Social Networks 40 (2015) 84–89

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Twitter users change word usage according to conversation-partner social identity

Nadine Tamburrini[a], Marco Cinnirella[b], Vincent A.A. Jansen[a], John Bryden[a,*]

[a] School of Biological Sciences, Royal Holloway University of London, Egham TW20 0EX, UK
[b] Department of Psychology, Royal Holloway University of London, Egham TW20 0EX, UK

ARTICLE INFO

Keywords:
Twitter
Community structure
Social identity
Language accommodation
Linguistic convergence
Social network analysis

ABSTRACT

This paper investigates how people express social identity at a large scale on a social network. We looked at communities of users on the Twitter website, and tested two established social-psychology theories that are usually performed at local scale. We found evidence of *Communication Accommodation Theory*, where community members vary their language characteristics depending on which community they are communicating with. We also found the level of linguistic variation correlated with how isolated a community was: evidence that there is *Convergence* between linked members. This demonstrates the power of methods which analyse subtle human behaviour on social networks.

1. Introduction

Social identity is that proportion of an individual's self-concept that derives from membership of a social group (Tajfel and Turner, 1979). Group affiliation has functions of enhancing cooperation (Boyd and Richerson, 2009) and allowing individuals to define others through the group they belong to, in the same way that the individual defines him or herself through the identity of their own group (Ashforth and Mael, 1989). Group members share behaviour and social norms. This shared behaviour in social groups is thought to be generated through processes on social networks such as convergence of behaviour due to social relationships (Hormuth, 1990; Ethier and Deaux, 1994).

The way we use language is strongly associated with our social identity (Scott, 2007). The convergence of behaviour, proposed by social identity theory, is often studied through the language used within social groups. This demonstrates how language is more than just a means of communication and sociolinguistic studies have shown that varieties of a language can be strongly associated with social or cultural groups (Gumperz, 1958; Labov, 1966; Carroll, 2008; Bryden et al., 2013).

By using language as a proxy for social behaviour, studies have been able to understand how expression of social identity is often strongly context dependent: people will behave differently depending on which social identity has the strongest salience in the current situation (Hogg and Reid, 2006). Studies show how this often manifests in the accommodation of language according to the social identity of the interlocutor (Giles, 1973; Gallois et al., 2005). Individuals negotiate the social distance between themselves and the person with whom they are conversing, and are therefore in control of its creation and maintenance (Shepard et al., 2001). For example, Iwasaki and Horie (2000) reported how Thai speakers would adjust their linguistic registers when interacting with strangers. These studies look at specific groups or social situations, but we do not know whether this behaviour can be found at a large scale across many groups where these groups are allowed to freely interact with one another.

Online social networking platforms are providing us with a large scale platform to study human behaviour. With over 200 million monthly active users (Costolo, 2013), the Twitter social network is particularly useful due to its publicly accessible nature (Virk, 2011) and network size. The analysis of large networks brings with it considerable statistical power that allows for the detection of patterns that in traditional, smaller scale network studies would be undetectable. Twitter functions as a micro-blogging website, working on the premise of users sharing their opinions and thoughts in brief messages (maximum 140 characters), which are referred to as "tweets". An investigation into the reasons why people post on the Twitter website by Java et al. (2007) found that about one eighth of posts were conversational messages rendering Twitter as a prime resource for public access to naturally occurring communication (Danescu-Niculescu-Mizil et al., 2011) making this public resource an excellent place to study the expression of social identity.

* Corresponding author. Tel.: +44 (0) 1784 414189.
E-mail address: john.bryden@rhul.ac.uk (J. Bryden).

http://dx.doi.org/10.1016/j.socnet.2014.07.004
0378-8733/

The study of how identity affects our use of language online is a growing field. There is evidence for communication accommodation between offline conversation partners (Danescu-Niculescu-Mizil et al., 2011) showing that syntax, pitch, gestures, word choice, length or form can differ according to interlocutor. Evidence for linguistic convergence online is mixed with studies finding evidence both for (e.g. Riordan et al., 2013) and against (e.g. Christopherson, 2011) the existence of convergence in online communication. The anonymity sometimes engendered in computer mediated environments can act to enhance the significance of social identity in contexts where a relevant shared group membership is salient to users (Postmes et al., 2000). Consequently, social identity can be heightened which explains why some group phenomena, such as polarisation of attitudes, and stereotyping, can seem enhanced in some online environments (e.g. Postmes et al., 2001). This is evident due to collective identity amongst communities of websites of environmental activists (Ackland and O'Neil, 2011). However, such studies of social identity in computer-mediated-communication are still in their relative infancy and this research aims to contribute to the further development of this field by looking to expressly link communication accommodation and convergence to social groups that have formed on Twitter.

In order to identify online groups, we look to the study of complex networks. In this field, the term communities is used to denote parts of the network that are more strongly linked within themselves than to the rest of the network, a phenomenon that has been observed in many human social networks (Porter et al., 2009). In this sense, communities are an emergent property of network structure. Much work has gone into developing methods to detect such groups from topological analysis (Fortunato, 2010), and the extent to which this is possible has been termed modularity (Newman, 2006). The communities found in this way are usually associated with groups of friends or acquaintances, or similarity in traits (Porter et al., 2009; Bryden et al., 2011; Traud et al., 2012) and have also been shown to share language features (Bryden et al., 2013). We hypothesise that communities found in online networks will share social identity and consequently we expect to find that they demonstrate communication accommodation and convergence.

In this study we focus on a specific aspect of behaviour that is strongly associated with social identity, asking whether individuals will shift their linguistic behaviour according to which social group they are messaging. The data of online communities that we used came from a previous study of the Twitter website (Bryden et al., 2013). We tested for communication accommodation by looking to see if users varied specific language characteristics according to whether they had sent conversational messages to members of the same community or to members from other communities. We tested for convergence by looking to see whether this level of language variation for a community correlated with how strongly linked a community is within itself.

2. Methods

The data upon which we did our tests was a network of 189,000 Twitter users. To identify users to download we used a snowball-sample where, for each user sampled, all their tweets which mentioned other users (using the '@' symbol) were recorded and any new users referenced added to a list of users from which the next user to be sampled was picked. Starting from a random user, conversational tweets, time-stamped between January 2007 to November 2009 were sampled from the Twitter website during December 2009, yielding over 200 million messages. The network was formed of bidirectional links, where both nodes had sent at least one message to one another, and weighted by the number of tweets sent between the two users linked. We ignored messages that were copies of other messages (so-called retweets, which are identified by a case-insensitive search for the text 'RT'). In total the network had 75 million messages (tweets) directed from users of the network to one another.

The network was partitioned into communities using a modularity maximisation algorithm (Blondel et al., 2008) and a partition of the network was found where 91% of the tweets were sent by users to other users within the same community. For each community, characteristic words were generated that were used more commonly in that group than the global average (see Supplementary information for characteristic words). These allowed us to identify English speaking groups and also qualitatively summarise shared characteristics of each group. For more information on how characteristic words were generated, and an argument that the network sampled was representative of the complete Twitter network, see Bryden et al. (2013).

To investigate changes in language characteristics, we divided messages into two collections: internal messages that were sent to other members of the same group, and external messages that were sent to members of different communities. For each group, we made sure that both collections were of the same size by discarding messages at random from the larger collection. The difference in word usage between the samples from the two classes was calculated.

To calculate differences between word usage between the two samples we used text similarity measures. We used two different text measures (Gomaa and Fahmy, 2013) to confirm that the result was not an artefact generated by one of the measures. For a word w we define numbers of usages of w in the internal and external samples as $\lambda_i(w)$ and $\lambda_e(w)$ respectively. The first measure was the Euclidean distance between relative word usage frequencies for each collection, given by,

$$\left[\sum_w \left(\frac{\lambda_i(w)}{\sum_v \lambda_i(v)} - \frac{\lambda_e(w)}{\sum_v \lambda_e(v)}\right)^2\right]^{1/2}. \tag{1}$$

The second measure was the quantitative version of the Jaccard distance measure (Gallagher, 1999) which is one minus the multiset intersection of the two samples divided by the multiset union. This is given by,

$$1 - \frac{\sum_w \min\left[\lambda_i(w), \lambda_e(w)\right]}{\sum_v \max\left[\lambda_i(v), \lambda_e(v)\right]}. \tag{2}$$

To look at other linguistic features that can be indicative of changes in linguistic style (see e.g. Bryden et al., 2013; Wagner et al., 2013), we also calculated differences between word-ending frequencies (using both Euclidean and Jaccard distances) and apostrophe frequencies. Differences between apostrophe frequencies were calculated by calculating the frequency of apostrophes per word used by each of the two collections and then calculating the absolute difference between these two values.

3. Results

The partitioning of the sample network of Twitter users yielded 414 groups, with 42 groups having more than 250 users. A variety of languages were found with different groups using different languages. To eliminate the effects of a user simply changing between different languages depending on which group they were speaking to, we did the study on the 24 groups (of a size greater than 250 users) that used the English language which were selected in a previous study ((Bryden et al., 2013), and see methods).

With these English-speaking groups, we formed collections of internal and external messages for each group, and then measured

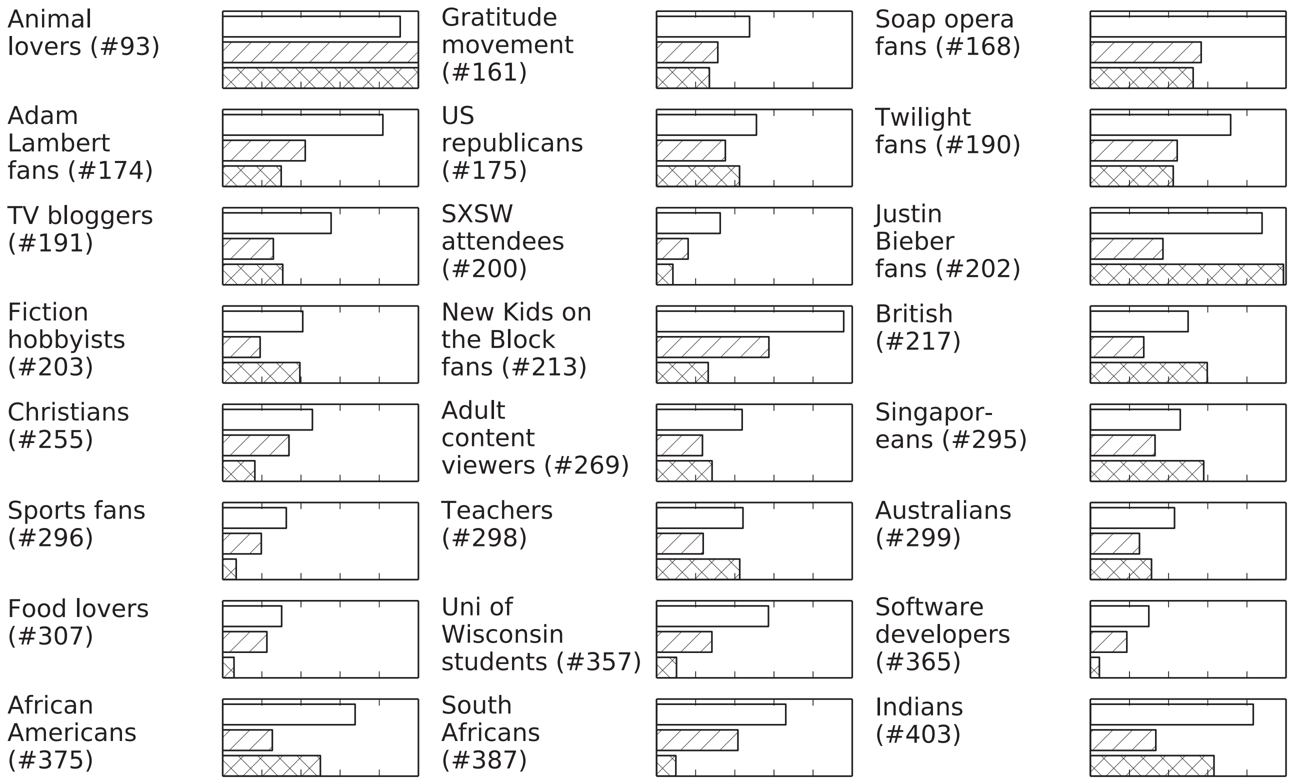

Fig. 1. A comparison of the 24 English speaking groups in the Twitter network showing the extent of linguistic variation between internal and external tweets. The bars show Euclidean distances on a group-by-group basis between internal and external tweets for the three measurements: word-usage frequencies (solid bars at the top of each plot), word-ending frequencies (slashed bars in the middle) and apostrophe usage (crossed bars at the bottom). For each measurement, all groups were scaled so that the values ranged between 0.0 and 1.0. Each group has a short description and a group number. The short description was generated by qualitatively inspecting unusual words generated for each group (see Supplementary information).

the Euclidean distance in word usage frequencies between the two collections. Since differences of word-usage frequencies can arise because users within a group communicate about one or a limited number of subjects, we also measured distances of word-ending usage frequencies and apostrophe usage frequencies to look at markers of linguistic style. We found a variety of distances between internal and external messages in all three measurements (Fig. 1).

There is a variety of distances between the internal and external word usages in Fig. 1. It is possible that these differences in word usage could have happened by random chance. To test this on a group-by-group basis we used a bootstrap by resampling (with replacement) new random pairs of collections of messages from the union of the original internal and external collections used to generate Fig. 1. By calculating linguistic distances between the newly sampled pairs of collections, we can confirm that the difference found between the original group did not happen by chance. Repeating this procedure 1000 times for each group, we calculated the p-value: the proportion of resampled collections for which a linguistic distance exceeded that of the original internal and external collections. In fact, using both the Euclidean and Jaccard measures, none of the distances between the word, or word-ending usages, of the resamples exceeded that of the original collections ($p \leq 0.001$). This showed that the users we studied do indeed change their word and word-ending usage according to whether they are messaging other members of the group or not. For distance between internal and external apostrophe usages, 17 of the 24 groups were significant ($p \leq 0.05$).

The difference between the language use of external and internal messages raises a question as to how much this change in language characteristics is due to the sender of a message conforming to the language use of the receiver. An alternative scenario may be that external messages may have their own language patterns. We investigated this by comparing, using both the Jaccard and the Euclidean measures, the external messages to and from a focal community against the internal messages of every community. We found that the most similar community in each case was the original focal community. This indicates that the change in language characteristics is indeed due to the sender of a message conforming to the language use of the receiver.

The groups of Twitter users analysed in this work were generated by partitioning the sampled network of Twitter users such that the proportion of messages sent within the groups was maximised: so called modularity maximisation (Blondel et al., 2008; Newman and Girvan, 2004). This generated closely interlinked groups that are relatively isolated from the rest of the network. We assessed whether there is any relationship between the level of isolation of a group, measured as the proportion of messages sent by that group to other members of the same group, and the amount of linguistic variation between internal and external messages. We found that the distances between word and apostrophe usage correlated significantly with the proportion of messages sent within the groups (Fig. 2). This indicates that the more a group was isolated from the rest of the network, the more it showed linguistic convergence.

We did not find a significant correlation for word-ending variation against the proportion of internal tweets (Fig. 2, panel b). A visual inspection of the figure reveals that one of the groups is an outlier from the rest across all three measurements of linguistic variation. This group (number 93) is made up of a network of people that organise online parties called ‘pawpawties’ to raise money for animal charities (Manning, 2009). It is intriguing that this group, which largely exists on Twitter, has much stronger language accommodation features compared to similar groups which appear

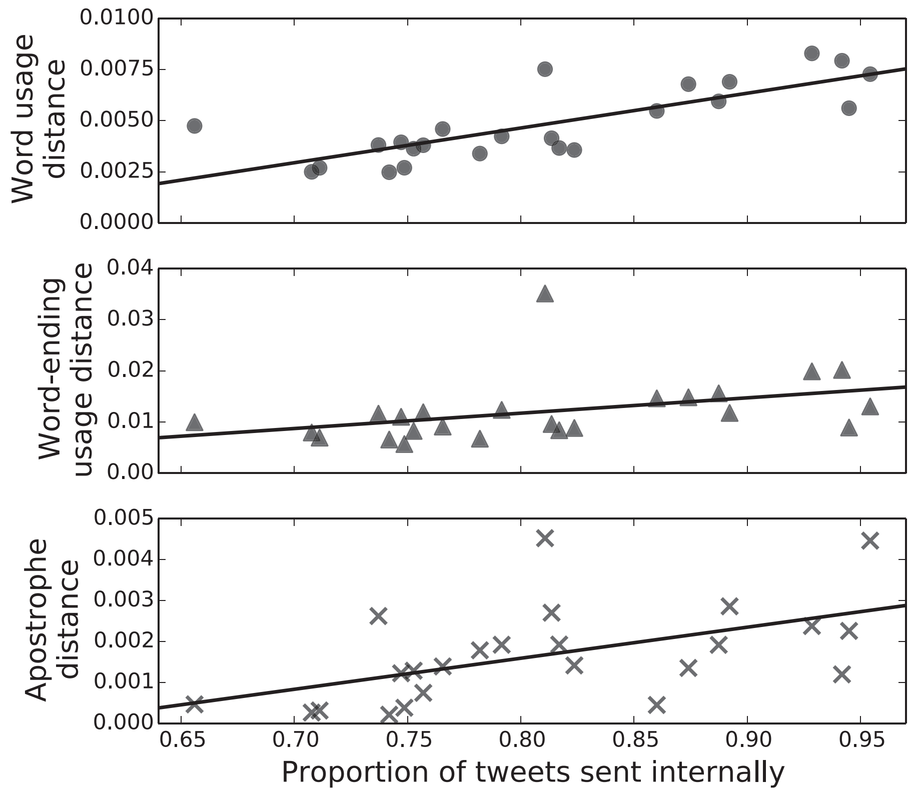

Fig. 2. Linguistic variation between internal and external tweets increases with the proportion of tweets sent within a group. (a) Distance between word-usage (circles with regression line, two-tailed $p = 5.6 \times 10^{-6}$), (b) distance between word-endings (triangles and regression line, two-tailed $p = 0.052$), and (c) distance between apostrophe usage (crosses and regression line, two-tailed $p = 0.0074$).

to have much stronger offline interaction. When we remove this outlier from the regression, we find that there is a significant correlation for word-ending variation against the proportion of internal tweets (two-tailed, $p = 0.00080$).

4. Discussion

Our work demonstrates how computational methods can be used to study social processes on large-scale social networks. Our study was done on an unrestricted large-scale sample of Twitter where individuals interact freely with one another. We used topological analysis to identify social groups in the network and then demonstrated how linguistic behaviour will change according to the group membership of the interlocutors. This shows how subtle trends in linguistic behaviour aggregate to form social identity through communication accommodation and linguistic convergence.

The work illustrates an important methodological tool for studying social processes on large scale social networks. Measurements of social behaviour, especially language features, rarely appear to conform to a normal distribution and are thus difficult to analyse with traditional statistical methods. In this work we use a bootstrap method which, through resampling our data, is independent of whichever distribution the original measurements might come from. The bootstrap is a simple, but powerful, tool for statistical analysis of subtle social processes at such a large scale (Efron and Tibshirani, 1993).

Our study has found evidence of behaviour on the Twitter social network that is consistent with theory on social identity. The results show that people are aware, either implicitly or explicitly, of the social identity of their interlocutor and change their language usage accordingly. This demonstrates that interaction networks with limited communication channels are still sophisticated enough to allow their members to express social identities. We have also found that the extent to which members change their language characteristics depend on how isolated their group is from the rest of the network. This shows that social convergence between several individuals is strongly related to the proportion of their total interaction that they spend within the group.

This study is compatible with other studies of linguistic variation within and between groups (e.g. Bell, 1984; Gregory and Carroll, 1978), and the idea that communities may develop unique linguistic styles which can become intertwined with, and markers of, their identity. Our finding of linguistic differences between ‘internal’ and ‘external’ tweets echoes sociolinguistic work on situational fluctuations in linguistic registers (e.g. Iwasaki and Horie, 2000) and supports a social identity perspective that views such linguistic variation as part of the process of social categorisation.

An important difference with previous studies is the scale at which this study took place. For instance, previous studies that have looked at convergence did not find significance with sample sizes of 30 conversations (Christopherson, 2011). Our approach surpasses the boundaries of survey or interview, and laboratory or field based investigations, with millions of conversations being analysed yielding significant statistical power. While the environment of Twitter is somewhat specific and does not relate to many other on- and offline environments, the fact that our results here were replicated for each community tested indicates that our result is likely to be generalisable.

The differences in word usage between the internal and external messages of each group may be due to each group sharing interests in certain subjects. To go beyond subject areas, we also looked at word endings and apostrophe usage. This is consistent with theory which shows how groups become associated with particular communication styles, members may reference those styles in their communicative acts as a means of claiming or expressing the identity in question (Rampton, 1995).

Our study was restricted to English language groups because a large proportion of the groups in our sample of Twitter used English. While there were groups that spoke other languages in our data, we did not have the quantity of data to adequately resolve sub-groups for non-English speaking Twitter users. We would expect, with more data, to be able to resolve sub-groups for non-English speaking users, and thus be able to test the theory across many different languages.

It is possible that the sampling algorithm that we originally used to sample the Twitter messages may have some introduced some biases which would mean that our sample is not representative of Twitter as a whole. A sampling process used can have some bias toward Twitter users that have had messages sent to them. To mitigate this, we made sure that unsampled users were only placed once on the list of users to be sampled, even if they have been messaged by several previously sampled users. The second issue is that there may be a bias toward certain communities – especially toward the community of the user first sampled. We cover this in more detail in a previous paper (Bryden et al., 2013) arguing that the sampler will move to random communities relatively quickly. We found that our sampling method detected a broad variety of communities and this indicates the sample is likely to be representative of the population.

Interesting future topics which are possible extensions of our work include theory on out-groups, where theory such as Communication Accommodation Theory and the Social Identity Model of Deindividuation predict divergence when interlocutors message certain external groups. We did not find any evidence of this in our study as we found that external messages for a particular group were still closer to the internal messages of the group than any other. Further investigations of how language characteristics converge and/or diverge over time may shed some light on this topic and be of interest in their own right. Finally, we may also be able to improve an algorithm that predicts the groups of individuals based on their language patterns (Bryden et al., 2013), by comparing an individual's language use against that of only the internal tweets of the groups.

Even though the conversations we studied on Twitter were made up of very short text messages which are publicly posted, these results indicate that many complex features of normal offline communication take place online. While such behaviour may not be evident at a small scale, the large quantities of data used in this study meant that we were able to identify these subtle patterns. This indicates that future studies on social identity, social behaviour and cooperation are likely to prove fruitful.

Acknowledgements

Thanks to Shaun Wright, Tim Harrison and the anonymous reviewers. This work was supported by the Economic and Social Research Council (grant ES/L000113/1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.socnet.2014.07.004.

References

Ackland, R., O'Neil, M., 2011. Online collective identity: the case of the environmental movement. Social Netw. 33 (3), 177–190 http://www.sciencedirect.com/science/article/pii/S0378873311000153

Ashforth, B.E., Mael, F., 1989. Social identity theory and the organization. Acad. Manage. Rev. 14 (1), 20–39 http://amr.aom.org/content/14/1/20.short

Bell, A., 1984. Language style as audience design. Lang. Soc. 13 (2), 145–204 http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=2990992

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech.: Theor. Exp. (10), P10008.

Boyd, R., Richerson, P.J., 2009. Culture and the evolution of human cooperation. Philos. Trans. R. Soc. B: Biol. Sci. 364 (1533), 3281–3288 http://rstb.royalsocietypublishing.org/content/364/1533/3281.short

Bryden, J., Funk, S., Geard, N., Bullock, S., Jansen, V.A., 2011. Stability in flux: community structure in dynamic networks. J. R. Soc. Interface 8 (60), 1031–1040 http://rsif.royalsocietypublishing.org/content/8/60/1031.short

Bryden, J., Funk, S., Jansen, V.A., 2013. Word usage mirrors community structure in the online social network Twitter. EPJ Data Sci. 2 (1), 1–9, http://dx.doi.org/10.1140/epjds15.

Carroll, K.S., 2008. Puerto Rican language use on MySpace.com. Centro J. 20 (1), 96–111 http://www.redalyc.org/articulo.oa?id=37720110

Christopherson, L., 2011. Can u help me plz?? Cyberlanguage accommodation in virtual reference conversations. Proc. Am. Soc. Inform. Sci. Technol. 48 (1), 1–9, http://dx.doi.org/10.1002/meet.2011.14504801080/full.

Costolo, R., 2013. Twitter, Inc.: Initial Public Offering. http://www.sec.gov/Archives/edgar/data/1418091/000119312513424260/d564001ds1a.htm

Danescu-Niculescu-Mizil, C., Gamon, M., Dumais, S., 2011. Mark my words!: linguistic style accommodation in social media. In: Proceedings of the 20th International Conference on World Wide Web, pp. 745–754 http://dl.acm.org/citation.cfm?id=1963509

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, New York.

Ethier, K.A., Deaux, K., 1994. Negotiating social identity when contexts change: maintaining identification and responding to threat. J. Pers. Soc. Psychol. 67 (2), 243 http://psycnet.apa.org/journals/psp/67/2/243/

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3), 75–174 http://www.sciencedirect.com/science/article/pii/S0370157309002841

Gallagher, E.D., 1999. COMPAH documentation. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.1334&rep=rep1&type=pdf

Gallois, C., Ogay, T., Giles, H., 2005. Communication accommodation theory: a look back and a look ahead. In: Gudykunst, W.B. (Ed.), Theorizing About Intercultural Communication. Sage, Thousand Oaks, CA, pp. 121–148 http://espace.library.uq.edu.au/view/UQ:72030

Giles, H., 1973. Accent mobility: a model and some data. Anthropol. Linguist. 15 (2), 87–105, http://dx.doi.org/10.2307/30029508.

Gomaa, W.H., Fahmy, A.A., 2013. A survey of text similarity approaches. Int. J. Comput. Appl. 86 (13), 13–18.

Gregory, M., Carroll, S., 1978. Language and Situation: Language Varieties and Their Social Contexts. Routledge and Kegan Paul, London/Henley/Boston http://www.getcited.org/pub/101782218

Gumperz, J.J., 1958. Dialect differences and social stratification in a north Indian village. Am. Anthropol. 60 (4), 668–682, http://dx.doi.org/10.1525/aa.1958.60.4.02a00050/abstract.

Hogg, M.A., Reid, S.A., 2006. Social identity, self-categorization, and the communication of group norms. Commun. Theor. 16 (1), 7–30, http://dx.doi.org/10.1111/j.1468-2885.2006.00003.x/full.

Hormuth, S.E., 1990. The Ecology of the Self: Relocation and Self-Concept Change. Cambridge University Press http://books.google.co.uk/books?hl=en&lr=&id=k9cn1y9iimMC&oi=fnd&pg=PR15&dq=hormuth+1990&ots=UWAaS2HDOn&sig=0RErtS-V_S9Umqk_pYJ25yBnraA

Iwasaki, S., Horie, P.I., 2000. Creating speech register in Thai conversation. Lang. Soc. 29 (04), 519–554 http://journals.cambridge.org/abstract_S0047404500004024

Java, A., Song, X., Finin, T., Tseng, B., 2007. Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, pp. 56–65 http://dl.acm.org/citation.cfm?id=1348556

Labov, W., 1966. The linguistic variable as structural unit. Wash. Linguist. Rev. 3, 4–22.

Manning, S., 2009. Animal lovers throw 'pawpawties' for charity. SFGate http://www.sfgate.com/bayarea/article/Animal-lovers-throw-pawpawties-for-charity-3285436.php#src=fb

Newman, M.E., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. U. S. A. 103 (23), 8577–8582 http://www.pnas.org/content/103/23/8577.short

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69 (2), 026113 http://pre.aps.org/abstract/PRE/v69/i2/e026113

Porter, M.A., Onnela, J.-P., Mucha, P.J., 2009. Communities in networks. Notices Am. Math. Soc. 56 (9), 1082–1097 http://www.ams.org/notices/200909/rtx090901082p.pdf

Postmes, T., Spears, R., Lea, M., 2000. The formation of group norms in computer-mediated communication. Hum. Commun. Res. 26 (3), 341–371, http://dx.doi.org/10.1111/j.1468-2958.2000.tb00761.x/abstract.

Postmes, T., Spears, R., Sakhel, K., De Groot, D., 2001. Social influence in computer-mediated communication: the effects of anonymity on group behavior. Pers. Soc. Psychol. Bull. 27 (10), 1243–1254 http://psp.sagepub.com/content/27/10/1243.short

Rampton, B., 1995. Crossing: Language and Ethnicity Among Adolescents. Longman, London http://www.getcited.org/pub/103194568

Riordan, M.A., Markman, K.M., Stewart, C.O., 2013. Communication accommodation in instant messaging an examination of temporal convergence. J. Lang. Soc. Psychol. 32 (1), 84–95 http://jls.sagepub.com/content/32/1/84.short

Scott, C.R., 2007. Communication and social identity theory: existing and potential connections in organizational identification research. Commun. Stud. 58 (2), 123–138, http://dx.doi.org/10.1080/10510970701341063.

Shepard, C.A., Giles, H., Le Poire, B.A., 2001. Communication accommodation theory. In: Robinson, W.P., Giles, H. (Eds.), The New Handbook of Language and Social Psychology. John Wiley, New York, pp. 33–56.

Tajfel, H., Turner, J.C., 1979. An integrative theory of intergroup conflict. In: Austin, W.G., Worchel, S. (Eds.), The Social Psychology of Intergroup Relations. Brooks/Cole, Monterey, CA, pp. 33–47 http://dtserv2.compsy.uni-jena.de/ss2009/sozpsy_uj/86956663/content.nsf/Pages/58BD3B477ED06679C125759B003B9C0F/$FILE/Tajfel%20Turner%201979.pdf

Traud, A.L., Mucha, P.J., Porter, M.A., 2012. Social structure of facebook networks. Physica A 391 (16), 4165–4180 http://www.sciencedirect.com/science/article/pii/S0378437111009186

Virk, A., 2011. Twitter: the strength of weak ties. Univ. Auckl. Bus. Rev. 13 (1), 19–21 http://www.thebookshelf.auckland.ac.nz/docs/UABusReview/2011_13_i01-5-twitter.pdf

Wagner, C., Asur, S., Hailpern, J., 2013. Religious politicians and creative photographers: automatic user categorization in twitter. In: ASE/IEEE International Conference on Social Computing (SocialCom2013), http://claudiawagner.info/publications/socialcom_userClassification_short_new.pdf