Culture, Networks, Twitter and foursquare: Testing a Model of Cultural Conversion with Social Media Data

Kenneth Joseph and **Kathleen Carley**
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA, 15213
kjoseph@cs.cmu.edu, kathleen.carley@cs.cmu.edu

Abstract

Social network research often takes the view that networks chiefly influence the spread of culture, with few reciprocal effects. While some network scholars have implied a coevolutionary relationship between the two, cultural sociologists have provided increasingly convincing evidence that it is in fact cultural preferences which mediate network structure, rather than the other way around. In the present work, we attempt to validate one such model of the conversion of cultural capital to network position. We use Twitter data to extract the ego networks of individuals and foursquare check-ins to understand their cultural preferences. Our results are indicative of the importance of considering sociological models in which culture influences network structure.

Introduction

Our understanding of the relationship between culture and social network structure is, at best, murky. Many network scholars believe that social connections drive cultural preferences, with little reciprocal influence of culture on network structure (Pachucki and Breiger 2010). Others have come to believe that the relationship between culture and networks is best viewed as symbiotic; cultural preferences coevolve with social interaction and the sharing of information (Carley 1991). Finally, cultural sociologists have recently reinvigorated the assertion that cultural preferences should be considered to play the causal role in the relationship between culture and social network structure, rather than the other way around (Vaisey and Lizardo 2010).

In the present work, we attempt to validate the claims of Lizardo's *cultural conversion model* (CCM) (Lizardo 2006; 2011), a sociological theory which meshes the symbiotic and "culture-first" perspectives. Lizardo argues that culture does not simply coevolve with network structure. Rather, he suggests that individuals are constantly *using* culture in particular ways with particular social ties. More specifically, people who have "passing knowledge" in many domains can use this *weak culture* (Schultz and Breiger 2010) to jump in at the fringes of many different social groups. In contrast, individuals who hold many varieties of *strong culture*, or deep knowledge within particular domains, can use it to form stronger bonds with like-minded individuals. The end result of this process is that ownership of more weak culture leads to less clustered personal (or "ego") networks, while ownership of more strong culture leads to more closed ego networks.

In order to perform an empirical study of the CCM, we collect data on 1,817 Twitter users who routinely post foursquare checkins as public tweets. We use these checkins as markers of the strong and weak cultural preferences that these individuals hold by manually re-coding foursquare venue categories into cultural preference domains used in a previous test of the CCM. We then extract the ego networks of these users by crawling their tweets, follower and followee relationships and the tweets of all others they have mentioned. Finally, we calculate linguistic characteristics of the user's tweets to both relate our efforts back to previous work and to examine the role that these lexical markers of culture may play in network evolution.

Armed with this data, we build regression models to test the following assertions of the CCM:

- The more strong (weak) cultural preferences one has, the more (less) closed one's ego network is
- The more total cultural preferences one has, the more total social ties one has
- The more strong cultural preferences one has, the more strong ties one has
- The more weak cultural preferences one has, the more weak ties one has

Our results support portions of the CCM, but in general seem to be better supported by alternative theoretical work that admits different levels of dynamism exist in both network and cultural structures (Patterson 2014).

Related Work

Several scholars have considered the extent to which various markers of an individual's topical and cultural preferences predict the number of followers she has (Wang and Kraut 2012; Hutto, Yardi, and Gilbert 2013). While this line of work provides useful methodological approaches that are utilized here, it is not clear that the sociotheoretic groundings of the CCM apply to studies of follower counts. This because while the CCM focuses on social ties, following relationships may be representative of "informational" connections rather than social ones (Ma, Sun, and Cong 2013). Scholars seeking to study distinctly social ties on Twitter thus have used various means to extract social relationships and ignore informational ones. The most frequent operationalizations of a *social* tie on Twitter make use of mutual following relationships, mutual retweets or mutual mentions. Though efforts have been made to calibrate better models of tie strength on Twitter (Gilbert 2012), measures of interaction frequency still seem to reliably predict relational strength in social media data (Jones et al. 2013).

Researchers have also considered how social relationships intertwine with various forms of culture on Twitter (Romero, Tan, and Kleinberg 2013; Quercia, Capra, and Crowcroft 2012) and foursquare (Silva et al. ; Joseph, Carley, and Hong 2014). These works provide us with confidence in the existence of an important, if broadly defined, relationship between cultural preferences and network structure in Twitter data. Our work compliments these efforts by making a distinction between two different forms of culture measured in previous work and considering both in a single model. We consider both lexical measures of culture, which have been shown to be relatively dynamic (Eisenstein et al. 2014), and culture as defined by interests in distinct topical domains, which empirical work suggests are far more stable (Lizardo 2006). These two measures of culture are related, but theoretically and thus operationally distinct.

Data and Methodology

The foursquare data we work with is a collection of approximately 12M foursquare check-ins posted publicly to Twitter, along with information on the category of the venues at which users checked in (e.g. "Airport")[1]. In order to extract cultural preferences from this data, we manually match venue categories from check-ins to the nine categories of cultural forms (e.g. Sports, Music, Science, etc.) studied by Lizardo (2011) in his empirical analysis of the CCM. Three human coders were shown a list of venue categories and were asked to label them as being from one of these nine categories, or a "none" category. Fleiss' kappa was 0.64, suggesting manual codings showed "substantial agreement" (Landis and Koch 1977). We then determined the "strength" of the preference for each cultural form for each user in our dataset using a hard threshold - users who had three or more check-ins in a specific cultural preference domain were deemed to have a "strong" preference for that domain. Users who had one or two check-ins in a domain had a "weak" preference for the domain. The hard threshold approach is also used by Lizardo (2011).

Having extracted the strong and weak cultural preferences of our users, we then constructed ego networks using additional data extracted from Twitter. As collecting this data for all users was computationally prohibitive, we a small subset of 1,817 reasonably active users (between 100 and 25K tweets overall, more than 50 tweets in 2014, more than 10 checkins and fewer than 5K followers) for our study. These users were drawn randomly from across the distribution of combined numbers of strong/weak preferences. However, we only considered only users with five or fewer strong and weak preferences, as data beyond this was too sparse.

For each user in our subsample, we collected follower/followee relationships and their full tweet timeline (up to their last 3200 tweets). We used this data to determine the social ties that made up each user's ego network. We considered a social tie to exist between two Twitter users if and only if they both followed each other and had mentioned each other at least once in a tweet sent *during 2014*. The strength of a tie between two users was computed as the minimum number of times one mentioned the other during 2014. For each user we have check-in data for, we completed the extraction of their first-order ego network by adding social ties between their alters where the relationship between the alters fit the definition of a social tie described here. This process required the collection of follower/followee relationship and all tweets for each of these individuals.

After extracting ego networks for our set of users, we then extracted three linguistic markers of their tweets that have been utilized in prior studies: proportion expected to contain informational content (as defined by Hutto, Yardi, and Gilbert (2013)), average number of hashtags per tweet (Hutto, Yardi, and Gilbert 2013) and the average pairwise cosine similarity of unigram representations of the user's tweets (Hutto, Yardi, and Gilbert 2013; Wang and Kraut 2012). These linguistic markers are extracted from only the users' tweets sent *before* 2014, and thus precede tweets used to construct ego networks. This is also true of the data used to extract users' cultural preferences, as data collection for the foursquare check-ins ended in 2012.

Using all of the data described above, we construct four negative binomial regression models (with the canonical logit link function), one to test each of the listed assertions of the CCM. Due to the number of implicit comparisons made, we use $\alpha = .01$ to determine significance; all coefficients discussed in the following section are significant with $p < .01$. Additionally, the models presented are parsimonious, as determined by starting with a full predictive model and then selectively excluding uninteresting variables using ANOVAs to compare nested models. All models discussed show a reliable ($p < .01$) fit to the data.

Before model selection, all four full models include our three linguistic variables as well as three controls-the logarithms of the number of a user's check-ins, the total number of mentions by the user and the total number of tweets by the user in 2014. In cases where the CCM predicts an effect of strong and/or weak cultural preferences, the full model includes both variables as predictors. In the single case where the CCM predicts an effect of the total number of cultural preferences, we use total preference counts as opposed to including both strong and weak counts as predictors. Finally, in the closure model, we follow Lizardo (2011) and include an offset term for the logarithm of the total number of possible connections (i.e. the number of ties squared).

All coefficients in all models are standardized by subtracting the mean and dividing by two standard devi-

[1]We thank Brendan O'Connor and Justin Cranshaw for providing the data

ations (Gelman 2008). Finally, we display results using the *Incident Rate Ratio* (IRR) of the outcome variables. The IRR can be interpreted as a multiplicative effect that a two standard deviation change in the independent variable has on the dependent variable. All code and data necessary to replicate our analysis are available at https://github.com/kennyjoseph/icwsm_lizardo.

Results

Figure 1 displays coefficients, excluding the intercept, for the most parsimonious models for predicting, from left to right, the number of total, strong and weak social ties of the Twitter users we study, as well as the tie closure models. The total tie and weak tie models provide support for two of the assertions we posed regarding the CCM. A two standard deviation increase in a users' total number of cultural preferences is associated with an 18.6% [7.8-30.4%] increase in the users' total number of social ties. Similarly, users with high levels of weak cultural preferences have, on average, almost 14% [4.6-22.9%] more weak ties than those with low levels of cultural preferences As the middle plot in Figure 1 shows, however, there is no significant effect of strong cultural preferences on the number of strong ties that a user has.

Figure 1 also shows that the only other variable appearing in each of the tie count models is the pairwise cosine similarity of a user's tweets prior to 2014. This variable is negatively associated with the number of strong, weak and henceforth total number of social ties for a user. Users with low levels of cosine similarity in their tweets prior to 2014 have, on average, around only 65% of the strong, weak and total ties that users with higher levels of linguistic similarity do. The only other variables we observed that were negative predictors of tie count were proportion of tweets containing informative content, which had a negative effect on total (15.1-22.0% decrease) and strong (36.0-48.4% decrease) tie counts, and the number of check-ins a user had, which had a weaker but reliable negative effect on the total number of ties an actor had (7.0-15.8% decrease).

The right-most plot in Figure 1 shows results for the network closure model. Due to the use of the offset variable, all coefficients are here interpreted relative to the possible number of connections between their social ties. We find no support for the claims of the CCM in our data, as neither strong nor weak cultural preferences emerge as significant predictors of network closure. The only predictors to remain in the parsimonious model of network closure are average hashtag usage, the number of tweets a user sent in 2014 and the lexical coherence of a user's tweets prior to 2014 as measured via cosine similarity.

Discussion

Space constraints limit a full exploration of results. However, relevant to the CCM, we find that weak cultural preferences determined using data from 2012 have a reliable effect on Twitter ego networks constructed from tweets sent *12-24 months* later. Our work thus adds novel empirical evidence to the increasingly popular sociological view that culture

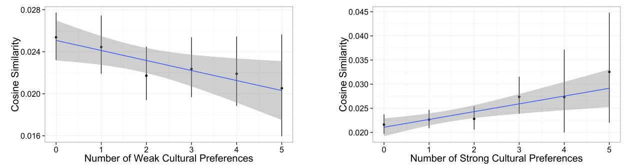

Figure 2: Left: Relationship between weak cultural preferences and cosine similarity of a user's tweets. Black vertical bars are 95% bootstrapped CIs. The blue line is a best-fit least-squares estimate of the linear relationship between the variables, with its own 95% CI in grey. Right: the same plot, except in comparison to strong cultural preferences

has a stable and profound effect on network structure. However, these findings must be qualified in two important ways. First, strong cultural preferences have no effect on strong social ties, nor on ego network closure. It is distinctly possible that this is a result of the "weak tie" nature of Twitter (Gilbert 2012; Hutto, Yardi, and Gilbert 2013) precluding the study of the true impact of strong cultural preferences on network structure. The second caveat is that weak cultural preferences do not decrease closure in a user's ego network. Instead, it is the cosine similarity of users' tweets that has the expected negative association with network closure. Cosine similarity of a users' tweets also predicts a strong decrease in their number of social ties. Neither of these findings can be remedied by the theoretical guidelines established by the CCM, which explicitly focuses on more stable cultural forms that exist beyond language.

Both of these findings are, however, consistent with the "symbiotic" theory from which the CCM draws. Specifically, Constructuralist theory (Carley 1991) predicts that an actor with a more restricted vocabulary should have both a smaller and more closed social network. There thus exists a causal story that posits some stable, external propensity of an actor to have a high level of consistency in their language, which in turn may lead to smaller, more clustered personal network. Though this is not directly implied by the CCM, this interpretation is consistent with a slightly more generic cultural conversion model in which stable cultural schemata influence the emission of more dynamic cultural artifacts, which in turn co-evolve with network structure.

If this were to be the case, we would thus expect that an increase in weak culture is associated with less linguistic similarity in users' tweets, while more strong cultural preferences are indicative of more consistent language. Figure 2 shows, on the left, a negative, significant ($p < .001$) association between lexical coherence and the number of weak cultural preferences one has. On the right, we observe a positive, significant ($p < .001$) association between lexical coherence and the number of strong cultural preferences one has. Our data thus support the idea that stable cultural preferences influence less stable linguistic markers of a user's cultural embeddings, which in turn exist within a symbiotic relationship with network structures.

Conclusion

The present work is motivated by the ongoing debate over the relationship between culture and networks (Pachucki and

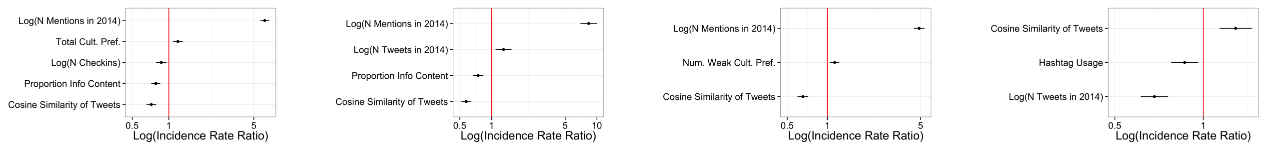

Figure 1: Regression coefficients with 95% CIs for the three tie count models and the closure model. From left to right, we display the total tie count model, the strong tie count model the weak tie count model and finally the tie closure model results. The red line at an IRR of 1 indicates the value at which the independent variable would have no effect on the dependent variable.

Breiger 2010). As with any study that uses social media data, myriad methodological issues may have hindered or played a mediating role in our results (Ruths and Pfeffer 2014). There are, however, issues specific to our efforts. Most importantly, while we feel the use of check-in data comes at least as close to the definition of cultural preferences provided by the CCM as the survey data Lizardo himself used, it is unclear how well foursquare checkins, or the way in which we divided them into strong and weak preferences, really detail the true cultural preferences of users.

Such limitations aside, however, our work provides interesting empirical insight into the ongoing debate over the relationship between culture and networks, furthering recent suggestions in the sociological literature that, as is so often the case, everyone is right. Our results are consistent with a world in which there are certain elements of culture that are highly stable and thus cannot be readily changed via social interaction. These stable cultural forms may have strong effects, in part through less stable cultural artifacts, on the structure of our evolving social networks. Our findings do not preclude the existence, however, of strong social ties which are themselves unaffected by cultural preferences, thus forming a backbone of sociality that deeply affects less stable cultural preferences. Finally, the exchange of the ephemeral elements of culture and the transitory nature of social ties combine to form a mezzo-level, symbiotic linkage between culture and network forms. In such a model, both cultural and network structures exist on a spectrum of dynamism, where more dynamic network elements are more amenable to mediation by more stable cultural elements as well as the other way around. This depiction of culture and networks falls in the spirit, if not in the precise assumptions as they are understood here, of Lizardo's cultural conversion model (Lizardo 2006; 2011). It has also been implied in several other recent discussions of the interplay of culture, cognition and networks (Patterson 2014).

Acknowledgements

This work was supported in part by the Office of Naval Research (ONR) through a MURI N000140811186 on adversarial reasoning and a MINERVA N000141310835.

References

Carley, K. 1991. A theory of group stability. *American Sociological Review* 56(3):331–354.

Eisenstein, J.; O'Connor, B.; Smith, N. A.; and Xing, E. P. 2014. Diffusion of lexical change in social media. *PLoS ONE* 9(11):e113114.

Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine* 27(15):2865–2873.

Gilbert, E. 2012. Predicting tie strength in a new medium. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, 1047–1056. New York, NY, USA: ACM.

Hutto, C. J.; Yardi, S.; and Gilbert, E. 2013. A longitudinal study of follow predictors on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 821–830.

Jones, J. J.; Settle, J. E.; Bond, R. M.; Fariss, C. J.; Marlow, C.; and Fowler, J. H. 2013. Inferring tie strength from online directed behavior. *PLoS ONE* 8(1):e52168.

Joseph, K.; Carley, K. M.; and Hong, J. I. 2014. Check-ins in "Blau Space": Applying Blau's macrosociological theory to foursquare check-ins from New York city. *ACM TIST* 5(3):46:1–46:22.

Landis, J. R., and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *biometrics* 159–174.

Lizardo, O. 2006. How cultural tastes shape personal networks. *American Sociological Review* 71(5):778–807.

Lizardo, O. 2011. Cultural correlates of ego-network closure. *Sociological Perspectives* 54(3):479–487.

Ma, Z.; Sun, A.; and Cong, G. 2013. On predicting the popularity of newly emerging hashtags in twitter. *Journal of the Am. Soc. for Info. Sci. and Tech.* 64(7):1399–1410.

Pachucki, M. A., and Breiger, R. L. 2010. Cultural holes: Beyond relationality in social networks and culture. *Annual Review of Sociology* 36(1):205–224.

Patterson, O. 2014. Making sense of culture. *Annual Review of Sociology* (0).

Quercia, D.; Capra, L.; and Crowcroft, J. 2012. The social world of twitter: Topics, geography, and emotions. In *ICWSM*.

Romero, D. M.; Tan, C.; and Kleinberg, J. 2013. On the interplay between social and topical structure. In *ICWSM*.

Ruths, D., and Pfeffer, J. 2014. Social media for large studies of behavior. *Science* 346(6213):1063–1064.

Schultz, J., and Breiger, R. L. 2010. The strength of weak culture. *Poetics* 38(6):610–624.

Silva, T. H.; de Melo, P. O.; Almeida, J.; Musolesi, M.; and Loureiro, A. You are what you eat (and drink): Identifying cultural boundaries by analyzing food & drink habits in foursquare. In *ICWSM '14*, ICWSM '14. AAAI.

Vaisey, S., and Lizardo, O. 2010. Can cultural worldviews influence network composition? *Social Forces* 88(4):1595–1618.

Wang, Y.-C., and Kraut, R. 2012. Twitter and the development of an audience: those who stay on topic thrive! In *CSCW*, 1515–1518.