Exploration of Gender Differences in COVID-19 Discourse on Reddit

Jai Aggarwal **Ella Rabinovich** **Suzanne Stevenson**

Department of Computer Science, University of Toronto

{jai,ella,suzanne}@cs.toronto.edu

Abstract

Decades of research on differences in the language of men and women have established postulates about preferences in lexical, topical, and emotional expression between the two genders, along with their sociological underpinnings. Using a novel dataset of male and female linguistic productions collected from the Reddit discussion platform, we further confirm existing assumptions about gender-linked affective distinctions, and demonstrate that these distinctions are amplified in social media postings involving emotionally-charged discourse related to COVID-19. Our analysis also confirms considerable differences in topical preferences between male and female authors in spontaneous pandemic-related discussions.

1 Introduction

Research on gender differences in language has a long history spanning psychology, gender studies, sociolinguistics, and, more recently, computational linguistics. A considerable body of linguistic studies highlights the differences between the language of men and women in topical, lexical, and syntactic aspects (Lakoff, 1973; Labov, 1990), and such differences have proven to be accurately detectable by automatic classification tools (Koppel et al., 2002; Schler et al., 2006; Schwartz et al., 2013). Here, we study the differences in male (M) and female (F) language in discussions of COVID-19[1] on the Reddit[2] discussion platform. Responses to the virus on social media have been heavily emotionally-charged, accompanied by feelings of anxiety, grief, and fear, and have discussed far-ranging concerns regarding personal and public health, the economy, and social aspects of life. In this work, we explore how established emotional and topical cross-gender differences are carried over into this pandemic-related discourse. Insights regrading these distinctions will advance our understanding of gender-linked linguistic traits, and may further help to inform public policy and communications around the pandemic.

Research has considered the emotional content of social media on the topic of the COVID pandemic (e.g., Lwin et al., 2020; Stella et al., 2020), but little work has looked specifically at the impact of gender on affective expression (van der Vegt and Kleinberg, 2020). Gender-linked linguistic distinctions across emotional dimensions have been a subject of prolific research (Burriss et al., 2007; Hoffman, 2008; Thelwall et al., 2010), with findings suggesting that women are more likely than men to express positive emotions, while men exhibit higher tendency to dominance, engagement, and control (although see Park et al. (2016) for an alternative finding). van der Vegt and Kleinberg (2020) compared the self-reported emotional state of male vs. female crowdsourced workers who contributed to the Real World Worry Dataset (RWWD, Kleinberg et al., in press), in which they were also asked to write about their feelings around COVID. However, because van der Vegt and Kleinberg (2020) restricted the affective analysis to the workers' emotional ratings, it remains an open question regarding whether, and how, the natural linguistic productions of males and females about COVID will exhibit detectably different patterns of emotion.

Topical analysis of social media during the pandemic has also been a focus of recent work (e.g., Liu et al., 2020; Abd-Alrazaq et al., 2020), again with few studies devoted to gender differences (Thelwall and Thelwall, 2020; van der Vegt and Kleinberg, 2020). Much prior work has found distinctions in topical preferences in spontaneous productions of the two genders (e.g., Mulac et al., 2001; Mulac, 2006; Newman et al., 2008), showing that men were more likely to discuss money-

[1] We refer to COVID-19 by 'COVID' hereafter.

[2] https://www.reddit.com/

and occupation-related topics, focused on objects and impersonal matters, while women preferred discussion on family and social life, topics related to psychological and social processes. In the recent context, Thelwall and Thelwall (2020) found these observations persisted in COVID-19 tweets, with a male focus on sports and politics, and female focus on family and caring. In the prompted texts of the RWWD, van der Vegt and Kleinberg (2020) also found the expected M vs. F topical differences, with men talking more about the international impact of the pandemic, as well as governmental policy, and women more commonly discussing social aspects – family, friends, and solidarity. Moreover, van der Vegt and Kleinberg (2020) further found differences between the elicited short (tweet-sized) and longer essays, revealing the impact of the goal and size of the text on such analyses. Again, an open question remains concerning the topical distinctions between M and F authors in spontaneous productions without artificial restrictions on length.

Here, we aim to address the above gaps in the literature, by performing a comprehensive analysis of the similarities and differences between male and female language collected from the Reddit discussion platform. Our main corpus is a large collection of spontaneous COVID-related utterances by (self-reported) M and F authors. Importantly, we also collect productions on a wide variety of topics by the same set of authors as a 'baseline' dataset. First, using a multidimensional affective framework from psychology (Bradley and Lang, 1994), we draw on a recently-released dataset of human affective ratings of words Mohammad (2018) to support the emotional assessment of male and female posts in our datasets. Through this approach, we corroborate existing assumptions on differences in the emotional aspects of linguistic productions of men and women in the COVID corpus. Moreover, our use of a baseline dataset enables us to further show that these distinctions are amplified in the emotionally-intensive setting of COVID discussions compared to productions on other topics. Second, we take a topic modeling approach to demonstrate detectable distinctions in the range of topics discussed by the two genders in our COVID corpus, reinforcing (to some extent) assumptions on gender-related topical preferences, in this natural discourse in an emotionally-charged context.[3]

[3]All data and code is available at https://github.com/ellarabi/covid19-demography.

2 Datasets

As noted above, our goal is to analyze emotions and topics in spontaneous utterances that are relatively unconstrained by length. To that end, our main dataset comprises a large collection of spontaneous, COVID-related English utterances by male and female authors from the Reddit discussion platforms. As of May 2020, Reddit had over 430M active users, 1.2M topical threads (subreddits), and over 70% of its user base coming from English-speaking countries. Subreddits often encourage their subscribers to specify a meta-property (called a 'flair', a textual tag), projecting a small glimpse about themselves (e.g., political association, country of origin, age), thereby customizing their presence within a subreddit.

We identified a set of subreddits, such as 'r/askmen' and 'r/askwomen', where authors commonly self-report their gender, and extracted a set of unique user-ids of authors who specified male or female gender as a flair.[4] This process yielded the user-ids for 10,421 males and 5,630 females (as self-reported). Using this extracted set of ids, we collected COVID-related submissions and comments[5] from across the Reddit discussion platform for a period of 15 weeks, from February 1st through June 1st. COVID-related posts were identified as those containing one or more of a set of predefined keywords: 'covid', 'covid-19', 'covid19', 'corona', 'coronavirus', 'the virus', 'pandemic'. This process resulted in over 70K male and 35K female posts spanning 7,583 topical threads; the male subcorpus contains 5.3M tokens and the female subcorpus 2.8M tokens. Figure 1 presents the weekly amount of COVID-related posts in the combined corpus, showing a peak in early-mid March (weeks 5–6).

Aiming at a comparative analysis between virus-related and 'neutral' (baseline) linguistic productions by men and women, we collected an additional dataset comprising a randomly sampled 10K posts per week by the same set of authors, totalling 150K posts for each gender. The baseline dataset contains 6.8M tokens in the male subcorpus and 5.3M tokens in the female subcorpus.

We use our COVID and baseline datasets for analysis of emotional differences as well as topical preferences in spontaneous productions by male

[4]Although gender can be viewed as a continuum rather than binary, we limit this study to the two most prominent gender markers in our corpus: male and female.

[5]For convenience, we refer to both initial submissions and comments to submissions as 'posts' hereafter.

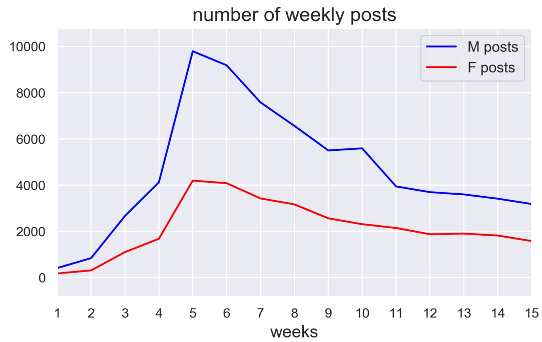

Figure 1: Weekly COVID-related posts by gender.

and female authors on Reddit. The ample size of the corpora facilitates analysis of distinctions in these two aspects between the two genders in their discourse on the pandemic, and as compared to non-COVID discussion.

3 Analysis of Emotional Dimensions

3.1 Methods

A common way to study emotions in psycholinguistics uses an approach that groups affective states into a few major dimensions, such as the Valence-Arousal-Dominance (VAD) affect representation, where *valence* refers to the degree of positiveness of the affect, *arousal* to the degree of its intensity, and *dominance* represents the level of control (Bradley and Lang, 1994). Computational studies applying this approach to emotion analysis have been relatively scarce due to the limited availability of a comprehensive resource of VAD rankings, with (to the best of our knowledge) no large-scale study on cross-gender language. Here we make use of the recently-released NRC-VAD Lexicon, a large dataset of human ratings of $20,000$ English words (Mohammad, 2018), in which each word is assigned V, A, and D values, each in the range $[0–1]$. For example, the word 'fabulous' is rated high on the valence dimension, while 'deceptive' is rated low. In this study we aim at estimating the VAD values of posts (typically comprising multiple sentences), rather than individual words; we do so by inferring the affective ratings of sentences using those of individual words, as follows.

Word embedding spaces have been shown to capture variability in emotional dimensions closely corresponding to valence, arousal, and dominance (Hollis and Westbury, 2016), implying that such semantic representations carry over information useful for the task of emotional affect assessment. Therefore, we exploit affective dimension ratings assigned to individual words for supervision in extracting ratings of sentences. We use the model introduced by Reimers and Gurevych (2019) for producing word- and sentence-embeddings using Siamese BERT-Networks,[6] thereby obtaining semantic representations for the $20,000$ words in Mohammad (2018) as well as for sentences in our datasets. This model performs significantly better than alternatives (such as averaging over a sentence's individual word embeddings and using BERT encoding (Reimers and Gurevych, 2019)) on the SentEval toolkit, a popular evaluation toolkit for sentence embeddings (Conneau and Kiela, 2018).

Next, we trained beta regression models[7] (Zeileis et al., 2010) to predict VAD scores (dependent variables) of words from their embeddings (independent predictors), yielding Pearson's correlations of 0.85, 0.78, and 0.81 on a 1000-word held-out set for V, A, and D, respectively. The trained models were then used to infer VAD values for each sentence within a post using the sentence embeddings.[8] A post's final score was computed as the average of the predicted scores for each of its constituent sentences. As an example, the post *'most countries handled the covid-19 situation appropriately'* was assigned a low arousal score of 0.274, whereas a high arousal score of 0.882 was assigned to *'gonna shoot the virus to death!'*.

3.2 Results and Discussion

We compared V, A, and D scores of male posts to those of female posts, in each of the COVID and baseline datasets, using Wilcoxon rank-sum tests. All differences were significant, and Cohen's d (Cohen, 2013) was used to find the effect size of these differences; see Table 1. We also compared the scores for each gender in the COVID dataset to their respective scores in the baseline dataset (discussed below). We further show, in Figure 2, the diachronic trends in VAD for M and F authors in the two sub-corpora: COVID and baseline.

First, Table 1 shows considerable differences between M and F authors in the baseline dataset for all three emotional dimensions (albeit a tiny effect size in valence), in line with established assumptions in this field (Burriss et al., 2007; Hoffman, 2008; Thelwall et al., 2010): women score higher in use of pos-

[6] We used the `bert-large-nli-mean-tokens` model, obtaining highest scores on a the STS benchmark.
[7] An alternative to linear regression in cases where the dependent variable is a proportion (in 0–1 range).
[8] We excluded sentences shorter than 5 tokens.

	COVID-related posts					Baseline posts				
	mean(M)	std(M)	mean(F)	std(F)	eff. size	mean(M)	std(M)	mean(F)	std(F)	eff. size
V	0.375	0.12	**0.388**	0.11	-0.120	0.453	0.14	**0.459**	0.14	-0.043
A	**0.579**	0.09	0.567	0.08	0.144	**0.570**	0.10	0.559	0.09	0.109
D	**0.490**	0.08	0.476	0.07	0.183	**0.486**	0.09	0.469	0.09	0.185

Table 1: Means of M and F posts for each affective dimension, and effect size of differences within each corpus. All differences significant at p<0.001. Highest mean score for each of V, A, D, in COVID and baseline, is boldfaced.

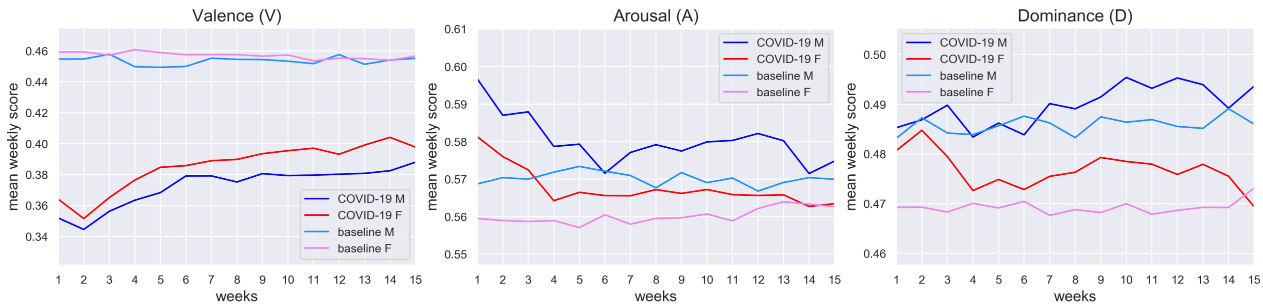

Figure 2: Diachronic analysis of valence (left), arousal (middle), and dominance (right) scores for Reddit data.

itive language, while men score higher on arousal and dominance. Interestingly, the cross-gender differences in V and A are amplified between baseline and COVID data, with an increase in effect size from 0.043 to 0.120 for V and 0.109 to 0.144 for A. By comparison, virtually no difference was detected in D between M and F authors in baseline vs. virus-related discussions. Thus we find that men seem to use more negative and emotionally-charged language when discussing COVID than women do – and to a greater degree than in non-COVID discussion – presumably indicating a grimmer outlook towards the pandemic. This finding is particularly interesting, given that van der Vegt and Kleinberg (2020) find that women self-report more negative emotion in reaction to the pandemic, and underlies the importance of analysis of implicit indications of affective state in spontaneous text.

COVID-related data trends (Figure 2) show comparatively low scores for valence and high scores for arousal in the early weeks of our analysis (February to mid-March). We attribute these findings to an increased level of alarm and uncertainty about the pandemic in its early stages, which gradually attenuated as the population learned more about the virus. As expected, both genders exhibit lower V scores in COVID discussions compared to baseline: Cohen's d effect size of -0.617 for M and -0.554 for F authors. Smaller, yet considerable, differences between the two sub-corpora exist also for A and D (0.095 and 0.047 for M, and 0.083 and 0.085, for F). These affective divergences from baseline show how emotionally-intensive is COVID-related discourse.

4 Analysis of Topical Distinctions

We study topical distinctions in male vs. female COVID-related discussions with two complementary analyses: (1) comparison of topics found by topic modelling over each of the M and F subcorpora separately, and (2) comparison of the distribution of dominant topics in M vs. F posts as derived from a topic model over the entire M+F dataset.

For each analysis, we used a publicly-available topic modeling tool (MALLET, McCallum, 2002). Each topic is represented by a probability distribution over the entire vocabulary, where terms more characteristic of a topic are assigned a higher probability.[9] A common way to evaluate a topic learned from a set of documents is by computing its *coherence score* – a measure reflecting its overall quality (Newman et al., 2010). We assess the quality of a learned model by averaging the scores of its individual topics – the *model* coherence score.

Analysis of Cross-gender Topics. Here we explore topical aspects of the productions of the two genders by comparing two topic models: one created using M posts, and another using F posts, in the COVID dataset. We selected the optimal number of topics for each set of posts by maximizing its model coherence score, resulting in 8 topics

[9] Prior to topic modeling we applied a preprocessing step including lemmatization of a post's text and filtering out stopwords (the 300 most frequent words in the corpus).

M-1	M-2	M-3	M-4	F-1	F-2	F-3	F-4
money	week	case	fuck	virus	feel	mask	week
economy	health	rate	mask	make	thing	hand	test
business	close	spread	claim	good	good	wear	hospital
market	food	hospital	news	thing	friend	woman	sick
crisis	open	week	post	vaccine	talk	food	patient
make	travel	month	comment	point	make	face	symptom
economic	supply	testing	call	happen	love	call	doctor
pandemic	store	social	article	human	parent	store	positive
lose	stay	lockdown	chinese	body	anxiety	close	start
vote	plan	measure	medium	study	read	stay	care

Table 2: Most coherent topics identified in male (**M-1**–**M-4**) and female (**F-1**–**F-4**) COVID-related posts.

	Topic	Keywords	Male	Female
1	**Economy**	money, business, make, month, food, economy, market, supply, store, cost	**0.17**	**0.10**
2	**Social**	feel, thing, live, good, make, friend, talk, love, hard, start	**0.07**	**0.26**
3	Distancing	close, social, health, open, plan, stay, travel, week, continue, risk	0.09	0.11
4	Virus	virus, kill, human, disease, study, body, spread, effect, similar, immune	0.11	0.07
5	Health (1)	mask, hand, stop, make, call, good, wear, face, person, woman	0.07	0.08
6	Health (2)	case, test, hospital, rate, spread, patient, risk, care, sick, testing	0.17	0.14
7	**Politics**	problem, issue, change, response, vote, policy, support, power, action, agree	**0.17**	**0.07**
8	Media	point, make, question, post, news, read, fact, information, understand, article	0.08	0.07
9	Misc.	good, start, thing, make, hour, stuff, play, pretty, find, easy	0.08	0.10

Table 3: Distribution of dominant topics in the COVID corpus. Entries in columns M(ale) and F(emale) represent the ratio of posts with the topic in that row as their main topic. Ratios are calculated for M and F posts separately (each of columns M and F sum to 1). Bolded topics indicate those with substantial differences between M and F.

for male and 7 topics for female posts (coherence scores of 0.48 and 0.46).

We examined the similarities and the differences across the two topical distributions by extracting the top 4 topics – those with the highest individual coherence scores – in each of the M and F models. Table 2 presents the 10 words with highest likelihood for these topics in each model; topics within each are ordered by decreasing coherence score (left to right). We can see that both genders are occupied with health-related issues (topics **M-3**, **F-1**, **F-4**), and the implications on consumption habits (topics **M-2**, **F-3**). However, clear distinctions in topical preference are also revealed by our analysis: men discuss economy/market and media-related topics (**M-1**, **M-4**), while women focus more on family and social aspects (**F-2**). Collectively these results show that the established postulates regarding gender-linked topical preferences are evident in spontaneous COVID-related discourse on Reddit.

Analysis of Dominance of Topics across Genders. We next performed a complementary analysis, creating a topic model over the combined male and female sub-corpora, yielding 9 topics.[10] We calculate, for the two sets of M and F posts, the distribution of dominant topics – that is, for each of topics 1–9, what proportion of M (respectively F) posts had that topic as its first-ranked topic.

Table 3 reports the results; e.g., row 1 shows that the economy is the main topic of 17% of male posts, but only 10% of female posts. We see that males tend to focus more on economic and political topics than females (rows 1 and 7); conversely, females focus far more on social topics than did males (row 2). Once again, these findings highlight cross-gender topical distinctions in COVID discussions on Reddit in support of prior results.

5 Conclusions

A large body of studies spanning a range of disciplines has suggested (and corroborated) assumptions regarding the differences in linguistic productions of male and female speakers. Using a large dataset of COVID-related utterances by men and women on the Reddit discussion platforms, we show clear distinctions along emotional dimensions between the two genders, and demonstrate that these differences are amplified in emotionally-intensive discourse on the pandemic. Our analysis of topic modeling further highlights distinctions in topical preferences between men and women.

[10] We used the model with the 2nd-best number of topics (9, coherence score 0.432) as inspection revealed it to be more descriptive than the optimal number of topics (2, score 0.450).

Acknowledgments

This research was supported by NSERC grant RGPIN-2017-06506 to Suzanne Stevenson, and by an NSERC USRA to Jai Aggarwal.

References

Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. 2020. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of Medical Internet Research*, 22(4):e19016.

Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.

Louisa Burriss, DA Powell, and Jeffrey White. 2007. Psychophysiological and subjective indices of emotion as a function of age and gender. *Cognition and emotion*, 21(1):182–210.

Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.

Alexis Conneau and Douwe Kiela. 2018. SentEval: An Evaluation Toolkit for Universal Sentence Representations. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, pages 1699–1704.

Martin L Hoffman. 2008. Empathy and prosocial behavior. *Handbook of emotions*, 3:440–455.

Geoff Hollis and Chris Westbury. 2016. The principals of meaning: Extracting semantic dimensions from co-occurrence models of semantics. *Psychonomic Bulletin and Review*, 23(6):1744–1756.

Bennett Kleinberg, Isabelle van der Vegt, and Maximilian Mozes. in press. Measuring Emotions in the COVID-19 Real World Worry Dataset. Association for Computational Linguistics.

Moshe Koppel, Shlomo Argamon, and Anat Rachel Shimoni. 2002. Automatically categorizing written texts by author gender. *Literary and linguistic computing*, 17(4):401–412.

William Labov. 1990. The intersection of sex and social class in the course of linguistic change. *Language variation and change*, 2(2):205–254.

Robin Lakoff. 1973. Language and woman's place. *Language in society*, 2(1):45–79.

Qian Liu, Zequan Zheng, Jiabin Zheng, Qiuyi Chen, Guan Liu, Sihan Chen, Bojia Chu, Hongyu Zhu, Babatunde Akinwunmi, Jian Huang, Casper J. P. Zhang, and Wai-Kit Ming. 2020. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. *Journal of Medical Internet Research*, 22(4):e19118. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

May Oo Lwin, Jiahui Lu, Anita Sheldenkar, Peter Johannes Schulz, Wonsun Shin, Raj Gupta, and Yinping Yang. 2020. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. *JMIR Public Health and Surveillance*, 6(2).

Andrew Kachites McCallum. 2002. MALLET: A machine learning for language toolkit.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184.

Anthony Mulac. 2006. *The gender-linked language effect: Do language differences really make a difference?* Lawrence Erlbaum Associates Publishers.

Anthony Mulac, James J Bradac, and Pamela Gibbons. 2001. Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. *Human Communication Research*, 27(1):121–152.

David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 100–108. Association for Computational Linguistics.

Matthew L Newman, Carla J Groom, Lori D Handelman, and James W Pennebaker. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3):211–236.

Gregory Park, David Bryce Yaden, H Andrew Schwartz, Margaret L Kern, Johannes C Eichstaedt, Michael Kosinski, David Stillwell, Lyle H Ungar, and Martin EP Seligman. 2016. Women are warmer but no less assertive than men: Gender and language on facebook. *PloS one*, 11(5).

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pages 3982–3992.

Jonathan Schler, Moshe Koppel, Shlomo Argamon, and James W Pennebaker. 2006. Effects of age and gender on blogging. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, volume 6, pages 199–205.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Massimo Stella, Valerio Restocchi, and Simon De Deyne. 2020. #lockdown: Network-Enhanced Emotional Profiling in the Time of COVID-19. *Big Data and Cognitive Computing*, 4(2).

Mike Thelwall and Saheeda Thelwall. 2020. Covid-19 tweeting in English: Gender differences. *El Profesional de la Información*, 29(3).

Mike Thelwall, David Wilkinson, and Sukhvinder Uppal. 2010. Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology*, 61(1):190–199.

Isabelle van der Vegt and Bennett Kleinberg. 2020. Women worry about family, men about the economy: Gender differences in emotional responses to covid-19. In *Proceedings of the 12th International Conference on Social Informatics*.

Achim Zeileis, Francisco Cribari-Neto, Bettina Grün, and I Kos-midis. 2010. Beta regression in r. *Journal of statistical software*, 34(2):1–24.