Making “fetch” happen: The influence of social and linguistic context on the success of lexical innovations

Ian Stewart and Jacob Eisenstein
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
{istewart6+jacobe}@gatech.edu

Abstract
In an online community, new words come and go: today’s lol may be replaced by tomorrow’s lulz. Changes in online writing are usually studied as a social process, with lexical innovations diffusing through a network of individuals in a speech community. But unlike other types of innovation, language change is shaped and constrained by the system in which it takes part. To investigate the links between social and structural factors in language change, we undertake a large-scale analysis of lexical innovation in the online community Reddit. We find that dissemination across many linguistic contexts is a sign of success: words that appear in more linguistic contexts grow faster and live longer. Furthermore, social and context dissemination are complementary. By combining them, it is possible to predict which innovations will stick, and to forecast when the others will begin to decline.

1 Introduction
Stop trying to make “fetch” happen! It’s not going to happen! – Regina George (Mean Girls, 2005)

With the fast-paced and ephemeral nature of online discourse, language change in online writing is both prevalent (Androutsopoulos, 2011) and noticeable (Squires, 2010). In social media, new slang words emerge constantly to replace even basic expressions such as laughter: yesterday’s lol is tomorrow’s lulz. The reasons for such changes are various: lexical innovations may be used to signal familiarity with the latest trends in a community (Bucholtz, 1999), they may represent the orthographic representation of an existing spoken vernacular (Eisenstein, 2015) or they may even result from an exogenous shock, such as the censorship of related terms (Chancellor et al., 2016). But while post hoc analysis of successful innovations can offer insights, we must also address a more basic set of questions. Why do some innovations succeed and extend themselves to new contexts, while others, like fetch, fail to catch on? Can the success of a lexical innovation be predicted from patterns of usage during its early days?

Language change in general, and lexical change in particular, can be treated like other social innovations, such as the spread of hyperlinks (Bakshy et al., 2011) or hashtags (Romero et al., 2011; Tsur and Rappoport, 2015). A key property for predicting the success of an innovation is its dissemination: is it used by many people, and in many social contexts? High dissemination enables words to achieve greater exposure among social groups (Altmann et al., 2011) and may help to convince potential adopters that the innovation is positively evaluated.

But while language change shares some properties with other social innovations, it is also bound by the constraints of the language’s grammar (D’Arcy and Tagliamonte, 2015). New words and phrases rarely change the rules of the game but must instead find their place in a competitive ecosystem with finely-differentiated linguistic roles, or “niches” (MacWhinney, 1989). These structural properties may play a crucial role in determining which an innovation will succeed or fail. We therefore posit a structural analogue to social dissemination.
tion, which we call context dissemination. Some innovations become valid in a broad range of linguistic contexts, while others remain bound to a small number of fixed expressions. We compare the fates of such innovations to determine how contextual and social dissemination each relate to the success of lexical innovations. In particular, we evaluate the following hypotheses:

- **H1**: Words with higher initial social dissemination are more likely to succeed. Previous work has found conflicting effects of social dissemination on innovation success: Garley and Hockenmaier (2012) report a negative correlation between dissemination and frequency change, while Altmann et al. (2011) report a positive correlation. Following the intuition that lexical innovations require a large social base to succeed, we hypothesize a positive correlation between social dissemination and innovations on Reddit.

- **H2**: Words with higher context dissemination in the early phase of their history are more likely to succeed. This follows from work in corpus linguistics showing that words and grammatical patterns with a higher diversity of collocations are more likely to be adopted (Ito and Tagliamonte, 2003; Partington, 1993).

- **H2a**: Words with higher context dissemination are more likely to succeed, even after controlling for social dissemination. This follows from the intuition that linguistic context and social context contribute differently to the success of lexical innovations.

We focus on the adoption of slang words in the popular online community Reddit between 2013 and 2016. Our study departs from previous work by comparing successful innovations with failing innovations, rather than looking at successful innovations alone. To address H2 and H2a, we develop a novel metric for characterizing linguistic context diversity, by comparing the observed number of n-gram contexts to the number of contexts that would be predicted based on frequency alone. Our analysis of the fate of innovations includes: correlation against frequency change (as in prior work); prediction of success versus failure, using matched pairs of words with similar early frequency; and survival analysis, to determine the factors that predict when a word’s popularity will begin to decline. Each of these analyses demonstrates that context diversity plays an important role in explaining the success and longevity of lexical innovations.

## 2 Related

### Lexical change online

Language changes constantly, and one of the most notable forms of change is the adoption of new words, or lexical innovations (Metcalf, 2004). Innovations may arise through the mutation of existing forms, including truncation (favorite to fave; Grieve et al., 2016) and blending (web+log to weblog to blog; Cook and Stevenson, 2010). The fast pace and interconnected nature of online communication is particularly conducive to lexical innovations, and social media provides a “birds-eye view” on the process of change (Androutsopoulos, 2011; Danescu-Niculescu-Mizil et al., 2013; Kershaw et al., 2016; Tsur and Rappoport, 2015). We use Reddit as an example online community to track the success and failure of lexical innovations.

### Social dissemination

Language changes as a result of transmission across generations (Labov, 2007) as well as diffusion across individuals and social groups (Bucholtz, 1999). One way of quantifying the degree of social diffusion is a metric known as social dissemination, equal to the count of social units (e.g., speakers, communities) who have adopted an innovation, normalized by the expected count under a null model in which the innovation is used with equal frequency across the entire population. Altmann et al. (2011) use dissemination of words across forum users and forum threads to predict the words’ change in frequency. They find a robust positive correlation between frequency change and both kinds of social dissemination. To contrast, Garley and Hockenmaier (2012) use the same metric to predict the success of English loanwords on German hip-hop forums and find that social dissemination is still significant but has less predictive power than expected. We seek to replicate these prior findings, test whether they hold even after accounting for context dissemination, and extend the definition of
social dissemination beyond forum users and threads to include sub-communities.

**Linguistic dissemination** In addition to social dissemination, it is important to consider the linguistic context of a lexical innovation: does a new word disseminate to multiple semantic contexts and find a unique “niche” (MacWhinney, 1989) or remain bound to a fixed expression? Work in historical linguistics suggests that the distribution of a new word or construction across lexical contexts can signal future success (Partington, 1993). Furthermore, variationist sociolinguistics work has highlighted the role of grammatical and lexical factors on the production of linguistic variants (Ito and Tagliamonte, 2003; Cacoullos and Walker, 2009), which can often provide more insight onto language change than social factors alone. Our study proposes a generalizable method of measuring the dissemination of a word across lexical contexts and jointly compares the social and linguistic dissemination as predictors of word adoption and abandonment.

### 3 Data

Our study examines the adoption of words on social media, and we focus on Reddit as a source of language change. Reddit is a social content sharing site separated into distinct sub-communities or “subreddits” that center around particular topics (Gilbert, 2013). Reddit is a diverse and dynamic online platform, making it an ideal environment for research on language change (Kershaw et al., 2016). In addition, the division of discussion among communities provides an additional lens on social dissemination beyond users alone: for instance, the dissemination among subreddits may provide a lexical innovation with more exposure than dissemination among users.

We analyze a set of public Reddit comments generated between 1 June 2013 and 31 May 2016, totalling $T = 36$ months of data (comments are grouped by month). To reduce noise in the data, we filter all comments generated by known bots and spam users and filter all comments created in well-known non-English subreddits. We also filter all comments that had been deleted by the time of collection (1 October 2016). The final data collected is summarized in Table 1.

To normalize the text and avoid data sparsity, we replace all references to subreddits and users (marked by the convention r/subreddit and u/user) with r/SUB and u/USER tokens. Similarly, we replace all hyperlinks with a URL token. We also reduce all repeated character sequences to a maximum length of three (e.g., loool to loool).

### 3.1 Finding successful innovations

Our work seeks to study the success of lexical innovations such as the intensifier *af* (“as fuck”). To detect such innovations, we compute the Spearman correlation coefficient $\rho$ between the time steps $\{1...T\}$ and each word’s frequency time series $f_{(1:T)}$, for all words in the top 100,000 words in the vocabulary. The Spearman correlation coefficient captures monotonic, gradual growth that characterizes the adoption of lexical innovations (Grieve et al., 2016; Kenter et al., 2015).

We first filter for words whose Spearman correlation coefficient exceeded the 95th percentile. From this set of words, we identify 1,451 successful lexi-

---

Table 1: Data summary statistics.

|                  | Total       | Monthly mean |
|------------------|-------------|--------------|
| Comments         | 1,625,271,269 | 45,146,424   |
| Tokens           | 56,674,728,199 | 1,574,298,006 |
| Subreddits       | 333,874     | 48,786       |
| Users            | 14,556,010  | 2,302,812    |
| Threads          | 102,908,726 | 3,079,780    |

---

1Data downloaded from http://files.pushshift.io/reddit/comments/ (Accessed 1 October 2016).

2The same list used in Tan and Lee (2015) available online: https://chenhaot.com/data/multi-community/README.txt (Accessed 1 October 2016).

3We randomly sampled 100 posts from the top 500 subreddits and labelled a subreddit as non-English if fewer than 90% of its posts were identified by langid.py as English.

4Because of the qualitative analysis required to identify innovations, we restricted our search to the top 100,000 words. This reduces the risk of data sparsity but may exclude rare innovations.
Figure 1: Detecting failure innovations by fitting a linear piecewise function (left) and a logistic function (right).

3.2 Finding failed innovations

To determine what makes the growth words successful, we need an unsuccessful vocabulary — words that are adopted and later abandoned. Although we do not know if these words were successful before the time period investigated (before 2013), we can assume that their abandonment makes them a useful comparison for the successful innovations. We find such “failed innovations” by fitting two parametric models to the word frequency over time.

**Piecewise linear** We fit a two-part piecewise linear regression on every word’s log-frequency time series \( f_{(1:T)} \), which splits the time series into \( f_{(1:\hat{t})} \) and \( f_{(\hat{t}+1:T)} \) where split point \( \hat{t} \) is a free parameter. The goal is to select a split point \( \hat{t} \) to minimize the sum of the squared error between observed frequency \( f \) and predicted frequency \( \hat{f} \):

\[
\hat{f}(m_1,m_2,b,t) = \begin{cases} 
    b + m_1 t & t \leq \hat{t} \\
    b + m_1 \hat{t} + m_2 (t - \hat{t}) & t > \hat{t},
\end{cases}
\]

where \( b \) is the intercept, \( m_1 \) is the slope of the first phase, and \( m_2 \) is the slope of the second phase. Failed innovations \( F_p \) (“piecewise failures”) display growth in the first phase (\( m_1 \geq 0 \)), decline in the second phase (\( m_2 \leq 0 \)), and a strong fit between observed and predicted data (\( R^2(f,\hat{f}) \) above the 95th percentile). An example word *wot* (“what”) and its piecewise fit are shown in the left panel of Figure 1.

**Logistic fit** To account for smoother growth-decline trajectories, we also fit the curve of a logistic distribution, which is a continuous unimodal distribution with support over the non-negative reals. We identify candidates \( F_l \) (“logistic failures”) as words with a strong fit to this distribution, as indicated by \( R^2 \) above the 95th percentile. An example word *iifym* (“if it fits your macros”) is shown in the right panel of Figure 1. The logistic failures partially overlap with the piecewise failures, because some words’ frequency time series show a strong fit to both the piecewise function and the logistic distribution.

**Combined set** We combine the sets \( F_p \) and \( F_l \) to produce a set of failure candidates. Next, we filter this combined set to only include words from the word categories outlined in § 3.1, yielding a total of 600 failed innovations in set \( F \). Each innovation is assigned a split point \( \hat{t} \) based on the estimated time of switch between the growth phase and the decline.

---

5 Word lists to be released upon publication.
4 Predictors

We now outline the predictors used to approximate the social and linguistic niche of the success and failure words.

4.1 Social dissemination

We consider several versions of the dissemination metric proposed by Altmann et al. (2011) to measure the degree to which a word occupies a specific social niche. Low dissemination implies that a word occupies a limited niche, while a high dissemination implies wide-scale social acceptance. We compute the dissemination of words across a particular social variable (user, thread, and subreddit) as follows. To compute user dissemination $D^U$ for word $w$ at time $t$, we first compute the number of individual users who used word $w$ at time $t$, written $\tilde{U}^{(w)}_t$. We then compare this with the expectation $\hat{U}^{(w)}_t$ under a model in which each user’s decision is completely independent. The user dissemination is the log ratio,

$$\log \frac{\hat{U}^{(w)}_t}{\tilde{U}^{(w)}_t} = \log \tilde{U}^{(w)}_t - \log \hat{U}^{(w)}_t. \quad (2)$$

We compute the expected count $\tilde{U}^{(w)}_t$ using the same equation as Altmann et al. (2011),

$$\tilde{U}^{(w)}_t = \sum_{u \in \mathcal{U}_t} (1 - e^{-f^{(w)}_t m^{(w)}_u}), \quad (3)$$

where $m^{(w)}_u$ equals the total number of words contributed by user $u$ in month $t$ and $\mathcal{U}_t$ is the set of all users active in month $t$. This corresponds to a model in which each token from a user has identical likelihood $f^{(w)}_t$ of being word $w$. In this way, we compute dissemination for all users ($D^U$), subreddits ($D^S$) and threads ($D^T$) for each month $t \in \{1...T\}$.

Examples of words with high and low average social dissemination are shown in Table 2. Highly disseminated words among users include the acronym $asap$ (average $D^U = 0.977$), a widely accepted form of $as\ soon\ as\ possible$, while low-dissemination words among subreddits include $crit$ ("critical hit", average $D^U = 0.453$), which is restricted to users interested in video games. Similarly, $D^T$ approximates the spread of a word among discussion threads, which is relevant to words such as $pvp$ ("player versus player", average $D^T = 0.330$), an acronym restricted to video game threads.

4.2 Context dissemination

Context dissemination captures the diversity of linguistic contexts in which a word appears, as measured by unique $n$-gram counts. We compute the unique count of trigram contexts for all words ($C^3$) using all possible trigram positions: in the sentence "that’s cool af haha", $af$ appears in three unique trigrams, that’s cool af, cool af haha, af haha <END>.\footnote{Pilot analyses with bigram contexts gave similar results.}

The unique number of trigram contexts is strongly correlated with word frequency ($\rho(C^3, f) = 0.904$). We therefore adjust this statistic by comparing with its expected value $\hat{C}$, as in social dissemination. At each timestep $t$, we fit a linear regression between log-frequency and log-unique ngram counts, and then compute the residual $D^C$ between the observed log count of unique trigrams and its expectation, $C^{(w)}_t - \hat{C}^{(w)}_t$.

The expected log-count $\hat{C}^{(w)}_t$ is predicted by a linear regression from log-frequency. This follows from the observation that the relationship between word frequency and contexts follows a roughly
log-log relationship, similar to Heaps’ law (Egghe, 2007). The residual $D^C$, or context dissemination, identifies words with a higher or lower number of lexical contexts than expected.

Examples of growth words with high and low context dissemination are shown in Table 2. High dissemination words include flexible acronyms (aka) and modifiers that can apply to a variety of contexts (ish). Words with low context dissemination include words that are often used in sentence initial or final position (e.g., yikes), and more topic-specific words (e.g., ooc, "out of character"). These differences are due in part to the grammatical aspects of context dissemination: for instance, adjectives (ingame) tend to have higher context diversity than interjections (yeah).

### Grammatical aspects of context dissemination

The grammatical aspects of linguistic context dissemination are confirmed with the distribution of $D^C$ across part of speech tags. These tags were obtained automatically from the CMU Twitter Part-of-Speech tagger (Gimpel et al., 2011). As shown in Figure 2, interjections have lower context dissemination, which follows from being restricted to sentence-initial or sentence-final position. In contrast, adjectives and adverbs have high context dissemination because they can appear throughout the sentence, often near lexical words such as nouns and verbs. But while these differences are real and in some cases substantial (one-way ANOVA between part-of-speech groups: $F = 822.6$, $p < 0.0001$), robustness checks in § 5.2 show that the role of context dissemination in the success of lexical innovations goes beyond part-of-speech category.

### 5 Results

We test the hypotheses about social and context dissemination using three analyses: correlations against frequency; binary classification of successful versus failed innovations; and a survival analysis.

#### 5.1 Correlating frequency change

We can first test the relative importance of the linguistic and social context on innovation success by correlating the context covariates with frequency change ($\Delta f_t = f_t - f_{t-k}$) across all successful innovations. This replicates the methodology in prior work by Altmann et al. (2011) and Garley and Hockenmaier (2012). Focusing on long-term change with $k = 12$ (one year) and $k = 24$ (two years), we compute the proportion of variance in frequency change explained by the covariates (Kruskal, 1987).

The results of the regression are shown in Table 3. Frequency itself is the strongest predictor, which is due to the fact that innovations with low initial frequency often show the most frequency change. In both short- and long-term prediction, context dissemination has a much higher relative importance than each of the social dissemination metrics, indicating the importance of contextual flexibility in predicting word success.

### Table 3: Percent of variance explained in frequency change, computed over all successful words $G$. The number of observations $N$ is equal to 34,848 for $k = 12$, and $N = 17,424$ for $k = 24$. 

| Variance explained | Lower, upper 95% |
|--------------------|------------------|
| $f_{t-12}$         | 10.0% [9.47%, 10.6%] |
| $D^C_{t-12}$       | 1.02% [0.81%, 1.24%] |
| $D^L_{t-12}$       | 0.257% [0.21%, 0.34%] |
| $D^S_{t-12}$       | 0.349% [0.26%, 0.47%] |
| $D^T_{t-12}$       | 0.153% [0.14%, 0.19%] |
| $f_{t-24}$         | 19.9% [18.9%, 20.8%] |
| $D^C_{t-24}$       | 2.13% [1.74%, 2.55%] |
| $D^L_{t-24}$       | 0.414% [0.36%, 0.51%] |
| $D^S_{t-24}$       | 0.881% [0.67%, 1.11%] |
| $D^T_{t-24}$       | 0.532% [0.41%, 0.69%] |

---

7https://github.com/brendano/ark-tweet-nlp (Accessed 17 June 2017).
found results of Garley and Hockenmaier (2012), who
with the other predictors, and with regard to the prior
variance in frequency change, both in comparison
line, match point $\hat{t}$ with matched success word
fanart

Figure 3: Example failure word fuckwit (“idiot”) with matched success word fanart, on $k = 1$ month of frequency data. Split point $\hat{t}$ marked at vertical line, match point $\hat{t} - k$ marked with $\oplus$.

User and thread dissemination explain much less variance in frequency change, both in comparison with the other predictors, and with regard to the prior results of Garley and Hockenmaier (2012), who found 1.5% of variance explained by $D^U$ and 1.9% for $D^F$ at $k = 24$. Our results were robust to the exclusion of the predictor $D^C$, which was not included by Garley and Hockenmaier (2012). One possible explanation relates to our focus on Reddit, where sub-community dissemination ($D^S$) is more relevant to the success of linguistic innovations. This is supported by a separate regression that omitted $D^S$ and found a similar amount of variance explained in the other dissemination predictors. In any case, all of the predictors have nonzero relative significance, according to a bootstrap method (Tonidandel et al., 2009).

5.2 Predicting success and failure

While correlation analysis can help explain the relationship between dissemination and frequency change for successful words, it does not explain what separates successful words from those that fail to stick. We therefore introduce a new analysis, which compares successful innovations with unsuccessful words that had similar initial frequency trajectories.

This analysis is based on a prediction task, differentiating successful innovations from a matched failed innovation, using $k$ months of training data. Each of the failed innovations $w_i$ is matched with a successful innovation $w_j$, based on frequency $f$ from $k$ months before the decline phase beginning at split point $\hat{t}$ for $w_i$. We optimize the matching by grouping the failure words by split point $\hat{t}$ and performing optimal matching within each group (Greevy et al., 2004). For each split point $\hat{t}$, we gather all failure words with the split point into set $F_\hat{t}$ of size $N_\hat{t}$ and use optimal matching (without replacement) to find the set of matched success innovations $\tilde{M}_\hat{t}$ with the best fit. The matching procedure attempts to minimize the Mahalanobis distance $D_m$ between matched word pairs as follows:

$$\min_{\mathcal{M}_\hat{t}} \sum_{(w_i, w_j) \in \mathcal{M}_\hat{t}} D_m(\tilde{t}_{\hat{t} - k; i}, \tilde{t}_{\hat{t} - k; j})$$

s.t. $\mathcal{M}_\hat{t} \in \text{matchings}(\mathcal{G}, F_\hat{t})$, \hspace{1cm} (5)

where $\text{matchings}(\mathcal{G}, F_\hat{t})$ is set of possible matchings between the successful words $\mathcal{G}$ and the failed innovations $F_\hat{t}$. An example match is shown in Figure 3, where the failed innovation fuckwit is matched with successful innovation fanart at $\hat{t} - k$, with $k = 1$.

For each matched pair $(w_i, w_j)$ (success, failure), we include $w_i$ with label $y = 1$, and $w_j$ with label $y = 0$. By design, the resulting dataset will have balanced labels, and equal aggregate frequency across classes at each $\hat{t}$. We then train a logistic regression classifier to predict $y$ (based on $\hat{t}$ months of data before $\hat{t}$), using the same predictors as in the correlation analysis: frequency, social dissemination and context dissemination. We compare the following sub-sets of the predictors: frequency-only ($f$), frequency plus context dissemination ($f + C$), frequency plus social dissemination ($f + S$) and all features ($f + C + S$). To address uncertainty in matching, we bootstrap sample from the failed innovations $F$ with replacement $B = 100$ times. Within each bootstrap sample, we use the matching procedure in Equation 4 to construct a dataset and report the 10-fold cross-validated accuracy.

Using $k = 1$ as our base case (predicting success from failure using only one month of data prior

---

8We use the optmatch package for optimal matching within each split point group: https://www.r-pkg.org/pkg/optmatch.
to the split point, we find that context dissemination and social dissemination both contribute to the likelihood of word success. The accuracies in Figure 4 demonstrate that the combination of context dissemination and social dissemination provides the most information on predicting success from failure. In the full model, the coefficients for $D^C$ are consistently positive (see Table 4), providing support for hypothesis H2 — higher context dissemination makes success more likely. The situation for the social dissemination predictors is more complex. $D^S$ and $D^U$ have positive coefficients, but the coefficient for $D^T$ (thread dissemination) is negative and insignificant. This contrasts with the findings on $D^T$ from Altmann et al. (2011), which suggests that on Reddit the dissemination of a word among threads is less important than dissemination among users, as compared to Usenet where thread and user dissemination are comparable.

### Robustness checks

We perform this same classification task for a range of history lengths, $k \in \{2...8\}$, and find similar accuracy results across the feature sets: $f+C+S$ outperforms $f+C$ and $f+S$, which outperform $f$. Next, based on the distribution of context dissemination across part of speech groups (see Figure 2), it may seem that context dissemination serves as a proxy for differentiation by part of speech. To address this concern, we included part-of-speech tags as predictors in the full model in place of context dissemination ($f+C+S+P$) in the prediction task outlined above and found that the model yielded a lower mean accuracy (64.0%) than the model with the rest of the predictors ($f+C+S$) ($t = -29.6, p < 0.0001$). We further test this hypothesis by grouping the success and failure words by part of speech and comparing the mean context dissemination of success versus failure words, plotted in Figure 5. Across the most frequent speech groups (verbs, common nouns and adjectives), failed innovations have significantly lower context diversity than successful innovations in the same category.

### 5.3 Survival analysis of failed innovations

We now focus on the factors that precede the decline phase in failed innovations, which can be viewed as
Figure 6: Survival curve of all failed and successful words.

the beginning of “word death” (Drouin and Dury, 2009) for many of these terms — though some may emerge again later. We use the failed words as “uncensored” data, words with an observed death date, and the successful words as “censored” data, words with an unobserved death date. The distribution of survivors is shown in Figure 6, which shows that most of the failed innovations begin to decline before the year mark ($t = 12$).

Predicting the time until a word’s decline can be framed as survival analysis (Clark et al., 2003), in which a word is said to survive until the beginning of its decline phase at split point $\hat{t}$. In the Cox proportional hazards model (Cox, 1972), the hazard of death at each time $t$, written $\lambda(t)$, is modeled as a linear function of some predictors $X$,

$$\lambda_i(t) = \lambda_0(t) \exp(\beta \cdot x_i),$$  \hspace{1cm} (6)

where $x_i$ is the vector of predictors for word $i$, and $\beta$ is the vector of coefficients. Each cell $x_{i,j}$ is set to the mean value of predictor $j$ for word $i$ over the training period $t = \{1...k\}$ where $k = 3$. We use frequency, social dissemination and context dissemination as predictors in a Cox regression model.

The estimated coefficients from the regression are shown in Table 5. We find a negative coefficient for context dissemination ($\beta = -0.323, p < 0.001$), which mirrors the results from the previous prediction task: higher $D^C$ leads to a lower hazard of word death, and therefore a higher likelihood of survival. We also find that higher subreddit dissemination leads to a lower likelihood of word death ($\beta = -0.213, p < 0.01$), which also reinforces the finding from the prediction task. Thread dissemination pointed in the opposite direction and user dissemination had no significant effect.

The role of each factor is also tested by comparing the goodness-of-fit for Cox regression models using different feature sets: frequency ($f$), frequency plus context dissemination ($f+C$), frequency plus social dissemination ($f+S$) and all factors ($f+C+S$). The results in Table 6 demonstrate that the model with dissemination and the model with context diversity each have significantly better-than-null fits than the null model. However, the all-factor model ($f+C+S$) does not have a significantly higher deviance than the context dissemination model ($f+C$) ($\chi^2 = 4.6, p = 0.80$), therefore adding social dissemination does not significantly improve the model fit. This points to the especially important role of context dissemination in predicting the success of lexical innovations over time.

To compare the predictive performance of the separate Cox models, we compute their concordance

| Predictor | $\beta$ | stderr | $Z$ | $p$ |
|-----------|---------|--------|-----|-----|
| $f$       | -0.230  | 0.0451 | -5.063 | *** |
| $D^C$     | -0.323  | 0.03459 | -9.350 | *** |
| $D^U$     | -0.0253 | 0.0583 | -0.545 | *  |
| $D^S$     | -0.2127 | 0.07010 | -3.034 | ** |
| $D^T$     | 0.1222  | 0.05606 | 2.180  | *  |

Table 5: Cox regression results for predicting word death with all predictors ($f + C + S$). *** indicates $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, otherwise $p > 0.05$.

| Model    | Deviance | d.f. | $\chi^2$ | $p$-value |
|----------|----------|------|----------|-----------|
| Null     | 4493.6   | 0    |          |           |
| $f$      | 4485.6   | 1    | 8.0      | ***       |
| $f+D$    | 4478.2   | 4    | 15.4     | ***       |
| $f+C$    | 4446.8   | 2    | 46.8     | ***       |
| $f+C+D$  | 4442.2   | 5    | 51.4     | ***       |

Table 6: Deviance values for relevant survival models. $\chi^2$ values based on the likelihood ratio test comparing each model to the null model. Higher deviance indicates better model fit. *** indicates $p < 0.001$. 

($\beta = -0.213, p < 0.01$), which also reinforces the finding from the prediction task. Thread dissemination pointed in the opposite direction and user dissemination had no significant effect.
scores using 10-fold cross-validation. As shown in Figure 7, the model incorporating context dissemination (f+C) consistently achieves higher concordance than the baseline frequency-only model (f), according to Welch’s t-test (t = 4.29, p < 0.001) and the model with all predictors f+C+D outperforms the model with social dissemination f+S (t = 4.64, p < 0.001).

6 Discussion

All three quantitative analyses find a strong role for context dissemination as a positive predictor in the success of lexical innovations: it was the strongest predictor of year-to-year frequency changes, the best differentiator of successful and failed innovations, and the most effective warning sign that the survival of an innovation is coming to an end. Overall, H2 and its stronger form, H2a, are well supported by these analyses. Context dissemination can be related to theories such as the FUDGE factors (Metcalf, 2004), in which the success of lexical innovations depends on a combination of frequency (F), unobtrusiveness (U), diversity of users and situations (D), generation of other forms and meanings (G), and endurance (E). Context dissemination provides an example of “diversity of situation,” because innovations with a higher context dissemination occur in more diverse lexical situations.

Regarding H1, we generally found a positive role for social dissemination as well, although these results were not consistent across all social dissemination predictions. Furthermore, the social dissemination features were relatively ineffective in the survival analysis. Overall, these findings are somewhat aligned with the conclusion from Garley and Hockenmaier (2012), who argued that social dissemination is less predictive of innovation success than Altmann et al. (2011) originally suggested. One possible explanation is the inclusion of word categories such as proper nouns in the analysis of Altmann et al. (2011); it is plausible that the dissemination of such terms responds to different social dynamics.

Future work We approximate lexical context dissemination using trigram counts, because they are easy to compute and generalize across word categories. In future work, a more linguistically sophisticated approach might estimate context dissemination with syntactic features such as appearance across different phrase heads (Kroch, 1989; Ito and Tagliamonte, 2003) or across nouns of different semantic classes (D’Arcy and Tagliamonte, 2015). However, the poor performance of automatic parsers on social media data (Eisenstein, 2013; Blodgett et al., 2016) and the limits of manual annotation may render this typical analysis difficult or impossible. Future work should also investigate the possibility of more semantically-aware definitions of context dissemination. The existence of semantic “neighbors” occurring in similar contexts (e.g., the influence of intensifier very on acronym af) may prevent a new slang word from reaching widespread popularity.

References

Eduardo Altmann, Janet Pierrehumbert, and Adilson Motter. 2011. Niche as a determinant of word fate in online groups. PLoS ONE, 6(5):1–12.

Jannis Androutsopoulos. 2011. Language change and digital media: a review of conceptions and evidence. In Kristiansen Tore and Nikolas Coupland, editors, Standard Languages and Language Standards in a Changing Europe, pages 145–159. Novus Press, Oslo.
Eytan Bakshy, Jake Hofman, Winter Mason, and Duncan Watts. 2011. Everyone’s an influencer: quantifying influence on Twitter. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pages 65–74.

Su Lin Blodgett, Lisa Green, and Brendan O’Connor. 2016. Demographic Dialectal Variation in Social Media: A Case Study of African-American English. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 1119–1130.

Mary Bucholtz. 1999. "Why be normal?": Language and identity practices in a community of nerd girls. Language in Society, 28:203–223.

Rena Torres Cacoullos and James Walker. 2009. The Present of the English Future: Grammatical Variation and Collocations in Discourse. Language, 85(2):321–354.

Stevie Chancellor, Jessica Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thighgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. Proceedings of the 2016 Conference on Computer-Supported Cooperative Work, 27:1201–1213.

T.G. Clark, M.J. Bradburn, S.B. Love, and D.G. Altman. 2003. Survival analysis part I: basic concepts and first analyses. British Journal of Cancer, 89(2):232.

Paul Cook and Suzanne Stevenson. 2010. Automatically identifying the source words of lexical blends in English. Computational Linguistics, 36(1):129–149.

David Cox. 1972. Regression models and life tables. Journal of the Royal Statistical Society, 34:187–220.

Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, and Christopher Potts. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. Proceedings of the 22nd International Conference on World Wide Web, pages 307–317.

Alexandra D’Arcy and Sali Tagliamonte. 2015. Not always variable: probing the vernacular grammar. Language Variation and Change, 27(3):255–285.

Patrick Drouin and Pascaline Dury. 2009. When terms disappear from a specialized lexicon: A semi-automatic investigation into necrology. ICAME journal, 35:19–34.

Leo Egghe. 2007. Untangling Herdan’s law and Heaps’ Law: Mathematical and informetric arguments. Journal of the American Society for Information Science and Technology, 58(5):702–709.

Jacob Eisenstein. 2013. What to do about bad language on the internet. In Proceedings of the 2013 Conference of the North American Association of Computational Linguistics, pages 359–369.

Jacob Eisenstein. 2015. Systematic patterning in phonologically-motivated orthographic variation. Journal of Sociolinguistics, 19:161–188.

Matt Garley and Julia Hockenmaier. 2012. Beefmoves: dissemination, diversity, and dynamics of English borrowings in a German hip hop forum. In Proceedings of the 50th Annual Meeting of the Association of Computational Linguistics, pages 135–139.

Eric Gilbert. 2013. Widespread Underprovision on Reddit. Proceedings of the 2013 Conference on Computer-Supported Cooperative Work, pages 803–808.

Kevin Gimpel, Nathan Schneider, Brendan O’Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah Smith. 2011. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 42–47.

Robert Greevy, Bo Lu, Jeffrey Silber, and Paul Rosenbaum. 2004. Optimal multivariate matching before randomization. Biostatistics, 5(2):263–275.

Jack Grieve, Andrea Nini, and Diansheng Guo. 2016. Analyzing lexical emergence in Modern American English online. English Language and Linguistics, 20(2):1–29.

Rika Ito and Sali Tagliamonte. 2003. Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. Language in Society, 32:257–279.

Tom Kenter, Melvin Wevers, Pim Huijnen, and Maurten de Rijke. 2015. Ad Hoc Monitoring of Vocabulary Shifts over Time. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pages 1191–1200.

Daniel Kershaw, Matthew Rowe, and Patrick Stacey. 2016. Towards Modelling Language Innovation Acceptance in Online Social Networks. Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pages 553–562.

Anthony S Kroch. 1989. Reflexes of grammar in patterns of language change. Language Variation and Change, 1(3):199–244.

William Kruskal. 1987. Relative importance by averaging over orderings. The American Statistician, 41(1):6–10.

William Labov. 2007. Transmission and Diffusion. Language, 83(2):344–387.

Brian MacWhinney. 1989. Competition and Lexical Categorization. Linguistic Categorization, pages 195–242.

Allan Metcalf. 2004. Predicting new words: The secrets of their success. Houghton Mifflin, New York.
Alan Partington. 1993. Corpus evidence of language change: The case of the intensifier. In Mona Baker, Gill Francis, and Elena Tognini-Bonelli, editors, Text and Technology: In Honour of John Sinclair, pages 177–192. John Benjamins Publishing, Philadelphia.

Daniel Romero, Brendan Meeder, and Jon Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In Proceedings of the 20th International Conference on World Wide Web, pages 695–704. ACM.

Lauren Squires. 2010. Enregistering internet language. Language in Society, 39:457–492.

Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In Proceedings of the 24th International Conference on World Wide Web, pages 1056–1066.

Scott Tonidandel, James LeBreton, and Jeff Johnson. 2009. Determining the statistical significance of relative weights. Psychological methods, 14(4):387.

Oren Tsur and Ari Rappoport. 2015. Don’t Let Me Be #Misunderstood: Linguistically Motivated Algorithm for Predicting the Popularity of Textual Memes. Ninth International AAAI Conference on Web and Social Media, pages 426–435.