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

Ian Stewart, Stevie Chancellor, Munmun De Choudhury and Jacob Eisenstein

SocialNLP 2017

Georgia Tech

Language variation online

- Online language is subject to profound variation and rapid **change** over time. (Androutsopoulos 2011)
- An individual's adoption of language change in a particular online community is related to their community **membership.** (Sebba 2009)

Orthographic change

- Prior studies in online language have focused mainly on **lexical** change. (Danescu-Niculescu-Mizil et al. 2013)
- Change at the level of **orthography** is also important but less well understood. (Herring 2012)

anorexia ⇨ anarexia ⇨ anarexyia

Research questions

RQ1: **Who** adopts new orthographic variants?

RQ2: Does a variant's **depth** influence its likelihood of **adoption** by these community members?

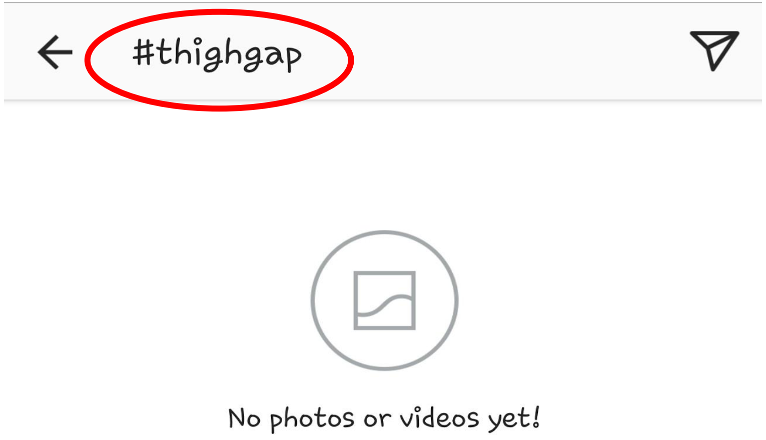
RQ3: Does a variant's **depth** influence its **social reception**?

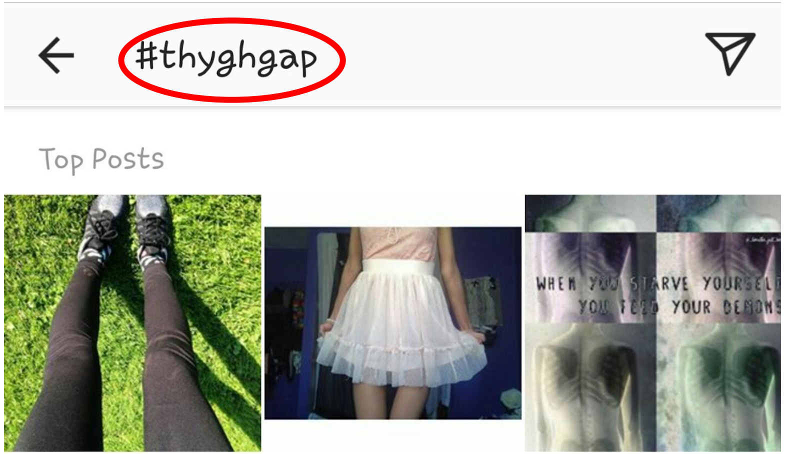
Warning: eating disorders

Pro-ED Instagram

- Community that “share[s] content, advice and provide[s] social support for **disordered or unusual eating choices.**” (Chancellor et al. 2016)
- Community of practice: “aggregate of people who come together around **mutual engagement in an endeavor.**” (Eckert & McConnell-Ginet 1992)

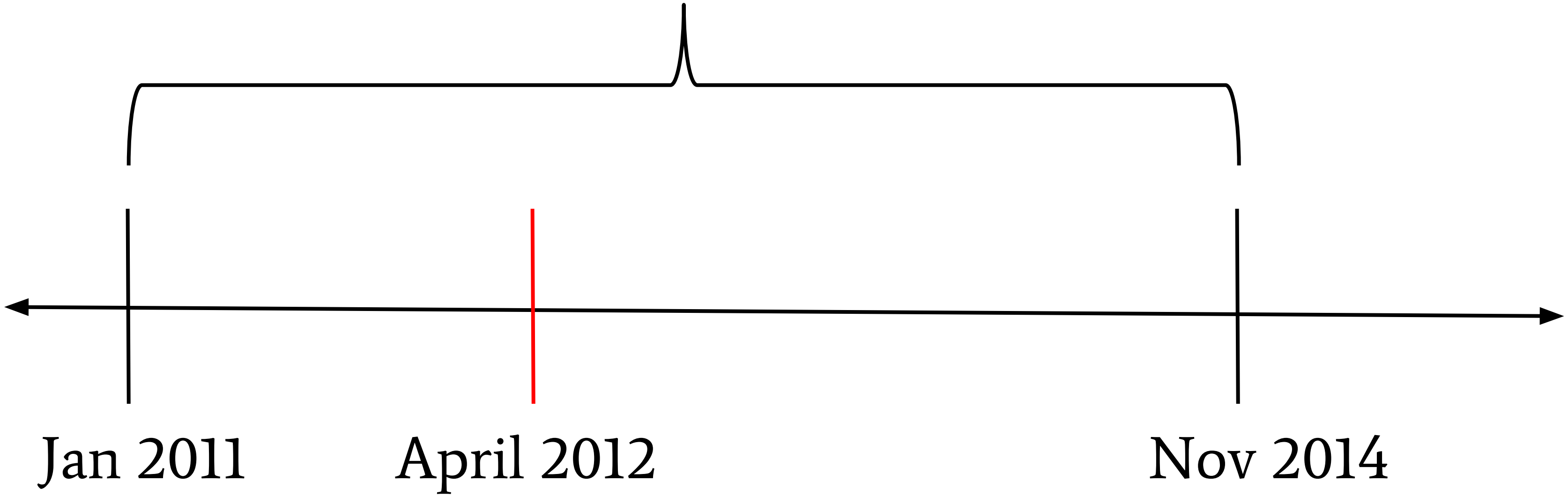
Content ban


#thighgap
No photos or videos yet!

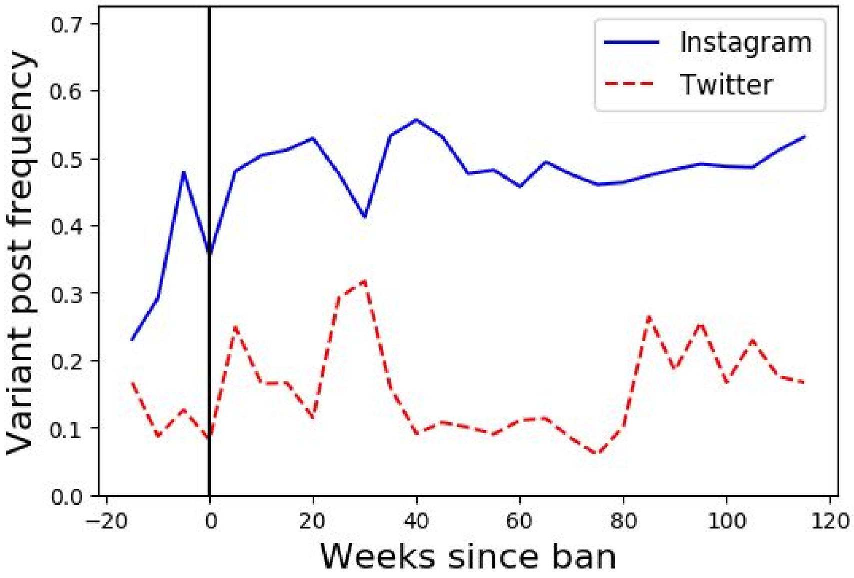

#thyghgap
Top Posts
WHEN YOU STARVE YOURSELF
YOU FEED YOUR DEMONS

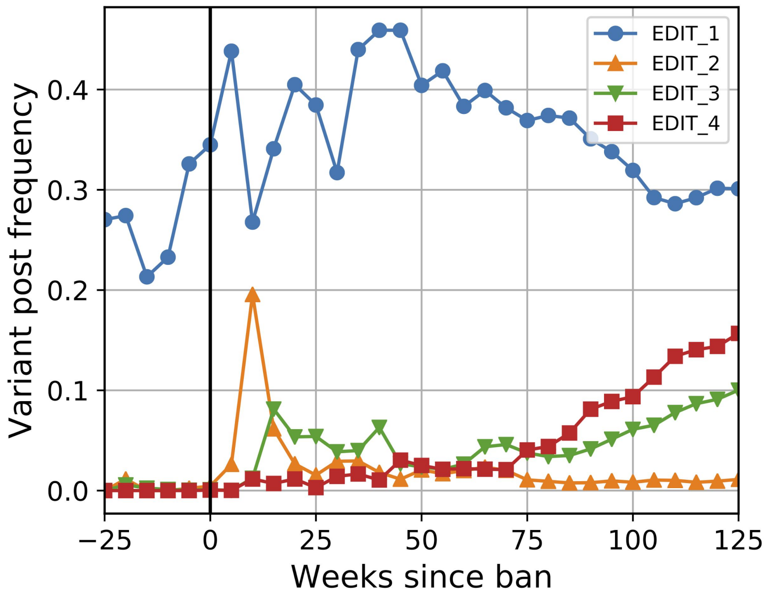
Pro-ED posts (Chancellor et al. 2016)

2.4 million pro-ED posts
17 **source** hashtags
673 **variant** hashtags



Variants grow more frequent, “deeper”





Who drives this change?

RQ1: Who adopts new orthographic variants?

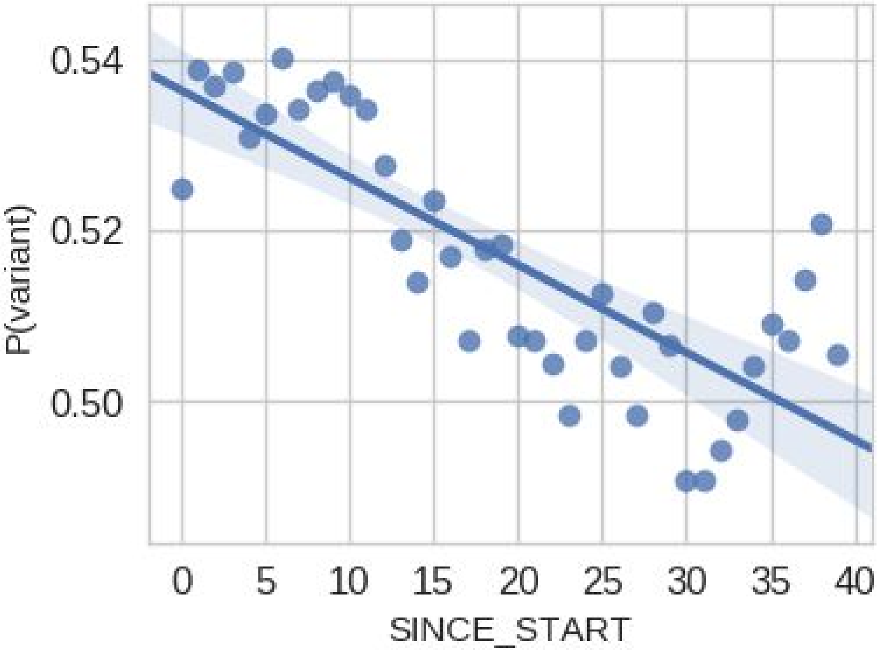
Differentiating community members

- Membership attributes:
  - Age
  - Lifespan

RQ1: Who adopts new orthographic variants?

- **Newcomers** and **committed** (long-lifespan) community members.

RQ1: Who adopts new orthographic variants?



RQ1: Who adopts new orthographic variants?

- Regression results
  - Predicting appearance of any variant in a post.

RQ1: Who adopts new orthographic variants?

- Regression results
  - Age: $\beta = -0.00456^{***}$, effect size = **-0.348**
  - Lifespan: $\beta = 0.00294^{***}$, effect size = **0.654**

  *** = $p < 0.001$

RQ1: Who adopts new orthographic variants?

- Conclusion: variants adopted more often by **newcomers** and **committed** members.

Research questions

RQ1: Which community members adopt more variants?

RQ2: Does a variant's depth influence its likelihood of adoption by these community members?

RQ3: Does a variant's depth influence the post's social reception?

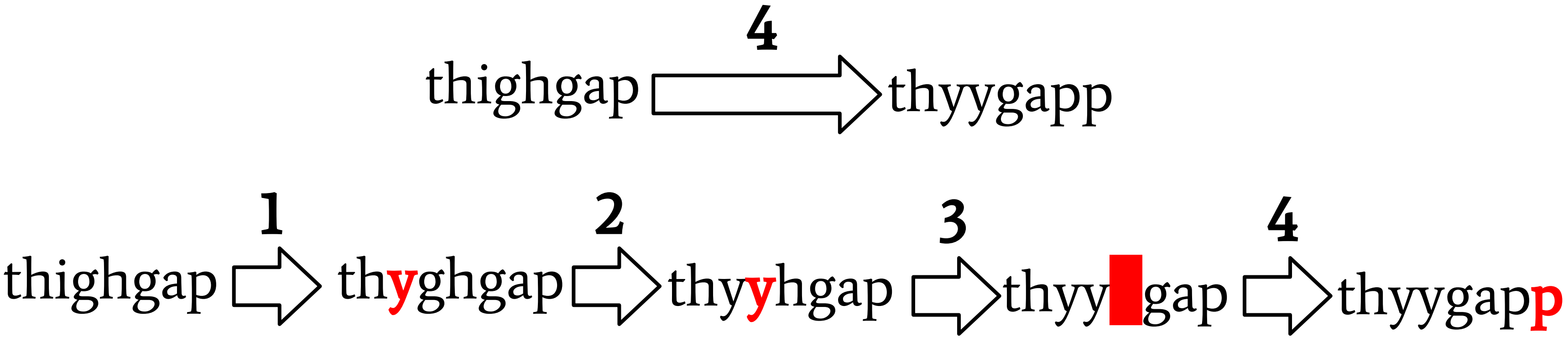
Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)

thighgap ⇒ thyygapp

Compute depth: edit distance

- Operations needed to transform source → variant hashtag
  - Used in dialectology (Nerbonne, Heeringa & Kleiweg 1999)



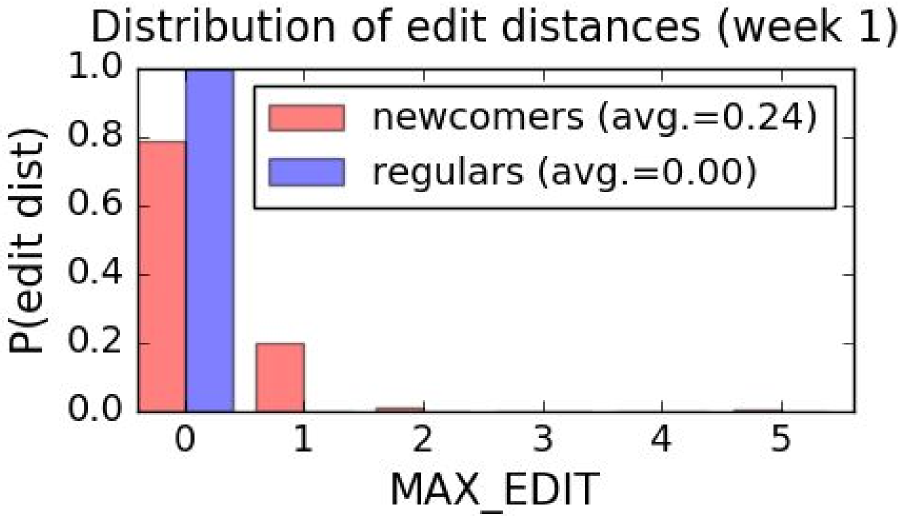
RQ2: Does a variant’s depth influence its likelihood of adoption?

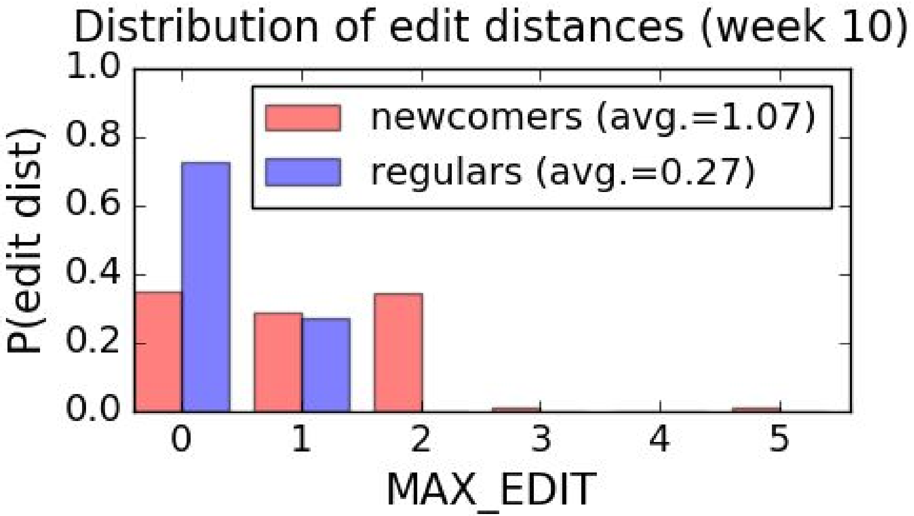
RQ2: Does a variant's depth influence its likelihood of adoption?

- Deeper variants are adopted by **newcomers** and **committed** members.

RQ2: Does a variant's depth influence its likelihood of adoption?

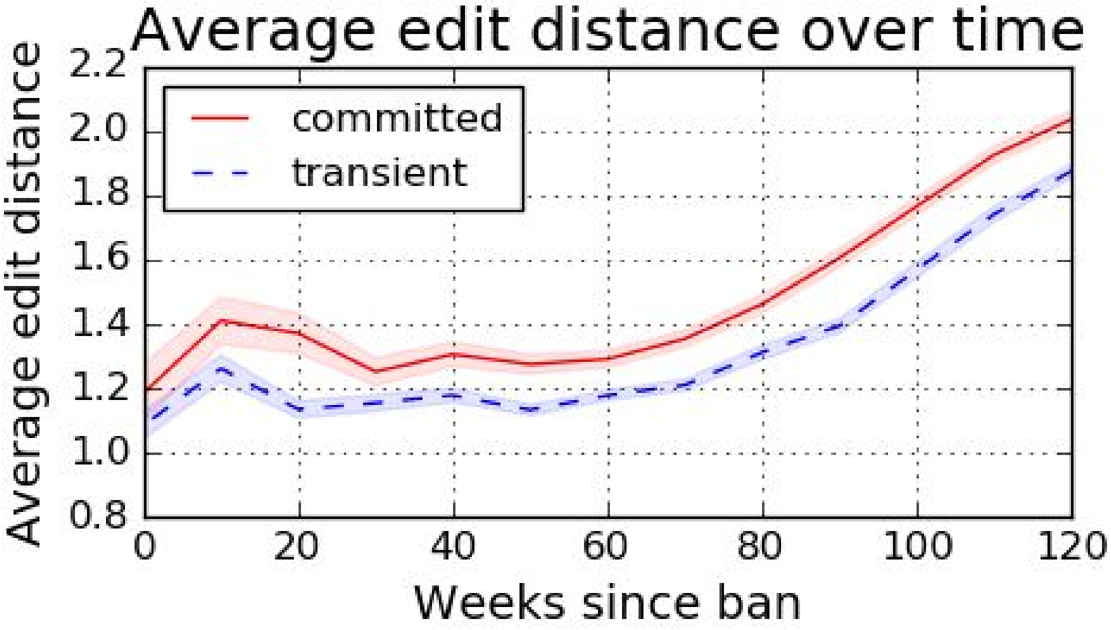
Newcomers > regulars





RQ2: Does a variant's depth influence its likelihood of adoption?

Committed > transient



RQ2: Does a variant's depth influence its likelihood of adoption?

- Regression
  - Predicting appearance of **shallow** (distance 1) and **deep** (distance 4) variant.

RQ2: Does a variant's depth influence its likelihood of adoption?

- **More extreme effects** for deeper variants.
- Distance 1
    - Age: $\beta = -0.00177^{***}$, effect size = **-0.097**
- Distance 4
    - Age: $\beta = -0.00450^{***}$, effect size = **-0.416**

RQ2: Does a variant's depth influence its likelihood of adoption?

- Conclusion: **deeper variants** are more often adopted by newcomers and committed members.

RQ3: Does a variant's depth influence the post's social reception?

- Social reception: **likes** and **comments**.

RQ3: Does a variant's depth influence the post's social reception?

- Regression results
  - Control for hashtags used, presence of any variant, and fixed-effect for each member.

RQ3: Does a variant's depth influence the post's social reception?

COMMENTS	LIKES
Edit distance: -3.72E-3	Edit distance: 0.0130***

*** = $p < 0.001$,
otherwise $p > 0.05$

RQ3: Does a variant's depth influence the post's social reception?

- Conclusion: greater orthographic depth implies **more likes**.

Implications

- Why do **newcomers** use more variants and deeper variation?
  - (1) Newcomers are **over-compensating** for perceived orthographic conventions.
  - (2) “Newcomers” are banned accounts whose behavior is worsening after returning from ban (Cheng, Danescu-Niculescu-Mizil, Leskovec 2014).
  - (3) Newcomers are generating new variants to **distinguish** from current community members (“flag-planting”).

Implications

- Why do newcomers use more variants and deeper variation?
  - (1) Newcomers are over-compensating for perceived orthographic conventions.
  - (2) “Newcomers” are banned accounts whose behavior is worsening after returning from ban (Cheng, Danescu-Niculescu-Mizil, Leskovec 2014).
  - (3) Newcomers are generating new variants to distinguish from current community members (“flag-planting”).
- Deeper variants → **stronger effects**.
  - Effect is **incremental**, not just binary.
  - Orthographic variation reveals hidden community dynamics.

References

Androutsopoulos, J. (2011). Language change and digital media: a review of conceptions and evidence. *Standard Languages and Language Standards in a Changing Europe*, 145–159.

Chancellor, S., Pater, J. A., Clear, T., Gilbert, E., & De Choudhury, M. (2016). #thyghgapp : Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. *19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1201–1213.

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2014). How Community Feedback Shapes User Behavior. In *ICWSM* (pp. 41--50).

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., & Potts, C. (2013). No Country for Old Members : User Lifecycle and Linguistic Change in Online Communities. *Proceedings of the 22nd International Conference on World Wide Web*, 307–317.

Eckert, P., & McConnell-Ginet, S. (1992). Think Practically and Look Locally: Language and Gender as Community-Based Practice. *Annual Review of Anthropology*, *21*(1992), 461–490.

Herring, S. C. (2012). Grammar and electronic communication. *The Encyclopedia of Applied Linguistics*, 1–9.

Nerbonne, J., Heeringa, W. and Kleiweg, P. (1999). "Edit distance and dialect proximity". In *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, 2nd edition, Edited by: Sankoff, D. and Kruskal, J. v–xv.

Sebba, M. (2009). Sociolinguistic approaches to writing systems research. *Writing Systems Research*, *1*(1), 35–49.

Acknowledgments

Thanks to Brendan O’Connor for the Twitter data.

This research was supported by Air Force Office of Scientific Research award FA9550-14-1-0379, National Institutes of Health award R01-GM112697, and National Science Foundation award 1452443.

Questions?

istewart6@gatech.edu

@alethioguy

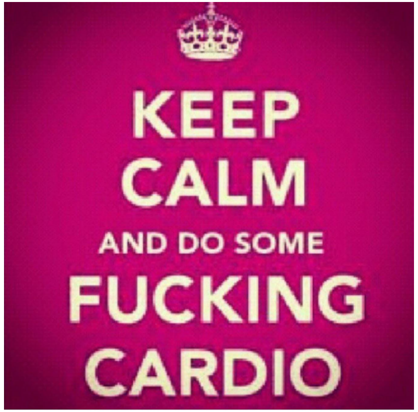
Backup slides

Pro-ED Instagram

♥ 7 likes

[redacted] Todays total : 906 calories 😟 today was horrible 😳😟 #Ana #anorexia #staystrong #fat #ugly #disgusting #nothappy #sad #strong #ed #eatingdisorder #girl #cutting #losingweight #weight #wishweight #clean #horribleday #horrible #selfharm

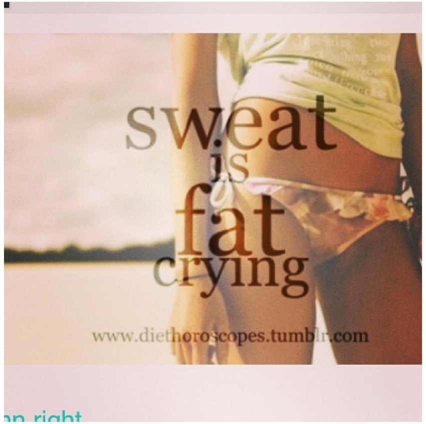
[redacted] Great



♥ 6 likes

[redacted] Feeling a little better..if i want to reach my goal i have to be patient and work harder. I hope you are doing the same :) #cardio #excercise #loseweight #workout #skinny #thin #thighggap #collarbones

[redacted] #thighgap



♥ 22 likes

[redacted] #ana #mia #ed #bones #bonsepo #fitspo #thygap #thynsperation

♥ 11 likes

[redacted] #dinner today was some whole grain spelt bread with sunflower seeds (yummy) topped with humous, tomatoes, dried tomato spread and some cucumber... I also had some leftover sauerkraut ❤ #ed #anorexia #bulimia #vegan #dinner #veganrecovery #thisorhospital #edfamily #anawho #2fab4ana #recover #edfighter #edwarriors #edwarrior #edsoldier #edfree #eatingdisorderrecovery

Pro-ED Instagram

[redacted] Feeling a little better..if i want to reach my goal i have to be patient and work harder. I hope you are doing the same :) #cardio #excercise #loseweight #workout #skinny #thin #thighggap #collarbones

[redacted] #thighgap

Pro-ED Instagram

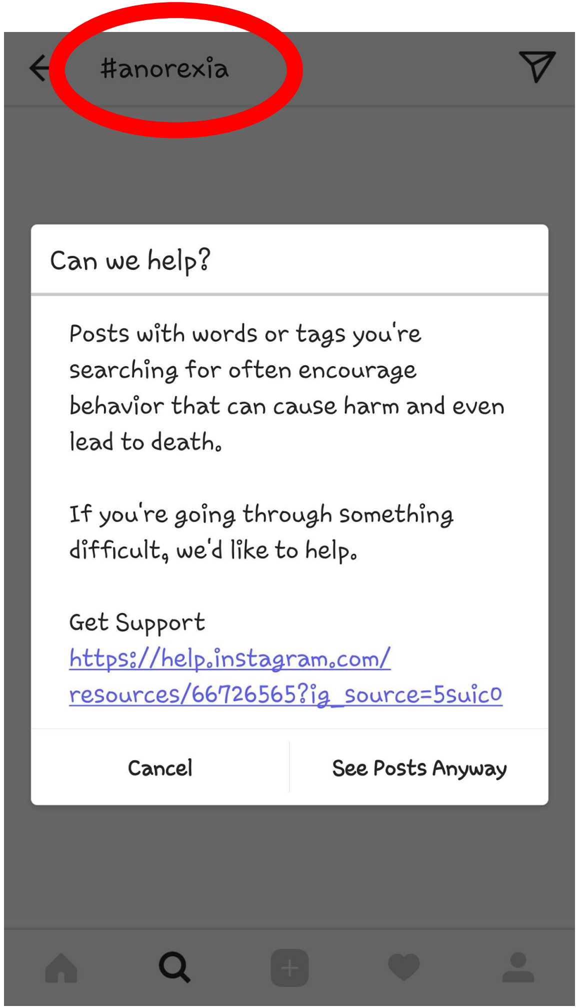
[redacted] Feeling a little better..if i want to reach my goal i have to be patient and work harder. I hope you are doing the same :) #cardio #excercise #loseweight #workout #skinny #thin #thighggap #collarbones

[redacted] #thighgap

thighgap

(space between thighs)

Ban effect


#anorexia
Can we help?
Posts with words or tags you're searching for often encourage behavior that can cause harm and even lead to death.
If you're going through something difficult, we'd like to help.
Get Support
https://help.instagram.com/resources/66726565?ig_source=5suic0
Cancel
See Posts Anyway

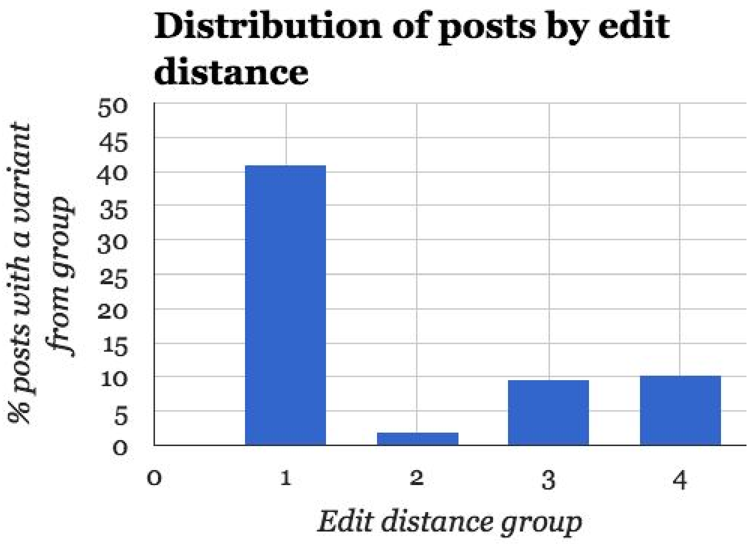
All source hashtags

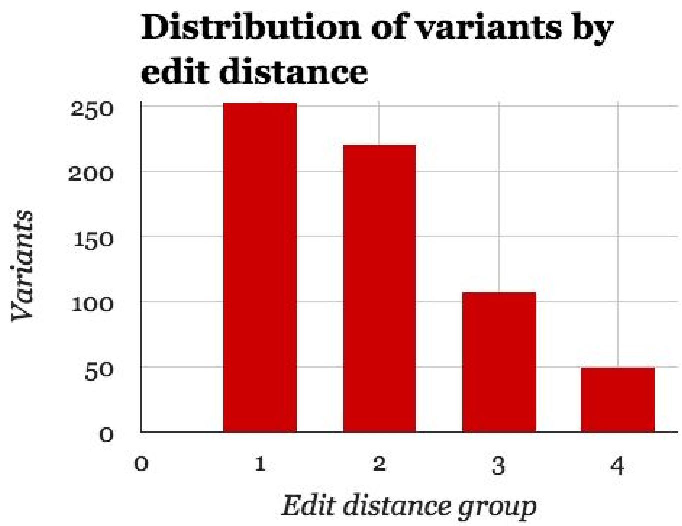
- ana
- anorexia
- anorexianervosa
- bonespo
- bulimia
- eatingdisorder
- mia
- proana
- proanorexia
- probulimia
- promia
- secretsociety123
- skinny
- thighgap
- thin
- thinspiration
- thinspo

Distribution of variants

Edit distance	Top 3 variants	Source hashtags	Variants	% posts with at least one variant from 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%

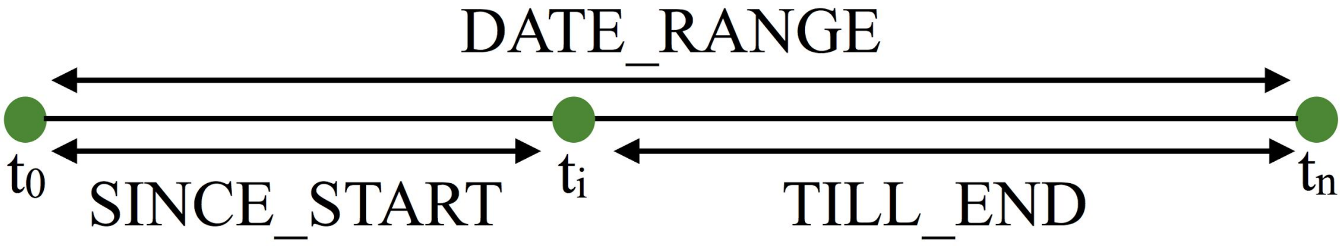
Distribution of variants



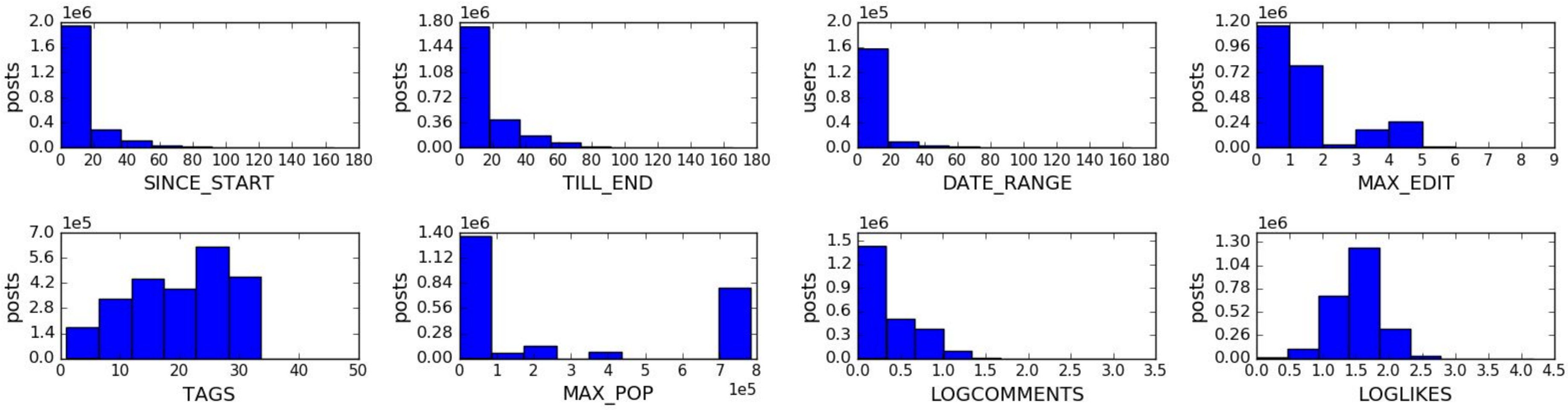


Differentiating community members

- Per post:
  - SINCE_START, TILL_END
- Per member:
  - DATE_RANGE



Variable distributions



Regression: predictors

	RQ1	RQ2	RQ3
Regression	Logistic	Logistic	Poisson
Predicted	VARIANT	DIST_1, DIST_4	COMMENTS, LIKES
Predictors	SINCE_START, TILL_END, DATE_RANGE, DATE	SINCE_START, DATE_RANGE, DATE	SINCE_START, DATE_RANGE, MAX_EDIT, VARIANT, MAX_POP, fixed effect for member

RQ1: Who adopts new orthographic variants?

- Regression results

	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

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

*** = $p < 0.001$

RQ3: Does a variant's depth influence the post's social reception?

COMMENTS

	β	SE
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	-3.72E-3	5.51E-3

LIKES

	β	SE
SINCE_START	-0.0319***	9.03E-4
TAGS	0.224***	1.47E-3
VARIANT	-0.00149***	1.98E-3
MAX_POP	-3.89E-3***	7.25E-4
MAX_EDIT	**0.0130*** **	3.16E-3

*** = $p < 0.001$,
* = $p < 0.05$,
otherwise $p > 0.05$