

#WhoAmI in 160 Characters? Classifying Social Identities Based on Twitter Profile Descriptions

Anna Priante
Public Administration
University of Twente
a.priante@utwente.nl

Djoerd Hiemstra
Database Group
University of Twente
d.hiemstra@utwente.nl

Tijs van den Broek
NIKOS
University of Twente
t.a.vandenbroek@utwente.nl

Aaqib Saeed
Computer Science
University of Twente
a.saeed@student.utwente.nl

Michel Ehrenhard
NIKOS
University of Twente
m.l.ehrenhard@utwente.nl

Ariana Need
Public Administration
University of Twente
a.need@utwente.nl

Abstract

We combine social theory and NLP methods to classify English-speaking Twitter users' online social identity in profile descriptions. We conduct two text classification experiments. In Experiment 1 we use a 5-category online social identity classification based on identity and self-categorization theories. While we are able to automatically classify two identity categories (Relational and Occupational), automatic classification of the other three identities (Political, Ethnic/religious and Stigmatized) is challenging. In Experiment 2 we test a merger of such identities based on theoretical arguments. We find that by combining these identities we can improve the predictive performance of the classifiers in the experiment. Our study shows how social theory can be used to guide NLP methods, and how such methods provide input to revisit traditional social theory that is strongly consolidated in offline settings.

1 Introduction

Non-profit organizations increasingly use social media, such as Twitter, to mobilize people and organize cause-related collective action, such as health advocacy campaigns.

Studies in social psychology (Postmes and Brunsting, 2002; Van Zomeren et al., 2008; Park and Yang, 2012; Alberici and Milesi, 2013; Chan, 2014; Thomas et al., 2015) demonstrate that *social identity* motivates people to participate in collective action, which is the joint pursuit of a common goal or interest (Olson, 1971). *Social identity* is an individual's

self-concept derived from social roles or memberships to social groups (Stryker, 1980; Tajfel, 1981; Turner et al., 1987; Stryker et al., 2000). The use of language is strongly associated with an individual's social identity (Bucholtz and Hall, 2005; Nguyen et al., 2014; Tamburrini et al., 2015). On Twitter, profile descriptions and tweets are online expressions of people's identities. Therefore, social media provide an enormous amount of data for social scientists interested in studying how identities are expressed online via language.

We identify two main research opportunities on online identity. First, online identity research is often confined to relatively small datasets. Social scientists rarely exploit computational methods to measure identity over social media. Such methods may offer tools to enrich online identity research. For example, Natural Language Processing (NLP) and Machine Learning (ML) methods assist to quickly classify and infer vast amounts of data. Various studies investigate how to predict individual characteristics from language use on Twitter, such as age and gender (Rao et al., 2010; Burger et al., 2011; Al Zamal et al., 2012; Van Durme, 2012; Ciot et al., 2013; Nguyen et al., 2013; Nguyen et al., 2014; Preotiuc-Pietro et al., 2015), personality and emotions (Preotiuc-Pietro et al., 2015; Volkova et al., 2015; Volkova and Bachrach, 2015), political orientation and ethnicity (Rao et al., 2010; Pennacchiotti and Popescu, 2011; Al Zamal et al., 2012; Cohen and Ruths, 2013; Volkova et al., 2014), profession and interests (Al Zamal et al., 2012; Li et al., 2014).

Second, only a few studies combine social theory and NLP methods to study online identity in

relation to collective action. One recent example uses the Social Identity Model of Collective Action (Van Zomeren et al., 2008) to study health campaigns organized on Twitter (Nguyen et al., 2015). The authors automatically identify participants' motivations to take action online by analyzing profile descriptions and tweets.

In this line, our study contributes to scale-up research on online identity. We explore automatic text classification of online identities based on a 5-category social identity classification built on theories of identity. We analyze 2633 English-speaking Twitter users' 160-character profile description to classify their social identities. We only focus on profile descriptions as they represent the most immediate, essential expression of an individual's identity.

We conduct two classification experiments: Experiment 1 is based on the original 5-category social identity classification, whereas Experiment 2 tests a merger of three categories for which automatic classification does not work in Experiment 1. We show that by combining these identities we can improve the predictive performance of the classifiers in the experiment.

Our study makes two main contributions. First, we combine social theory on identity and NLP methods to classify English-speaking Twitter users' online social identities. We show how social theory can be used to guide NLP methods, and how such methods provide input to revisit traditional social theory that is strongly consolidated in offline settings.

Second, we evaluate different classification algorithms in the task of automatically classifying online social identities. We show that computers can perform a reliable automatic classification for most social identity categories. In this way, we provide social scientists with new tools (i.e., social identity classifiers) for scaling-up online identity research to massive datasets derived from social media.

The rest of the paper is structured as follows. First, we illustrate the theoretical framework and the online social identity classification which guides the text classification experiments (Section 2). Second, we explain the data collection (Section 3) and methods (Section 4). Third, we report the results of the two experiments (Section 5 and 6). Finally, we discuss our findings and provide recommendations for future research (Section 7).

2 Theoretical Framework: a 5-category Online Social Identity Classification Grounded in Social Theory

We define *social identity* as an individual's self-definition based on social roles played in society or memberships of social groups. This definition combines two main theories in social psychology: identity theory (Stryker, 1980; Stryker et al., 2000) and social identity, or self-categorization, theory (Tajfel, 1981; Turner et al., 1987), which respectively focus on social roles and memberships of social groups. We combine these two theories as together they provide a more complete definition of identity (Stets and Burke, 2000). The likelihood of participating in collective action does increase when individuals both identify themselves with a social group and are committed to the role(s) they play in the group (Stets and Burke, 2000).

We create a 5-category online social identity classification that is based on previous studies of offline settings (Deaux et al., 1995; Ashforth et al., 2008; Ashforth et al., 2016). We apply such classification to Twitter users' profile descriptions as they represent the most immediate, essential expression of an individual's identity (Jensen and Bang, 2013). While tweets mostly feature statements and conversations, the profile description provides a dedicated, even limited (160 characters), space where users can write about the self-definitions they want to communicate on Twitter.

The five social identity categories of our classification are:

(1) Relational identity: self-definition based on (reciprocal or unreciprocal) relationships that an individual has with other people, and on social roles played by the individual in society. Examples on Twitter are “*I am the father of an amazing baby girl!*”, “*Happily married to @John*”, “*Crazy Justin Bieber fan*”, “*Manchester United team is my family*”.

(2) Occupational identity: self-definition based on occupation, profession and career, individual vocations, avocations, interests and hobbies. Examples on Twitter are “*Manager Communication expert*”, “*I am a Gamer, YouTuber*”, “*Big fan of pizza!*”, “*Writing about my passions: love cooking traveling reading*”.

(3) **Political identity:** self-definition based on political affiliations, parties and groups, as well as being a member of social movements or taking part in collective action. Examples on Twitter are “*Feminist Activist*”, “*I am Democrat*”, “*I’m a council candidate in local elections for []*”, “*mobro in #movember*”, “*#BlackLivesMatter*”.

(4) **Ethnic/Religious identity:** self-definition based on membership of ethnic or religious groups. Examples on Twitter are “*God first*”, “*Will also tweet about #atheism*”, “*Native Washingtonian*”, “*Scottish no Australian no-both?*”.

(5) **Stigmatized identity:** self-definition based on membership of a stigmatized group, which is considered different from what the society defines as normal according to social and cultural norms (Goffman, 1959). Examples on Twitter are “*People call me an affectionate idiot*”, “*I know people call me a dork and that’s okay with me*”. Twitter users also attach a stigma to themselves with an ironic tone. Examples are “*I am an idiot savant*”, “*Workaholic man with ADHD*”, “*I didn’t choose the nerd life, the nerd life chose me*”.

Social identity categories are not mutually exclusive. Individuals may have more than one social identity and embed all identities in their definition of the self. On Twitter, it is common to find users who express more than one identity in the profile description. For example, “*Mom of 2 boys, wife and catholic conservative, school and school sport volunteer*”, “*Proud Northerner, Arsenal fan by luck. Red Level and AST member. Gamer. Sports fan. English Civic Nationalist. Contributor at HSR. Pro-#rewilding*”.

3 Data Collection

We collect data by randomly sampling English tweets. From the tweets, we retrieve the user’s profile description. We remove all profiles (i.e., 30% of the total amount) where no description is provided.

We are interested in developing an automatic classification tool (i.e., social identity classifier) that can be used to study identities of both people engaged in online collective action and general Twitter users. For this purpose, we use two different sources to collect our data: (1) English tweets from two-year (2013 and 2014) Movember cancer aware-

ness campaign¹, which aims at changing the image of men’s health (i.e., prostate and testicular cancer, mental health and physical inactivity); and (2) English random tweets posted in February and March 2015 obtained via the Twitter Streaming API. We select the tweets from the UK, US and Australia, which are the three largest countries with native English speakers. For this selection, we use a country classifier, which has been found to be fairly accurate in predicting tweets’ geolocation for these countries (Van der Veen et al., 2015). As on Twitter only 2% of tweets are geo-located, we decide to use this classifier to get the data for our text classification.

From these two data sources, we obtain two Twitter user populations: Movember participants and random generic users. We sample from these two groups to have a similar number of profiles in our dataset. We obtain 1,611 Movember profiles and 1,022 random profiles. Our final dataset consists of 2,633 Twitter users’ profile descriptions.

4 Methods

In this study, we combine qualitative content analysis with human annotation (Section 4.1) and text classification experiment (Section 4.2).

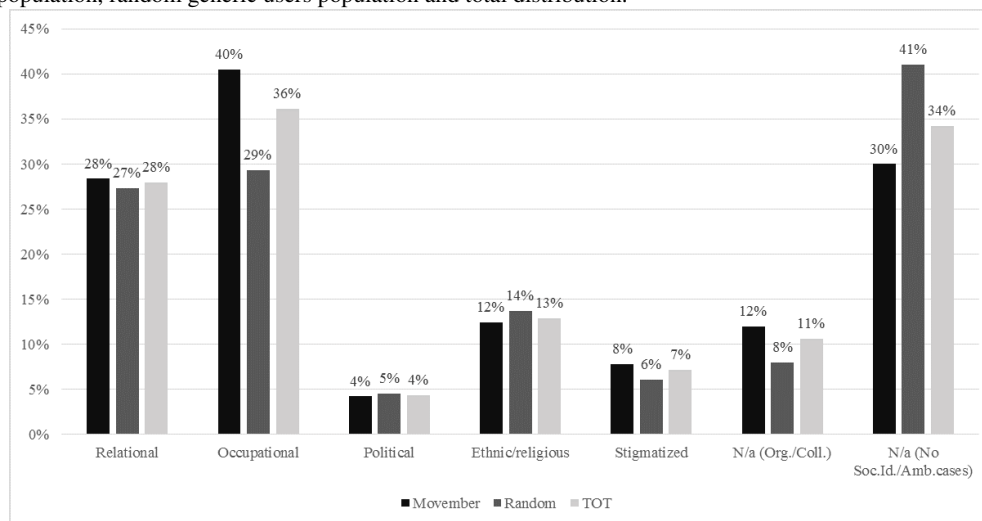
4.1 Qualitative Content Analysis with Human Annotation

We use qualitative content analysis to manually annotate our 2,633 Twitter users’ profile descriptions. Two coders are involved in the annotation. The coders meet in training and testing sessions to agree upon rules and build a codebook² that guides the annotation. The identity categories of our codebook are based on the 5-category social identity classification described in Section 2. In the annotation, a Twitter profile description is labeled with “Yes” or “No” for each category label, depending on whether the profile belongs to such category or not. Multiple identities may be assigned to a single Twitter user (i.e., identity categories are not mutually exclusive). We calculate the inter-rater reliability

¹This data was obtained via a Twitter datagrant, see <https://blog.twitter.com/2014/twitter-datagrants-selections>

²The codebook, code and datasets used in the experiments are available at <https://github.com/annapriante/identityclassifier>

Figure 1: Distributions (in %) of social identity categories over the total amount of annotated profiles (N=2,633): Movember participant population, random generic users population and total distribution.



bility using Krippendorff’s alpha, or Kalpha³ (Krippendorff, 2004) based on 300 double annotations. Kalpha values are very good for all categories (Relational=0.902; Occupational=0.891; Political=0.919; Ethnic/Religious=0.891; Stigmatized=0.853).

The definition of social identity is applicable only to one individual. Accounts that belong to more than one person, or to collectives, groups, or organizations (N=280), are annotated as “Not applicable”, or “N/a” (Kalpha=0.8268). Such category also includes individual profiles (N=900) for which: 1) no social identity category fits (e.g., profiles contain quote/citations/self-promotion; or individual attributes descriptions with no reference to social roles or group membership); and 2) ambiguous or incomprehensible cases⁴.

Looking at the distributions of social identity categories in the annotated profile descriptions provides an overview of the types of Twitter users in our data. We check if such distributions differ in the two populations (i.e., Movember participants and random generic users). We find that each identity category

is similarly distributed in the two groups (Figure 1). We conclude that the two populations are thus similar in their members’ social identities.

Figure 1 shows the distributions of social identity categories over the total amount of annotated profiles (N=2,633). N/a profile descriptions are the 45% (N=1180) of the total number of profiles: organizations/collective profiles are 11% (N=280), whereas no social identity profiles/ambiguous cases are 34% (N=900). It means that only a little more than a half, i.e., the remaining 55% profiles (N=1,453), of the Twitter users in our dataset have one or more social identities. Users mainly define themselves on the basis of their occupation or interests (Occupational identities=36%), and social roles played in society or relationships with others (Relational identities=28%). By contrast, individuals do not often describe themselves in terms of political or social movement affiliation, ethnicity, nationality, religion, or stigmatized group membership. Political, Ethnic/Religious and Stigmatized identities categories are less frequent (respectively, 4%, 13% and 7%).

4.2 Automatic Text Classification

We use machine learning to automatically assign predefined identity categories to 160-character Twitter profile descriptions (N=2,633), that are manually annotated as explained in Section 4.1. For each identity category we want to classify whether the profile description belongs to a category or not. We thus

³We use Krippendorff’s alpha as it is considered the most reliable inter-coder reliability statistics in content analysis.

⁴We keep N/a profiles in our dataset to let the classifiers learn that those profiles are not examples of social identities. Such choice considerably increases the number of negative examples over the positive ones that are used to detect the identity categories. However, we find that including or excluding N/a profiles does not make any significant difference in the classifier performance.

treat the social identity classification as a binary text classification problem, where each class label can take only two values (i.e. yes or no).

We use automatic text classification and develop binary classifiers in two experiments. Experiment 1 is based on the 5-category social identity classification explained in Section 2. In Experiment 1, we compare the classifiers performance in two scenarios. First, we use a combined dataset made by both Movember participants and random generic users. Profiles are randomly assigned to a training set (Combined(1): N=2338) and a test set (Combined(2): N=295). Second, we use separated datasets, i.e., random generic users as training set (Random: N=1022) and Movember participants as test set (Movember: N=1611), and vice versa.

Experiment 2 is a follow-up of Experiment 1 and we use only combined data⁵. We test a merger of three social identity categories (i.e., Political, Ethnic/religious and Stigmatized) for which we do not obtain acceptable results in Experiment 1.

4.2.1 Features Extraction

We use TF-IDF weighting (Salton and Buckley, 1988) to extract useful features from the user’s profile description. We measure how important a word, or term, is in the text. Terms with a high TF-IDF score occur more frequently in the text and provide the most of information. In addition, we adopt standard text processing techniques, such as Lowercasing and Stop words, to clean up the feature set (Sebastiani, 2002). We use the Chi Square feature selection on the profile description term matrix resulted from the TF-IDF weighting to select the terms that are mostly correlated with the specific identity category (Sebastiani, 2002).

4.2.2 Classification Algorithms

In the automatic text classification experiments, we evaluate four classification algorithms. First, we use Support Vector machine (SVM) with a linear kernel, which requires less parameters to optimize and is faster compared to other kernel functions, such as Polynomial kernel (Joachims, 1998). Balanced mode is used to automatically adjust

⁵We conduct Experiment 2 only on the combined set because in Experiment 1 we find that classifiers trained on the combined data performs better than trained on separated sets.

weights for class labels. Second, Bernoulli Naïve Bayes (BNB) is applied with the Laplace smoothing value set to 1. Third, Logistic Regression (LR) is trained with balanced subsample technique to provide weights for class labels. Fourth, the Random Forest (RF) classifier is trained with 100 trees to speed up the computation compared to a higher number of trees, for which no significant difference has been found in the classifier performance. Balanced subsample technique is used to provide weights for class labels.

4.2.3 Evaluation Measures

Experimental evaluation of the classifiers is conducted to determine their performance, i.e., the degree of correct classification. We compare the four classification algorithms on the training sets using Stratified 10-Fold Cross Validation. This technique seeks to ensure that each fold is a good representative of the whole dataset and it is considered better than regular cross validation in terms of bias-variance trade-offs (Kohavi and others, 1995). In feature selection, we check for different subsets of features (i.e., 500, 1000, 2000, 3000, 4000 and 5000) with the highest Chi Square from the original feature set, which consists of highly informative features. We find that 1000 features are the most informative.

Furthermore, we calculate precision (P), recall (R) and F-score to assess the accuracy and completeness of the classifiers. The classification algorithm that provide the best performance according to F-score in the Stratified 10-Fold Cross Validation is then tested on the test sets to get better insight into the classification results.

5 Classification Experiment 1

In this section, we present the results of Experiment 1 on automatically identifying 5 online social identities based on the annotated Twitter profile descriptions. In Section 5.1, we show the results of the Stratified 10 Fold Cross Validation in three training sets, i.e., Combined(1), Movember and Random. In Section 5.2, we illustrate and discuss the results of the best classification algorithm on the test sets.

Table 1: Relational and Occupational identities. Stratified 10 Fold Cross Validation in three training sets: precision (P), recall (R) and F-score.

<i>Classifier</i>	<i>Training Set</i>	RELATIONAL			OCCUPATIONAL		
		P	R	F	P	R	F
SVM	Combined(1)	0.764	0.705	0.723	0.827	0.793	0.804
	Movember	0.792	0.709	0.729	0.822	0.788	0.797
	Random	0.742	0.624	0.634	0.845	0.715	0.742
BNB	Combined(1)	0.855	0.635	0.652	0.848	0.769	0.788
	Movember	0.848	0.616	0.619	0.846	0.780	0.791
	Random	0.793*	0.524	0.471*	0.859	0.605	0.599
LR	Combined(1)	0.760	0.708	0.724	0.823	0.788	0.800
	Movember	0.786	0.718	0.735	0.817	0.789	0.796
	Random	0.717	0.627	0.637	0.848	0.721	0.748
RF	Combined(1)	0.803	0.660	0.682	0.842	0.780	0.797
	Movember	0.836	0.671	0.692	0.817	0.774	0.783
	Random	0.789	0.583	0.577	0.857	0.706	0.733

Table 2: Political, Ethnic/religious and Stigmatized identities. Stratified 10 Fold Cross Validation in three training sets: precision (P), recall (R) and F-score.

<i>Classifier</i>	<i>Training Set</i>	POLITICAL			ETHNIC/RELIGIOUS			STIGMATIZED		
		P	R	F	P	R	F	P	R	F
SVM	Combined(1)	0.646*	0.548	0.563*	0.750	0.594	0.619	0.713*	0.551	0.573*
	Movember	0.680*	0.529	0.541*	0.740	0.585	0.609	0.825	0.592	0.629
	Random	0.528*	0.510	0.505*	0.784	0.581	0.602	0.520*	0.507	0.498*
BNB	Combined(1)	0.482	0.500	0.491	0.572*	0.506	0.483*	0.478	0.500	0.488
	Movember	0.479*	0.500	0.489*	0.664	0.512	0.491	0.561*	0.507	0.494*
	Random	0.478	0.500	0.488	0.432	0.500	0.463	0.470	0.500	0.484
LR	Combined(1)	0.662	0.540	0.554	0.724	0.600	0.626	0.781	0.564	0.593
	Movember	0.655	0.536	0.550	0.720	0.603	0.628	0.742	0.589	0.621
	Random	0.528	0.509	0.505	0.751	0.592	0.613	0.52	0.506	0.498
RF	Combined(1)	0.633*	0.524	0.532*	0.856*	0.526	0.523*	0.654	0.519	0.524*
	Movember	0.479*	0.500	0.489*	0.848*	0.551	0.560*	0.884*	0.585	0.623*
	Random	0.478*	0.500	0.488*	0.672	0.524	0.508	0.470*	0.500	0.484*

5.1 Stratified 10 Fold Cross Validation Results on Five Social Identity Categories

Relational identity. All classifiers provide very precise results ($P > 0.700$) for the Relational identity category in the all three training sets (Table 1). The most precise classification algorithm is BNB in the combined set ($P = 0.855$). By contrast, recall is quite low ($0.500 < R < 0.700$) in all classifiers in each training set, thus affecting the final F-scores. The classification algorithm with the highest recall is LR in the Movember set ($R = 0.708$). According to F-scores, all classifiers provide from acceptable ($0.400 < F < 0.690$) to good/excellent ($F > 0.700$) results. Classifiers trained on the Movember set provide the highest F-scores, except for BNB where F-score is higher in the combined set. By contrast, the

Random set provides the lowest performances in all cases. Overall, LR is the most precise and complete classifier in all three training sets (combined: $F = 0.724$; Movember: $F = 0.735$; Random: $F = 0.637$).

Occupational identity. All classifiers provide very high precision ($P > 0.800$) and recall ($R > 0.750$) for the Occupational identity category (Table 1). The most precise classification algorithm is BNB in the Random set ($P = 0.859$), whereas the classification algorithm with the highest recall is SVM in the combined set ($R = 0.793$). According to F-scores, all classifiers provide good and excellent performances ($F > 0.700$), except for BNB in the Random set ($F = 0.599$). Classifiers trained on the combined set provide the highest F-scores, except for BNB where F-score is higher in the Movember set. By contrast, the Random set provides the lowest per-

formances. Overall, SVM and LR provide the best F-scores in all three training set.

Political, Ethnic/religious and Stigmatized identities. Classifiers perform less well in automatically classifying Political, Ethnic/religious and Stigmatized identities than in Relational and Occupational ones (Table 2). Both precision and recall are almost acceptable ($0.400 < P, R < 0.690$) in all three training sets. When training SVM, BNB and RF, we get ill-defined precision and F-score, which are consequently set to 0.0 in labels with no predicted samples (in Table 2, these values are marked with a *). As we noticed earlier in Figure 1, the low number of positive examples of Political, Ethnic/religious and Stigmatized identities in the data may cause this outcome. Classifiers trained on combined and Movember sets provide similar results, whereas the Random set provides the lowest performance. Overall, LR classifier provide the best F-scores for each category in all training sets.

5.2 LR Classifier Testing

Stratified 10 Fold Cross Validation show that the optimal classification algorithm for each identity category is LR. The LR classifier is evaluated on the test sets in order to get better insight into the classification results. Since we use three training sets, we evaluate the classifier on three different test sets as explained in Section 4.2.

According to the F-scores (Table 3), we are able to automatically classify Relational and Occupational identities in all three test sets. LR trained and tested on combined data provides the best results (Relational: $F=0.699$; Occupational: $F=0.766$). Although in the Stratified 10 Fold Cross Validation the classifier trained on the Random set has lower performance than trained on the Movember set, in the final testing the classifier performs better when we use Random as training set and Movember as test set (Relational: $F=0.594$; Occupational: $F=0.737$).

Final training and testing using LR on Political, Ethnic/religious and Stigmatized identities (Table 4) is affected by the low number of positive examples in the test sets, as these identities are less frequent in our annotated sample. Classifying Political identities is the most difficult task for the classifier in all three test sets and the performance is very low (Combined(2): $F=0.300$; Random: $F=0.266$;

Movember: $F=0.098$). Regarding Ethnic/religious and Stigmatized identities, LR provides almost acceptable F-scores only on the combined data (Ethnic religious: $F=0.543$; Stigmatized: $F=0.425$).

5.3 Discussion: Merging Identity Categories

In Experiment 1 we show that a classifier trained on the combined data performs better than a classifier trained on only Movember profiles or Random profiles. Our results are of sufficient quality for Relational and Occupational identities on the combined set, and thus we are able to automatically classify such social identities on Twitter using LR. Experiment 1 also shows that automatically classifying Political, Ethnic/religious and Stigmatized identities may be a challenging task. Although LR provides acceptable F-scores in the Stratified 10 Fold Cross Validation, the classifier is not able to automatically classify those three identities. This may be due to unbalanced distributions of identity categories in our data, that thus affect the text classification experiment.

Despite of the unsatisfactory classifier performances in detecting Political, Ethnic/religious and Stigmatized identities, we conduct a second experiment to find an alternative way to classify such identities because of their importance in the study of collective action. Therefore, we find that using NLP methods invites us to go back to theory and revisit our framework.

People with strong Political, Ethnic/religious and/or Stigmatized identities are often more engaged in online and offline collective action (Ren et al., 2007; Spears et al., 2002). These identities have a collective, rather than individualistic, nature as they address individual membership to one or multiple social groups. By sharing a common identity with other group members, individuals may feel more committed to the group's topic or goal. Consequently, they may engage in collective action on behalf of the group, even in cases of power struggle, i.e., individuals have a *politicized identity*, see (Klandermans et al., 2002; Simon and Klandermans, 2001). Political, Ethnic/religious and/or Stigmatized identities are indeed action-oriented (Ren et al., 2007), rather than social statuses as for Relational and Occupational identities (Deaux et al., 1995). Thus, the collective, action-oriented nature

Table 3: LR Classifier Testing on Relational and Occupational identities: precision (P), recall (R) and F-score.

<i>Training set</i>	<i>Test set</i>	RELATIONAL			OCCUPATIONAL		
		P	R	F	P	R	F
Combined(1)	Combined(2)	0.757	0.648	0.699	0.743	0.791	0.766
Movember	Random	0.649	0.491	0.559	0.722	0.693	0.707
Random	Movember	0.638	0.555	0.594	0.814	0.673	0.737

Table 4: LR Classifier Testing on Political, Ethnic/religious and Stigmatized identities: precision (P), recall (R) and F-score.

<i>Training set</i>	<i>Test set</i>	POLITICAL			ETHNIC/RELIGIOUS			STIGMATIZED		
		P	R	F	P	R	F	P	R	F
Combined(1)	Combined(2)	0.600	0.200	0.300	0.661	0.460	0.543	0.958	0.273	0.425
Movember	Random	0.571	0.173	0.266	0.531	0.300	0.383	0.360	0.145	0.206
Random	Movember	0.307	0.058	0.098	0.364	0.250	0.296	0.444	0.126	0.197

of certain Political, Ethnic/religious and Stigmatized identities show how such identities may often overlap and consequently influence human behaviors and actions.

Following these theoretical arguments, we decide to merge Political, Ethnic/religious and Stigmatized identities in one category, called PES identity (N=556). In this way, we also provide more positive examples to the classifiers. In Experiment 2, we train and test again the four classification algorithms on the PES identity using the combined data. In the next section, we present the results of this second experiment and show that by combining these identities we can improve the predictive performance of the classifiers.

6 Classification Experiment 2

Table 5 shows value of precision, recall and F-score in the Stratified 10 Fold Cross Validation on the training set (i.e., Combined (1): N=2338) to select the optimal classifier. Overall, all classifiers provide quite acceptable performances for the PES identity category ($0.500 < F < 0.650$). Only when validating the BNB classifier, we obtain an ill-defined F-score (in Table 5, this value is marked with a *). RF is the most precise classification algorithm (P=0.758), whereas LR has the highest recall (R=0.608). As in Experiment 1, LR is the optimal classifier with the highest F-score (F=0.623).

LR classifier is evaluated on the test set (i.e., Combined (2): N=295) to get better insight into the classification results. The classifier is highly precise in identifying PES identities (P=0.857). By contrast, recall is quite low (R=0.466), thus affecting

Table 5: PES identity. Stratified 10 Fold Cross Validation on combined data: precision (P), recall (R) and F-score.

<i>Classifier</i>	P	R	F
SVM	0.664	0.583	0.595
BNB	0.750	0.524	0.504*
LR	0.678	0.608	0.623
RF	0.758	0.543	0.540

final F-score (F=0.604). In conclusion, only if we merge political, religious and stigmatized identities, the classifier performance is acceptable.

7 Final Discussion and Conclusions

In this study, we explore the task of automatically classifying Twitter social identities of Movember participants and random generic users in two text classification experiments. We are able to automatically classify two identity categories (Relational and Occupational) and a 3-identity category merger (Political, Ethnic/religious and Stigmatized). Furthermore, we find that a classifier trained on the combined data performs better than a classifier trained on one group (e.g. Random) and test on the other one (e.g. Movember).

We make two main contributions from which both social theory on identity and NLP methods can benefit. First, by combining the two we find that social theory can be used to guide NLP methods to quickly classify and infer vast amounts of data in social media. Furthermore, using NLP methods can provide input to revisit traditional social theory that is often strongly consolidated in offline settings.

Second, we show that computers can perform a reliable automatic classification for most types of

social identities on Twitter. In NLP research there is already much earlier work on inferring demographic traits, therefore it may not be surprising that at least some of these identities can be easily inferred on Twitter. Our contribution is in the second experiment, where we show that merged identities are useful features to improve the predictive performance of the classifiers. In such way, we provide social scientists with three social identity classifiers (i.e., Relational, Occupational and PES identities) grounded in social theory that can scale-up online identity research to massive datasets. Social identity classifiers may assist researchers interested in the relation between language and identity, and identity and collective action. In practice, they can be exploited by organizations to target specific audiences and improve their campaign strategies.

Our study presents some limitations that future research may address and improve. First, we retrieve the user' profile description from randomly sampled tweets. In this way, people who tweet a lot have a bigger chance of ending up in our data. Future research could explore alternative ways of profile description retrieval that avoid biases of this kind.

Second, our social identity classifiers are based only on 160-characters profile descriptions, which alone may not be sufficient features for the text classification. We plan to test the classifiers also on tweets, other profile information and network features. Furthermore, the 160-character limitation constrains Twitter users to carefully select which identities express in such a short space. In our study, we do not investigate identity salience, that is, the degree or probability that an identity is more prominent than others in the text. Future research that combine sociolinguistics and NLP methods could investigate how semantics are associated to identity salience, and how individuals select and order their multiple identities on Twitter texts.

Third, in the experiments we use standard text classification techniques that are not particularly novel in NLP research. However, they are simple, effective ways to provide input for social theory. We plan to improve the classifiers performance by including other features, such as n-grams and cluster of words. Furthermore, we will explore larger datasets and include more training data for further experimentation with more complex tech-

niques (e.g., neural networks, World2Vec).

Acknowledgments

Thanks to Balsam Awlad Wadair for the valuable help in the manual annotation; Dong Nguyen, Robin Aly, Fons Wijnhoven, Fons Mentink, Koen Kuijpers and Mike Visser for helpful discussions and feedback; and Twitter for providing part of the tweets used in this study through the Twitter DataGrant. Thanks also to the anonymous reviewers for their helpful comments.

References

- Faiyaz Al Zamal, Wendy Liu, and Derek Ruths. 2012. Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. *ICWSM*, 270.
- Augusta Isabella Alberici and Patrizia Milesi. 2013. The influence of the internet on the psychosocial predictors of collective action. *Journal of Community & Applied Social Psychology*, 23(5):373–388.
- Blake E. Ashforth, Spencer H. Harrison, and Kevin G. Corley. 2008. Identification in organizations: An examination of four fundamental questions. *Journal of Management*, 34(3):325–374.
- Blake E. Ashforth, Beth S. Schinoff, and Kristie M. Rogers. 2016. i identify with her,i identify with him: Unpacking the dynamics of personal identification in organizations. *Academy of Management Review*, 41(1):28–60.
- Mary Bucholtz and Kira Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse studies*, 7(4-5):585–614.
- John D. Burger, John Henderson, George Kim, and Guido Zarrella. 2011. Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1301–1309. Association for Computational Linguistics.
- Michael Chan. 2014. Social identity gratifications of social network sites and their impact on collective action participation. *Asian Journal of Social Psychology*, 17(3):229–235.
- Morgane Ciot, Morgan Sonderegger, and Derek Ruths. 2013. Gender inference of twitter users in non-english contexts. In *EMNLP*, pages 1136–1145.
- Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on twitter: It's not easy! In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*.

- Kay Deaux, Anne Reid, Kim Mizrahi, and Kathleen A Ethier. 1995. Parameters of social identity. *Journal of personality and social psychology*, 68(2):280.
- Erving Goffman. 1959. The presentation of self in everyday life. *Garden City, NY*.
- Michael J. Jensen and Henrik P. Bang. 2013. Occupy wall street: A new political form of movement and community? *Journal of Information Technology & Politics*, 10(4):444–461.
- Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer.
- Bert Klandermans, Jose Manuel Sabucedo, Mauro Rodriguez, and Marga De Weerd. 2002. Identity processes in collective action participation: Farmers’ identity and farmers’ protest in the netherlands and spain. *Political Psychology*, 23(2):235–251.
- Ron Kohavi et al. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145.
- Klaus Krippendorff. 2004. Reliability in content analysis. *Human communication research*, 30(3):411–433.
- Jiwei Li, Alan Ritter, and Eduard H Hovy. 2014. Weakly supervised user profile extraction from twitter. In *ACL (1)*, pages 165–174.
- Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. 2013. ”how old do you think i am?” a study of language and age in twitter. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, ICWSM 2013, Cambridge, MA, USA*, pages 439–448, Palo Alto, CA, USA, July. AAAI Press.
- Dong Nguyen, Dolf Trieschnigg, A. Seza Dogruoz, Rilana Gravel, Mariet Theune, Theo Meder, and Franciska de Jong. 2014. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, page 19501961. Association for Computational Linguistics, Dublin, Ireland, August 23-29 2014.
- Dong Nguyen, Tijs van den Broek, Claudia Hauff, Djoerd Hiemstra, and Michel Ehrenhard. 2015. #supportthecause: Identifying motivations to participate in online health campaigns. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal*, pages 2570–2576, New York, USA, September. Association for Computational Linguistics.
- Mancur Olson. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups, Second printing with new preface and appendix (Harvard Economic Studies)*. Harvard economic studies, v. 124. Harvard University Press, revised edition.
- Namkee Park and Aimei Yang. 2012. Online environmental community members intention to participate in environmental activities: An application of the theory of planned behavior in the chinese context. *Computers in human behavior*, 28(4):1298–1306.
- Marco Pennacchiotti and Ana-Maria Popescu. 2011. A machine learning approach to twitter user classification. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 11(1):281–288.
- Tom Postmes and Suzanne Brunsting. 2002. Collective action in the age of the internet mass communication and online mobilization. *Social Science Computer Review*, 20(3):290–301.
- Daniel Preotiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age and gender in tweeting about mental illnesses. *NAACL HLT 2015*, page 21.
- Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pages 37–44. ACM.
- Yuqing Ren, Robert Kraut, and Sara Kiesler. 2007. Applying common identity and bond theory to design of online communities. *Organization studies*, 28(3):377–408.
- Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523.
- Fabrizio Sebastiani. 2002. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47.
- Bernd Simon and Bert Klandermans. 2001. Politicized collective identity: A social psychological analysis. *American psychologist*, 56(4):319.
- Russell Spears, Martin Lea, Rolf Arne Corneliusen, Tom Postmes, and Wouter Ter Haar. 2002. Computer-mediated communication as a channel for social resistance the strategic side of side. *Small Group Research*, 33(5):555–574.
- Jan E. Stets and Peter J. Burke. 2000. Identity theory and social identity theory. *Social psychology quarterly*, pages 224–237.
- Sheldon Stryker, Timothy Joseph Owens, and Robert W White. 2000. *Self, identity, and social movements*, volume 13. University of Minnesota Press.
- Sheldon Stryker. 1980. *Symbolic interactionism: A social structural version*. Benjamin-Cummings Publishing Company.
- Henri Tajfel. 1981. *Human groups and social categories: Studies in social psychology*. CUP Archive.

- Nadine Tamburrini, Marco Cinnirella, Vincent AA Jansen, and John Bryden. 2015. Twitter users change word usage according to conversation-partner social identity. *Social Networks*, 40:84–89.
- Emma F. Thomas, Craig McGarty, Girish Lala, Avelie Stuart, Lauren J. Hall, and Alice Goddard. 2015. Whatever happened to kony2012? understanding a global internet phenomenon as an emergent social identity. *European Journal of Social Psychology*, 45(3):356–367.
- John C. Turner, Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell. 1987. *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.
- Han Van der Veen, Djoerd Hiemstra, Tijs van den Broek, Michel Ehrenhard, and Ariana Need. 2015. Determine the user country of a tweet. *arXiv preprint arXiv:1508.02483*.
- Benjamin Van Durme. 2012. Streaming analysis of discourse participants. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 48–58. Association for Computational Linguistics.
- Martijn Van Zomeren, Tom Postmes, and Russell Spears. 2008. Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. *Psychological bulletin*, 134(4):504.
- Svitlana Volkova and Yoram Bachrach. 2015. On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. *Cyberpsychology, Behavior, and Social Networking*, 18(12):726–736.
- Svitlana Volkova, Glen Coppersmith, and Benjamin Van Durme. 2014. Inferring user political preferences from streaming communications. In *ACL (1)*, pages 186–196.
- Svitlana Volkova, Yoram Bachrach, Michael Armstrong, and Vijay Sharma. 2015. Inferring latent user properties from texts published in social media. In *AAAI*, pages 4296–4297. Citeseer.