General Online Research Conference
GOR 19
6 to 8 March 2019, TH Köln – University of Applied Sciences, Cologne, Germany

M. Rohangis Mohseni, TU Ilmenau
Johannes Breuer, GESIS
Julian Kohne, GESIS

Methods and Tools for the Automatic Sampling and Analysis of YouTube Comments

Contact: Rohangis.Mohseni@tu-Ilmenau.de

Suggested citation: Mohseni, M.R., Breuer, J., & Kohne, J. (2019). Methods and Tools for the Automatic Sampling and Analysis of YouTube Comments. Presented at the General Online Research (GOR) Conference, Cologne.

Agenda

- **Relevance**
- Automatic vs. Manual Sampling
- Automatic vs. Manual Analysis
- Emojis
- Outlook

Relevance of YouTube

- Largest / most important online video platform
  (Alexa Traffic Ranks, 2018; Konijn, Veldhuis, & Plaisier, 2013)
- Esp. popular among adolescents who use YouTube to watch movies & shows, listen to music, and retrieve information
  (Feierabend, Plankenhorn, & Rathgeb, 2016)
- For adolescents, YouTube partly replaces TV
  (Defy Media, 2016)

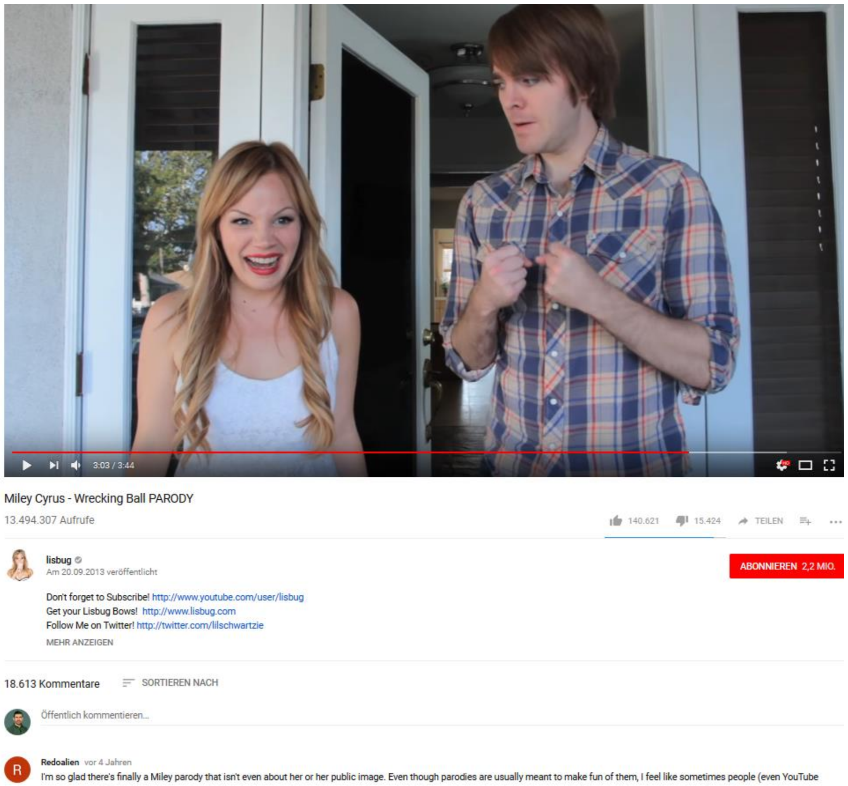
Comments on YouTube

- Useful for research on media content, communicators, and user interaction
- Data publicly available
- Relatively easy to retrieve via YouTube API

gesis Leibniz Institute for the Social Sciences

TECHNISCHE UNIVERSITÄT ILMENAU

Research Example



e.g. Lisa Schwartz (lisbug)	e.g. Shane Dawson (ShaneDawsonTV)
what a bitch !!!!	Now most lesbians look like fat dudes.
lisbug come fuck me in the ass	I want ur dick

Research Overview

- **What** do people write (content)?
  - Sexist Online Hate Speech
    (Döring & Mohseni, 2018, in press, under revision; Wotanis & McMillan, 2014)
  - Comment characteristics
    (Thelwall, Sud, & Vis, 2012)
  - Subtopics, sentiments, & gender differences
    (Thelwall, 2017)
- **Who** writes it (communicator)?
  - User experiences
    (Defy Media, 2016; Lange, 2007; Moor, Heuvelman, & Verleur, 2010; Oksanen, Hawdon, Holkeri, Näsi, & Räsänen, 2014; Szostak, 2013; Yang, Hsu, & Tan, 2010)

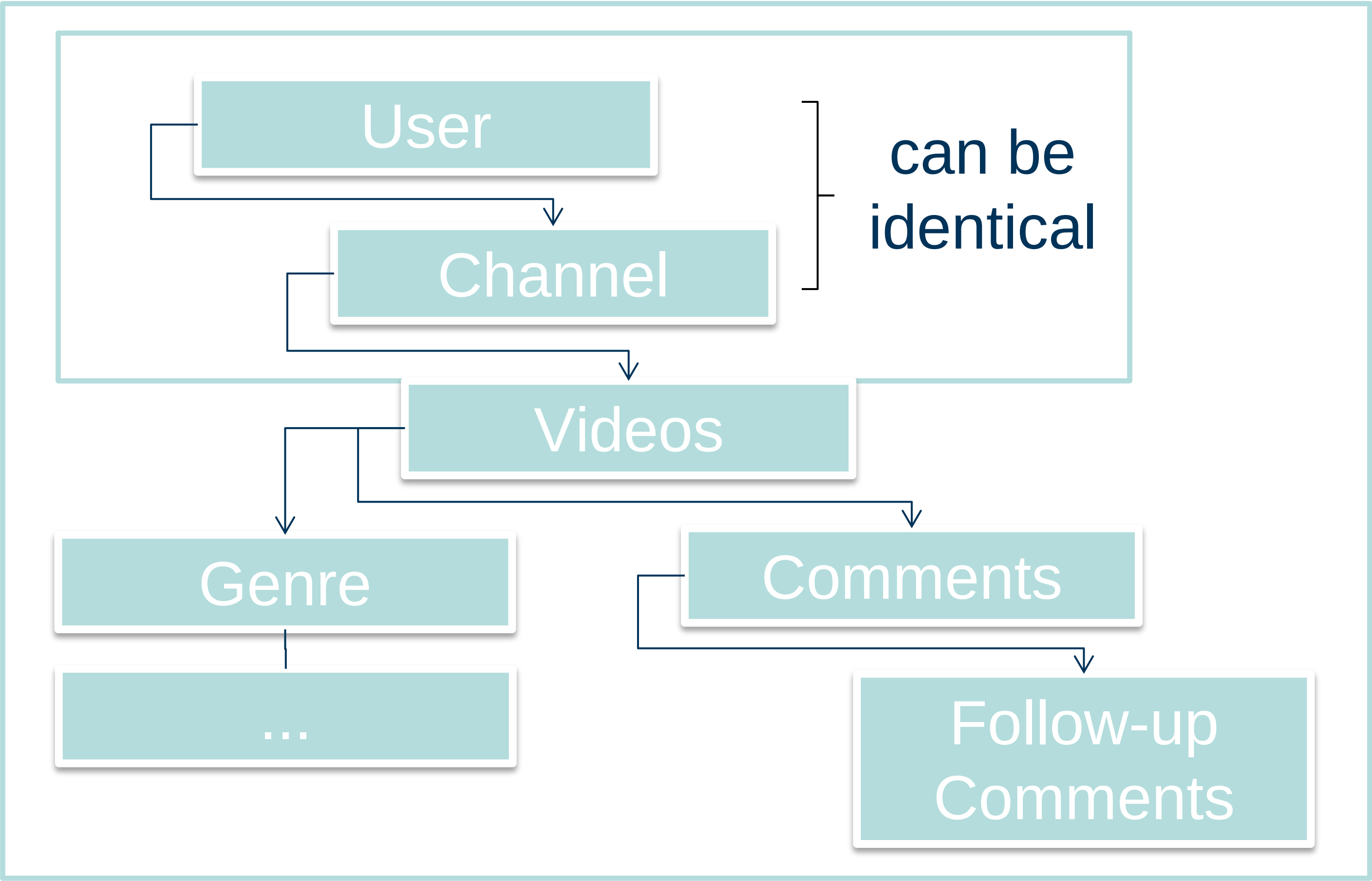
Understanding YouTube Comments

- Unusual spelling: V e r t i c a l
  - S p a c e d
  - CAPITALISED
  - 13375p34k
- Neologism: Emoji = e (picture) + moji (character)
- Slang: Ur effin hot
- Irony: This is great! 💩 💩 💩
- Emojis: cultural, intraindividual, contextual, and platform differences

Agenda

- Relevance
- **Automatic vs. Manual Sampling**
- Automatic vs. Manual Analysis
- Emojis
- Outlook

YouTube data structure



Options for YouTube comment scraping

Method	Manual	Webometric Analyst	YouTube Data Tools	Tuber
Type	n/a	Program	Web service	Package for R
Platforms	All	Win	All	Win, Mac, Linux, Unix
Collected Features	Depends on coding scheme	Channel Info, Video Info, Comments, Video Search	Channel Info, Video Info, Comments, Video List	Channel Info, Video Info, Comments, Subtitles, All searches
Scoping	Depends on coding scheme	100 most recent or all comments	All comments	20-100 most recent or all comments

Short tutorials on how to use the tools are available in our GitHub repo:
https://github.com/JuKo007/YouTubeComments

Tool Comparison

Method	Manual	Webometric Analytics	YouTube Data Tools	Tuber
Need API-Key	No	Yes	No	Yes
Disadvantages	Time-consuming	Only first 5 follow-up comments; No error feedback; Undetectable Time-outs	Lacking flexibility; Less infos	Only first 5 follow-up comments (Issue on GitHub open)
Ease of Use	High	Low	High	Low
License	n/a	Free for n/c	Open Source	Open Source
Example: Dayum Video (22.2.@2p.m.)	47.163	44.828	47.153	44.810

Webometric Analytics

VideoID	CommentID	Comment Published	Comment Updated	CommentText Display
DcJFdCmN98s	UgyPhXB5E7Fwip6y_OF4AaABAg	2019-02-22T 08:41:41.000Z	2019-02-22T 08:41:41.000Z	2019 Lets go

CommentAuthor Name	CommentAuthorURI	Comment CanReply	CommentTotalReply Count
kevork gamal	http://www.youtube.com/channel/UC6aYwoJeQAIt_Nmu3fjy9pg	true	0

Comment isPublic	Comment LikeCount	Comment VewerRating	IsReply	CommentPosterInfo
true	0	none	0	

YouTube Data Tools

id	reply Count	like Count	publishedAt	authorName	text
UgyPhXB5E7Fwip6y_OF4AaABAg	0	0	22.02.2019 08:41	kevork gamal	2019 Lets go

authorChannel Id	authorChannelUrl	isReply	isReplyTo	isReplyTo Name
UC6aYwoJeQAIt_Nmu3fjy9pg	http://www.youtube.com/channel/UC6aYwoJeQAIt_Nmu3fjy9pg	0		

Tuber

	authorDisplayName	authorProfileImageUrl	authorChannelUrl	authorChannelId.value
2	kevork gamal	https://yt3.ggpht.com/-NjWx9Ys1hFE/AAAAAAAAAAI/AAAAAAAAAAA/XVIU23CLC2w/s28-c-k-no-mo-rj-c0xffffff/photo.jpg	http://www.youtube.com/channel/UC6aYwoJeQAIt_Nmu3fjy9pg	UC6aYwoJeQAIt_Nmu3fjy9pg

videoId	textDisplay	textOriginal	canRate	viewerRating	likeCount
DcJFdCmN98s	2019 Lets go	2019 Lets go	TRUE	none	0

publishedAt	updatedAt	moderationStatus	id	parentId
2019-02-22T08:41:41.000Z	2019-02-22T08:41:41.000Z	NA	UgyPhXB5E7Fwip6y_OF4AaABAg	NA

Interactive Jupyter Notebook & R script with examples for scraping and simple sentiment analysis of comment text and emojis can be found in our GitHub repo: https://github.com/JuKo007/YouTubeComments

Agenda

- Relevance
- Automatic vs. Manual Sampling
- **Automatic vs. Manual Analysis**
- Emojis
- Outlook

Automatic vs. Manual Analysis

- Manual Content Analysis
  - Two or more humans code text units
- Automatic Content Analysis
  - From frequencies of letters to meaning of text units (NLP)
  - Supervised vs. unsupervised
- Automatic Sentiment Analysis
  - Typically bag-of-words approach
  - Dictionaries or ML
- Automatic Topic Modelling
  - Frequencies & networks of topics

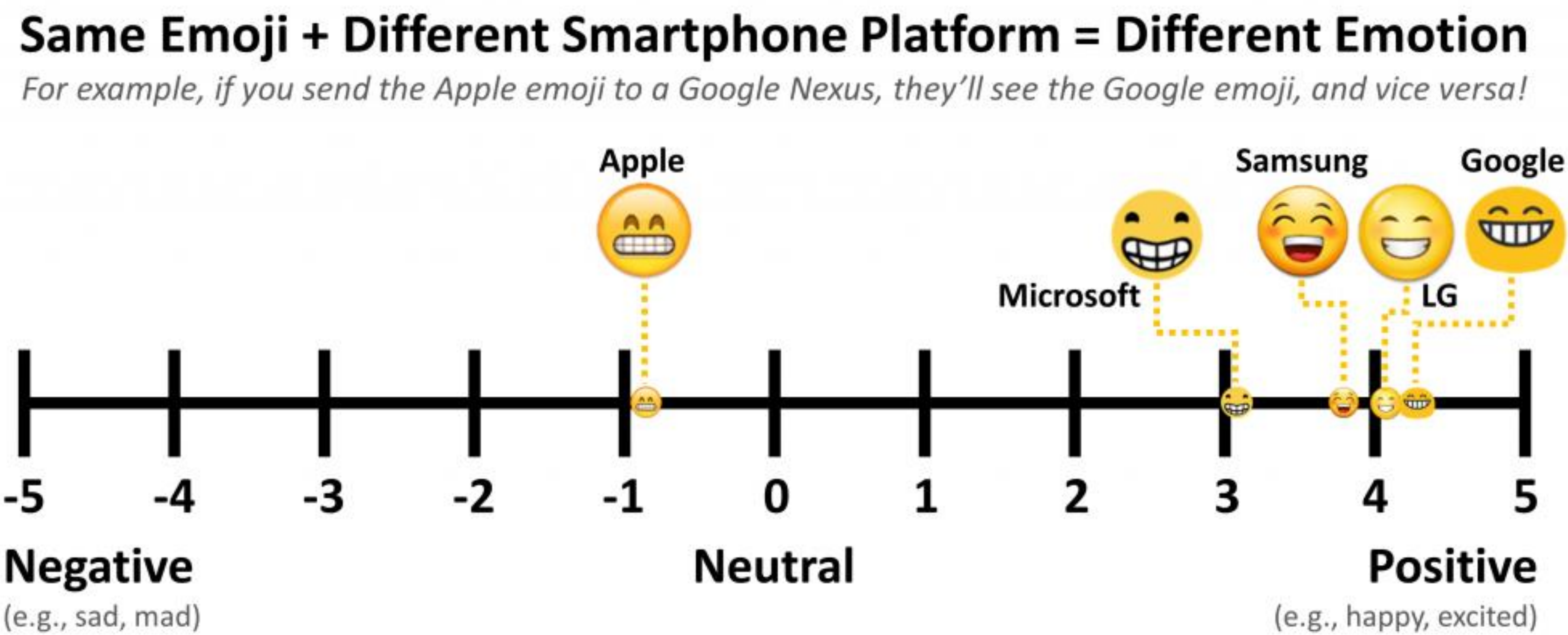
Agenda

- Relevance
- Automatic vs. Manual Sampling
- Automatic vs. Manual Analysis
- **Emojis**
- Outlook

Understanding Emojis

- Interpreting emojis
  - Emoji ≈ verbal descriptor
  - Multiple meanings: 🙏 = clapping, praying, blessed
  - Context-dependent: 🍆 = eggplant or penis
  - Culture-dependent
  - Person-dependent
  - Platform-dependent: see next slide
- Interpreting emojis in (con-)text
  - Meaning of emoji can depend on text: 🙏 🍆
  - Emojis can stress meaning of text: Hot!🔥🚒
  - Emojis can change meaning of text: Great! 💩⍰

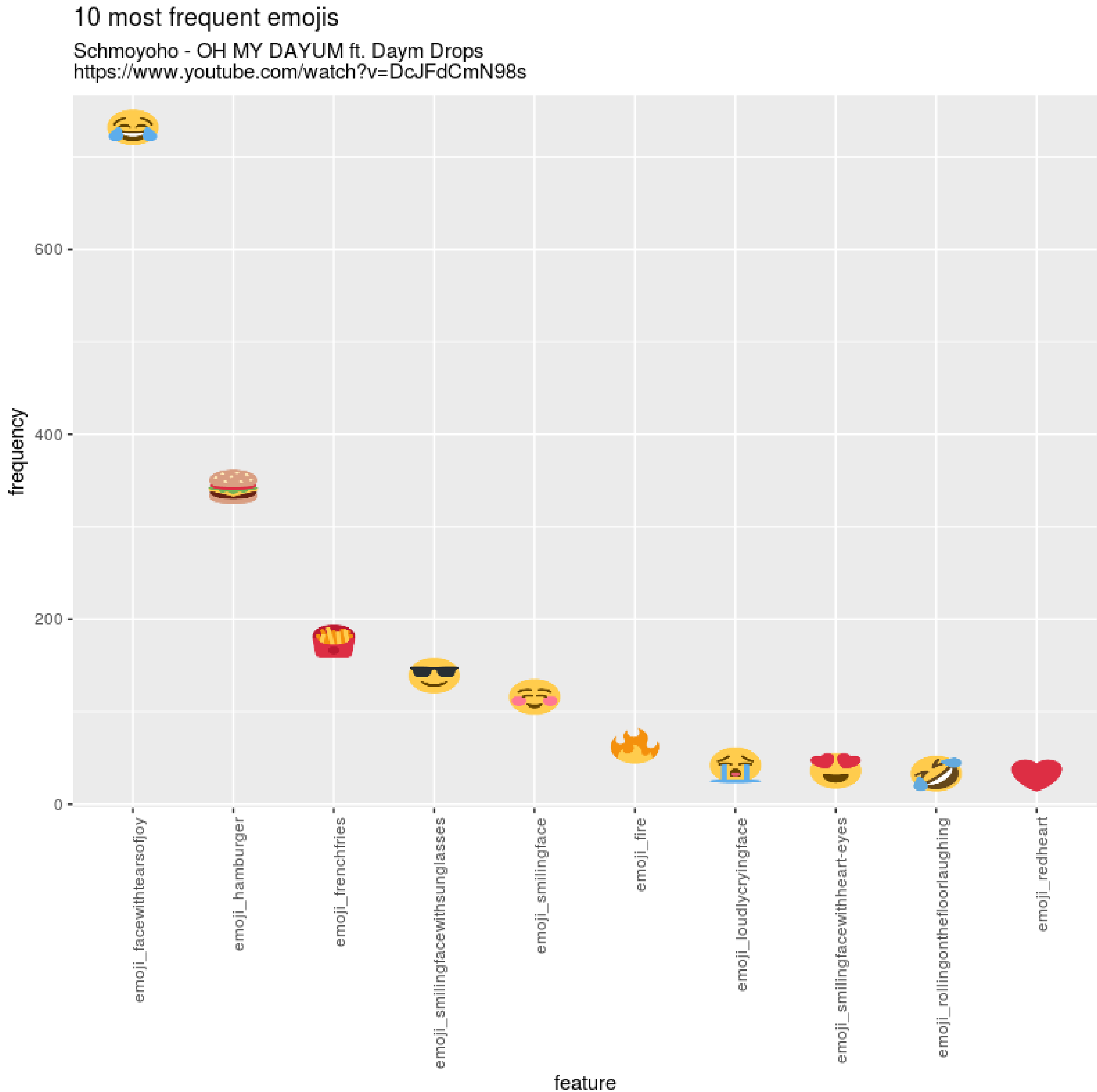
Platform Dependence: Emojis

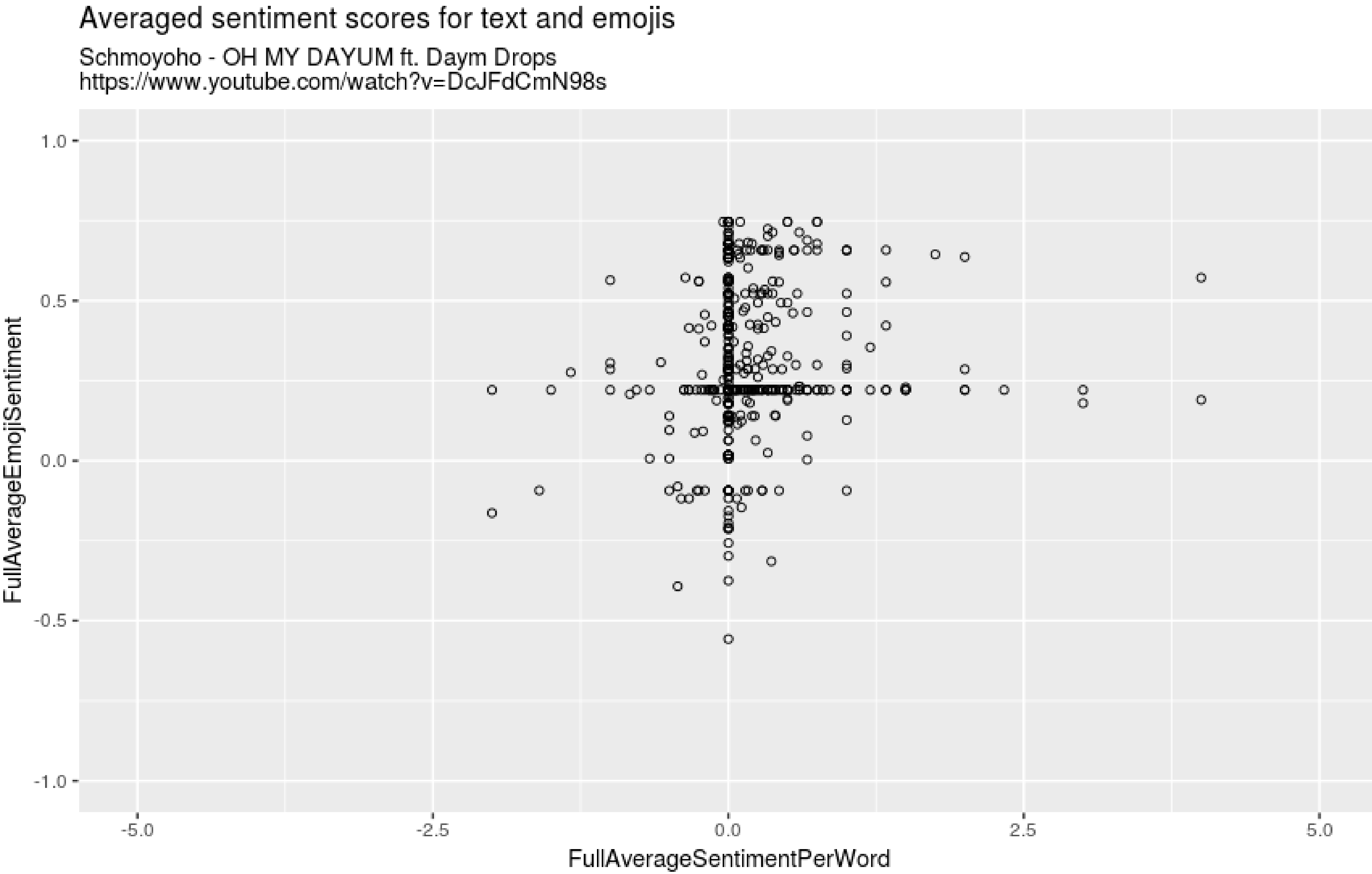


Miller et al., 2016

Parsing Emojis

- Parsing emojis
  - Manual emoticons ≠ emojis
  - Encoding can differ 😂
    - Hex: <f0><U+009F><U+0098><U+0082>
    - Unicode: U+1F602
    - HTML: &#128514
    - Most Problems when in Windows because UTF-8 encoding is interpreted as Win 1252: 😂 -> ðŸ˜‚
  - Constant dictionary updates needed for new emojis
    - 230 new emojis in 2019


10 most frequent emojis
Schmoyoho - OH MY DAYUM ft. Daym Drops
https://www.youtube.com/watch?v=DcJFdCmN98s
frequency
600
400
200
0
emoji_facewithtearsofjoy
emoji_hamburger
emoji_frenchfries
emoji_smilingfacewithsunglasses
emoji_smilingface
emoji_fire
emoji_loudlycryingface
emoji_smilingfacewithheart-eyes
emoji_rollingonthefloorlaughing
emoji_redheart
feature


Averaged sentiment scores for text and emojis
Schmoyoho - OH MY DAYUM ft. Daym Drops
https://www.youtube.com/watch?v=DcJFdCmN98s
1.0
0.5
0.0
-0.5
-1.0
FullAverageEmojiSentiment
-5.0
-2.5
0.0
2.5
5.0
FullAverageSentimentPerWord

Limitations of Emoji Sentiment Analysis

- Dictionary does not include all emojis
- Many emojis with small/neutral valence ratings
- Differences in interpretation due to multiple meanings, (con-)text, culture, person, & platform

Limitations of Bag-Of-Word Sentiment

Senti-ment	Comment
>+10	Schmoyoho, we're not really **entertained** by you anymore. You're sort of like Dane Cook. At first we thought, "**Wow**! Get a **load** of this channel! It's **funny**!" But then we realized after far too long, "**Wow**, these guys are just a one trick pony! There is absolutely nothing I **like** about these people!" You've run your course. The shenanigans, the "songifies".. we get it. It's just not that **funny** man. We don't really **like** you.
<-10	**Fucking** hilarious! And that guy could either do commercials or be an actor, I've never, in my entire life, heard anyone express themselves that strongly about a **fucking** hamburger. And now all I know is I have never eaten one of those but **damned** if I won't have it on my list of **shit** to do tomorrow! **Hell** of a job by schmoyoho as well, whoever said this should be a commercial **hit** it on the head.

Agenda

- Relevance
- Automatic vs. Manual Sampling
- Automatic vs. Manual Analysis
- Emojis
- **Outlook**

Outlook

- Systematic sampling of videos
- Going beyond bag-of-words/-emojis
  - N-grams
  - Word/emoji embeddings
- Machine Learning
  - Instead of dictionary-based sentiment analysis
- Topic Models with emojis
  - Using textual description of emojis
- Automated text analysis using network techniques (see, e.g., https://github.com/cbail/textnets) + emojis

General Online Research Conference
GOR 19
6 to 8 March 2019, TH Köln – University of Applied Sciences, Cologne, Germany

M. Rohangis Mohseni, TU Ilmenau
Johannes Breuer, GESIS
Julian Kohne, GESIS

Methods and Tools for the Automatic Sampling and Analysis of YouTube Comments

Contact: Rohangis.Mohseni@tu-Ilmenau.de

Suggested citation: Mohseni, M.R., Breuer, J., & Kohne, J. (2019). Methods and Tools for the Automatic Sampling and Analysis of YouTube Comments. Presented at the General Online Research (GOR) Conference, Cologne.

References

- Alexa Traffic Ranks (2018, March 12). How popular is youtube.com? Retrieved from http://www.alexa.com/siteinfo/youtube.com
- Defy Media. (2016). *Youth Video Diet: Acumen Report*. USA. Retrieved from http://www.defymedia.com/acumen/acumen-report-youth-video-diet
- Döring, N. & Mohseni, M. R. (in press). Gendered Hate Speech in YouTube Comments: Results of a Content Analysis. *Studies in Communication and Media*.
- Döring, N. & Mohseni, M. R. (2018). Male Dominance and Sexism on YouTube: Results of Three Content Analyses. *Feminist Media Studies*. Advance online publication. doi:10.1080/14680777.2018.1467945
- Feierabend, S., Plankenhorn, T., & Rathgeb, T. (2016). *JIM 2016: Jugend, Information, (Multi-) Media. Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland*. Stuttgart. Retrieved from https://www.mpfs.de/studien/jim-studie/2016/

References

- Konijn, E. A., Veldhuis, J., & Plaisier, X. S. (2013). YouTube as a research tool: Three approaches. *Cyberpsychology, Behavior and Social Networking, 16*(9), 695–701. https://doi.org/10.1089/cyber.2012.0357
- Lange, P. G. (2007, March). *Commenting on comments: Investigating responses to antagonism on YouTube*. Paper presented at the Society for Applied Anthropology Conference, Tampa, Florida, USA. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.170.3808&rep=rep1&type=pdf
- Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying Interpretations of Emoji. In Association for the Advancement of Artificial Intelligence (Ed.), *Proceedings of ICWSM*. Retrieved from https://grouplens.org/site-content/uploads/Emoji_Interpretation.pdf
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior, 26*, 1536–1546. https://doi.org/10.1016/j.chb.2010.05.023
- Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. In M. N. Warehime (Ed.), *Sociological Studies of Children and Youth: Vol. 18. Soul of society: A focus on the lives of children & youth* (Vol. 18, pp. 253–273). Bingley, U.K: Emerald. https://doi.org/10.1108/S1537-466120140000018021

References

- Pfaffenberger, F. (2016). *Twitter als Basis wissenschaftlicher Studien. Eine Bewertung gängiger Erhebungs- und Analysemethoden der Twitter-Forschung*. Berlin: Springer.
- Szostak, N. (2013). Girls on YouTube: Gender politics and the potential for a public sphere. *The McMaster Journal of Communication, 8*, 46–58.
- Thelwall, M. (2017). Social media analytics for YouTube comments: Potential and limitations. *International Journal of Social Research Methodology, 21*(3), 303–316. https://doi.org/10.1080/13645579.2017.1381821
- Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube videos: From guatemalan rock to El Big Bang. *Journal of the American Society for Information Science and Technology, 63(3*), 616–629. https://doi.org/10.1002/asi.21679
- Wotanis, L., & McMillan, L. (2014). Performing gender on YouTube. *Feminist Media Studies, 14*(6), 912–928. https://doi.org/10.1080/14680777.2014.882373
- Yang, C., Hsu, Y.-C., & Tan, S. (2010). Predicting the determinants of users' intentions for using YouTube to share video: Moderating gender effects. *Cyberpsychology, Behavior and Social Networking, 13*(2), 141–152. https://doi.org/10.1089/cyber.2009.0105