#anorexia, #anarexia, #anarexyia: Characterizing Online Community Practices with Orthographic Variation

Ian Stewart, Stevie Chancellor, Munmun De Choudhury, and Jacob Eisenstein
Georgia Institute of Technology
Atlanta, GA
{istewart6,schancellor3,munmun.choudhury,jacob.eisenstein}@gatech.edu

Abstract—Distinctive linguistic practices help communities build solidarity and differentiate themselves from outsiders. In an online community, one such practice is variation in *orthography*, which includes spelling, punctuation, and capitalization. Using a dataset of over two million Instagram posts, we investigate orthographic variation in a community that shares pro-eating disorder (pro-ED) content. We find that not only does orthographic variation grow more frequent over time, it also becomes more profound or "deep," with variants becoming increasingly distant from the original: as, for example, *#anarexyia* is more distant than *#anarexia* from the original spelling *#anorexia*. We find that the these changes are driven by newcomers, who adopt the most extreme linguistic practices as they enter the community. Moreover, this behavior correlates with of engagement with the community: the newcomers that adopt deeper variant orthography tend to remain active for longer in the community, and posts with deeper variation receive more positive feedback in the form of "likes." Previous work has linked community membership change with language change, and our work casts this connection in a new light, with newcomers driving an evolving practice, rather than adapting to it. We also demonstrate the utility of orthographic variation as a new lens to study sociolinguistic change in online communities, particularly when the change results from an exogenous force such as a content ban.

1 Introduction

Online communities are defined by their membership and the shared practices of their members. A member of a community with strictly civil practices, like thanking someone for answering a question, would likely have trouble adapting to the language of a community like 4chan [1]. The adoption of such practices can differentiate new members from regular community members, as new members must learn the community's practices in order to be considered a regular community participant [2], [3]. Among community practices, language plays a particularly important role as a signal of shared identity [4]. In the online setting, non-standard **orthography** such as "leet speek" can differentiate community newbies or "noobs" from accepted members [5]. As important as language practices are, they are subject to constant change as a result of exogenous and endogenous events [6], [7]. Who in a community drives these changes? If changing practices are not adopted by all community members, then what characterizes the members who accept and advance these changes?

The social meaning of language change in online communities can be better understood by linking language change to community membership dynamics, i.e., the progression of individual community members from new to regular member. For example, studies have shown that the adoption of slang words and jargon online follows predictable temporal patterns, both at the community level and over the lifespan of individual community members [6], [8]. This lifecycle pattern mirrors the generational aspect of language change by which children acquire a dialect from their parents and peers, and then retain the dialect into adulthood (the "adult language stability assumption") [4].

However, language change may also result from non-generational social forces, such as a content ban in an online community [9], [10]. In 2012, Instagram banned hashtags that promoted eating disorder behaviors, or pro-ED content, such as *#thinspo* [11]. In response, members of the pro-ED community adopted orthographic variations of hashtags to circumvent the ban. Over time, these hashtags grew more popular and more complex, becoming increasingly distant from the original spellings.

This paper outlines a novel approach to measuring change in community practices via orthographic variation. We present the following three research questions to explore in the pro-ED Instagram community:

- **RQ1**: Who uses orthographic variants?
- **RQ2**: Is depth of variation affected by membership attributes (i.e. age and lifespan)?
- **RQ3**: Does orthographic variation affect social reception (via likes and comments) of pro-ED content?

We first address the correlation of orthographic variation to the behavior of pro-ED community members and then the social reception of such variation. In RQs 1 and 2, we focus on two variables that define community membership: **age** in the community and **lifespan**. Prior work has highlighted the role of member age as a factor in the adoption of practices: newcomers can drive adoption of new words within a community but may become more resistant to change as they spend more time in the community [6]. Furthermore, member lifespan, or total duration of time spent in the community, can impact adoption of community practices [12]. RQ3 addresses the social relevance of orthographic variation, which can help explain its adoption within the community.

To address these questions, we analyze over two million Instagram posts and nearly 700 identified orthographic variants of pro-ED hashtags on Instagram. We find that in this community, orthographic variation is driven primarily by newcomers, especially those who will become long-term participants: these individuals are more likely to use orthographical variants, particularly deep variants that are far from the original spellings. The depth of orthographic variation is also correlated with community engagement: messages containing deeper orthographic variants receive more "likes." Lastly, we compare the prevalence of pro-ED hashtags on Instagram and Twitter and find that orthographic variation on Twitter is limited, providing more evidence for the influence of Instagram's content ban.

2 Related Work

Our work draws on research on the adoption of community practices and research in language variation that focuses on orthographic variation.

2.1 Adoption of community practices

The process of knowledge transfer and community growth can be viewed within the framework of **communities of practice**. A community of practice is a group of people who share a set of problems and who demonstrate their expertise in the area through the development of consistent practices [13]. Communities of practice relate to the theory of Legitimate Peripheral Participation (LPP) [3], under which newcomers learn community practices from older members to become full participants. LPP has been frequently employed in qualitative research on online communities to understand the establishment of practices [2], [14].

Under LPP, new community members begin at the periphery and become a full participant through the adoption of community practices [3]. Community practices often receive enforcement from members with more authority or experience [15], such as regular and well-connected members [7]. However, other studies have shown that newcomers are early adopters of ongoing changes; these individuals then become conservative, maintaining the practices that were innovative at the time when they joined the community [6]. Community practices may also be adopted differently depending on member lifespan: for instance, transient, or short-lifespan, members often invest less in community practices than committed, or long-lifespan, members [12].

In our study of changing language practices, we argue that the Instagram members who have adopted pro-ED hashtags form a community of practice. Shared practices include the use of these variants, which make it possible to share pro-ED content in defiance of Instagram's efforts to moderate. Prior work has shown that pro-ED content online is produced by a consistent set of users with similar posting practices and a shared set of problems [9], [16], [17], which further supports our framing of these Instagram users as a community of practice.

2.2 Language Variation

Language variation refers to structured and consistent differences in language use across communities, individuals, or situations. Variation can reveal important social distinctions in attitudes and personal identities [4], [18]. Furthermore, language can vary over time, as when a new generation of speakers learns the language of the previous generation and advances a language change in progress ("transmission") [19]. Sociolinguists have largely focused on variation in spoken language, but written text also exhibits variation, especially in online settings in which traditional language norms may be relaxed [20], [21].

Our work examines orthographic variation, the deliberate use of alternative spellings and other character-level features [22]. This includes phenomena ranging from phonologically-motivated spellings [23] to purely typographical practices such as leet speak [15] and alternative capitalization schemes [24]. Orthographic variation online has been tied to a variety of social behaviors such as identity expression [5], [25], stylistic creativity [26], and community membership [27], [28]. However, there has been little work demonstrating how orthographic variation arises and spreads through online communities. Our work therefore breaks new ground in three important ways: (1) by tracking the use of orthographic variation over each user's lifespan in the community, (2) by linking orthographic variation to signals of social reception, and (3) by explicitly differentiating the frequency and depth of variation.

3 Data

Our dataset has over two million Instagram posts whose content promotes and glorifies disordered eating and exercise behaviors [9]. This dataset is unique in that it has manual labels of hundreds of orthographic variants of pro-ED hashtags, linking variant spellings to their source forms. Furthermore, knowing the original spelling for each variant (e.g., that #*anarexyia* is related to #*anorexia*) allows us to quantify the distance between the variant and source, and thus to quantify the depth of variation.

3.1 Data Collection

We received a dataset of over two million posts and 689 labeled hashtags with 672 unique orthographic variations of seventeen source tags, originally used by Chancellor *et al.* [9]. The details of data collection can be found in that paper; we summarize only the most relevant points here.

3.1.1 Dataset Collection

In late 2014, Chancellor *et al.* used the public Instagram API to create a dataset of pro-ED posts. Because many hashtags could not be queried directly due to the Instagram bans, they identified a set of nine non-banned "seed tags" related to eating disorders. They gathered posts on those seed hashtags for 30 days, and identified the 222 most popular hashtags related to pro-ED behaviors. They manually removed hashtags that were ambiguous (e.g. #*fat*) or related to eating disorder recovery (e.g. #*anorexiarecovery*). This resulted in a set of 72 hashtags, which they used to gather a large dataset. After removing posts with recovery hashtags, the dataset contained 6.5 million posts, dating between January 2011 and November 2014.

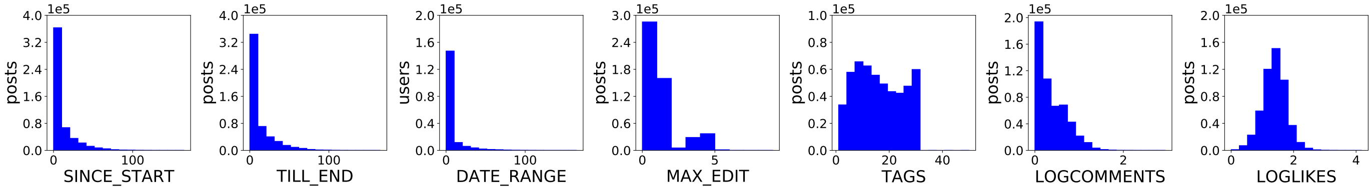

Fig. 1: Summary histograms for all variables of interest, including relative time (e.g. DATE_RANGE), linguistic (MAX_EDIT) and social variables (LOGCOMMENTS).

3.1.2 Finding Source Hashtags and Variant Hashtags

From these 6.5 million posts, Chancellor *et al.* manually checked the top 200 most popular hashtags to see how many were banned by Instagram or placed on a "content advisory" [11]. They found seventeen source hashtags (e.g. #*thighgap* or #*anorexia*) that underwent some form of Instagram intervention. They then developed a set of regular expressions (e.g. $an{*}a{*}$ for #*ana*) to extract semantically similar yet orthographically variant hashtags from the source hashtags. The manual rating yielded 672 unique *orthographic variants*, and seventeen source hashtags, totaling 689 hashtags, which we study here.

Our dataset has 2,416,259 posts from January 2011 to November 2014, each of which contains at least one orthographic variant or source hashtag. Of these, 51% contain at least one variant and no source hashtags.

3.2 Data and metadata extraction

For our analysis, we extracted and calculated the following features from every post, and the post's associated user, relevant to our research questions:

- Real time of post, measured in weeks since Instagram instituted a ban on several pro-ED hashtags[1] (DATE).
- Number of weeks since user's initial post in the data, measuring the user's **age** (SINCE_START).
- Number of weeks until user's final post in the data (TILL_END).
- The total duration (in weeks) of a user's activity, measuring the user's **lifespan** (DATE_RANGE).
- The appearance (binary) of any variant in the post (VARIANT).
- The appearance (binary) of a variant with a specified edit distance in the post (EDIT_DIST_1, etc.; see § 4.2 for a description of how edit distance is computed).
- Maximum orthographic edit distance out of all variants in the post (MAX_EDIT); set to 0 when no variants were in post.
- Total number of all hashtags (variant and non-variant) per post (TAGS).
- Number of comments (COMMENTS) and likes (LIKES) on a post, counted at time of data collection in 2014; log-transformed to adjust for the distributions' long tails.

The distributions of all scalar variables are shown in Figure 1. All of the temporal variables have long-tail distributions, indicating that most member lifespans are short. We also note the bimodal distribution for MAX_POP, showing that the overwhelming majority of posts either contain at least one very popular hashtag, or no popular hashtags at all.

1. This date is not reported by Instagram, but is estimated to be April 1, 2012 [9].

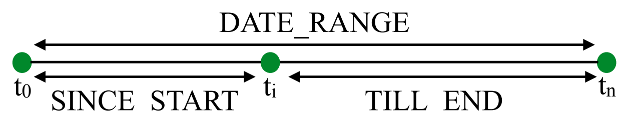

Fig. 2: Example timeline of user posts at times t_0 (first), t_i and t_n (final) that shows age with statistics SINCE_START and TILL_END, and showing lifespan with DATE_RANGE.

4 Methods

We now outline the methods used in our analysis, including operational definitions for key terms, the edit distance metric used to quantify orthographic variation, and the statistical approaches to address our research questions.

4.1 Definitions

For convenience, we provide definitions for the key concepts in our study:

- **Age**: for a given post and the associated user, the length of time between the post at time t_i and the first pro-ED post created by the user time t_0. Age is quantified as the variable SINCE_START, which is equal to the number of weeks since the user's first pro-ED post (SINCE_START $= t_i - t_0$). The variable TILL_END equals the number of weeks until the user's final pro-ED post at time t_n (TILL_END $= t_n - t_i$). These statistics are shown in Figure 2. We define a **newcomer** as a user who, at time of posting, has spent less than ten weeks in the community, and a **regular** as a user who, at time of posting, has spent at least ten weeks in the community.
- **Lifespan**: for a given user, the length of time between a user's first and final pro-ED post. Lifespan is quantified as the variable DATE_RANGE, which is equal to the number of weeks between the user's first and final pro-ED post (DATE_RANGE $= t_n - t_0$). This statistic is shown in Figure 2. We define a **transient** member as having a lifespan less than ten weeks in length, and a **committed** member as having a lifespan of at least ten weeks.
- **Source**: any pro-ED hashtag that was banned in April 2012 and has at least one documented orthographic variant; e.g., #*anorexia*.
- **Variant**: any orthographically-varied hashtag that can be associated with a source hashtag; e.g., #*anoreksya*.
- **Depth**: the linguistic distance between a source and its variant: e.g., the variant #*anoreksya* has a depth 3 from its source #*anorexia* (see § 4.2).

Edit distance	Top 3 variants	Source hashtags	Unique variants	% posts with at least one variant in group
1	*anarexia, bulimic, eatingdisorders*	17	253	41.1%
2	*anarexyia, thinspooo, thynspoo*	15	221	2.07%
3	*secretsociety123, thinspoooo, thygap*	15	108	9.60%
4	*secret_society123, secretsociety_123, thinspooooo*	10	50	10.4%

TABLE 1: Summary of orthographic variants grouped by edit distance. The edit distance 1 group has the greatest variety of source hashtags and unique variants, while the edit distance 4 group has the lowest variety. We restrict our study to variant hashtags with edit distance at or below 4, due to data sparsity above edit distance 4.

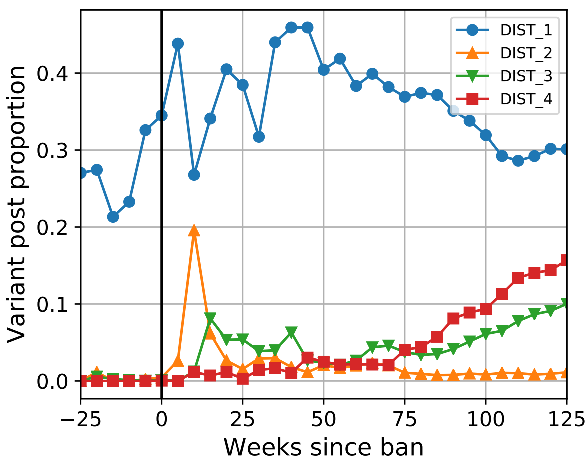

Fig. 3: Frequency of variants over time, grouped by edit distance: e.g., DIST_1 tracks the normalized frequency of all posts with at least one variant with edit distance 1, such as #*anorexiaa*.

We acknowledge that the variables SINCE_START, TILL_END, and DATE_RANGE only capture a slice of each community member's behavior, because a pro-ED hashtag user's actual first post on Instagram may be unrelated to pro-ED (and thus unobservable).

4.2 Measuring Orthographic Variation: Edit Distance

We quantify the depth of orthographic variation by calculating each variant's Levenshtein edit distance from its original form [29]. We count the minimum number of insertion, deletion and substitution operations necessary to convert an source hashtag to its variant form. For example, transforming *anorexia* to *anoreksya* requires two substitutions ($x \rightarrow k$ and $i \rightarrow y$) and an insertion ($\varnothing \rightarrow s$), thus an edit distance of $1 \times 2 + 1 = 3$. Although in some cases it is useful to design a customized edit distance cost function [30], in this study we weight all operations equally for simplicity.[2]

We group orthographic variants by edit distance in Table 1 and provide summary statistics for each group, showing the uneven distribution across groups. We also display the frequency of variants grouped by their edit distance in Figure 3 and note that the overall frequency of orthographic variants increases over time, particularly for the deeper variants at edit distances 3 and 4. Our study examines which community members drive this increase in the frequency and depth of variation over time.

4.3 Statistical Models

We use logistic and Poisson regressions as models for their ease of interpretability, since our RQs concern the relative importance of the temporal, social and linguistic variables of interest. We choose a Poisson regression to address the dependent variables (LIKES and COMMENTS in RQ3), because they are count variables with high dispersion and non-normal distributions [31]. For regressions in which the predictors have multiple unit types, we scale the independent variables between 0 and 1 for consistency.[3] For interpretability, the coefficients are shown in log-odds-per-week. The specific regression models for each RQ are described below:

RQ1: Who uses orthographic variants? We use logistic regressions to predict whether a variant spelling appears in a post (dependent variable), using the following membership attributes as independent variables: (1) the post author's relative age (SINCE_START and TILL_END) and (2) the post author's lifespan (DATE_RANGE), as well as the absolute time DATE for both variables.

RQ2: Is depth of variation affected by membership attributes? We measure depth of variation as the edit distance of a variant from its original form. We consider as a dependent variable the presence of a variant of a specified edit distance from the original tag (e.g., any variant with edit distance 4). We again perform a set of logistic regressions, using the same independent variables as in RQ1 to determine the importance of membership attributes.

RQ3: Does orthographic variation affect social reception? We use Poisson regressions to predict the number of likes and comments that a post receives (dependent variable), using as independent variables the membership attribute variables of the post author (DATE_RANGE and SINCE_START) as well as the post's language content (TAGS, MAX_EDIT, VARIANT).

In all regressions, we remove duplicate posts by members who contribute more than one post for each date to avoid overfitting to the most active members. For logistic regressions, we randomly subsample the data ($n = 200,000$) and include an equal number of positive and negative labeled instances (for class balance). We demonstrate the

2. Preliminary tests with an experimental weighted edit distance showed little difference from the tests with the unweighted edit distance.

3. Following [32], we convert logistic regression coefficients to normal equivalent deviates by dividing by the standard deviation of the standard logistic distribution, $\frac{\pi}{\sqrt{3}} \approx 1.81$.

Variable	β	SE	Effect size
SINCE_START	-0.00456***	2.97E-4	-0.348
TILL_END	0.00294***	2.88E-4	0.654
DATE	0.00529***	1.77E-4	0.746

TABLE 2: Regression results for variant appearance in a post, as predicted by relative time variables. *** = $p < 0.0001$. β indicates regression coefficient, SE indicates the standard error, and "effect size" indicates the standard effect size computed from the odds ratio [32].

relative goodness of fit of models using the metric *deviance*, which is a measure of the lack of fit to data (lower values are better). A model's deviance is calculated by comparing the model with the saturated model, which we define as the "null model." To interpret the relative importance of the variables in the above regression models, we report the non-standardized coefficients of the regression, p-values (computed through the Wald test, adjusted for Bonferroni correction), and standardized effect sizes [32]. All regressions are performed using the Generalized Linear Model code from the `statsmodels` Python package.[4]

5 Results

We address our RQs by analyzing (§ 5.1) the attributes of users who adopt orthographic variants, (§ 5.2) the correlation between orthographic depth and membership attributes, and (§ 5.3) the correlation between orthographic depth and social reception. We also include a comparison with Twitter data to determine the prevalence of orthographic variation across social media platforms (§ 5.4).

5.1 RQ1: Who uses orthographic variants?

Our first task is to determine whether a specific subgroup, such as newcomers [6], appears to drive the community-level tendency toward more orthographic variation.

We display the results of the age regression in Table 2 and the results of the lifespan regression in Table 3. The date coefficient is consistently positive across regressions, reflecting a community-level trend toward more variants over real time (see Figure 3). We also see consistent coefficients for SINCE_START and TILL_END (negative and positive), showing a coherent member-level trend away from variants over the user's lifespan in the community. Taken together, these regressions indicate that orthographic variation is perpetuated by newcomers who bring in the new variants and abandon them over the course of their lifespan. The positive coefficient for DATE_RANGE shows that members who will participate or have participated for longer are more likely to use a variant, suggesting that committed members are more prone to participating in the community change.

Both models achieve a better fit than null. The deviance of the null model and the deviance of both models approximately follow a χ^2 distribution, with degrees of freedom equal to the number of additional variables in the latter model: for age, $\chi^2(3, N = 100,000) = 277,258 - 276,280 = 978$, $p < 10^{-5}$ and for lifespan, $\chi^2(2, N = 100,000) = 277,258 - 276,334 = 924$, $p < 10^{-5}$.

4. http://statsmodels.sourceforge.net/stable/glm.html

Variable	β	SE	Effect size
DATE_RANGE	0.00294***	2.89E-4	0.654
DATE	0.00541***	1.77E-4	0.746

TABLE 3: Regression results for variant appearance in a post, as predicted by the length of a member's lifespan (observed activity period). *** = $p < 0.0001$.

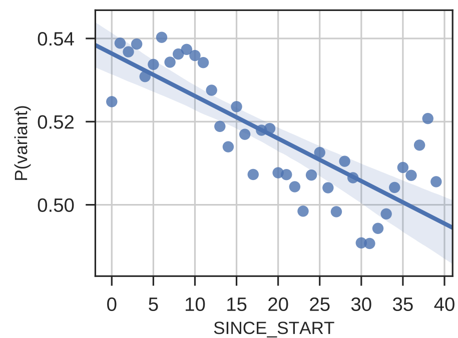

Fig. 4: Probability of using a variant versus a member's age (weeks since first pro-ED post)

This analysis uncovers a split between community-level and member-level variant adoption: as time passes, the pro-ED community is more likely to use variants; but as individual members grow older (especially if the committed members), they are less likely to use variants. We emphasize this trend with Figure 4 which shows the probability of using any variant compared with a member's age, which shows a steady decrease over relative time. Furthermore, members who use variants tend to have longer lifespans posting pro-ED content: the committed (longer lifespan) members are 4.33% more likely to use a variant than transient members ($t = 30.9, p < 0.001$). This difference holds up for the intersection of the two variables, such that committed newcomers are 5.09% more likely to use a variant than transient newcomers ($t = 25.4, p < 0.001$). Overall, we conclude that **committed** members and **newcomers** are the main contributors to the change toward more frequent orthographic variants.

5.2 RQ2: Is depth of variation affected by membership attributes?

Having determined a correlation between membership attributes and variant frequency, we examine whether the social correlates of orthographic variation are stronger for variants that are further from the original spelling. This is done by grouping variants by edit distance, and measuring the strength of association with membership attributes for low and high edit-distance spellings. We first compare these associations directly through univariate analysis, and then perform a set of multivariate regressions.

5.2.1 Univariate analysis

To examine the effect of member age, we plot the relative frequency of posts containing a variant of specified

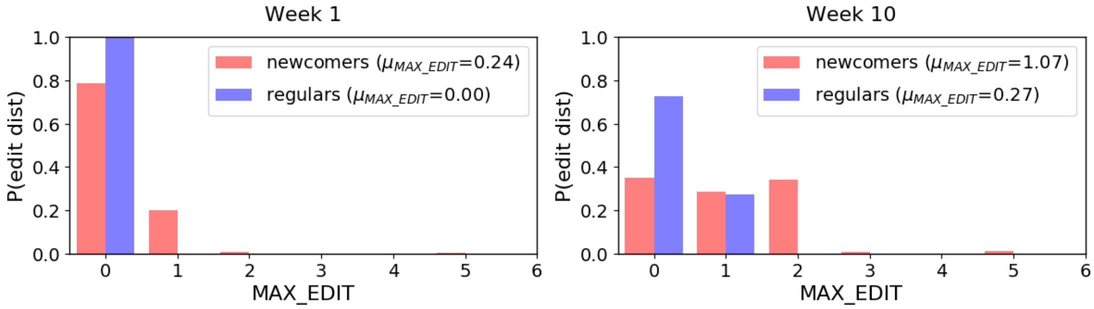

Fig. 5: Distribution of maximum edit distances across all posts of specified member group (regular versus newcomer) at one week and 10 weeks after the ban (including average edit distance for each group). The newcomers use orthographic variants with consistently higher edit distances than the regulars.

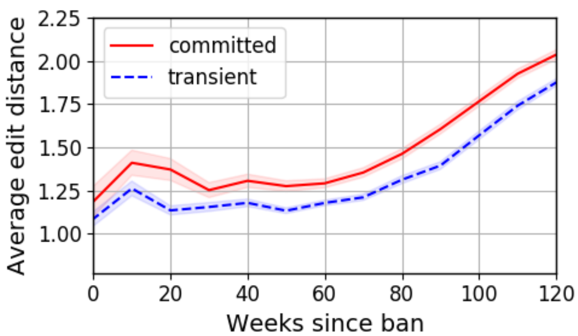

Fig. 6: Average edit distance over time, binned by DATE and DATE_RANGE and including 95% confidence intervals.

edit distance in Figure 5, which shows the distributions of maximum edit distances per post, at one week and ten weeks after the ban. We split the distributions according to relative age of the members who posted a variant, newcomers versus regulars. The newcomers clearly outpace the regular community members in adopting orthographic variants with higher edit distance.

To address member lifespan, we compare the average edit distance in posts from committed and transient members, shown in Figure 6. The y-axis shows the average maximum edit distance variant per post, binned into weekly snapshots. We see a clear separation between the member groups, as committed members consistently use orthographic variants with deeper variation than the transient members. Both transient and committed members follow the same community level trend toward using variants with higher edit distance over time. Interestingly, the separation between transient and committed members remains robust even two years after the ban. To confirm the difference between transient and committed members, we tested all split points in the range from 8-12 weeks and found similar results, suggesting that member lifespan can be reliably correlated with orthographic variation.

5.2.2 Multiple logistic regression

Controlling for confounding factors, we use logistic regression to predict a post's probability of using a variant of specified edit distance (using edit distances 1 and 4), with member age and lifespan as predictors. The results in Table 4 show that effect sizes are larger for the higher edit distance variants: these variants are more quickly adopted by newcomers, more quickly abandoned by older users, and more strongly favored by committed community members. Social differences therefore correlate not only with the frequency of orthographic variation, but also the depth; conversely, these deeper orthographic variables are better indicators of each member's position in the community.

All edit distance regression models achieve a fit significantly better than the null model: e.g., for the edit distance 4 age regression model, the difference of its deviance from that of the null model approximately follows a χ^2 distribution: $\chi^2(3, N = 2,416,259) = 277,258 - 253,808 = 23,450$, $p < 10^{-5}$.

5.3 RQ3: Does orthographic variation affect social reception?

Finally, we investigate how orthographic variation is received by the community using likes and comments received on a post. Although Chancellor *et al.* find that posts with a variant receive more social engagement, it remains to be seen whether this effect is strengthened with deeper edit distance. Since the community norm moves towards variants with deeper edit distance, we expect that posts containing deeper variants would achieve higher engagement in the form of both likes and comments.

To predict the social reception on a given post, we use a fixed-effects Poisson regression to account for the fixed effect of a user's social status on the reception of the post.[5] In this regression, the target predictor is the maximum edit distance of the variants in the post (MAX_EDIT). We also include several control predictors: absolute time (DATE), member age (SINCE_START), presence of hashtag variant in post (VARIANT), number of hashtags per post (TAGS), and presence of a source hashtag or one of its variants (e.g., a post with *#ana* and a post with *#anaa* each have a 1 for feature ANA). The hashtag-source variables partly control for post topic, since posts about a more popular topic like anorexia might also garner more social reception.

As shown in Table 5, posts with deeper variants (higher MAX_EDIT) are positively associated with social engagement through "likes." This complements the earlier finding from

5. We use the **plm** package in R [33] to implement the fixed effects regression: https://cran.r-project.org/web/packages/plm/.

Model type	Dependent variable: EDIT_DIST_1			DEPENDENT VARIABLE: EDIT_DIST_4		
	β	SE	Effect size	β	SE	Effect size
Age						
SINCE_START	-0.00177***	2.98E-4	-0.097	-0.00450***	3.12E-4	-0.416
TILL_END	0.00311***	2.85E-4	0.250	0.0133***	4.00E-4	1.22
DATE	-0.00149***	1.75E-4	-0.127	0.0410***	2.84E-4	3.85
Lifespan						
DATE_RANGE	0.00311***	2.88E-4	0.250	0.0133***	4.03E-4	1.22
DATE	-0.00149*	1.76E-4	-0.127	0.0344***	2.88E-4	3.85

TABLE 4: Logistic regressions to predict the appearance of a variant with a specified edit distance, as predicted by (1) age and (2) lifespan. *** = $p < 0.0001$, * = $p < 0.05$.

	β	SE
Dependent variable: LOGCOMMENTS		
SINCE_START	5.27E-3*	1.57E-3
TAGS	0.110***	2.57E-3
VARIANT	-7.89E-3	3.44E-3
MAX_POP	-2.33E-3	1.26E-3
MAX_EDIT	-3.716E-3	5.51E-3
Dependent variable: LOGLIKES		
SINCE_START	-0.0319***	9.03E-4
TAGS	0.224***	1.47E-3
VARIANT	-1.14E-3	1.98E-3
MAX_POP	-3.89E-3***	7.25E-4
MAX_EDIT	0.0130***	3.16E-3

TABLE 5: Poisson fixed-effects regression for number of comments (above) and likes (below) on a post, as predicted by membership and language variables (hashtag coefficients omitted for brevity). *** = $p < 0.0001$, otherwise $p > 0.05$. Both models achieve a weak fit: the LOGCOMMENTS regression has R^2=6.82E-3 ($F = 107, p < 0.001$) and the LOGLIKES has R^2=0.0902 ($F = 1550, p < 0.001$).

Chancellor *et al.* that posts with variants received more social attention: we see that the increased social attention varies with the depth of variation. It also suggests a success for the pro-ED community: despite Instagram's bans, community members were able to generate variants that attracted more social attention. However, edit distance is not significantly correlated with comments received, which suggests that posts with especially deep variant hashtags do not elicit the more expensive social signal of a comment (as opposed to the passive "like" signal). This may be due to the relatively high proportion of posts with no comments (heavy left-tail of LOGCOMMENTS in Figure 1).

The effect sizes for MAX_EDIT are relatively small, particularly in comparison with the strongest control predictor TAGS. As expected, posts with more tags tend to receive more engagement: such posts are easier to find, using Instagram's hashtag-based search functionality. We also find that community members tend to gain fewer likes as they "age" (positive SINCE_START), which may be a novelty factor that attracts the community's attention to newer members but wears off over time.

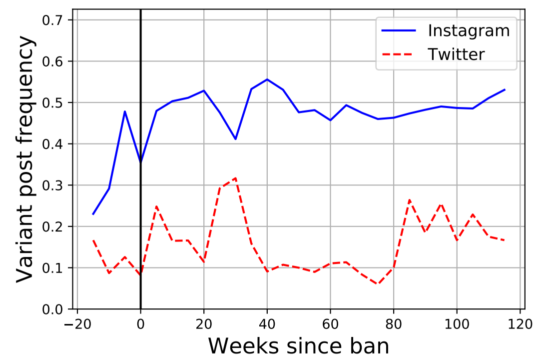

Fig. 7: Frequency of posts containing at least one hashtag variant, on Instagram and Twitter.

5.4 Comparison with Twitter

To better understand how the proliferation of orthographic variation relates to Instagram's restrictions on pro-ED hashtags, we compare with Twitter, which did not implement any restrictions on pro-ED hashtags. Twitter is a useful comparison because, like Instagram, users employ hashtags to index their posts by topic, which helps other users search for posts of interest. We extracted a sample of 4043 tweets containing at least one of the banned or variant hashtags, spanning from January 2012 to June 2014.

The proportion of variants in this sample is significantly lower than in our Instagram data: 51.9% of the pro-ED Instagram posts contain at least one variant, while on Twitter only 15.0% of pro-ED posts contain at least one variant ($Z = 46.8, p < 0.001$). Furthermore, on Twitter there was no noticeable change in posting behavior after Instagram's actions in April 2012, as shown in Figure 7. The difference in mean variant post frequency before and after the ban is significant in the Instagram data (difference = 13.1%, $Z = 16.4, p < 0.001$) but not significant in the Twitter data (difference = 3.4% $Z = -1.64, p > 0.05$). The orthographic variation on Twitter was also considerably less diverse: of the 608 tweets containing a variant hashtag, only 23 tweets contained a variant hashtag with edit distance greater than one. None of those 23 tweets reached Twitter's 140 character limit (mean character count 70.3), suggesting that the lack of orthographic variation on Twitter was not due to reaching the character limit.

This cross-community comparison demonstrates that the pro-ED hashtags were more likely a result of the content ban on Instagram than a result of an overall trend toward more variation across social media platforms (although there may be a difference in user demographics; see § 6.2).

6 Discussion

6.1 Theoretical and Practical Implications

Our main finding — that committed newcomers lead language change — shows that changing community practices can be tied to the users' progression from newcomers to regulars. It is important to relate this result to the earlier finding that newcomers adopt innovative language practices, and then retain these practices even as they become outdated with respect to the rest of the community [6], [34]. The "old-timers" in the beer forums studied by Danescu-Niculescu-Mizil *et al.* [6] are merely conservative, clinging to the linguistic habits of their youth. In contrast, pro-ED Instagram users were regressive: they began with innovative practices, but they abandoned these practices and returned to standard spellings — even as the overall community change was driven by subsequent waves of newcomers toward ever more frequent and deeper orthographic variation.

We also need to contextualize our findings within the scope of pro-ED behavior online. Since pro-ED behavior is stigmatized [9], newcomers in the community might be pressured to hide their early activities with especially deep variants and only later feel encouraged to post without variants. Under this view, the newcomers are using orthographic variation as a tool to hide their content rather than using it to signal community membership, because the newcomers may not even be aware of the broader pro-ED community when they begin posting.

Our work used orthographic variation as a lens to measure an individual's linguistic distance from an established standard language, a distance which reflects how members of a community can distinguish themselves from others [18]. This adds another method to the toolkit of language analysis and provides an interesting path for future work in social computing: can the behavior of members of a community be characterized along a continuum, by their linguistic distance from standard language? For example, in a community with relatively standard writing practices, the use of excessive capitalization and lengthening (e.g., *duuuuudddde*) may be viewed as a non-conformist position towards the community, moreso than mild examples of expressive lengthening (e.g., *duude*). The opposite can also apply: in a freewheeling community like 4chan, purposeful misspellings may be more common and the use of conventional orthography might be viewed as deviation from the community standard [1].

6.2 Limitations and future work

Because Instagram's content ban prevented us from collecting the data directly (e.g. querying for banned terms), we may have missed some orthographic variants. Furthermore, Instagram's API prevented us from querying for additional user information, such as the date at which each user joined the site instead of the first date at which they used a pro-ED hashtag. This information would complement our analysis and allow us to differentiate newcomers from regulars based on their actual first post date. Having more detailed user information would also provide a better perspective on the correlation between orthographic variation and social reception: for example, we would be able to test for a connection between social network structure and orthographic variation. This would help to control for the user-specific effects of social reception, for which we would be unable to account without more rigorous matching techniques. With respect to the cross-platform comparison, we also acknowledge the differences in user demographics between platforms: as of 2014 Instagram skewed slightly more toward women than Twitter [35], which could result in different aggregate behaviors between the platforms.

Future work should dig deeper into the unusual finding about newcomers, looking into three possible explanations. First, the new pro-ED community community members may adopt the most extreme practices to signal legitimacy in the community, which represents an extreme version of the Community of Practice model in which members gain legitimacy through adoption of social and linguistic practices [3]. Second, the adoption of more extreme hashtag variants may represent a form of "flag-planting," by which a newcomer attempts to claim a particular hashtag as their own with an especially extreme variant. Third, the supposed "newcomer" members could actually be new accounts created as a result of being banned, who then adopt more extreme variants to avoid being banned again. This third possibility is especially relevant in the face of prior findings that moderation of deviant behavior online may cause the deviant user's practices to become more extreme [36].

7 Conclusion

Our study has demonstrated the utility of orthographic variation as a means of characterizing community-level change and differentiating users by social role. We determine that the community level change toward more orthographic variation is driven by committed newcomers, who later abandon their use of variants and accordingly receive more social response. We also find that the depth of orthographic variation can differentiate members by age and lifespan, and can weakly predict the level of social response that a post receives. These results have the potential to push social computing research to consider a wider range of language variation, outside of typical change such as adoption of slang, when characterizing an online community and the behavior of its members. Employing non-lexical metrics like orthographic edit distance can help researchers capture linguistic change that may otherwise not be apparent.

Acknowledgments

We thank the anonymous reviewers for their feedback, and the audience at the Diversity and Variation in Language conference for their feedback on an early version of this work. We also thank Brendan O'Connor for providing the Twitter data for comparison in § 5.4. This work was supported by ... [@Jacob and @Munmun need to fill this in -I]

References

[1] M. Bernstein, A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. Vargas, "4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community," **Proceedings of ICWSM**, pp. 50–57, 2011.

[2] S. L. Bryant, A. Forte, and A. Bruckman, "Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia," in **Proceedings of GROUP**, 2005, pp. 1–10.

[3] J. Lave and E. Wenger, **Situated learning: Legitimate peripheral participation**. Cambridge University Press, 1991.

[4] W. Labov, **Principles of Linguistic Change: Social Factors**. Wiley-Blackwell, 2001, vol. 2.

[5] J. Androutsopoulos, "Language change and digital media: a review of conceptions and evidence," in **Standard Languages and Language Standards in a Changing Europe**. Novus Press, 2011, pp. 145–159.

[6] C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, and C. Potts, "No country for old members: User lifecycle and linguistic change in online communities," in **Proceedings of WWW**, 2013, pp. 307–318.

[7] F. Kooti, H. Yang, M. Cha, P. K. Gummadi, and W. A. Mason, "The emergence of conventions in online social networks." in **Proceedings of ICWSM**, 2012.

[8] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, "Diffusion of lexical change in social media," **PLoS ONE**, vol. 9, no. 11, 2014.

[9] S. Chancellor, J. A. Pater, T. Clear, E. Gilbert, and M. De Choudhury, "#thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities," in **Proceedings of CSCW**, 2016, pp. 1201–1213.

[10] C. Hiruncharoenvate, Z. Lin, and E. Gilbert, "Algorithmically Bypassing Censorship on Sina Weibo with Nondeterministic Homophone Substitutions," in **Proceedings of ICWSM**, 2015, pp. 150–158.

[11] H. Hasan, "Instagram bans thinspo content," **Time Newsfeed**, April 2012.

[12] Y. Ren, R. Kraut, S. Kiesler, and P. Resnick, "Encouraging commitment in online communities," in **Evidence-based social design: Mining the social sciences to build online communities**. MIT Press, 2011, pp. 77–125.

[13] L. Dubé, A. Bourhis, and R. Jacob, "Towards a typology of virtual communities of practice," **Interdisciplinary Journal of Information, Knowledge, and Management**, vol. 1, no. 1, pp. 69–93, 2006.

[14] M. Schlager, J. Fusco, and P. Schank, "Evolution of an online education community of practice," in **Building virtual communities: Learning and change in cyberspace**. Cambridge University Press, 2002, pp. 129–158.

[15] K. Blashki and S. Nichol, "Game geek's goss: linguistic creativity in young males within an online university forum," **Australian Journal of Emerging Technologies and Society**, vol. 3, no. 2, pp. 71–80, 2005.

[16] G. Cobb, ""This is not pro-ana": Denial and disguise in pro-anorexia online spaces," **Fat Studies**, vol. 5, no. 2, pp. 1–17, 2016.

[17] L. Manikonda and M. De Choudhury, "Modeling and characterizing visual attributes of mental health disclosures in social media," **Proceedings of CHI**, 2017.

[18] P. Eckert and S. McConnell-Ginet, "Think Practically and Look Locally: Language and Gender as Community-Based Practice," **Annual Review of Anthropology**, vol. 21, pp. 461–490, 1992.

[19] W. Labov, "Transmission and diffusion," **Language**, vol. 83, no. 2, pp. 344–387, 2007.

[20] J. Eisenstein, "What to do about bad language on the internet," in **Proceedings of HLT-NAACL**, 2013, pp. 359–369.

[21] S. C. Herring, "Grammar and electronic communication," **The Encyclopedia of Applied Linguistics**, pp. 1–9, 2012.

[22] A. Jaffe, J. Androutsopoulos, M. Sebba, and S. Johnson, **Orthography as Social Action: Scripts, Spelling, Identity and Power**. Berlin: Walter de Gruyter, 2012.

[23] J. Eisenstein, "Systematic patterning in phonologically-motivated orthographic variation," **Journal of Sociolinguistics**, vol. 19, pp. 161–188, 2015.

[24] O. Tsur and A. Rappoport, "Don't let me be #misunderstood: Linguistically motivated algorithm for predicting the popularity of textual memes," in **Proceedings of ICWSM**, 2015, pp. 426–435.

[25] A. Zelenkauskaite and S. C. Herring, "Television-mediated conversation: Coherence in Italian iTV SMS chat," in **Proceedings of HICSS**, 2008, pp. 145–145.

[26] T. Schnoebelen, "Do you smile with your nose? Stylistic variation in Twitter emoticons," **University of Pennsylvania Working Papers in Linguistics**, vol. 18, no. 2, p. 14, 2012.

[27] M. Back and M. Zepeda, "Performing and positioning orthography in Peruvian CMC," **Journal of Computer-Mediated Communication**, vol. 18, no. 2, pp. 119–135, 2013.

[28] C. Leigh, "Lurkers and lolcats: An easy way from out to in," **Journal of Digital Research and Publishing**, vol. 2, pp. 131–139, 2009.

[29] V. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," in **Soviet Physics Doklady**, vol. 10, 1966, pp. 707–710.

[30] W. Heeringa, P. Kleiweg, C. Gooskens, and J. Nerbonne, "Evaluation of string distance algorithms for dialectology," in **Proceedings of ACL**, 2006, pp. 51–62.

[31] W. Gardner, E. P. Mulvey, and E. C. Shaw, "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models," **Psychological bulletin**, vol. 118, no. 3, pp. 392–404, 1995.

[32] S. Chinn, "A simple method for converting an odds ratio to effect size for use in meta-analysis," **Statistics in medicine**, vol. 19, no. 22, pp. 3127–3131, 2000.

[33] Y. Croissant, G. Millo **et al.**, "Panel data econometrics in r: The plm package," **Journal of Statistical Software**, vol. 27, no. 2, pp. 1–43, 2008.

[34] S. A. Tagliamonte and A. D'Arcy, "Frequency and variation in the community grammar: Tracking a new change through the generations," **Language Variation and Change**, vol. 19, no. 02, pp. 199–217, 2007.

[35] M. Duggan, N. Ellison, C. Lampe, A. Lenhart, and M. Madden, "Social media update 2014. pew research center, 2015," 2015.

[36] J. Cheng, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Antisocial Behavior in Online Discussion Communities," in **Proceedings of ICWSM**, 2015.