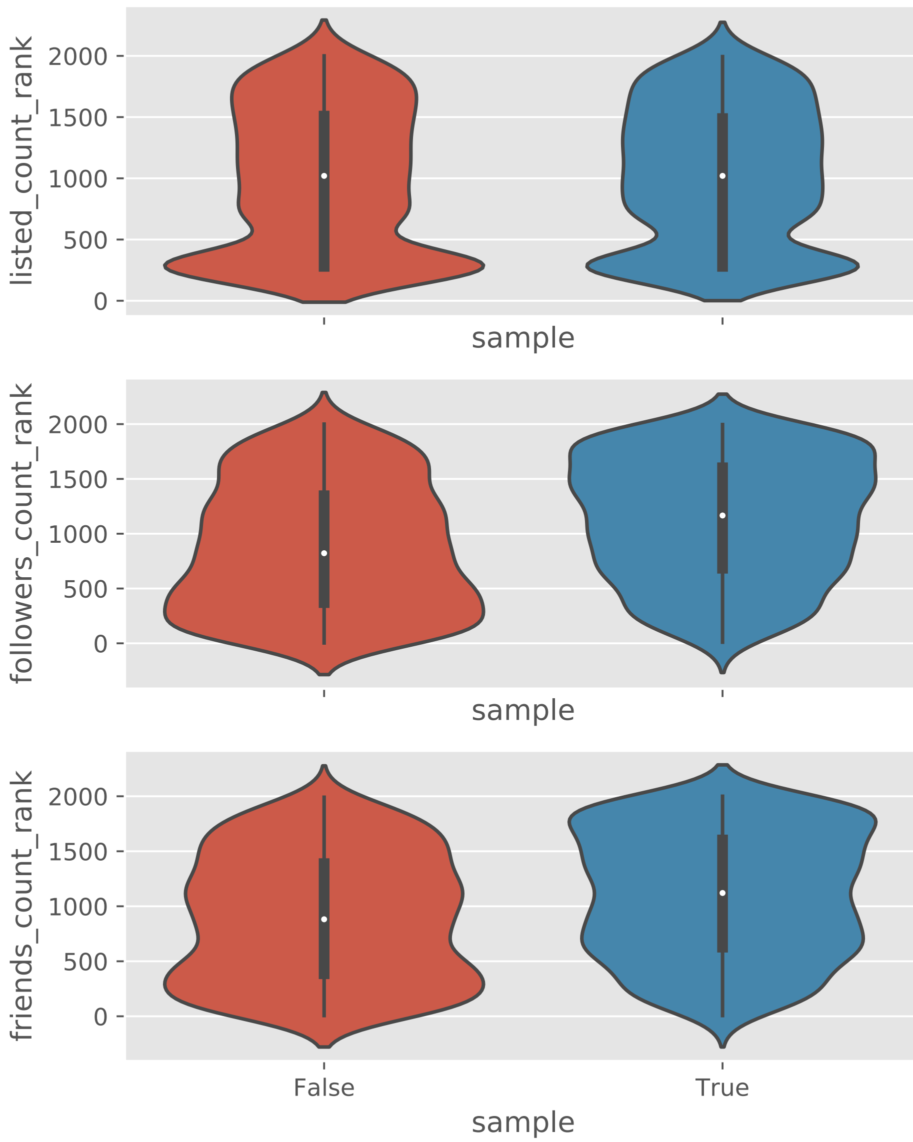

Fig. 1. Violin plot for the three variables in the sample analysis.

Table 4. A comparison of the number of accounts using emoji for the users followed by accounts in the sample and control groups for the most-used emoji.

Emoji	Total Count	Sample Following	Control Following
✨	26895	7153	2386
🇺🇸	24072	15776	8883
👑	12985	3199	1126
❌	11830	8660	4784
🌹	11396	3329	1582
🌊	10740	6791	4876
❤️	9584	2882	1268
🌻	7643	2075	892
🎄	7417	2314	1191
💜	7371	1809	494

Table 5. Matrix representing the number of accounts with each emoji (columns) following accounts with other emoji (rows).

	✨	❄️	❌	❤️	⭐	🇺🇸	🌊	🌹	👑	💜
✨	274	35	14	0	4	39	9	7	3	30
❄️	18	96	4	0	0	4	94	2	3	4
❌	3	1	969	0	331	2678	1	1	1	0
❤️	95	31	101	0	44	281	29	9	9	14
⭐	7	4	494	0	256	1377	3	0	3	4
🇺🇸	5	245	1434	0	531	3572	246	2	10	2
🌊	14	655	0	0	2	12	665	4	30	0
🌹	32	14	36	0	15	91	15	78	1	4
👑	109	12	18	0	1	19	8	2	14	4
💜	16	76	16	0	4	33	17	2	2	5

4 DISCUSSION

We used our dataset to carry out analysis on our hypothesis with both statistical and graph models. From our statistical models we found that users who use emoji on their profiles tend to connect with more users than users without emojis (R1). We found this by measuring the followers count, friends count, and listed count columns for each user in our first two datasets. For the secondary user side, we found users form cliques with similar emoji users, which can be measured by counting emojis per secondary user and sorting by the total most popular emojis over the third dataset (R2). Unsurprisingly, this applies to both followers and friends of the primary user from the first two datasets.

With our graph in Figure 2, the nodes are emojis and the edges are weighted by the number of connections between emoji. From our graph model we found that visually identifying strong ties between certain emojis is trivial (R3). It furthermore presents a validation method for the theory based on analysis of the homogeneity of Twitter networks.

4.1 Limitations

The small sample size (N = 1000) limits the ability to compare attributes between users with different kinds of emoji. Further study could reveal the differences between populations with specific emoji (say 🌹 verses 🌊), but the sample size we used was too small for each emoji to do this kind of analysis.

The samples were not stratified by location, account age, activity levels, or popularity. Further study could reveal if the increased activity levels associated with emoji use might be better explained by other factors.

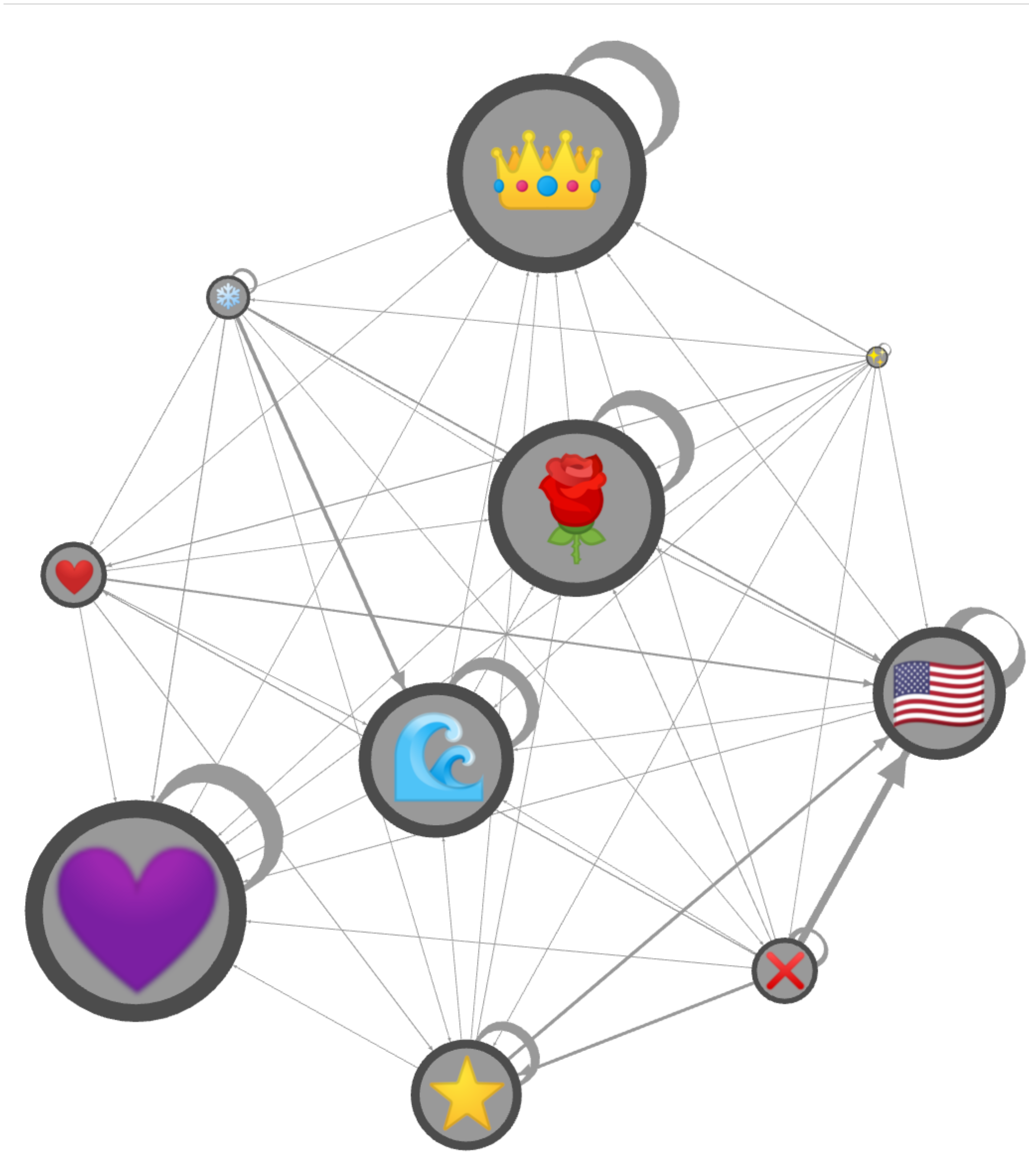

Fig. 2. The network of emojis are weighted by count. As shown by thick cyclic lines, secondary users tend to follow users with similar emojis. Interestingly, some emojis like the red X and American flag have strong associations.

Nonetheless, the results from R2 and R3 indicate that the use of emoji on Twitter profiles is influenced by group identity and indicate this could be a good area for further study.

REFERENCES

[1] Alice E Marwick and Danah Boyd. “I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience”. In: *New media & society* 13.1 (2011), pp. 114–133.

[2] Danah Michele Boyd. *Taken out of context: American teen sociality in networked publics*. University of California, Berkeley, 2008.

[3] Twitter Inc. *Flying Around With Hovercards*. https://blog.twitter.com/official/en_us/a/2010/flying-around-with-hovercards.html. [Online; accessed 17-Oct-2018]. 2010.

[4] Benedict Anderson. *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books, 2006.

[5] The Unicode Consortium. *Unicode® Technical Standard #51*. https://unicode.org/reports/tr51/#Flags.

[6] Tom Boellstorff. "Against state straightism: Five principles for including lgbt Indonesians". In: *E-International Relations* 21 (2016), pp. 181–207.

[7] M Rothenberg. "Emojitracker: realtime emoji use on Twitter". In: *Retrieved April* 28 (2013), p. 2017.