Classifying Group Identity on Twitter and Exploring its Impact on Information Flow

We each have groups we identify with. In a given moment we may see ourselves as Democrats, or members of IU Cognitive Science, or Americans, Germans, Fathers, Mac People, Women, etc. The construct of a "group identity"—a group we categorize ourselves with—has significant explanatory power for human behavior. Decades of research in social psychology demonstrate that group identities, when made salient, can influence behavior toward others, self-perceptions, and emotional responses to events (Kameda, 1997; Haslam, 1995, 1996).

Though thousands of experiments have manipulated group identity in the lab, little work has explored how group identities impact the flow of information in social networks. Taking for example the occupational identity of a teacher: Does information about new legislation impacting teachers flow faster in a network of highly identified teachers than less identified ones? Is information that critiques the job teachers are doing less likely to find its way into such a network? In general, how does the content of information shared by highly identified group members differ from that shared by people who are less identified? Are, for instance, highly identified people more likely to spread information that ridicule's opposing groups? In a related vein, how does the flow of group identity-relevant information differ for different kinds of identities? (e.g. National identity vs. Sports teams vs. Political)

Even ten years ago, answering such questions was not feasible for most researchers. We didn't have cost-effective, efficient ways to track information flows across social networks. Fortunately, the landscape has changed with publicly available data from the social networking site Twitter. With over 300 million users distributed across the globe, Twitter is a potential gold mine for psychological inquiry in an ecologically valid environment.

Despite the demonstrated explanatory power of the group identity construct in a host of domains, few Twitter studies have utilized group identity as a factor of interest. While there has been some work on classifying users from their behavior on Twitter, such as content from URL's posted in a given network (Herdagdelen et al 2012), labeled lists the users have been put into (Wu et al, 2010), or through laborious human coding (Yardi & boyd), none of these studies have directly asked people about their own group identities. Hence, previous work has lacked any "ground truth" to assess the accuracy of classification methods. For such *ground truth*, survey data with explicit labels of self-reported group identities is necessary.

The current proposal seeks to extend previous research in several ways. First, we will build a classifier of group identity that applies a number of data-mining and natural language processing methods in combination with survey responses that provide the ground truth of group identity. Crucially, unlike existing work, we aim to not only classify which discrete group categories (e.g. Democrat vs. Republican) people belong to, but to generate predictions about the *strength* of identification (e.g. strongly identified vs. weakly identified). To achieve this we will look at factors like the sentiment of tweets on topics related to group identity and the proportion of tweets that are related to group-identity. We hypothesize that the more frequent and emotional group identity-relevant tweets are, the stronger the identification.

Upon building a classifier, we will explore how group identities impact the flow of information in social networks. As outlined in the Methods section, we plan to do this by looking at both organically generated content and utilizing an experimental method (novel for Twitter research) where we plant specific content in networks and track its dissemination. We hypothesize that group-related information will flow more quickly through homogenous, highly identified networks. We will also explore whether different group identities (e.g. Political vs. Sports-related) show differential patterns in information flow.

Though our ultimate goal is to determine how group identities affect the flow of information in social networks, by building an effective classifier of group identity we open the door for numerous other analyses. For instance, one might expect that networks of highly identified group members would have larger clustering coefficients (e.g. there should be more mutual links between group members). We can see if this assumption holds for different types of identities. Another important application of this classifier involves looking at the role of time-dependence on the salience of group identity. For example, if someone posts something related to their political identity, in theory, this identity is temporarily made salient. Hence, we would expect this person to have a higher probability of posting something else political during a short subsequent time window. Finally, the classifier would also help us to explore the sharing of group emotions in response to group-relevant events, such as the outcome of a presidential election, or a sports team's victory or defeat.

Methods

Note: For the sake of concreteness, all following descriptions use political identity as the motivating example. We plan to utilize the same methods to study other types of identity such as sports team allegiance, cultural heritage, or religious identity and to compare these in our analyses.

Phase 1: Finding a Seed Population

Initially we will find Twitter users who appear to be interested in politics and, hence, are likely to identify with a particular party. Individuals who tweet about things that could be classified as political will be randomly selected. A number of methods will be employed for this such as analyzing people's use of politically related keywords (often marked by hash-tags) or URLs they have posted or that have been posted among close connections in their network. Phase 2 will outline other techniques that can be used for this task if necessary.

Once politically interested users have been identified, the next critical step involves having them answer questions about their group identities. One reason why such survey methodology has almost never been incorporated into Twitter research is that Twitter makes it difficult to directly solicit users. In order to directly contact someone, they need to "follow you", i.e. choose to see the content you post in advance.

However, there is an indirect method we plan to employ. It is possible to reference someone in a Tweet (a "mention" in Twitter terminology) and the user will be informed and get a chance to see what was posted about them. Hence, we plan to mention users of interest and in the mention ask them if they would be willing to fill out a survey with a posted link. Those users who agree to fill out the survey will be asked to also participate in the experimental component of the study, where they would be asked to Tweet particular group related content at specific times to allow us to track the flow of this information across their network. Finally, we would encourage respondents to "retweet" the link to the survey, so that we can also learn something about the group identities of their local network.

Though we expect a relatively low percentage of people we contact will fill out the survey and even fewer will participate in the experimental portion, we can efficiently target many individuals by automating the tweets mentioning them, to allow for a sufficient sample size.

Phase 2: Creating the Classifier

Once we have a seed population of individuals of whom we know their group identities, we can begin to use their Twitter content to develop an effective classifier. The classifier will essentially treat all their twitter content as a large feature vector. We will then use a variety of methods, outlined below, to model their identities.

Hashtag analysis: It's an increasingly common convention to label tweets on specific topics with a hashtag. Hence, one predictor we will use is the topical content of tweets as indicated by hashtags.

List Methodology: Twitter users have the option of putting users in lists and naming the list anything the users chooses. Work by Winter Mason and colleagues used this feature to infer whether users belonged to one of four categories—bloggers, media outlets, celebrities and organizations—but this technique could be used to find many other kinds of group identities for many users. Additionally, they found that around 50% of all content shared on Twitter, through URL's and re-posts, originates from around 20,000 elite users in those four categories. The particular elite users someone follows could also be used to glean their group identity (e.g. if someone follows Steven Colbert or Jon Stewart, there is a higher probability that they are a Democrat).

Classifying Tweet Topics Using Natural Language Processing: Currently Winter Mason and colleagues at Stevens Institute are refining an algorithm that classifies Tweets based on how similar they are to Wikipedia pages (Genc et al, 2011). We plan to apply this and potentially some other techniques that have seen success at categorizing Tweets, such as Latent Dirichlet Allocation (LDA; Ramage et al). We can then use the classified tweets to infer the identities of the user.

Analysis of URL posts: We will use similar natural language processing techniques to analyze the content of URLs posted by Twitter users.

In addition to looking at the Twitter content from individuals we are attempting to classify, we can apply the same methods to people in their networks to see the extent to which one's identity can be predicted by the identity of members of one's network, similar to Jernigan & Mistree (2009). In cases where there is very limited Twitter content for a given user (e.g. someone doesn't tweet frequently), this may be the most effective way to predict their group identities.

Since we will have ground truth data on group identity, from Phase1, we will be able see how effective our classification methods are and progressively refine them using a host of techniques.

Phase 3: Exploring How Group Identity Impacts Information Flow

After creating a classifier, we can look for groups of connected individuals with similar group identities. For example, we might find a connected group comprised of highly identified Republicans, or moderately identified Democrats. We can then look at the relative frequency of group-related topic posts and discussions across these different networks. We plan to use two methods of doing this. The first method involves coding politically related content that organically appears in the network. Such content could be a posted link, a meme, or a statement. We would look at the rate at which this content is reposted and referenced within the network.

In a more ambitious vein, we would like to perform a quasi-experiment on Twitter, which has never been done. Here we would ask certain individuals who agreed to participate in the experiment during phase 1 to post specific content related or unrelated to their group identity. This would allow us to experimentally test how group identity interacts with content in the flow of information through social networks. It would also be an important demonstration of how experimental methods can be employed on Twitter.

Since we will create classifiers for multiple kinds of identities, we would like to repeat similar methods and analytical tools to compare information flow for different group identities.

Intellectual Merit

This project would provide valuable knowledge about how group identity impacts the flow of different kinds of information through a social network. This would reveal a new dimension to the voluminous literature on the diffusion of information in social networks. It would also extend research on the relationship between social influence and group identity, situating the process of social influence and group identity in the wider framework of social networks.

This proposal includes many significant methodological advances. It would be one of the first to employ survey methodology in combination with Twitter data mining. It would be the first to combine URL analysis, content analysis using Wikipedia mapping, and the List methodology to create a classifier. This would also be the first study to experimentally manipulate tweets on Twitter.

The classifier we create can be used in other studies and by other researchers. Such a classifier could stand as a good demonstration of how survey data, data-mining and natural language processing techniques can be incorporated to predict what people identify with based on the content of their posts on social networking sites.

Broader impacts

Aside from helping the research community, this work could be used in applications that allow people on Twitter to connect with users of similar group identities and interests beyond what is possible with current search techniques on Twitter. Hence, activists could better link up with other activists; teachers could more efficiently find each other and share resources and reflections, etc. Such work could also be used to recommend content to users based on the simple principle that people with similar group identities likely enjoy similar content related to that identity. In another vein, this work could allow the general public to understand how their group identities are revealed through the messages they post and how these identities might influence the types of information they see on their network as well how readily such information is disseminated. This is valuable knowledge for every citizen in an era when our interactions are increasingly occurring on social networking sites. The research might also see fruitful applications for NGOs and charitable organizations looking to find which users on social networking sites might represent their key demographic based on group identities.

Host Choice

Though there are several researchers at IU who work with Twitter, none of them are social psychologists. Dr. Mason is unique in that he both has extensive knowledge and experience in data-mining, because of his time at Yahoo! Research, as well as survey and experimental methods in psychology. He has the distinction of being the only other person I know, perhaps the first in IU history, to double major in Cognitive Science and Social Psychology. The kind of interdisciplinary work he conducts in online environments is the kind of work I would like transition into. Being able to work directly with Dr. Mason for a summer and learn data-mining skills hands-on with an exciting, innovative project would be a tremendous opportunity.

References

Y Genc, Y Sakamoto, J Nickerson (2011). Discovering context: Classifying tweets through a semantic transform based on Wikipedia. *Foundations of Augmented Cognition* 2011 – Springer

S Haslam, J Turner (1995). Context-dependent variation in social stereotyping 3: Extremism as a self-categorical basis for polarized judgment. *European Journal of Social Psychology*

S Haslam, P Oakes, J Turner, C McGarty (1996). Social categorization and group homogeneity: Changes in the perceived applicability of stereotype content as a function of comparative context and trait favourableness. *British Journal of Social Psychology*

Kameda, Tatsuya, Ohtsubo, Yohsuke,Takezawa, Masanori (1997). Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. *Journal of Personality and Social Psychology*

Herdagdelen, W. Zuo, Gard-Murray, Bar-Yam (2012): An exploration of social identity: The geography and politics of news-sharing communities in twitter. arXiv:1202.4393v1 [physics.soc-ph] 20 Feb 2012.

Ramage, Dumais, Liebling (2010). Characterizing Microblogs with Topic Models. *ICWSM 2010.*

S Yardi and D Boyd. "Dynamic Debates: An Analysis of Group Polarization over Time on Twitter." Special Issue on Persistence and Change in Social Media. In Bulletin of Science. *Technology and Society* 30(5), Oct 2010.

Wu, S., Hofman, J., Mason, W., & Watts, D. J. (2011). Who Says What to Whom on Twitter. *International Conference on the World Wide Web (WWW), 2011, Hyderabad, India*