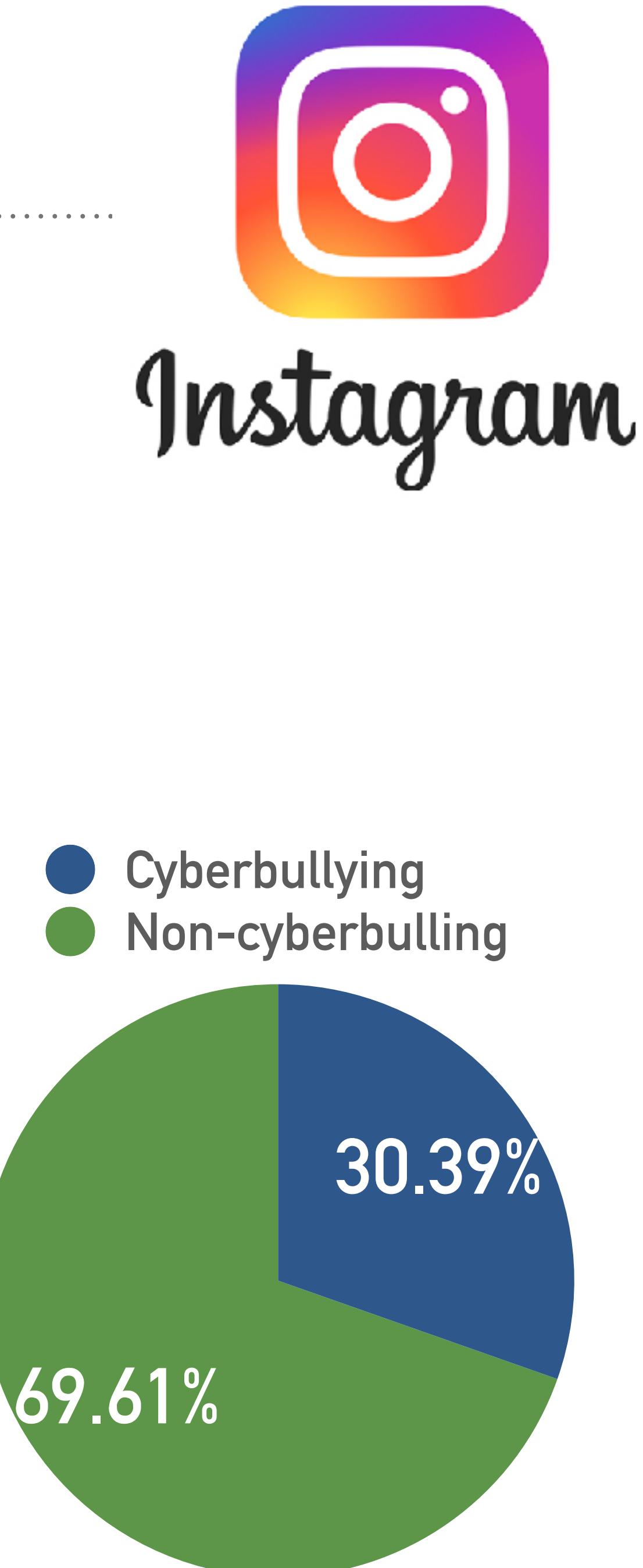
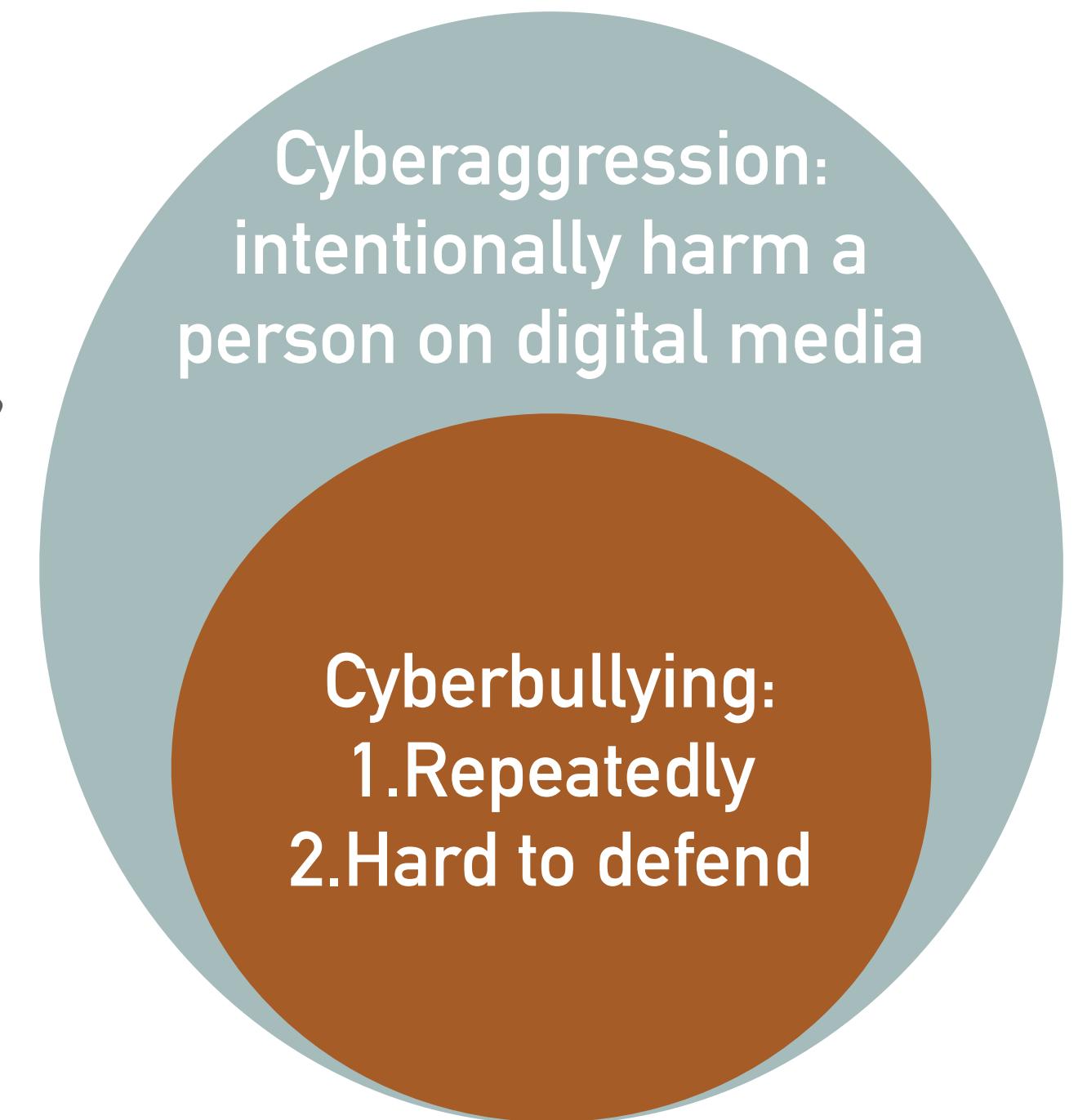


CYBERBULLYING IN INSTAGRAM

Reham Al Tamime, Qinyu E, Xindi Wang, Sijia Yang

INTRODUCTION

- Cyberbullying vs Cyberaggression
- Data
 - 2,218 Instagram images with comments,
 - with labels of
 - Image content (Person, dog, etc.)
 - Whether there is cyberbullying/ cyberaggression considering all comments



[1] HosseiniMardi, Homa, et al. "Detection of cyberbullying incidents on the instagram social network." arXiv preprint arXiv:1503.03909 (2015).

[2] HosseiniMardi, Homa, et al. "Prediction of cyberbullying incidents in a media-based social network." Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Press, 2016.

RQ1: IMAGE LABEL TO DETECT CYBERBULLYING?



IMAGE LABEL TO DETECT CYBER BULLYING?

- Supervised learning predicting cyberbullying using image label:
- “Certain image contents such as *drug* are strongly related with cyberbullying, while some other image contents such as *bike*, *food*, etc. have a very low relationship with cyberbullying.”(HosseiniMardi et al. 2015)
- Various method: Linear Regression, kNN, SVM, Random Forest, etc., and **Random Forest is the best!**

Random Forest Result

Accuracy	0.6
95% CI	(0.5485, 0.65)
Kappa	0.1935
P-Value (Acc>NIR)	0.001104
Mcnemar's Test P-Value	0.017892

IMAGE LABEL TO DETECT CYBER BULLYING?

- Improve with AI object recognition
- DeepMask
- SharpMask
- ...

APP \ MOBILE \ FACEBOOK

Facebook is giving away the software it uses to understand objects in photos

DeepMask and SharpMask are now open source

By Nick Statt | @nickstatt | Aug 25, 2016, 12:00pm EDT

SHARE

MOST READ

- Optimus unveils BlackBerry Evolve devices with massive batteries and displays
- Leaked Galaxy Note 9 video touts 1TB of storage
- Epic confirms it won't be bringing Fortnite for Android to Google Play

Facebook Prineville Data Center | Vjeran Pavic

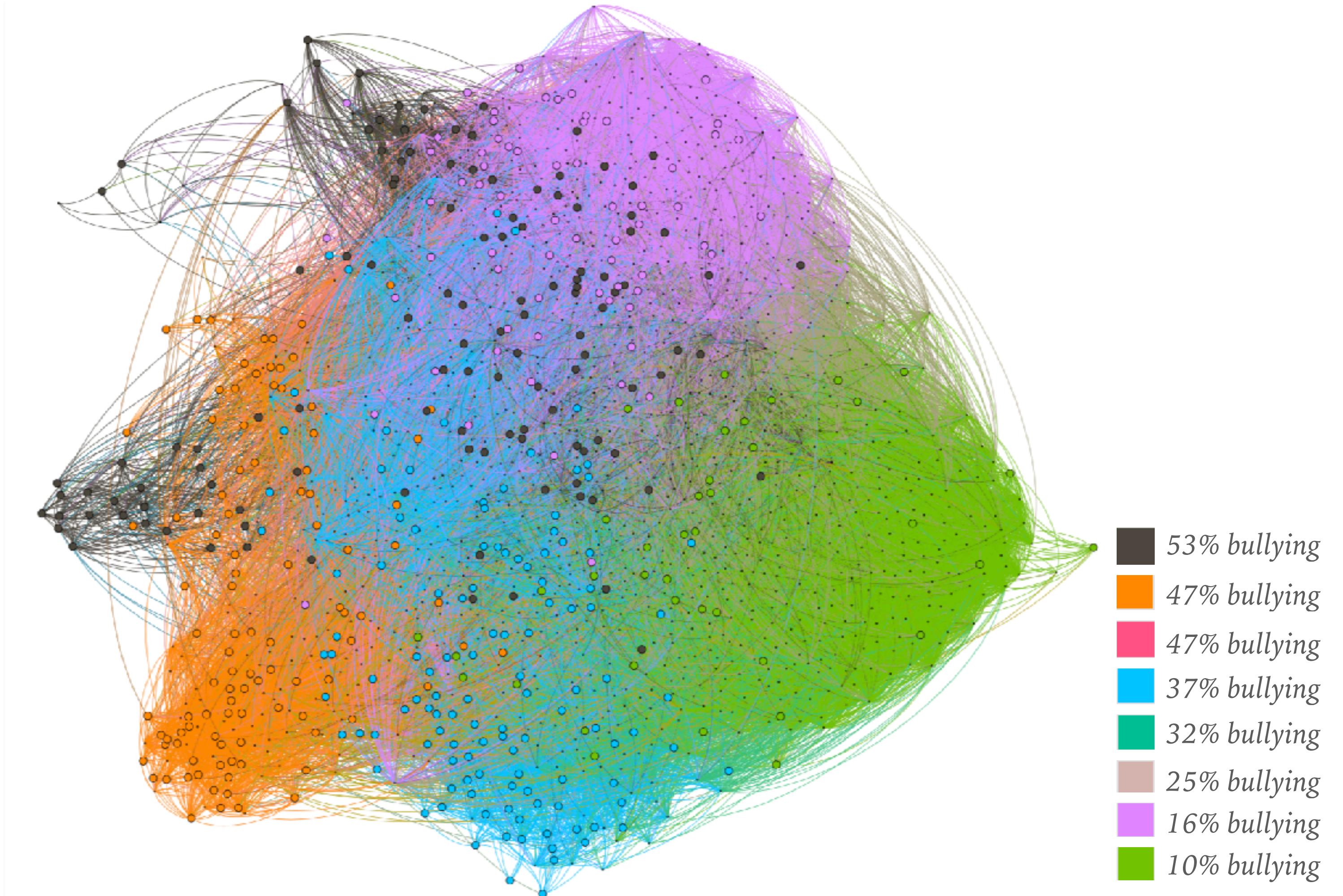


RQ2: CYBERBULLYING MORE OFTEN IN CERTAIN IMAGE CLUSTER?



CYBERBULLYING MORE OFTEN IN CERTAIN IMAGE CLUSTER?

- Bipartite network of image and words in comments
- Likelihood ratio test:
 $\chi^2(14) = 258.61, p < .001$



RQ3: PROMINENT TOPICS UNDER CYBERBULLYING THREAD?



STRUCTURAL TOPIC MODELING RESULTS

fuck

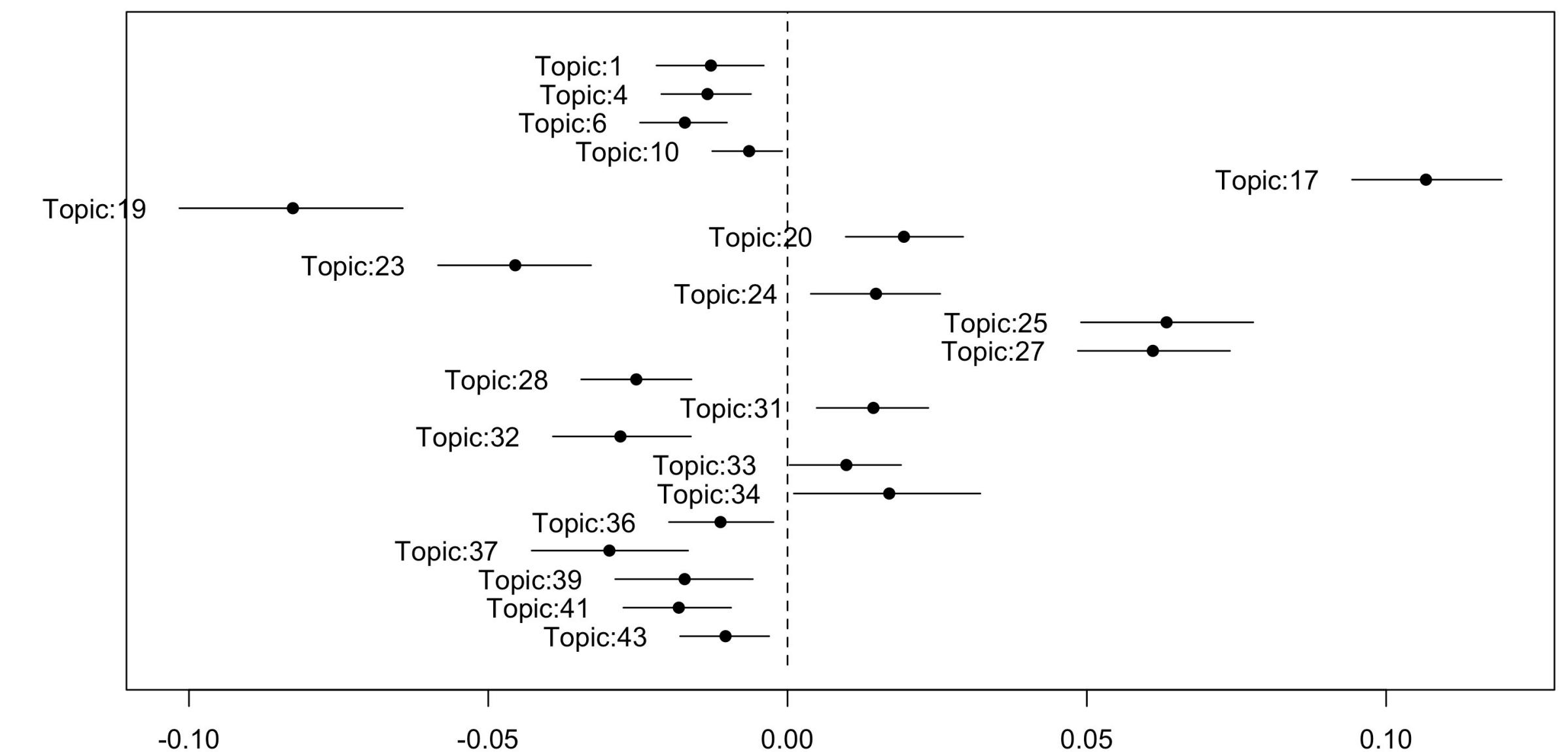
your
lmao
pic
mean
caus
haha
face
someth
asshol
piec
dumbass
didnt
now
beat
dude
gay
ill
shut
spell
shit
tell
idiot
littl
sure
boy
say
just
cuz
got
back
tri
take
lol
probabl
stfu
look
make
friend
dick
cant
even
hell
time
guy
one
man
can
fat
talk
fuckin
care
fag
dumb
year
pictur
old
comment
gonna
retard
read
call
suck
mother
give
funni
said
kid
stupid
get
hes
cunt
bitch
ass
still

dont
idiot
say
assault
usa
america
just
love
use
start
support
make
famili
men
ignor
right
never
tri
also
someon
see
comment
world
talk
believ
ppl
doesnt
thing
first
thank
pictur
live
armi
media
one
agree
know
think
feel
post
tell
guy
happen
back
cant
soldier
kill
can
stop
protect
peopl

birthday
handsom
omg
much
great
can
say
realli
world
smile
now
live
video
bab
pic
stop
see
hot
luv
sexi
selena
kik
voic
cri
ilysm
get
comment
that
nice
one
wanna
meet
favorit
amaz
pleas
beauti
perf
look
hair
god
pictur
perf
perfect
gorgeous
pretti
will
ever
day
mean
dream
lol
gim
hey
best
back
notic
thank
fan
ili
cant
awesom
come
danc
like
your
follow
happi
cute
make
crazi
good
new
wait
alway
marri
cut
face
cool
know
just
soo
like
photo
miss
babe

love

Topic Prevalence Differences in Bullying vs. Non-bullying Commentary

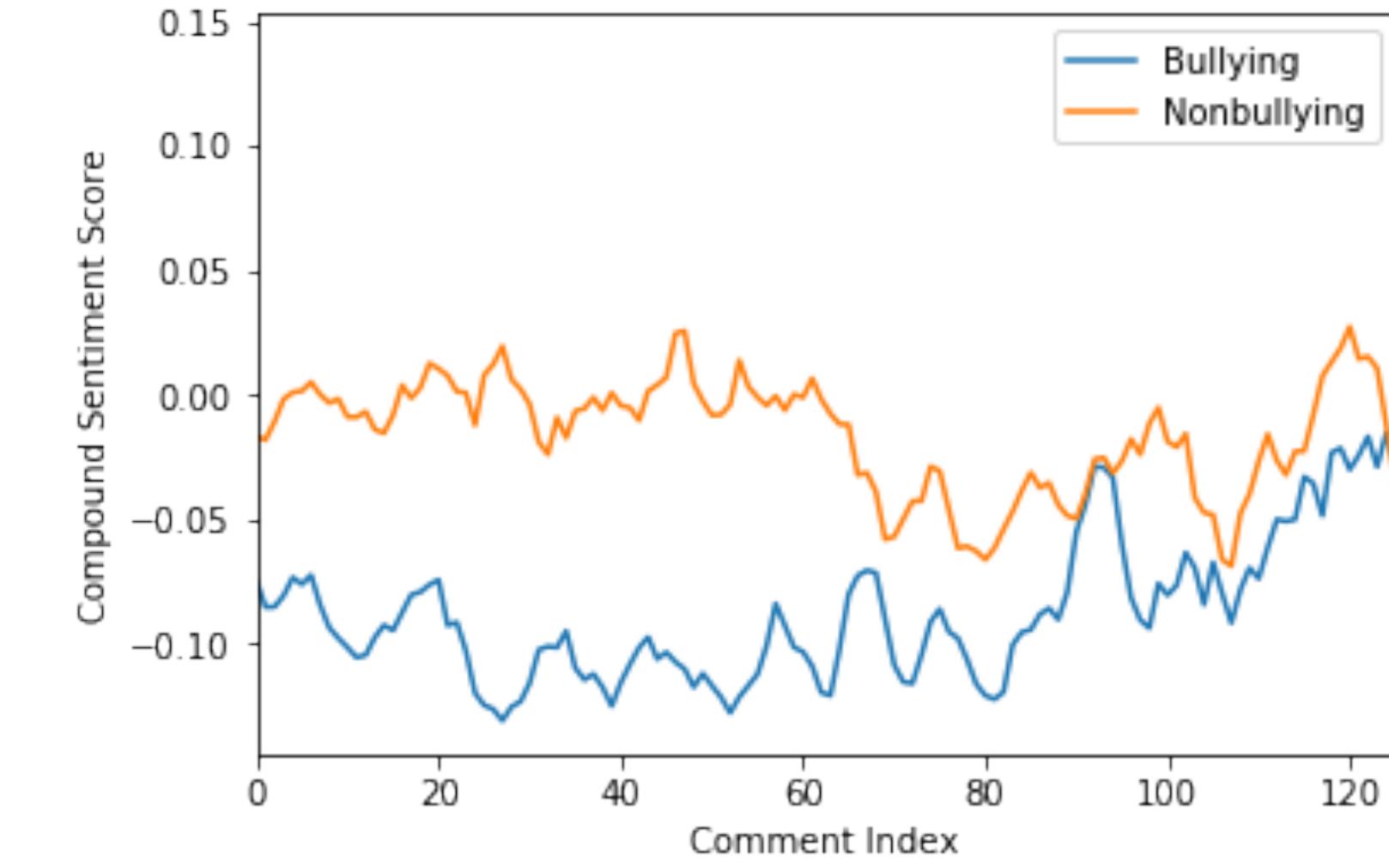
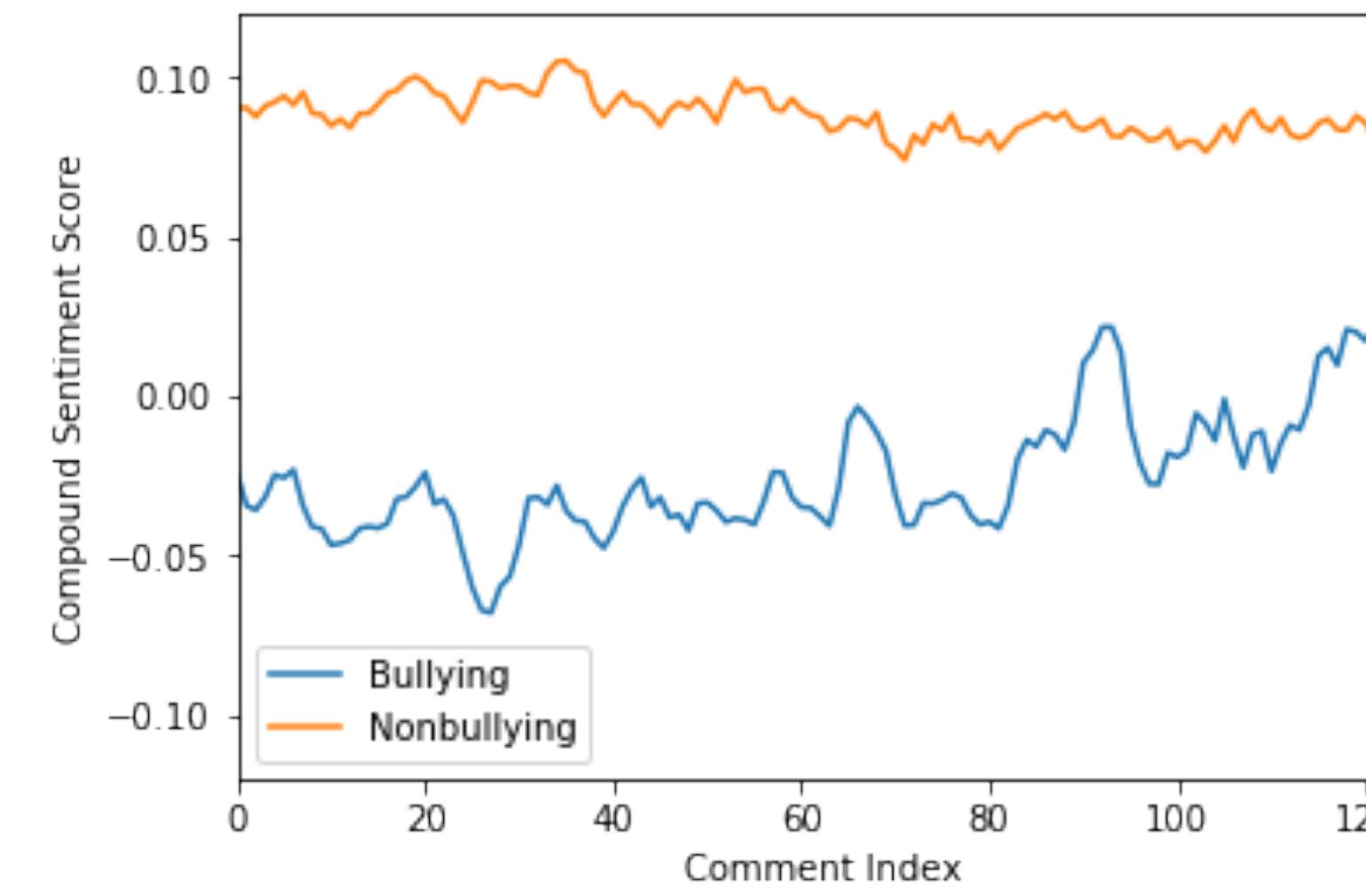
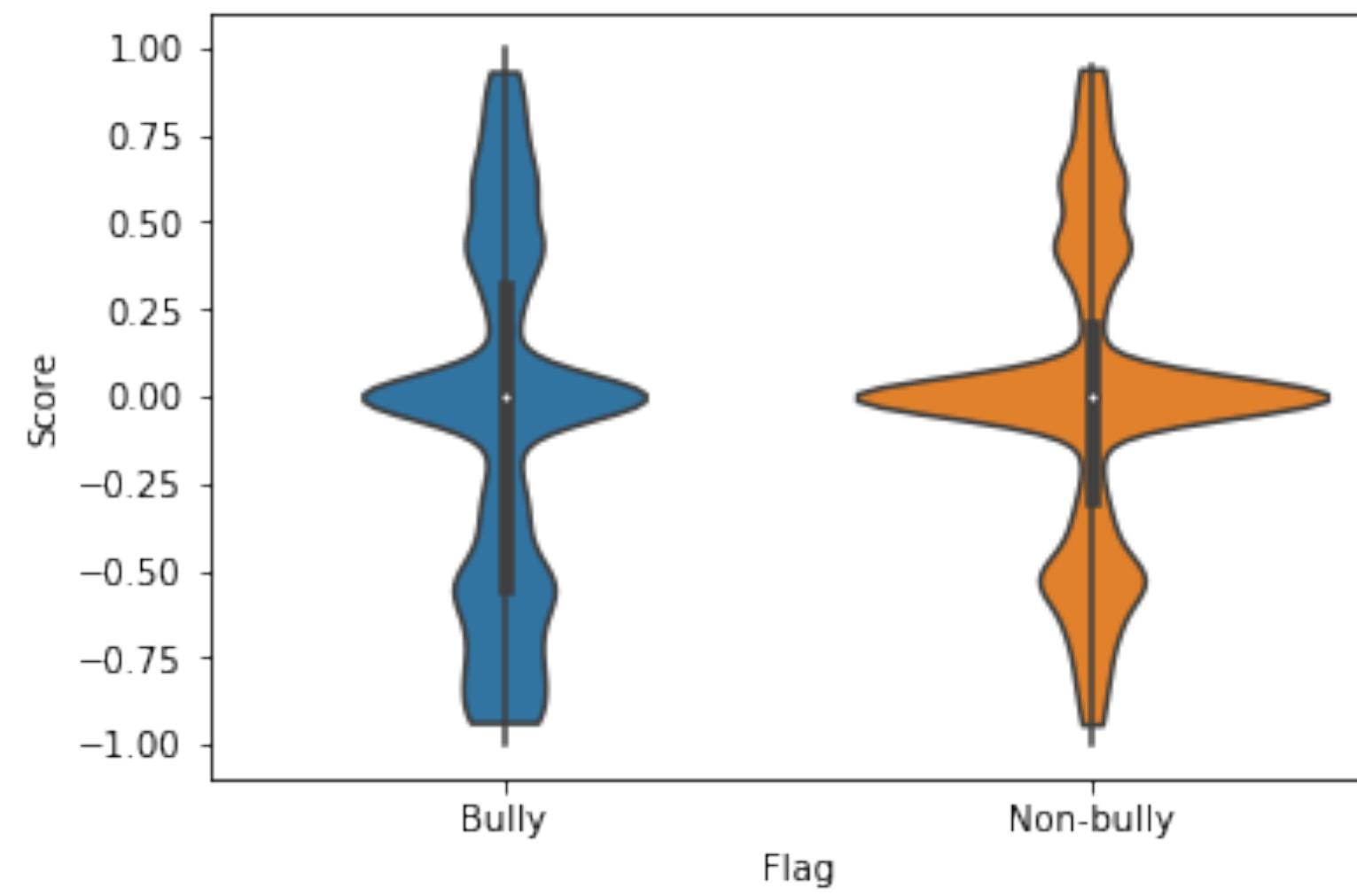


RQ4: SENTIMENT TREND OF CYBERBULLYING THREAD



SENTIMENT TREND OF CYBERBULLYING/NONCYBERBULLYING THREAD

- VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis on comments
- Using the “compound score” to represent positiveness/negativeness
- Analyze average sentiment trend for bullying conversations and non-bullying conversations



All

40% + negative words

RQ5: BYSTANDER SUPPORTIVE MESSAGES



BYSTANDER SUPPORTIVE MESSAGES

- Comments posted by bystanders after the initial bullying comment to express support for the bullied or counter bullying posters.
- Randomly selected 100 images labeled cyberbullying: 71.0% has supportive messages.



FUTURE WORK

- Causal inference
 - Impacts of image/comment-level features?
 - Would bystander intervention improve the victim's mental well-being?
- Sentiment analyses
 - Sensitive to slangs?
 - Formal time series analysis
- Intervention
 - Platform: automatic detection algorithm
 - Individuals: online media literacy
 - Online community: collective interventions





Homa Hosseini Mardi

[cs.SI] 12 Mar 2015

Detection of Cyberbullying Incidents on the Instagram Social Network

Homa Hosseini Mardi, Sabrina Arredondo Mattson, Rahat Ibn Rafiq, Richard Han, Qin Lv, Shivakant Mishra

Computer Science Department
University of Colorado Boulder
Boulder, Colorado

Abstract

Cyberbullying is a growing problem affecting more than half of all American teens. The main goal of this paper is to investigate fundamentally new approaches to understand and automatically detect incidents of cyberbullying over images in Instagram, a media-based mobile social network. To this end, we have collected a sample Instagram data set consisting of images and their associated comments, and designed a labeling study for cyberbullying as well as image content using human labelers at the crowd-sourced Crowdflower Web site. An analysis of the labeled data is then presented, including a study of correlations between different features and cyberbullying as well as cyberaggression. Using the labeled data, we further design and evaluate the accuracy of a classifier to automatically detect incidents of cyber-

2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Prediction of Cyberbullying Incidents in a Media-based Social Network

Homa Hosseini Mardi, Rahat Ibn Rafiq, Richard Han, Qin Lv, Shivakant Mishra

Department of Computer Science, University of Colorado Boulder
rado, USA 80309
richard.han, qin.lv, mishras}@colorado.edu

Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network

Homa Hosseini Mardi^{1(✉)}, Sabrina Arredondo Mattson², Rahat Ibn Rafiq¹,
Richard Han¹, Qin Lv¹, and Shivakant Mishra¹

¹ Computer Science Department, University of Colorado Boulder, Boulder, CO, USA
{homa.hosseini mardi, rahat.ibnrafiq,
richard.han, qin.lv, shivakant.mishra}@colorado.edu

² Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, USA
sabrina.mattson@colorado.edu

Abstract. Cyberbullying is a growing problem affecting more than half of all American teens. The main goal of this paper is to study labeled cyberbullying incidents in the Instagram social network. In this work, we have collected a sample data set consisting of Instagram images and their associated comments. We then designed a labeling study and employed human contributors at the crowd-sourced CrowdFlower website to label these media sessions for cyberbullying. A detailed analysis of the labeled data is then presented, including a study of relationships between cyberbullying and a host of features such as cyberaggression, profanity, social graph features, temporal commenting behavior, linguistic content, and image content.

computational resources more efficiently to focus on the most likely discussions that may be prone to cyberbullying, then this can substantially reduce the cost of cyberbullying detection. Cyberbullying prediction provides the ability to estimate in advance those users or media sessions whose discussions may result in cyberbullying. Therefore, we can efficiently focus our computational resources on these most vulnerable users or media sessions, rather than applying a brute force classification approach to all comments.

Cyberbullying prediction is further useful for identifying in advance users who may be the most vulnerable victims of cyberbullying. As a result, such vulnerable users may be forewarned to protect themselves from potentially negative incoming comments. Also, if the vulnerable users are minors, then their parents may be warned *a priori* that their children

may be more likely victims of cyberbullying. Other resources

A photograph of a large audience from behind, looking towards a stage. The stage is mostly dark, but bright lights illuminate the audience's raised hands and the tops of their heads. The scene suggests a live performance or a public gathering.

**Thank you!
Questions?**