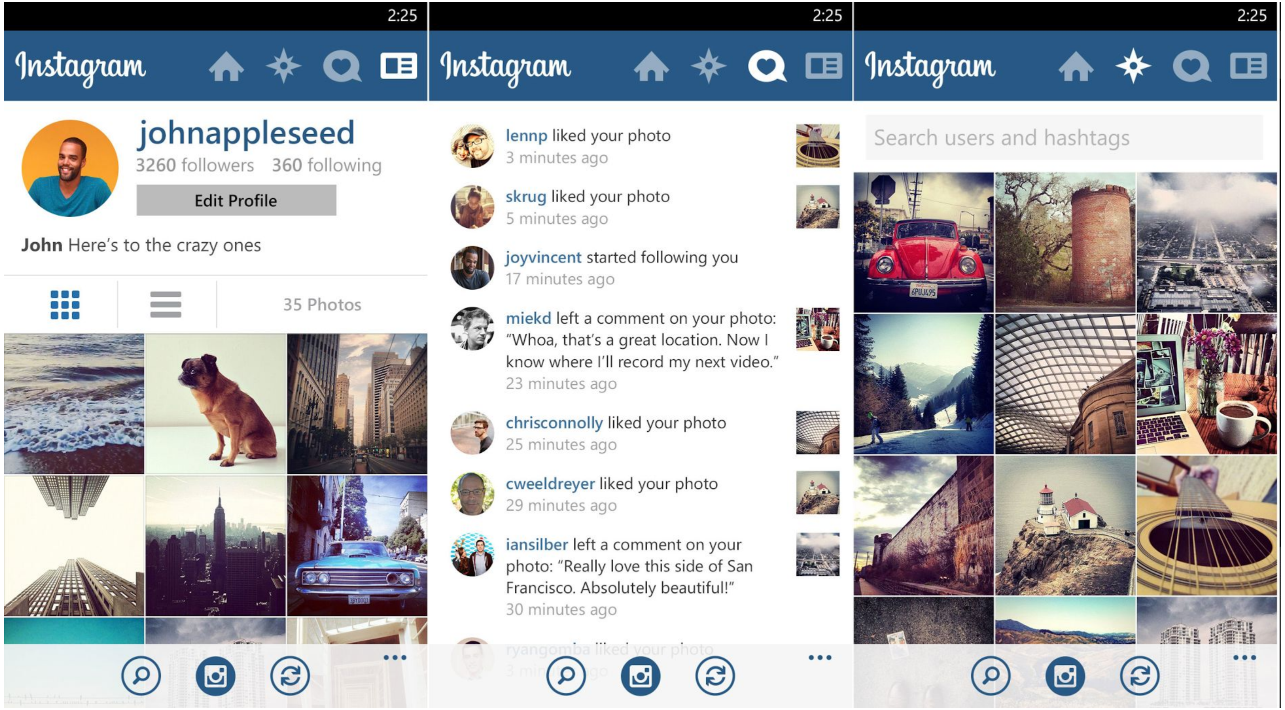
#thighgap to #thyghgapp: Incrementation of orthographic variation on Instagram

Ian Stewart and Jacob Eisenstein
Georgia Institute of Technology

Content warning: eating disorders

Instagram



http://aworldofapps.com/wp-content/uploads/2015/07/Profile-how-to-use-Instagram.jpg

Pro-ED Instagram

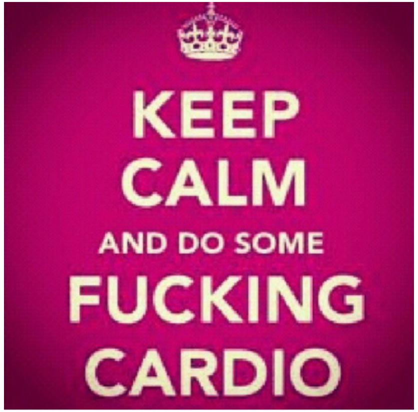
- Community that “share[s] content, advice and provide[s] social support for disordered or unusual eating choices” (Chancellor et al. 2015)

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

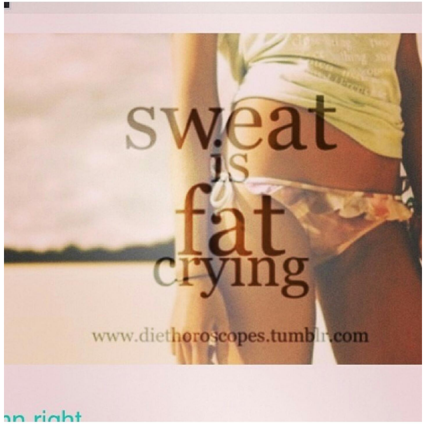
[redacted] Great



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)

Orthographic variation

- “Represent spoken and vernacular forms, simulate prosody or shorten the message” (Androutsopoulos 2011)
- Tied to social differentiation, identity marking (Sebba 2009)
  - Community may only allow certain variants (Herring 2012)

Phonetic	<just> → <jus>
Typographic	<leet speak> → <1337 5934K>
Syllabograms	<before> → <b4>

Dynamics of variation

- Writing conventions evolve over time (Sebba 2009)
- Communities are dynamic
  - Language change mirrors social dynamics (Danescu-Nicelscu-Mizil et al. 2013)
  - Locally-defined social categories: newcomers vs. regulars
- Changing practices of pro-ED community
  - Community of practice: “aggregate of people who come together around mutual engagement in an endeavor” (Eckert & McConnell-Ginet 1992)

Community change: hashtag ban

SOCIETY

Instagram Bans Thinspo Content

Instagram is the latest social media platform to ban thinspiration content. But are these policies effective?

By Heba Hasan @Heba__H | April 26, 2012

Share | Like 57 | Tweet | G+1 18 | Share 1 | Pin it | Read Later

Thinspo content will no longer be welcome on Instagram. Following in the footsteps of Pinterest and Tumblr, Instagram is the latest social media site to ban "thinspiration" photos — images that are meant to provide motivation for those who want to lose weight and which health experts say often contribute to eating disorders.

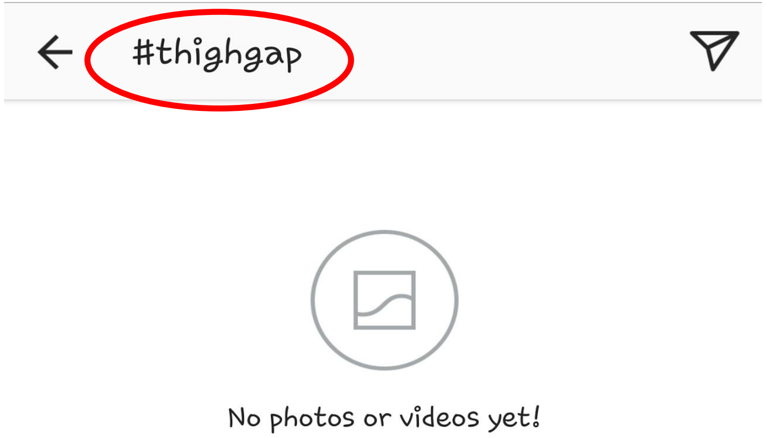
Instagram's new policy doesn't come as a surprise. The app came under scrutiny last week when celebrity and Instagram user Alexa Chung posted a photo of herself and was attacked by users for being too skinny.

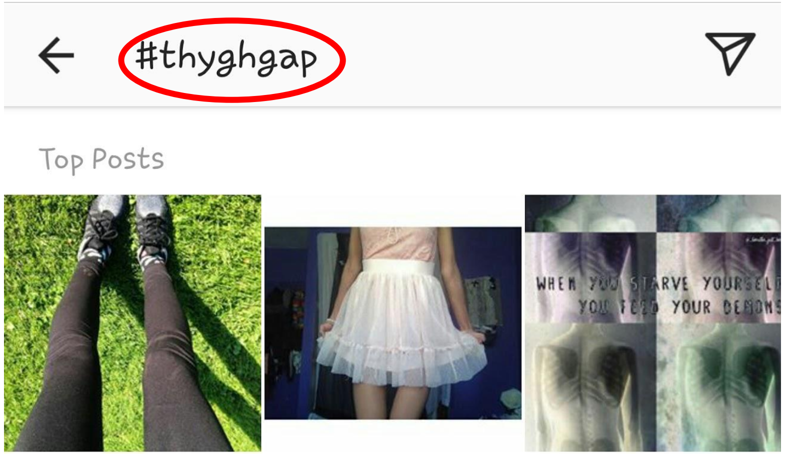
Michaela Begsteiger/Getty

Instagram bans thinspo content

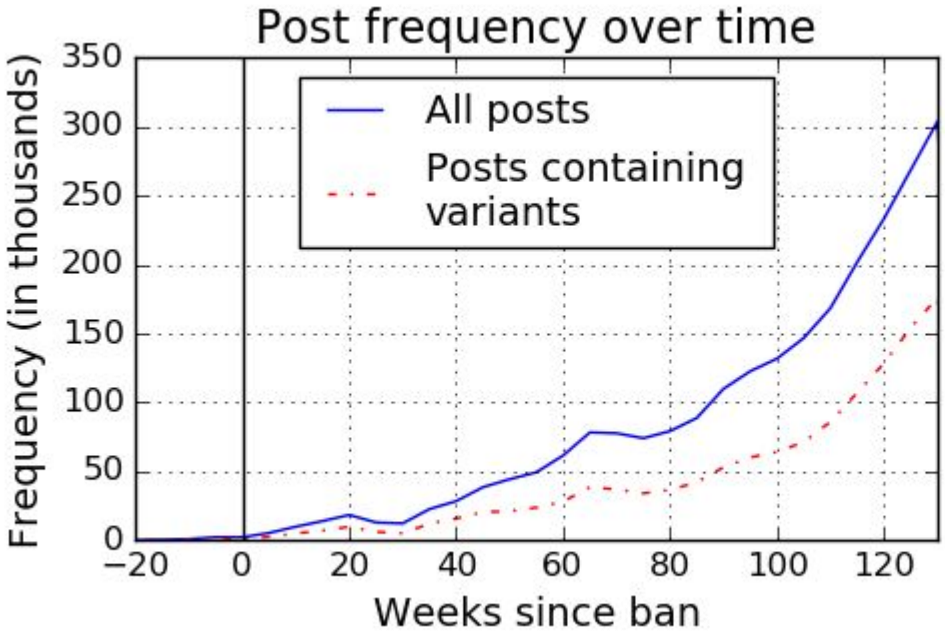
newsfeed.time.com/2012/04/26/instagram-bans-thinspo-content

Ban effect

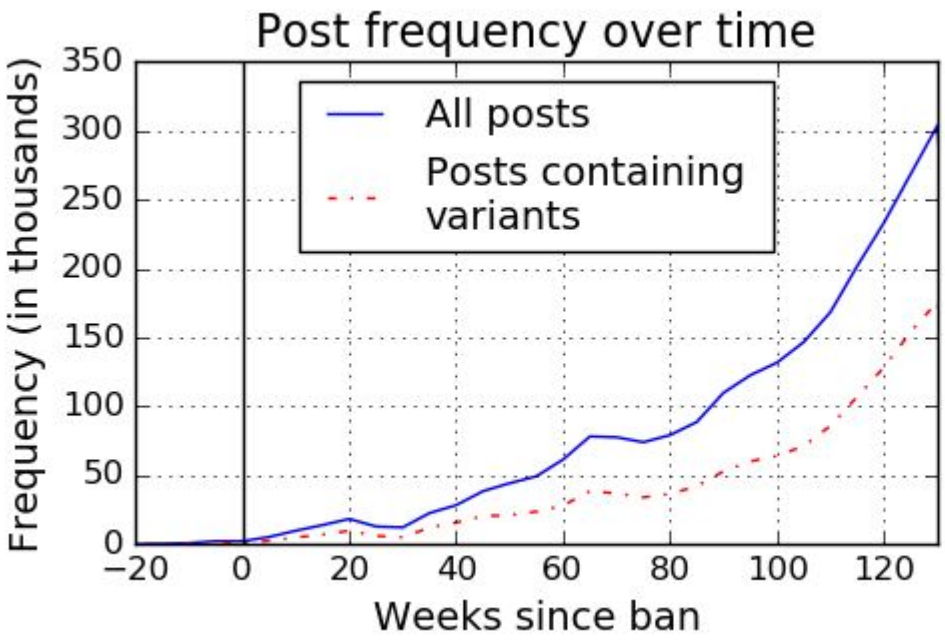


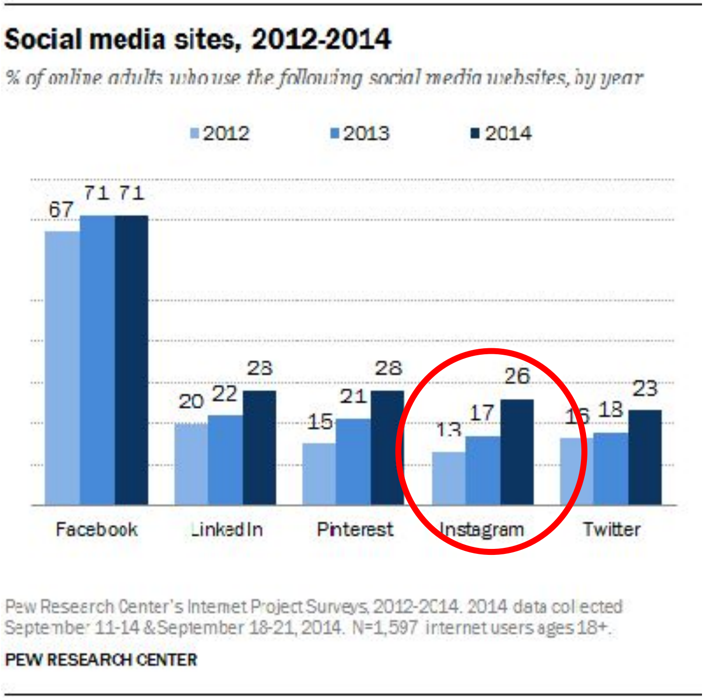


Ban response



Ban response





http://www.socialmediaexaminer.com/wp-content/uploads/2015/03/sd-pew-adult-user-stats.png

Research questions

RQ1: Which community members adopt more variants?

Example variants

thighgap

Example variants

	thyghgap	thyghgapp	thyygap	thyygapp
thighgap	thghgap	thiigap	thighgaappp	thygaps
	thightgap	thightgrap	thightpag	thygsp

Example variants

	thyghgap	thyghgapp	thyygap	thyygapp
thighgap	thghgap	thiigap	thighgaappp	thygaps
	thightgap	thightgrap	thightpag	thygsp

Depth

Example variants

	thyghgap	thyghgapp	thyygap	thyygapp
thighgap	thghgap	thiigap	thighgaappp	thygaps
	thightgap	thightgrap	thightpag	thygsp

Depth

Incrementation of variation

- Orthographic variation as continuum
- Similar to phonetic incrementation
  - “Successive cohorts and generations of children advance a change beyond the level of their caretakers and role models” (Labov 2001)
- Do community members adopt deeper variants differently than shallow ones?

Research questions

RQ1: Which community members adopt more variants?

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

Methods

- Data collection
- Compute orthographic depth (language variables)
- Compute membership attributes (community variables)
- Building regression models

Methods

- **Data collection**
- Compute orthographic depth (language variables)
- Compute membership attributes (community variables)
- Building regression models

Data collection (Chancellor et al. 2015)

- Collected in November 2014
  - Ban in April 2012
- 2.4 million posts
  - January 2011 to November 2014

Data collection (Chancellor et al. 2015)

Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram
anorexia, ed, bulimia

Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Manually identify 17 banned source hashtags

ana, thighgap, thinspo

Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Manually identify 17 banned source hashtags

ana, thighgap, thinspo

Extract 673 variants with regular expressions

th*nspo* => thynspoo

Data collection (Chancellor et al. 2015)

Identify pro-ED seed terms (not banned), mine Instagram

anorexia, ed, bulimia

Filter for pro-ED content, identify top 200 hashtags

Manually identify 17 banned source hashtags

ana, thighgap, thinspo

Extract 673 variants with regular expressions

th*nspo* => thynspoo

2.4 million posts
176,000 users
51% variant posts
673 variants
17 sources

Methods

- Data collection
- **Compute orthographic depth (language variables)**
- Compute membership attributes (community variables)
- Building regression models

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)

thighgap ⇒ thyygapp

thighgap

Compute depth: edit distance

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

thighgap $\Rightarrow$ thyygapp

thighgap $\xRightarrow{1}$ th**y**ghgap

Compute depth: edit distance

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

thighgap $\Rightarrow$ thyygapp

thighgap $\xRightarrow{1}$ th**y**ghgap $\xRightarrow{2}$ thy**y**hgap

Compute depth: edit distance

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

thighgap ⇒ thyygapp

thighgap $\xRightarrow{1}$ thyghgap $\xRightarrow{2}$ thyyhgap $\xRightarrow{3}$ thyy gap

Compute depth: edit distance

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

thighgap ⇒ thyygapp

thighgap $\xRightarrow{1}$ thyghgap $\xRightarrow{2}$ thyyhgap $\xRightarrow{3}$ thyy gap $\xRightarrow{4}$ thyygapp

Compute depth: edit distance

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

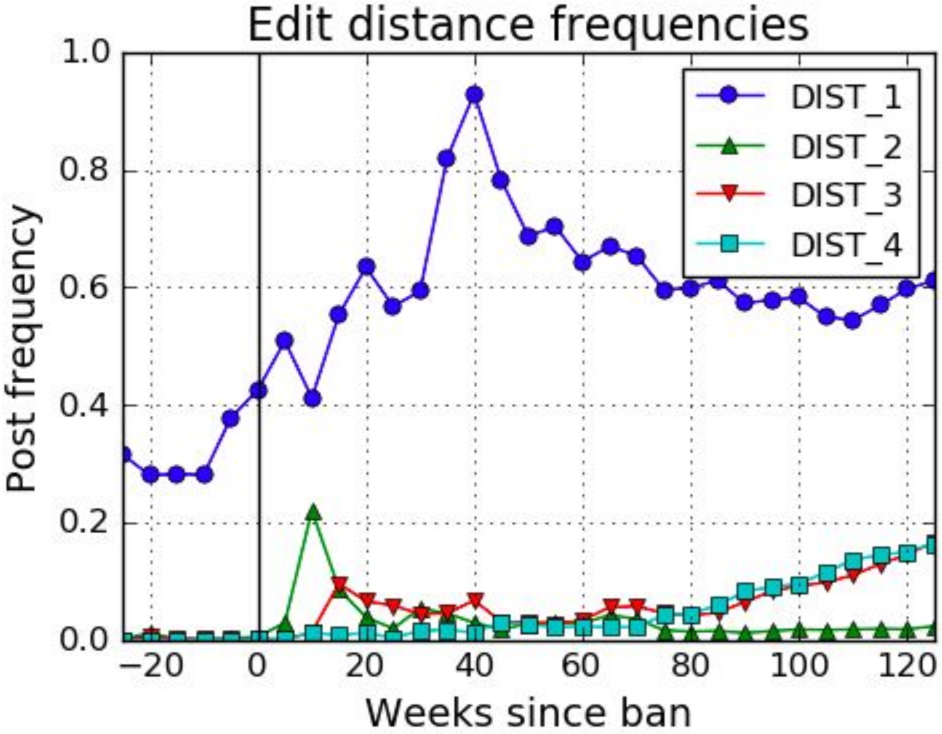
thighgap $\xrightarrow{4}$ thyygapp

thighgap $\xrightarrow{1}$ thyghgap $\xrightarrow{2}$ thyyhgap $\xrightarrow{3}$ thyy gap $\xrightarrow{4}$ thyygapp

Edit distance: Distribution of variants

Edit distance	Variants	Top 3 variants
1	253	anarexia, bulimic, eatingdisorders
2	221	anarexyia, thinspooo, thynspoo
3	108	secretsociety123, thinspoooo, thygap
4	50	secret_society123, secretsociety_123, thinspooooo

Edit distance: Adoption over time

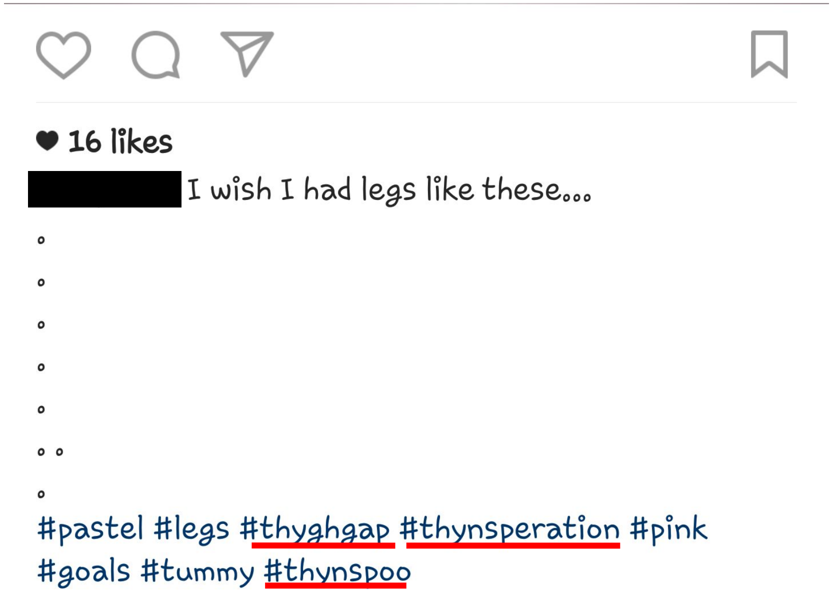


Language variables

- per post:
  - TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4

Language variables

- per post:
  - TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4



Language variables

- per post:
    - TAGS, VARIANT, MAX_EDIT, DIST_1, DIST_4

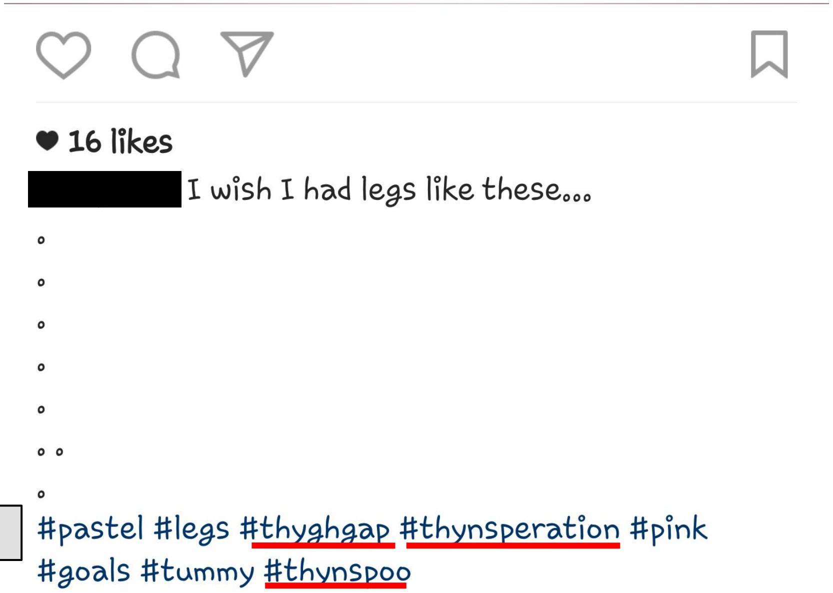
TAGS=3

VARIANT=1

MAX_EDIT=2

DIST_1=1

DIST_4=0



Methods

- Data collection
- **Compute orthographic depth (language variables)**
- **Compute membership attributes (community variables)**
- Building regression models

Community data: membership attributes

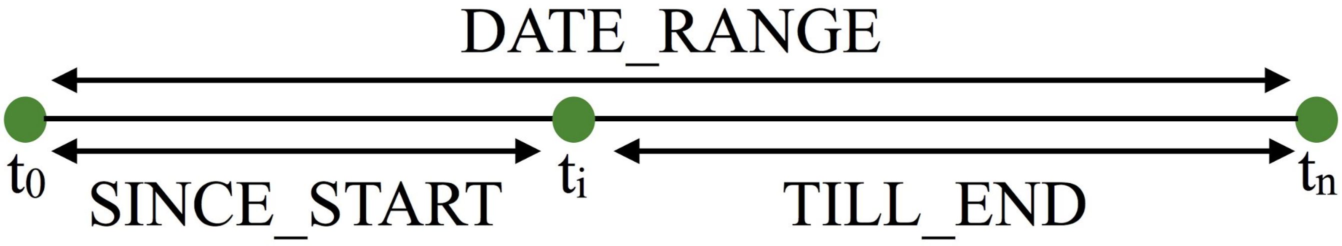
- Locally-defined variables (within pro-ED community):
  - relative age
  - lifespan

Community data: membership attributes

- per post:
  - SINCE_START, TILL_END
- per user:
  - DATE_RANGE

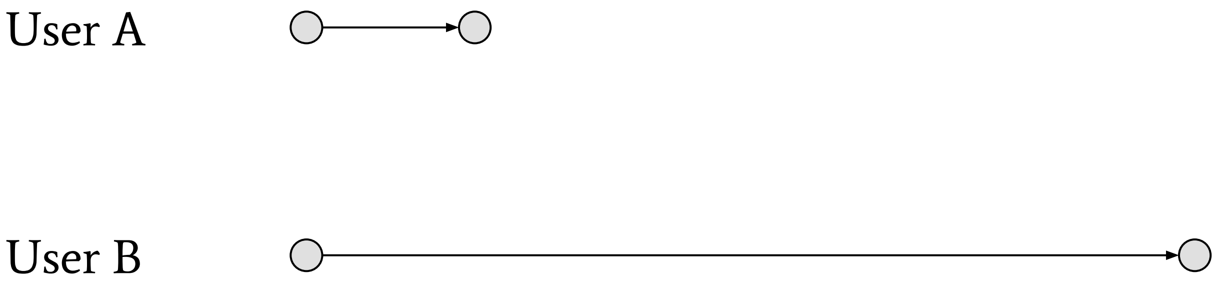
Community data: membership attributes

- per post:
    - SINCE_START, TILL_END
- per user:
    - DATE_RANGE



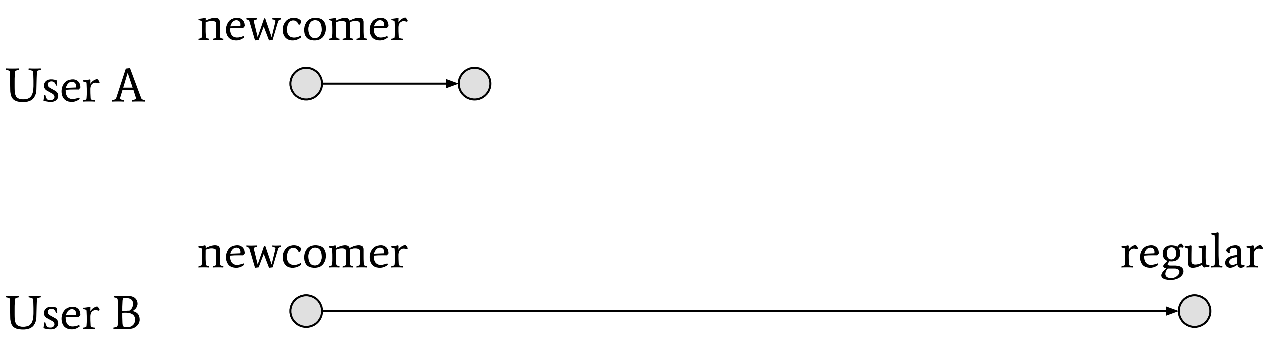
Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)



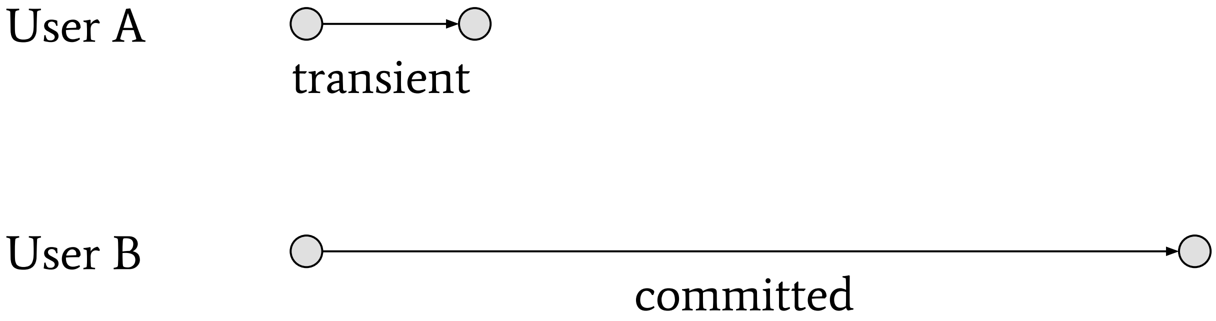
Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)



Community data: membership attributes

- newcomer = low SINCE_START (< 10 weeks)
- committed user = high DATE_RANGE (≥ 10 weeks)



Recap: all variables

- per post:
  - **VARIANT**, **DIST_1**, DIST_4, **MAX_EDIT**, TAGS, SINCE_START, TILL_END, DATE
- per user:
  - DATE_RANGE

Methods

- Data collection
- **Compute orthographic depth (language variables)**
- Compute membership attributes (community variables)
- **Building regression models**

Questions

RQ1: Which community members adopt variants?

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

Regression: predictors

	RQ1	RQ2
Regression	Logistic	Logistic
Predicted	VARIANT	DIST_1 DIST_4
Predictors	**SINCE_START** **TILL_END** **DATE_RANGE**	**SINCE_START** **TILL_END** **DATE_RANGE**

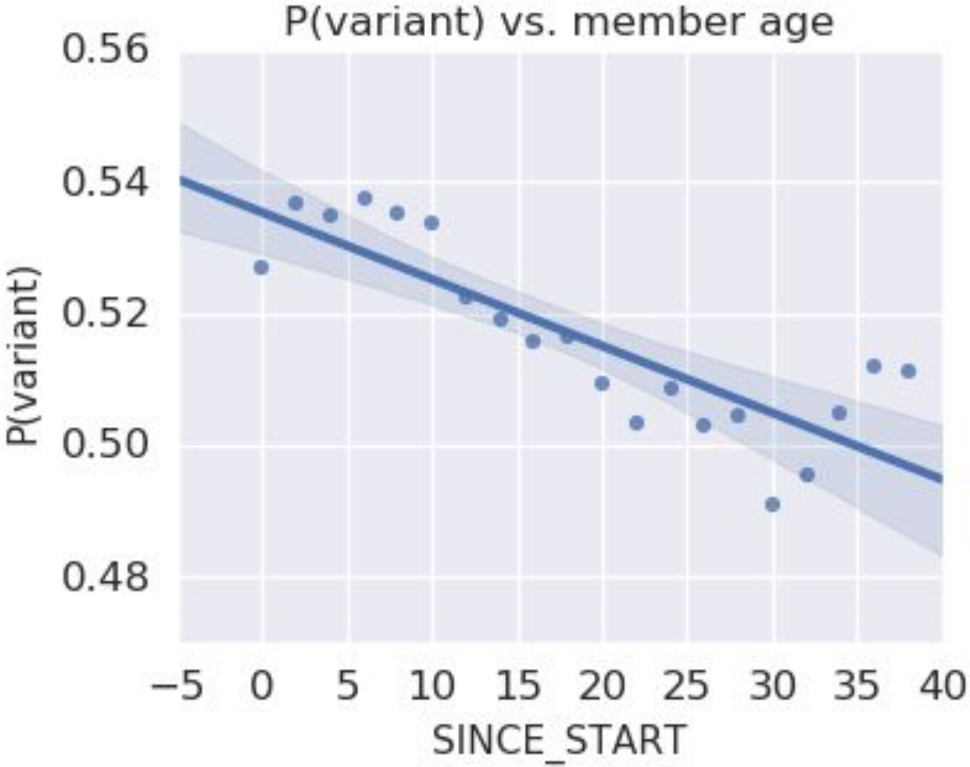
Results

RQ1: Which community members adopt variants?

RQ1: Which community members adopt variants?

- Newcomers and committed (long-lifespan) users

RQ1: Which community members adopt variants?



RQ1: Which community members adopt variants?

- Regression results

RQ1: Which community members adopt variants?

- Regression results
- Predicting VARIANT
    - SINCE_START negatively correlated ($\beta = -0.00456$, $p < 0.001$)
    - TILL_END positively correlated ($\beta = 0.00294$, $p < 0.001$)
    - DATE_RANGE positively correlated ($\beta = 0.00294$, $p<0.001$)

RQ1: Which community members adopt variants?

- Regression results
- Predicting VARIANT
  - SINCE_START negatively correlated ($\beta = -0.00456$, $p < 0.001$)
  - TILL_END positively correlated ($\beta = 0.00294$, $p < 0.001$)
  - DATE_RANGE positively correlated ($\beta = 0.00294$, $p<0.001$)
- Conclusion: variants adopted more often by newcomers and committed members

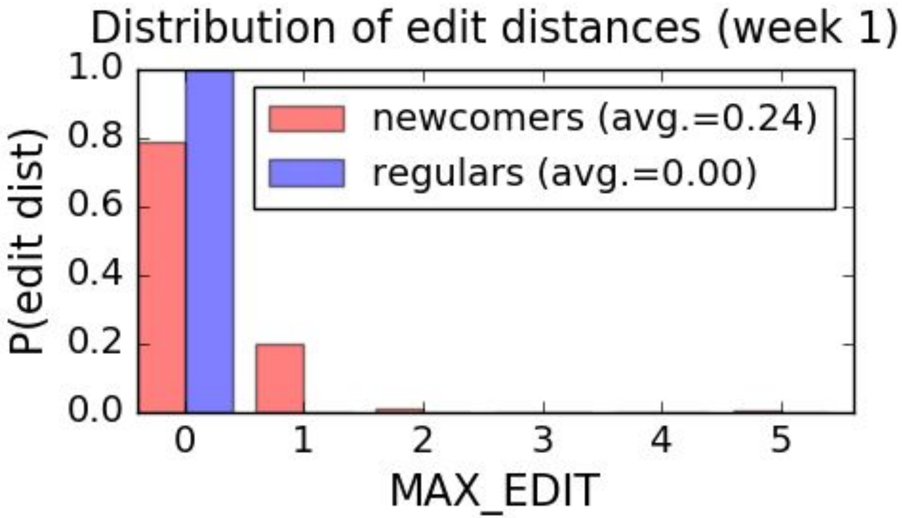
RQ2: Does a variant's depth influence its likelihood of adoption?

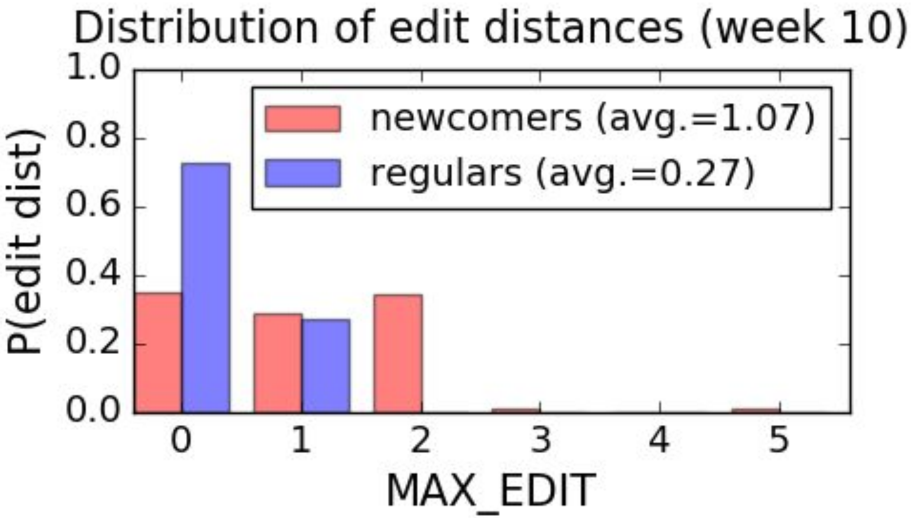
RQ2: Does a variant's depth influence its likelihood of adoption?

- Deeper variants associated with newcomers and committed members

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

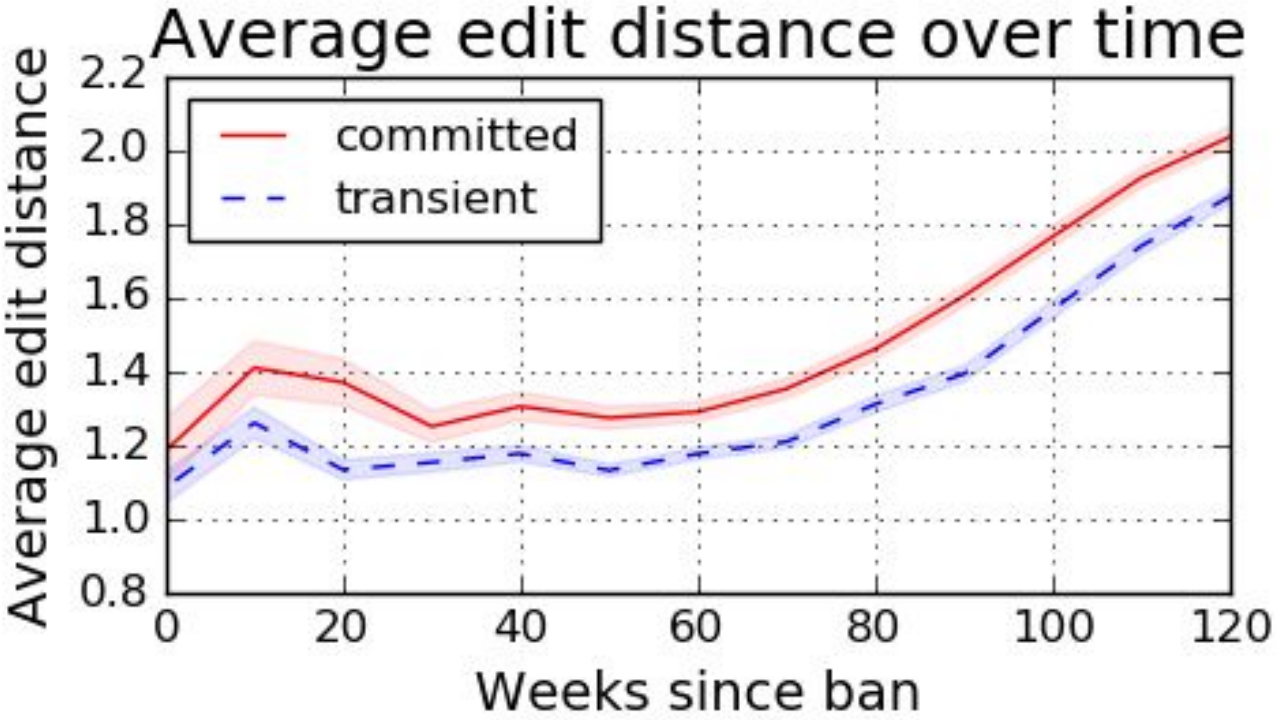
Newcomers versus regulars





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

Committed versus transient



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

- Regression results

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

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, ($p < 0.001$)
  - TILL_END $\beta = 0.00311$ ($p < 0.001$)

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

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, ($p < 0.001$)
  - TILL_END $\beta = 0.00311$ ($p < 0.001$)
- Predicting DIST_4
  - SINCE_START $\beta = -0.00450$ ($p < 0.001$)
  - TILL_END $\beta = 0.0133$ ($p < 0.001$)

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

- Regression results
- Predicting DIST_1
  - SINCE_START $\beta = -0.00177$, ($p < 0.001$)
  - TILL_END $\beta = 0.00311$ ($p < 0.001$)
- Predicting DIST_4
  - SINCE_START $\beta = -0.00450$ ($p < 0.001$)
  - TILL_END $\beta = 0.0133$ ($p < 0.001$)
- Conclusion: depth of variation correlates more strongly with adoption by newcomers and committed members

Summary of findings

- Newcomers use more variants, deeper variation
    - Supports prior findings (Danescu-Niculescu-Mizil et al. 2013)
- Committed members also use more/deeper variants
- Deeper variants → stronger effects
    - Depth may influence orthographic perception in pro-ED community
- Additional: unclear social reception
    - Mixed results (likes ≠ comments)

Implications and future work

- Implications
  - Online communities provide useful setting to study large-scale, long-term language variation
  - Orthographic variation as incrementation
  - Sociotechnical effect on language variation
- Future work
  - Different processes of orthographic variation: deletion, lengthening, metathesis

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. (2015). #thyghgapp : Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. *19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1201–1213.

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.

Eisenstein, J. (2015). Systematic patterning in phonologically‐motivated orthographic variation. *Journal of Sociolinguistics*, *19*(2), 161-188.

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

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social Factors*. Blackwell Publishers.

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.

Thank you!

Special thanks: Stevie Chancellor and Munmun De Choudhury (Georgia Institute of Technology)

Questions?