Emoticons and Phrases: Status Symbols in Social Media

Simo Tchokni
The Computer Laboratory
University of Cambridge
UK

Diarmuid Ó Séaghdha
The Computer Laboratory
University of Cambridge
UK

Daniele Quercia
Yahoo Labs
Barcelona
Spain

Abstract

There is a sociolinguistic interest in studying the social power dynamics that arise on online social networks and how these are reflected in their users' use of language. Online social power prediction can also be used to build tools for marketing and political campaigns that help them build an audience. Existing work has focused on finding correlations between status and linguistic features in email, Wikipedia discussions, and court hearings. While a few studies have tried predicting status on the basis of language on Twitter, they have proved less fruitful. We derive a rich set of features from literature in a variety of disciplines and build classifiers that assign Twitter users to different levels of status based on their language use. Using various metrics such as number of followers and Klout score, we achieve a classification accuracy of individual users as high as 82.4%. In a second step, we reached up to 71.6% accuracy on the task of predicting the more powerful user in a dyadic conversation. We find that the manner in which powerful users write differs from low status users in a number of different ways: not only in the extent to which they deviate from their usual writing habits when conversing with others but also in pronoun use, language complexity, sentiment expression, and emoticon use. By extending our analysis to Facebook, we also assess the generalisability of our results and discuss differences and similarities between these two sites.

1 Introduction

A large part of our social relationships are taking place online and an increasing number of researchers have turned to studying these interactions. Social networks like Twitter enable us to engage with people that can be socially far removed from us. Rich social interactions take place on Twitter, where users frequently exchange information with news outlets, celebrities and other socially prominent accounts (Kwak et al. 2010).

Although past research has focused primarily on the graph-theoretic aspects of social influence, a growing number of studies have identified ways in which social status is mediated through use of language. Linguistic style accommodation, for example, has been used to predict which of a pair of individuals is more powerful in the domains of Wikipedia and U.S. Supreme Court proceedings (Danescu-Niculescu-Mizil et al. 2012). Their attempt to extend this to dyadic Twitter conversations, however, "rendered relatively poor results" (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011). Twitter is a breeding ground for idiosyncratic uses of language since the 140 character limit on messages forces users to find new ways of expressing themselves. These aspects make it highly interesting for the study of social status: there is potential for the discovery of new or modified ways in which language mediates interactions between users. Aside from the sociolinguistic interest of such a study, there are also practical uses for the identification of powerful or influential actors online, for example, for social media marketing or political campaigning. In both of these areas, since we are increasingly using the Internet as a source of news and opinions, it would be helpful to learn how to be more influential online.

In this paper, we investigate how social power is related to language use and communication behaviour on Twitter by focusing on two different aspects of status, *popularity* and *social influence*. Furthermore, we look at status in two different ways. Firstly, the User Predictor (Section 3) predicts social power on an individual basis on Twitter and helps us investigate how a person's use of language online is connected to their social status. Secondly, we explore how social power *differentials* between Twitter users are reflected in the way the converse. The Conversation Predictor (Section 4) predicts which is the higher status user in a dyadic conversation. In building these two predictors, we make the following novel contributions:

Emphasis on prediction. Previous work has largely computed within-sample correlations between social power metrics and linguistic indicators. Because we perform out-of-sample evaluation, our results are more generalisable. Furthermore, we compare Twitter to Facebook by performing prediction experiments on a Facebook dataset.

A lexicon of phrases associated with social power on Twitter. Gilbert produced such a list for the domain of corporate email (Gilbert 2012). However, as mentioned above, Twitter is a very different medium and its social power relationships are not as clear-cut as a company hierarchy. We use the SVM weights of bag-of-*n*-gram features to pro-

duce a ranked list of phrases, which we present in Section 3.5.

New findings on how emoticons are related to social status. We look not only at lexical features but also at emoticons, which allows us to describe the relationship between emoticon use and social power. We discuss these findings in Section 3.5.

Successful prediction of social power differentials in Twitter conversations. Existing work has focused on finding correlations between status and linguistic features in email, Wikipedia discussions, and court hearings. However, discussions in these domains are goal-oriented (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011). Our Conversation Predictor is the first to look at a broader set of features and achieve good prediction results on this task.

2 Related Work

Sociolinguistic studies provide the basis for the underlying assumption of this study, namely that individuals with higher social power differ from low status individuals in their use of language. Such research suggests that people with low social status are more likely to use first person pronouns, whereas powerful individuals tend to use fewer first person pronouns but more second person pronouns (Chung and Pennebaker 2007) (Dino, Reysen, and Branscombe 2009), thereby suggesting that low status is characterised by increased egocentricity.

Furthermore, emails from employees ranking lower in a company hierarchy are perceived as more polite due to the use of linguistic devices like hedging, subjunctives, minimization and apologising (Morand 2000). There is also evidence that social power is linked to language complexity: researchers found that high status users in online forums and message boards are more likely to use large words (6 letters or more) than low status users (Dino, Reysen, and Branscombe 2009) (Reysen et al. 2010).

Moreover, studies link an individual's propensity to accommodate to their social status. The theory of accommodation states that people engaged in dyadic conversations tend to *unconsciously* mimic each other's communicative behaviour (Giles, Coupland, and Coupland 1991). The effects of accommodation have measured in discussions between Wikipedia editors and arguments before the U.S. Supreme Court (Danescu-Niculescu-Mizil et al. 2012). Linguistic style accommodation can also be observed on Twitter (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011).

With the exception of (Danescu-Niculescu-Mizil et al. 2012) and (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011), most of the studies above were studied in small-scale contexts. The emergence of social networks, however, has enabled more large-scale studies of the effect of social power on language. Twitter has been a popular choice, with work on unsupervised modelling of dialogue acts (Ritter, Cherry, and Dolan 2010), modelling participation behaviour in Twitter group chats (Budak and Agrawal 2013), examining how Twitter users influence others on a topic level (Liu et al. 2010) and on how different types of users vary in their use of language (Quercia et al. 2011). Quercia *et al.* analyse a set of 250K Twitter user's tweets and present correlations between dimensions of linguistic style (e.g., pronoun use and sentiment) and different proxies for popularity, such as number of followers and Klout. Most recently, Hutto *et al.* found high within-sample correlations between follower growth and positive sentiment as well as the use of large words (Hutto, Yardi, and Gilbert 2013). Negative sentiment and using self-referencing pronouns caused follower numbers to decrease. The task we put forward here is different from theirs in that they compute within-corpus correlations, while we are attempting to build classifiers that can make more generalisable out-of-sample predictions.

In the context of other online media, recent research examined the relationship between politeness and social power on Stack Exchange[1] and Wikipedia and found that admins tend to be less polite than non-admins (Danescu-Niculescu-Mizil et al. 2013). The Enron email corpus, a corpus of emails sent and received by Enron employees collected as part of the CALO Project (Klimt and Yang 2004), has been used to build a classifier that identifies how two individuals are positioned relative to each other in the company hierarchy using a combination of *n*-gram and POS-tag features extracted from the emails they exchanged (Bramsen et al. 2011). Gilbert compiled and published a set of phrases that signal social hierarchy within the Enron corpus (Gilbert 2012) and a recent study analyses the dynamics of workplace gossip (Mitra and Gilbert 2013). Since corporate emails are very different from typical interactions on social platforms where the hierarchy is less clear, our aim is to investigate how well this task can be solved on such networks.

3 User Predictor

The User Predictor addresses the task of predicting a single Twitter user's level of social power based on a sample of their tweets. There have been several studies that look at notions of influence on social networks and at how language is related to social status online. A central question is which metrics can be used to represent an individual's social power online. On Twitter, Cha *et al.* find that while number of followers represents a user's popularity, it does not say much about social influence (Cha et al. 2010). The latter is better measured by how many times a user's tweets are retweeted and by how often others mention them. In past research on Twitter, indegree, retweets and mentions, as well as Klout[2] have been used as proxies for social power (Quercia et al. 2011), (Cha et al. 2010), (Romero et al. 2011). Klout employs network-based features including following count, follower count, retweets and unique mentions to produce an online influence score between 1 and 100. In this work, we thus use number of followers/friends on Twitter as a measure of popularity and Klout to represent social influence. It is important to note that we cannot always expect these measures to reflect real life social power. Thus, while we expect to see similarities, the language of social power online may also show some differences to that of verbal communication.

[1] `http://stackexchange.com/about`

[2] `http://www.klout.com`

3.1 Hypotheses

Starting from prior work in sociolinguistics and psychology, we derive four hypotheses as to how high-status individuals differ in their use of language from those with low status. Following our discussion of previous findings on pronoun use (Chung and Pennebaker 2007) (Dino, Reysen, and Branscombe 2009), we can put forward the following hypotheses:

H_1: High status users use more second-person pronouns than low status users.
H_2: Low status users use more first-person pronouns than high status users.

Our third hypothesis is derived from research on language complexity (Dino, Reysen, and Branscombe 2009) (Reysen et al. 2010):

H_3: High status users use more large words than low status users.

Finally, Quercia *et al.*'s analysis used the "Linguistic Inquiry Word Count" (LIWC) (Pennebaker, Francis, and Booth 2001), which maps words and word stems to a set of categories representing emotional and cognitive processes as well as linguistic dimensions, to analyse the users' tweets. They find that popularity is correlated with the expression of positive emotion whereas influential users express more negative emotion and conclude that, in general, greater emotivity is associated with greater social power. From this, we derive a fourth conjecture:

H_4: High status users express more emotions than low status users.

3.2 Task and Datasets

Our Twitter dataset contains about 258,895 different English - speaking Twitter users and their tweets, adding up to about 31.5M tweets in total. This dataset was assembled by crawling the Twitter public API between September and December 2010, starting with a small seed set of popular London-based seed profiles of UK-based news outlets. We restricted ourselves to UK profiles to avoid conflating different culture-specific uses of language. We chose the Metro, a free newspaper with a readership of some 3.5 millions; The Independent, a center-left newspaper with a circulation of around 651,000 a day; and The Sun, a tabloid selling about 3 million copies daily. All of the profiles belonging to the seed profiles' followers were crawled and at most 200 of each user's tweets were downloaded.

We cast the problem of predicting social status to a *classification* task, where each user is assigned to one of two classes: **low** social power and **high** social power. This approach has often been taken in the domain of sentiment analysis of online reviews, where star ratings are mapped onto 'positive', 'negative' and sometimes also 'neutral' sentiment classes (Pang, Lee, and Vaithyanathan 2002). Our initial attempt at a regression task, whereby the system learns to predict an absolute number corresponding to a user's popularity or influence, produced poor results.

	FOLLOWERS	KLOUT
low cut-off	≤ 87	≤ 16.85
high cut-off	> 1113	> 46.25
Minimum	0	1
Maximum	6,520,279	100
Total **low** users	65,054	43,818
Total **high** users	64,711	43,692
Total users	129,765	87,510
Messages per user	111.6	143.9

Table 1: Characteristics of the FOLLOWERS and KLOUT datasets. The cut-off values are based on the top and bottom quartiles of each dataset.

For each power metric, **low** and **high** users were determined by assigning all users in the bottom quartile of the dataset to **low**, and all users in the top quartile to **high**. The resulting cut-off values for number of followers, Klout and number of friends are given in following section, where each dataset is discussed in more detail. This yields two smaller datasets: a FOLLOWERS dataset containing **low** and **high** users as determined by follower count, and a KLOUT dataset containing **low** and **high** users as determined by Klout score. See Table 1 for additional details about the FOLLOWERS and KLOUT datasets.

3.3 Features

Unigrams and Bigrams The value of each unigram or bigram feature is its L1-normalized frequency across all of a user's tweets. Tweets are tokenized around whitespace and common punctuation and hashtags, usernames, numbers and URLs were removed. All remaining words are lowercased. In a first experiment, symbols like ™, © and currency symbols were often picked up as highly informative by the classifier. Since these are difficult to interpret in a meaningful way, we also excluded all unigrams and bigrams containing non-ASCII characters. This yielded 2,837,175 unigrams and 42,296,563 bigrams on Twitter.

Dictionary-based Features These measure the degree to which an individual uses words that fall into certain dictionary categories. We use two dictionaries: the LIWC dictionary (Pennebaker, Francis, and Booth 2001) and the NRC Emotion dictionary (Mohammad and Turney 2012).

The version of the LIWC dictionary used for this project was adapted from the original LIWC dictionary, by combining certain categories and leaving out others. It restricts the matched word categories to the 8 style and 2 sentiment dimensions shown in Table 2. The LIWC has often been used for studies on variations in language use across different people.

The NRC Emotion Lexicon is a crowd-sourced word-emotion association lexicon (Mohammad and Turney 2012) and maps words onto 10 emotion dimensions, as presented in Table 3. Since previous findings indicate that emotional expression interacts with social status, we thought this lexicon could be helpful for our task.

Dimension	Example words
first person	*I, my, me ...*
second person	*you, your ...*
third person	*she, he, they ...*
cognitive	*believe, choice, apparently ...*
time	*anytime, new, long ...*
past	*arrived, asked, ended ...*
present	*begin, do, want ...*
future	*gonna, may, might ...*
posemo	*nice, outgoing, original ...*
negemo	*no, offend, protest ...*

Table 2: Dimensions of the 10-Category LIWC Dictionary.

Dimension	Example words
anger	*punch, reject, ruined ...*
anticipation	*punctual, savor, romantic ...*
disgust	*abuse, prejudiced, sickening ...*
fear	*abandon, rifle, scarce ...*
joy	*blessed, romantic, score ...*
negative	*misery, oversight, quit ...*
positive	*mate, nap, plentiful ...*
sadness	*blue, shatter, starvation ...*
surprise	*coincidence, catch, secrecy ...*
trust	*scientific, save, toughness ...*

Table 3: Dimensions of the NRC Emotion Lexicon.

A user's tweets are scored against the 20 lexical categories given above, yielding 20 features. Let $f_{D_c}(u)$ represent the value of the feature for category c of dictionary D, for a given user u. $f_{D_c}(u)$ is a value between 0 and 1 and is given by:

$$f_{D_c}(u) = \frac{w_{D_c}(u)}{N_D(u)}$$

where $w_{D_c}(u)$ is the total number of words matching category D_c across u's tweets and $N_D(u)$ is the total number of words matching any category in D across u's tweets. Additionally, two features represent the total fraction of categorized words for each dictionary. Let $N(u)$ represent the total number of words across all of u's tweets. Then they take on the values $\frac{N_{LIWC}(u)}{N(u)}$ and $\frac{N_{NRC(u)}}{N(u)}$.

Emoticon Use The following features relating to emoticon use were included: average number of emoticons per tweet and fraction of positive/negative emoticons used. We also use 5 binary features to bin the average number of emoticons per tweet into 5 intervals. An emoticon's sentiment is determined using an "Emoticon Sentiment Dictionary". We created it by manually labelling the emoticons found in our datasets as positive or negative, guided by Wasden's Internet Lingo Dictionary (Wasden 2010). The resulting dictionary contains 78 positive emoticons and 57 negative emoticons. Some examples of positive tweets are (-:, (': and :p, whereas)':,)= and :-@ express negative sentiment.

Tweet and Word Length Previous research has shown that high status users were found to use more large words than low status users. We also conjectured that a difference in average tweet length could exist between high and low status users. This is reflected in our choice of the following features: average word length, average tweet length, number of large words used as a fraction of total words and a binary feature indicating whether average word length is greater than 6 or not.

Spelling One feature was used to represent the fraction of misspelled words across all of a user's tweets. Since standard spell checker dictionaries may not have enough coverage to work well on tweets where abbreviations abound, words are checked against a list of common misspellings downloaded from Wikipedia.[3] The value of the spelling feature is the fraction of words that match a mispelling on this list.

Punctuation Use of punctuation is encoded by two features, namely the fraction of tweets containing at least one question mark and the fraction of tweets containing at least one exclamation mark.

Word Elongation Some users are prone to elongating words through character repetition, e.g., by writing *cooool* instead of *cool*. Brody & Diakopoulos find that this phenomenon is common on Twitter and that subjective terms in particular are lengthened in this way, presumably to intensify the expressed sentiment (Brody and Diakopoulos 2011). Word elongation may thus be indicative of emotivity, which we hypothesised could be linked to high popularity or influence. Elongating words can also be taken as an indication of a lack of formality. We thus record the fraction of words that a user elongates in this way. Since three or more identical, consecutive letters are very unusual in English, a word is considered elongated if the same character is repeated consecutively at least three times.

Mentioning Others and Retweeting We measure a user's level of engagement with others through the fraction of a user's tweets that are retweets (as indicated by the string *RT*) and the fraction of tweets that are addressed to other users (as indicated by the @ character).

3.4 User Prediction Task

We train Support Vector Machines (SVMs) with default settings on the features in Section 3.3, using the implementation provided by Liblinear (Fan et al. 2008). To assess the relative power of the different features, we trained separate classifiers on each of the feature sets from the previous section. Additionally, all features are combined to obtain a final classifier for each dataset. We evaluate using 10-fold cross-validation and compare performance to random guessing, giving a baseline accuracy of 50% (see Table 4).

For separate feature sets, unigram features reach the highest accuracies of 81.38% on FOLLOWERS and 80.43% on KLOUT. Bigrams do significantly worse than unigrams at

[3] http://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings

Features Used	FOLLOWERS	KLOUT
Baseline	50.00	50.00
unigrams	81.38***	80.43***
bigrams	80.59***	77.26***
NRC	64.30***	59.95***
LIWC	65.42***	65.11***
emoticons	66.46***	61.06***
tweet and word length	63.17***	58.98***
spelling	48.79	61.67
word elongation	49.02**	50.07**
punctuation	63.53**	54.11**
mentioning others	60.24***	57.95***
retweeting	70.02***	64.87***
All features	**82.37*****	**81.28*****

Table 4: 10-fold cross-validation accuracies on FOLLOWERS and KLOUT. (***), (**) and (*) indicate statistical significance with respect to the baseline at two-tailed p-values of $p < 0.0001$, $p < 0.01$, and $p < 0.05$, respectively. The highest achieved accuracies are shown in bold.

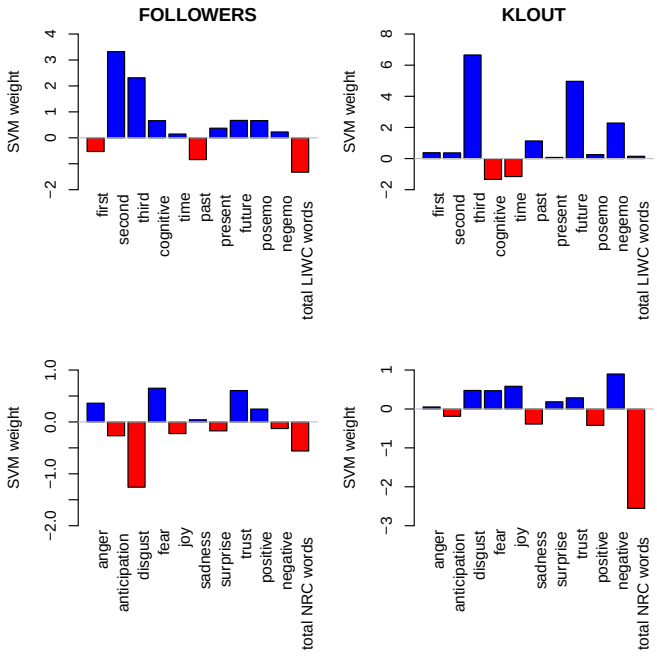

Figure 1: SVM weights of the LIWC and NRC features. High status is indicated by blue and low status by red bars.

$p < 0.0001$. Given the short length of Twitter messages, bigrams do not add much information and cause an explosion of the feature space. Note, however, that even purely stylistic features such as the NRC and LIWC dictionaries produce good accuracies that vary between 59.95% and 65.42%. Tweet and word length as well as punctuation features perform comparably and, perhaps most surprisingly, so do emoticon features despite their relatively narrow informational content. With accuracies of around 70% on FOLLOWERS and around 64% on KLOUT, retweet behaviour is also good indicator of social status.

Training a model on all features results in improvements of about 1%. These improvements are statistically significant on KLOUT ($p < 0.05$) but not on FOLLOWERS.[4]

3.5 User Prediction Results

In order to gain an insight into how low and high status users differ in their use of language and to evaluate the hypotheses given in 3.1, we examine the weight vectors produced by the SVM when trained on the full FOLLOWERS and KLOUT datasets.

Hypothesis Evaluation To test our initial hypotheses, we trained separate models on the LIWC, NRC, mentioning others and tweet and word length features.

Pronoun use: H_1 and H_2 are supported on FOLLOWERS (Figure 1), albeit the association between low status and first person pronouns is weak. On KLOUT, the associations between first and second person pronouns and high status are both weakly positive. Instead, third person pronouns are highly related to social influence. Nevertheless, we found that mentioning others, which it can be argued is similar to using the second person, is associated with both high numbers of followers and high Klout scores. A possible reason for the strong weight of third person pronouns could be that influentials point to what takes place in the world around them in order to motivate others to change or engage with it.

Large words: The weights confirm that high-power users employ more large words than low-power users. Large words have been linked to linguistic complexity.

Emotivity: Using many emotion words (*total NRC words* in Figure 1) is associated with low status on all datasets, which contradicts H_4. However, sentiment *polarity* also has an effect on a user's status. Positive emotion seems to be conducive to popularity while influentials write more negatively.

We investigated whether instead of emotivity, the *diversity* of emotions expressed could be related to high status. Indeed, training a classifier on the Shannon entropy of a user's distribution of NRC categories achieved good performance on FOLLOWERS and KLOUT, with accuracies of 65.36% and 62.38% respectively (both significant at $p < 0.0001$). On both datasets, the feature weight shows that powerful users tend to express a more varied range of emotions.

Differences in Word Choice The n-gram features allow us to assess the general assumption that differences in social power are expressed through language. We rank unigrams and bigrams according to how indicative of **high** or **low** social power they are using their SVM model weights. Tables 5 and 6 show the 30 highest ranking n-grams for each class for FOLLOWERS, KLOUT and FRIENDS, respectively.

The Twitter rankings include words like *in la* or *in nyc*. These can reliably indicate status because famous people tend be in these places but not because using these partic-

[4]Training a Linear Regression classifier on all features produced comparable results.

FOLLOWERS			
Unigrams		Bigrams	
low	**high**	**low**	**high**
surely	rts	avatar now	in la
shame	backstage	at work	the rt
;p	cc	well done	rt my
bloody	washington	an iphone	:) rt
whilst	questions	bring on	headed to
uni	nope	just seen	white house
cameron	hollywood	managed to	rt i
wondering	nyc	loving the	u s
yg	tells	for following	you're welcome
thinks	dm	bank holiday	you missed
gutted	bloggers	roll on	lindsay lohan
babeeee	headed	the follow	thanks so
rubbish	shows	oh dear	talks about
mum	sorry	come on	w the
preparing	toronto	you dauntons	rt just
twittering	powerful	the welcome	thank u
debra	y'all	back from	your favorite
boring	announced	the train	in nyc
luck	thx	this space	sorry i
pub	gracias	just watched	wall street

Table 5: Top 20 unigrams and bigrams for each class on the FOLLOWERS dataset.

ular words makes one more to gain a following. However, note that variations of the phrase *thank you* (e.g., *thanks so*, *thanks u*) and phrases referring to others (e.g., *you missed*) also appear in the **high** columns. This matches our findings in the previous section regarding pronoun use. Furthermore, on KLOUT, **high** *n*-grams include more instances of *I* than the corresponding columns for FOLLOWERS, a further indication that H_2 does not hold for influential users.

The *n*-grams further suggest that low status users are more likely to tweet about their daily lives (e.g., *bored*, *at work*) while high status individuals talk about events or issues that are of wider interest (e.g., *the pope*, *commonwealth games*, *tea party*).

A drawback of these rankings is the Twitter dataset is geographically skewed: most powerful users are from the United States whereas the low status users are British. We thus see that *rubbish* and *bloody* are associated with **low** whereas *white house* and *tea party* appear in the **high** columns. To generate more location-neutral *n*-grams, we trained separate SVM models on only UK and only US Twitter profiles. Performance remained comparable to using the full datasets and we found no strong differences between the British and American *n*-gram lists.

Emoticon Use The emoticon features achieved high performance, suggesting that there is a strong link between emoticon use and social power. Powerful users tend to use emoticons often and high Klout is strongly associated with positive emoticons (Figure 2), though we saw above that they often employ negative words. Low popularity is linked to negative emoticons. Indeed, a study on emoticon usage on Twitter found that these are usually used in positive contexts and rarely appear in angry or anxious tweets (Park et al. 2013). Perhaps breaking this social norm shows poor "internet literacy" and thus something powerful users would not do. Furthermore, influential users' may prefer negative words over negative emoticons because the former are more meaningful when expressing an opinion.

Additionally, emoticons appear among the top 20 *n*-grams on both FOLLOWERS and KLOUT. The emoticons ;p and

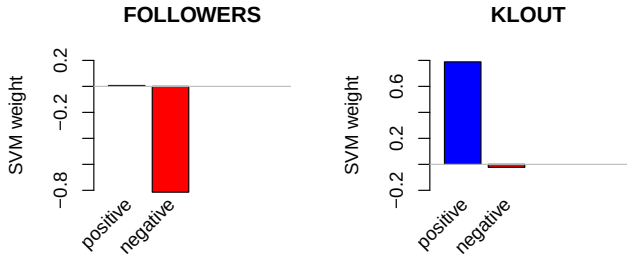

Figure 2: SVM weights of the Emoticon Sentiment features. High status is indicated by blue bars and low status by red bars.

:-(appear in the **low** column and are used to indicate joking and sadness respectively, whereas the :) emoticon indicates **high** social power. The latter is one of the most frequently used emoticons on Twitter (Park et al. 2013) and also the most basic. We take this to indicate that the socially powerful tend to be more conservative in their use of emoticons.

Perhaps counter-intuitively, emoticons seem better at predicting status than sentiment words. Sentiment polarity may be more clearly indicated by emoticons than by the occurrence of a sentiment word, since determining sentence sentiment goes beyond counting words.

Cross-Domain Analysis To assess the generalisability of our method, we repeated this classification task on a dataset of 121,823 different Facebook users and a subset of their English comments (with an average of 5.97 messages per user), using number of friends as a popularity measure.[5] This

[5]Obtained from the Facebook application myPersonality (`https://apps.facebook.com/mypersonality/`). Language was detected using the Java Language Detection Library (`https://code.google.com/p/language-detection/`).

KLOUT			
Unigrams		Bigrams	
low	high	low	high
easter	rts	new year	rt i
april	nope	the snow	rt my
snow	pope	swine flu	com o
wondering	oct	back from	the pope
swine	cc	for following	:) rt
june	yes	twitter directory	ed miliband
march	bro	to twitter	of course
bored	that's	the sun	in nyc
cameron	talks	the follow	commonwealth games
brown	sept	at work	with and
christmas	fall	just joined	rt just
twittering	miliband	loving the	i'm not
following	october	looking for	you missed
loving	she's	this site	i don't
looking	cuts	new website	thanks for
gordon	there's	check this	tea party
myself	questions	would like	yes i
trying	miners	the twitter	i know
preparing	dm	check it	thank u
website	nyc	checking out	you too

Table 6: Top 20 unigram and bigram rankings for each class on the KLOUT dataset.

also allows us to compare the linguistic symbols of status on these two networks.

We achieve close to 60% classification accuracy on Facebook, which is encouraging given that that the dataset is much smaller than for Twitter. Again, the emoticon features produced high performance, which bolsters our claim that there is a strong link between emoticons and social power. They appear among the top 20 *n*-grams for both sites but they are used differently: popular Facebook users use a more varied set of emoticons (e.g., :), :od and :')). These imply a certain familiarity which would not exist between a popular Twitter user and their followers. We also find that the first person is associated with popularity on Facebook, whereas the opposite is the case on Twitter. Since Facebook users tend to know one another personally, they perhaps do not need to build a sense of community and are more likely to reference themselves. Nevertheless, high status users of both networks use more other-oriented speech than less powerful individuals. Although some aspects of social power thus seem to be quite different on Facebook and Twitter, status indicators like second person pronouns and emoticons are reliably informative on both domains.

4 Conversation Predictor

We have successfully predicted status on an individual level. However, social status is always defined with respect to other people. The Conversation Predictor thus presents our investigation of social power *relationships*. On Twitter, users can address others using the @ symbol and reply to tweets. Based on the messages exchanged during dyadic Twitter conversations, we try to predict which of the two users is more popular, using number of followers as a proxy. In the following, we call this task *conversation prediction*.

4.1 Background

The theory of accommodation states that conversation partners unconsciously imitate the other along both verbal and non-verbal dimensions (Giles, Coupland, and Coupland 1991). This can be observed in the dimension of linguistic style (Niederhoffer and Pennebaker 2002) and a number of psycholinguistic studies have linked this phenomenon to social status (Giles, Coupland, and Coupland 1991), (Street and Giles 1982), (Infante, Rancer, and Womack 1993), (Giles 2008), one hypothesis being that we accommodate in order to gain the other's social approval. Low-power individuals would thus accommodate more toward more powerful people than the other way round. Recently, it was confirmed that linguistic style accommodation takes place on Twitter but the attempt to use linguistic style accommodation to perform conversation *prediction* was not very successful (Danescu-Niculescu-Mizil, Gamon, and Dumais 2011). Here, we thus define features that capture some of the aspects of accommodation but do not restrict ourselves to linguistic style. We also supplement them with other features taken from the User Predictor.

4.2 Task and Dataset

The dataset of Twitter conversations used for this experiment was collected over a period of 4 months from November 2012 to February 2013. We used the Twitter API to retrieve a random sample of users, crawl their tweets and reconstruct conversations using the `reply_to` information included with each tweet. After eliminating non-English[6] conversations and those that included self-replies, we were left with 2,158 conversations between 1,511 different users. These are typically very short, with an average of 2.9 turns per conversation.

For purposes that will become clear when we discuss our features in Section 4.3, we also downloaded a sample of additional tweets for each user in the dataset, which we call *background tweets*. Table 7 summarises the characteristics of this Twitter conversation dataset.

We define the conversation prediction task as follows: for a given conversation between two users and a set of their background tweets, decide which one has the higher number

[6]We used the Java implementation JTCL of the language guessing library libTextCat (`http://textcat.sourceforge.net/`) with libTextCat models trained on Twitter (Carter, Weerkamp, and Tsagkias 2013).

Twitter Conversations	
Number of conversations	2,158
Number of different pairs	1,353
Number of different users	1,511
Mean turns per conversation	2.9
Number of background tweets per user	25

Table 7: Characteristics of the Twitter conversation dataset.

of followers. Note that accuracies should be expected to remain well below those obtained for the User Predictor, given that we have access to significantly less data when making a prediction.

4.3 Features

We introduce three new feature types for this experiment, namely conversation start, deviation and echoing, and describe in more detail below. Hereafter, let (x, y) represent a pair of users engaged in a conversation C, T_x and T_y stand for x's tweets and y's tweets in this conversation and B_x and B_y stand for x's and y's background tweets, respectively.

Conversation Start It is reasonable to conjecture that an 'ordinary' user is less likely to successfully start a conversation with a celebrity than the other way round. We thus use a binary feature to record which user started the conversation.

Accommodation-based Features We devised two different metrics, *deviation* and *echoing*, which reflect some of the behaviour associated with accommodation and which we discuss in the following.

Deviation represents how much x deviates from their usual way of writing when talking to y and vice-versa, which we expect to happen if accommodation takes place. In order to measure x's deviation, we use the tweets in B_x and compare them to those in T_x along a set of *deviation dimensions* given by a dictionary D. For each dimension, we measure it's frequency in B_x and T_x. We can then calculate x's deviation on a given dimension D_c as follows:

$$Dev_{D_c}(C, x) = |f_{D_c}(B_x) - f_{D_c}(T_x)|$$

with $f_{D_c}(T) = \frac{w_{D_c}(T)}{N_D(T)}$ and where $w_{D_c}(u)$ is the total number of words matching D_c across the set of tweets T and N_D is the total number of words matching any category in D across all of the tweets in T. We also calculate x's total deviation $Dev_D(C, x) = \sum_c Dev_{D_c}(C, x)$. Given $Dev_{D_c}(C, x)$, $Dev_{D_c}(C, y)$, $Dev_D(C, x)$ and $Dev_D(C, y)$ we define binary features indicating which user deviates more on each dimension, as well as who deviates more overall.

Echoing measures a user's tendency to re-use words falling into certain dimensions given by a dictionary D after their conversation partner has used them. For each category D_c of the dictionary, we record whether x uses D_c for the first time after y has used it and vice-versa.

Of course, x re-using y's words does not necessarily mean that x was influenced by y – it could just be that x and y's use of language is similar in general. The coupling of echoing features with deviation features reflects two aspects of accommodation: diverging from one's usual habits and converging with those of the other.

The *style deviation* and *style echoing* features are captured by 27 LIWC dimensions, including pronouns, verbs, articles, prepositions and cognitive processes such as tentativeness and certainty. The NRC Emotion dictionary provides the *emotion deviation* and *emotion echoing* features. Lastly, we use unigrams (i.e. each word functions as a separate dimension) in order to measure *word choice deviation* and *word choice echoing*.

Based on the findings on accommodation presented in Section 4.1 and the fact that we expect deviation and echoing to behave similarly, we put forward the following hypotheses:

H_5: High status users use exhibit lower overall deviation than users with lower status.

H_6: High status users tend to echo their conversation partner's language less than users with lower status.

User Predictor Features We borrowed all User Predictor features except bigrams, retweets and mentions and defined binary features that record which user achieves a higher score for a given feature. For example, for the spelling feature, we use two features: one is true if and only if x makes more spelling mistakes than y and the other is true if and only if y makes more spelling mistakes than x. These only take into account the tweets in T_x and T_y and not x's and y's background tweets. We adapt the tweet and word length features by replacing average tweet length by the total number of words in all of a user's replies throughout the conversation.

4.4 Conversation Prediction Task

SVM classifiers with default settings were trained on the features listed in Section 4.3, both separately and combined. We report 10-fold cross-validation accuracies in Table 8 and compare results to random guessing. The cross-validation folds were constructed in such a way that all conversations between the same pair of users were placed in the same fold.

Of the features that only take into account the current conversation (lines 7 through 15), only conversation start and unigrams do significantly better than random guessing. The unigram features only provide a 3.96 point improvement over the baseline but the conversation start feature is more useful, reaching 58.27% accuracy.

The feature sets in lines 1 to 6 make use of background data and show better results. Word choice deviation achieves the maximum accuracy of 71.56%. Style deviation and emotion deviation also do significantly better than the baseline at 56.88% and 53.58%. Although this doesn't seem particularly high, style deviation accuracy is similar to what was achieved using stylistic features on Wikipedia discussions by Danescu-Niculescu-Mizil *et al.* (Danescu-Niculescu-Mizil et al. 2012). For a given pair of Wikipedia users, they predict the one with higher status based on *all* conversations exchanged between them. Using SVMs they achieve their highest performance of 59.2% with simple LIWC-based stylistic features and 57.7% using stylistic accommodation. The performance of our word choice deviation features thus greatly improves over existing results.

	Feature Set	Accuracy
	Baseline	50.00
(1)	style deviation	56.88**
(2)	emotion deviation	53.68**
(3)	word choice deviation	**71.56*****
(4)	style echoing	48.96*
(5)	emotion echoing	50.07*
(6)	word choice echoing	49.28
(7)	conversation start	58.27***
(8)	unigrams	53.96*
(9)	NRC	51.64
(10)	LIWC	50.35
(11)	emoticons	49.98
(12)	tweet and word length	53.50
(13)	spelling	49.70
(14)	word elongation	48.49
(15)	punctuation	47.34
(16)	All features	71.33***

Table 8: 10-fold cross-validation accuracies on the Twitter Conversations datasets for separate feature sets. The highest achieved accuracy is shown in bold.

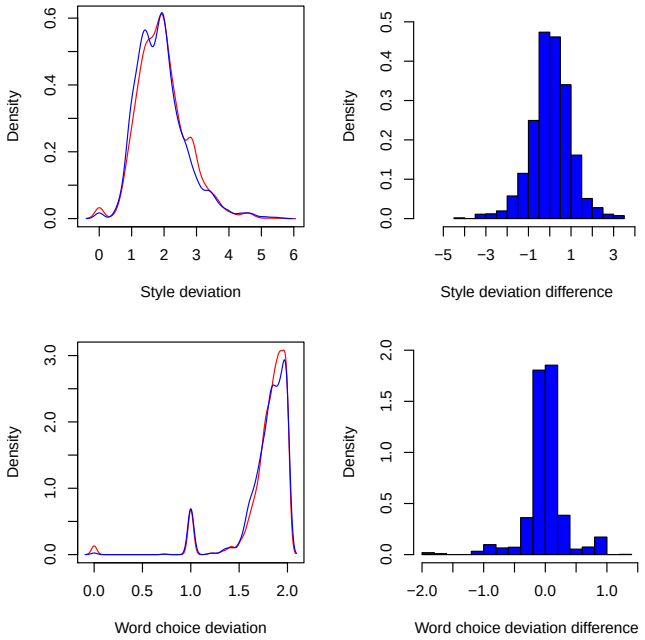

Figure 3: On the left, density plots of deviation for low-power users in red and high-power users in blue. On the right, histograms of $Dev_D(C, x)$ - $Dev_D(C, y)$ for a given dictionary D.

4.5 Conversation Prediction Results

As expected, conversation prediction on Twitter is a difficult task. Due to the shortness of Twitter conversations, little can be predicted without access to a user's background data. When this data is available, measuring the way each user deviates from their usual style and word choice achieved good results. The only feature that does well without access to background data is the conversation start feature. The SVM weights for this feature indicate that popular Twitter users are more likely to successfully start a conversation with someone less popular than vice-versa. Indeed, it may be more probable for a powerful user to receive a reply from a less powerful individual than the opposite.

Unfortunately, the echoing features were not useful and so we are not able to contradict or confirm H_6. Since many conversations are very short, one user using a certain word or LIWC/NRC dimension before the other is probably not indicative of influence but mostly due to chance. However, by looking at the SVM weights produced for the style deviation features we can confirm H_5, namely that popular users deviate less than low status users. Figure 3 compares the probability density of style and word deviation for the less powerful user and the more powerful user across all conversations in our dataset. We can see that although both conversation partner deviate, the low-power users (in red) show higher deviation. A histogram of the quantity $Dev_{Style}(C, x)$ - $Dev_{Style}(C, y)$ (where x is the less popular conversation partner) is shown in Figure 3 and is a further illustration of this phenomenon. The distribution has negative skew (-0.25) meaning that the probability mass lies on the side where x deviates more than y. The corresponding distribution for word choice deviation also has negative skew (-0.80).

It is important to note that, unlike accommodation, deviation doesn't take into account the temporal interaction between the users' replies (we do not capture whether deviation occurs in response to the conversation partner's behaviour) and does not measure linguistic style (or word choice) correlation between the users. Deviation only measures to what extent interacting with someone leads a user to change their usual way of expressing themselves. However, despite using fully-defined accommodation, Danescu-Niculescu-Mizil *et al.* showed that predictions on Wikipedia discussions resulted in an accuracy below 60%, in line with what we have found on Twitter.

5 Conclusion

This study of social power on Twitter and Facebook has shown that it is possible to make robust out-of-sample predictions of popularity and influence based on linguistic features of user messages. Of particular interest is that emoticon use is a powerful predictor of social status on both Twitter and Facebook despite being a rather simplistic way of conveying emotion. Individuals who use emoticons often (and positive emoticons in particular) tend to be popular or influential on Twitter. Since emoticons only occur in text-based communication, their role as a signal of social power is also very specific to the web. Furthermore, our study of Twitter conversations follows similar studies in other domains such as corporate email and Wikipedia. By looking at a broader range of features than have been explored in the past, we can reach above 70% accuracy and improve over previous attempts at conversation prediction. We find that the user who strays the

most from their usual word choice or linguistic style tends to have lower social status. We can predict which conversation partner has higher status with reasonable accuracy even without measuring specific interactions between the two users.

In the future, it would be interesting to explore conversation prediction on Twitter in a more large-scale study in order to further investigate the deviation features we introduce. Furthermore, a cross-platform analysis of conversations could shed light as to the generalisability of our findings on deviation.

References

Bramsen, P.; Escobar-Molano, M.; Patel, A.; and Alonso, R. 2011. Extracting social power relationships from natural language. In *Proceedings of ACL*.

Brody, S., and Diakopoulos, N. 2011. Cooooooooooooooollllllllllll!!!!!!!!!!!!!! using word lengthening to detect sentiment in microblogs. In *Proceedings of EMNLP*.

Budak, C., and Agrawal, R. 2013. On participation in group chats on twitter. In *Proceedings of the 22nd international conference on World Wide Web*, 165–176. International World Wide Web Conferences Steering Committee.

Carter, S.; Weerkamp, W.; and Tsagkias, E. 2013. Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text. *Language Resources and Evaluation Journal* 47(1).

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, K. 2010. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of ICWSM*.

Chung, C. K., and Pennebaker, J. W. 2007. The psychological function of function words. In Fiedler, K., ed., *Social communication: Frontiers of social psychology*. Psychology Press.

Danescu-Niculescu-Mizil, C.; Lee, L.; Pang, B.; and Kleinberg, J. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of WWW*.

Danescu-Niculescu-Mizil, C.; Sudhof, M.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. A computational approach to politeness with application to social factors. In *Proceedings of ACL*.

Danescu-Niculescu-Mizil, C.; Gamon, M.; and Dumais, S. 2011. Mark my words!: Linguistic style accommodation in social media. In *Proceedings of WWW*.

Dino, A.; Reysen, S.; and Branscombe, N. R. 2009. Online interactions between group members who differ in status. *Journal of Language and Social Psychology* 28(1):85–93.

Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; and Lin, C.-J. 2008. Liblinear: A library for large linear classification. *Journal of Machine Learning Research* 9:1871–1874.

Gilbert, E. 2012. Phrases that signal workplace hierarchy. In *Proceedings of CSCW*.

Giles, H.; Coupland, J.; and Coupland, N. 1991. Accommodation theory: Communication, context, and consequences. In *Contexts of accommodation: developments in applied sociolinguistics*. Cambridge University Press.

Giles, H. 2008. Communication accommodation theory. In *Engaging Theories in Interpersonal Communication: Multiple Perspectives*. Sage Publications.

Hutto, C.; Yardi, S.; and Gilbert, E. 2013. A longitudinal study of follow predictors on twitter. In *Proceedings of CHI*.

Infante, D.; Rancer, A.; and Womack, D. 1993. *Building communication theory*. Waveland Press.

Klimt, B., and Yang, Y. 2004. Introducing the enron corpus. In *First Conference on Email and Anti-Spam (CEAS)*.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *Proceedings of WWW*.

Liu, L.; Tang, J.; Han, J.; Jiang, M.; and Yang, S. 2010. Mining topic-level influence in heterogeneous networks. In *Proceedings of CIKM*.

Mitra, T., and Gilbert, E. 2013. Analyzing gossip in workplace email. *ACM SIGWEB Newsletter* Winter:5.

Mohammad, S. M., and Turney, P. D. 2012. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*.

Morand, D. A. 2000. Language and power: An empirical analysis of linguistic strategies used in superior-subordinate communication. *Journal of Organizational Behavior* 21(3).

Niederhoffer, K. G., and Pennebaker, J. W. 2002. Linguistic style matching in social interaction. *Journal of Language and Social Psychology* 21(4):337–360.

Pang, B.; Lee, L.; and Vaithyanathan, S. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of EMNLP*.

Park, J.; Barash, V.; Fink, C.; and Cha, M. 2013. Emoticon style: Interpreting differences in emoticons across cultures. In *Proceedings of ICWSM*.

Pennebaker, J. W.; Francis, M. E.; and Booth, R. J. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*.

Quercia, D.; Ellis, J.; Capra, L.; and Crowcroft, J. 2011. In the mood for being influential on twitter. In *Proceedings of SocialCom*.

Reysen, S.; Lloyd, J. D.; Katzarska-Miller, I.; Lemker, B. M.; and Foss, R. L. 2010. Intragroup status and social presence in online fan groups. *Computers in Human Behavior* 26(6).

Ritter, A.; Cherry, C.; and Dolan, B. 2010. Unsupervised modeling of Twitter conversations. In *Proceedings of NAACL*.

Romero, D.; Galuba, W.; Asur, S.; and Huberman, B. 2011. Influence and passivity in social media. In *Proceedings of WWW*.

Street, R., and Giles, H. 1982. Speech accommodation theory. In *Social Cognition and Communication*. Sage Publications.

Wasden, L. 2010. Internet lingo dictionary: A parent's guide to codes used in chat rooms, instant messaging, text messaging, and blogs. Technical report, Idaho Office of the Attorney General.