Culture, Networks, Twitter and foursquare: Testing a model of cultural conversion using social media data

Kenny Joseph
k joseph@cs.cmu.edu

Kathleen M. Carley
kathleen.carley@cs.cmu.edu
Overview

• Parts of the theory hold up very nicely
• Other parts don’t
• Results point to a “dynamically stable” view of networks and culture (Patterson, 2014)
The “network centric” view

Culture spreads through networks
(White, 1979)
The “culture centric” view

Networks form because of shared culture (Vaisey, 2010)
The Constructuralist model

Networks and culture co-evolve (Carley, 1991)
Lizardo’s Cultural Conversion Model (CCM)

We use culture in particular ways with particular social ties to obtain particular social positions (Lizardo, 2006, 2011)
Strong/weak culture for strong/weak ties

Weak culture for weak ties

Strong culture for strong ties
Hypotheses from Lizardo’s CCM

**H1:** More total cultural preferences, more total ties (2006)

**H2:** More weak cultural preferences, more weak ties (2006)

**H3:** More strong cultural preferences, more strong ties (2006)

**H4:** More strong (weak) cultural preferences, more (less) closed one’s network is (2011)
I mimic Lizardo’s data using social media

GSS data
2006
Where have you been?
(e.g. movies, opera)

2011
What websites have you visited?
(e.g. sport, science)

Strong/weak tie generator

2006

2011
Connections between Strong ties

SM data

foursquare

@jpmon_food
@MikeBerkun @louisstromberg

Carnegie Mellon Institute for Software Research
Specifics on the foursquare data

- ~12M check-ins from 120K users from 2010-2012
- Manually labeled foursquare categories to Lizardo’s website categories (kappa = .64)
- Strong/weak preferences decided via simple threshold

Foursquare (~200 categories)
- Food
  - Burgers
  - Italian
- Sports
  - Stadium
  - Soccer Field
- ... (9 categories)
- Manually Recode
- Sport
- Science
- Humor
- Hobby
- ...
- Threshold to determine strong/weak
  - User 1
    - [0 3 1 ...]
  - User N
    - [10 0 3 ...]
  - = Strong pref
  - = Weak pref
Specifics on Twitter network data

- **N=1817** (various controls)
- Collected full timeline of ego and everyone ego mentioned
- Constructed ego network from 2014 data
- Links are mutual mention, follow

Threshold for strong, weak

\[
\text{min(ment)} = 1
\]

\[
\text{min(ment)} > 1
\]

Compute ties, closure

**User 1:**
- 2 strong ties
- 3 weak ties
- 2 alter ties
Recap

Test Hypotheses
Hypotheses and dependent variables

H1/H2/H3: More total/weak/strong cultural preference, more total/weak/strong ties

Dependent Variable:  
# total/weak/strong ties

H4: More strong (weak) cultural preferences, more (less) closed one’s network is

Dependent Variable:  
# ties between alters
Independent Variables

- Culture variables
  - # strong preferences ($H_3, H_4$)
  - # weak preferences ($H_2, H_4$)
  - # total preferences ($H_1$)

- Control variables
  - Log(# checkins)
  - Log(# mentions) (2014)
  - Log(# tweets) (2014)

- Linguistic variables
  - Info content (Hutto et al., 2013)
  - Avg. number of hashtags (Hutto et al., 2013)

- Offset term for $H_4$
  - Log(# possible connections)
Regression Model

• Used Negative Binomial Regression
  – GLM for overdispersed count data
  – Canonical (logit) link

• Model selection by hand
  – Models shown are parsimonious \( (\alpha = .01) \)
  – Fit visually assessed at each selection

• Variables centered, scaled by 2 s.d.

• Results show Incidence Risk Ratio (IRR)
  – IRR=1: no effect
  – IRR=.5: 2 s.d. increase in IV is 50% decrease in DV
Results for H1 (total ties)

The more total cultural preferences one has, the more total social ties one has.

Data shows support for **H1**

18.6% increase
Results for H2 (weak ties)

The more weak cultural preferences one has, the more weak social ties one has.

Data shows support for H2

14% increase
Results for H3 (strong ties)

The more strong cultural preferences one has, the more strong social ties one has.

Data shows no support for H3.
Results for H4 (closure)

The more strong/weak cultural preferences one has, the more/less closed one’s ego network is

No support for H4
Overview of results

• Support for weak tie hypotheses, none for strong tie/closure hypotheses
  – Methodological/data issues/differences (paper)
  – Twitter is a weak tie platform (e.g. Gilbert, 2012)

• Cosine similarity correlated with smaller, more closed networks
  – Consistent with networks and language coevolving
  – Strong culture -> more cosine similarity?
  – Weak culture -> less cosine similarity?
Is language associated with culture vars?

Language coherence *positively* correlated with *strong* cultural preferences.
Is language associated with culture vars?

Language coherence negatively correlated with weak cultural preferences.
An interesting takeaway…

Culture, networks are “dynamically stable” (Patterson 2014, pg. 22)
Conclusion

• Tested theory of networks and culture using social media data
  – Yay for sociological computation!
• Theory didn’t really hold up
  – Twitter is a weak tie platform (e.g. Gilbert, 2012)
  – Methodological/data issues/differences (paper)
• Future work in formalizing “dynamically stable” networks and culture
Thanks!!

- Poster paper @ ICWSM ’15
- Replication data+code, full version of paper at
  https://github.com/kennyjoseph/icwsm_lizardo
- Me:
  Kenny Joseph
  cs.cmu.edu/~kjoseph
  @_kenny_joseph
The “network centric” view

Culture spreads through networks (White, 1979)
The “culture centric” view

Networks form because of shared culture (Vaisey, 2010)