

Body Shame Twitter Movement Text Mining and Data Analysis by Applying MDCOR

Xinning Li^{1,a,*}, Ruiming Li^{2,b}, and Yihan Zheng^{3,c}

¹*Shanghai World Foreign Language Academy, Shanghai, 200233, China*

²*Faculty of Social Sciences, The University of Hong Kong, Hong Kong, China*

³*Shanghai Qibaodwight High School, Shanghai, 201101, China*

a. lixinning2030@163.com, b. Liruiming0103@gmail.com, c. 545773563@qq.com

**corresponding author*

Abstract: As one of the most heated-discussed movements on Twitter these years, tweets in #bodyshame has been complicated and debatable. To explore the alternation of attitudes toward #bodyshame and how it has changed in the past decade, the study analyzes the data collected from three-time frames: 2012, 2017, and 2022, respectively. The increasing popularity and awareness of #bodyshame of the public are shown in this work. By applying MDCOR, an open-ended responses classification framework to data analysis, this paper was able to conclude the gradually alternating trend of Twitter users' attitudes toward body shame. Social media data analysis helps the understanding of the general development trend of #bodyshame movement, which ultimately provides a comprehensive overview of people's acceptance and opinions towards various body types. This study's data-oriented research on attitudes towards body shame is profound in meaning.

Keywords: body shame, Twitter, MDCOR, text mining, data analysis

1. Introduction

People always define body shame as the act of discounting or mocking others' bodies. These commentaries are towards aspects such as but not limited to a person's size, age, hair, clothes, food, hair, or level of perceived attractiveness [1].

Though anti-body shame was a positive movement at the beginning, now it shows its toxic and extreme side [2]. There are multi-factors affecting it: behaviors of celebrities, changes in beauty standards, and transformation in the definition of body shaming, etc... For example, Adele, an English singer, songwriter, and anti-body-shaming influencer, constantly suffered from cyberbullies for losing weight herself.

As for the group of people who hold a negative attitude toward this movement, they simply take the losing weight and disseminating the benefits of keeping fit as equivalent to exacerbating body shaming.

The views and attitudes tend to be complicated and controversial. This study uses network analysis to explore the alternation of attitudes toward the body-shaming movement on social media (Twitter) and how it has changed in the past decade.

2. Methodology

2.1. Network Analysis (NA)

Network analysis (NA) is a set of integrated techniques to depict relations among actors and to analyze the social structures that emerge from the recurrence of these relations [3].

Network analysis meets its systematical development on the basis of the concept “social network”. It was in the 1930s that Jacob Moreno and Helen Jennings introduced basic analytical methods [4]. And it was not until 1954 that John Arundel Barnes started using the term systematically to denote patterns of ties, encompassing concepts traditionally used by the public and those used by social scientists: bounded groups (e.g., tribes, families) and social categories (e.g., gender, ethnicity). Today, social network analysis has been extensively used in various academic disciplines as well as practical applications such as countering money laundering, security applications and terrorism [5].

2.2. Data Collection

This paper collects text data under the Twitter hashtag (keyword) #bodyshame from three timeframes, from January 1st to March 31st, 2012, from January 1st to March 31st, 2017, and from January 1st to March 31st, 2022. The paper attempts to generate 500,000 tweets for each timeframe, but there are only 23 tweets relevant to #bodyshame in the 2012 timeframe, 4890 tweets from 2017 are collected, and 12,356 tweets from 2022. This paper uses R Console (64-bit) to generate text analysis data from Academic Twitter.

2.3. Machine Driven Classification of Open-ended Responses (MDCOR)

MDCOR is an analytic framework and software application driven by machine learning and text classification. It runs locally and does not require internet access. Thus, it is suitable to analyze restricted or protected data.

Datasets should contain a column of response ID and a column of original texts to be run successfully in MDCOR. After loading the dataset, the program can start to conduct text mining and give the “initial text mining output”. Researchers can then choose to remove the most common words according to the ‘text mining output’. Machine learning sampling parameters selection is required before executing metrics for optimal number of code selection. Researchers may select the optimal number of codes based on the metrics assessment. The final step is to execute MDCOR of the selected number of most representative responses per code.

It effectively reduces subjectivity, human coding errors and classification errors, and yields time-saving and reliable outputs. At the same time, MDCOR provides researchers with fully classified datasets that could possibly be merged with the original quantitative datasets to apply traditional quantitative analyses [6].

3. Result

3.1. Dataset: Bodyshaming1

The word “bodyshame” occupies the top of the most frequent words list. However, it is simply the hashtag used to obtain tweets. Since it does not contribute to the understanding, the study trims “bodyshame”. As for this dataset, 500 burn-in samples and 1000 MCMC/Gibbs resampling were chosen. Based on the principal of minimizing CaoJuan2009 point and maximizing Griffiths2004 point, the number of codes to executed the MDCOR is adjusted 5(Figure1). The MDCOR output shows 5 bubbles that are far apart from each other, no overlap. “Please” and “share” occupy the most relevant terms for Topic 1(Figure2). Thus, it is deduced that the topic is asking people to share their

experience of being body shamed.

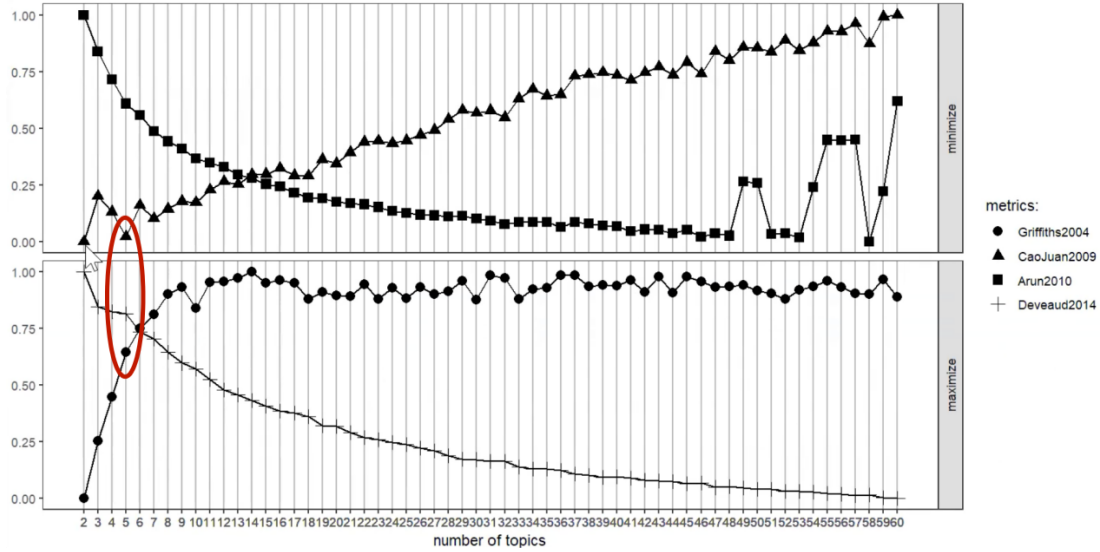


Figure 1: Metrics Plot for #bodyshame in 2012.



Figure 2: Summary statistics of latent code 1 for #bodyshame in 2012.

3.2. Dataset: Bodyshaming2

The size of the dataset expands significantly in 2017, indicating the rising awareness of the public and growing discussion of this movement. Some of the most frequently used words are really, look, want and honestly (Figure3). These words appear to be positive as they show a phenomenon that people who had once been plagued by body shame were trying to confront their body features, including those “drawbacks”. One term worth noting is “bully”. According to the context, people believed that appearance act as a dominant factor when judging and evaluating others. Many people were bullied, humiliated, and cyberbullied because of their appearance.

```
<<DocumentTermMatrix (documents: 2177, terms: 294)>>
Non-/sparse entries: 12515/627523
Sparsity : 98%
Maximal term length: 24
Weighting : term frequency (tf)
  words freq
1   really 302
2    look 301
3    want 298
4  honestly 260
```

Figure 3: Initial text mining output for #bodyshame in 2017.

The most frequent word pair in topic 2 is shame and female, indicating women were more likely to perceive body shame or to be the victims themselves (Figure4). “Fandom” also appears in this list, suggesting that body shame has its relevance to celebrities. The beauty standards and fashion trend is usually set or lead by celebrities due to fandom and medias. In Topic 3, the most frequent words are “nothing”, “judge”, and “find”. These words suggest that people were encouraged to take judgments or comments of others toward their bodies as “nothing”. In Topic 4, the most frequent words are “slutshame” and “cheap”. This suggests that people believe poverty is also cause of body shame.

“Really”, “look”, “want” are the top 3 most frequent terms in Topic 5. There exists a trend that people resist body shaming through their actions. They spontaneously showed their appearances and did not care about others’ opinions.

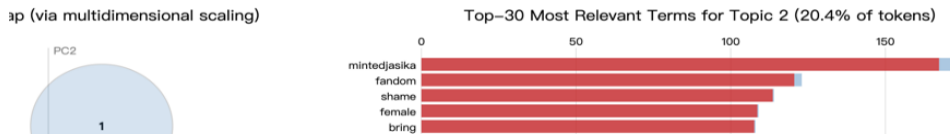


Figure 4: Summary statistics of latent code 2 for #bodyshame in 2017.

3.3. Dataset: Bodyshaming3

Initially, the text mining output directly generated through MDCOR provides basic information about the hashtag #bodyshame in 2022, including the total times it appeared on Twitter, relevant words, maximal term length, and term frequencies shown in Figure 5. The number of data collected within the 2022-time frame displays a sharp contrast with 2017 and 2012. In addition, words related to the body shame topic revolve around "people," "look," and "like," revealing a highly anthropic and societal concern within the issue.



Figure 5: Initial text mining output for #bodyshame in 2022.

When the number of topics is 6, an equilibrium between minimization and maximization is achieved. Identifying optimal codes allows the framework to generate a visual representation of the most frequent words in each code. Such a visualization in Figure 6 helps interpret the codes more clearly. It shows that code 6 includes words that are more generally related to body shame topics such as “sexuality” and “dream”, which reflects people’s general concerns while talking about the issue.

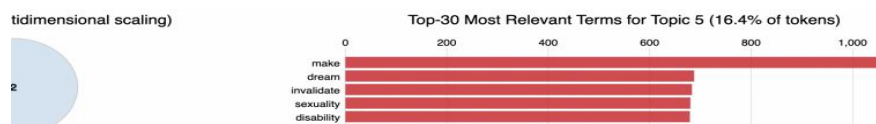


Figure 6: Summary statistics of latent code 6 for #bodyshame in 2022.

Conversely, code 3 contains words with more emotional content (Figure 7). Words such as “good”, “feel”, “right”, “disgust”, and “wrong” display people’s emotional reactions towards body shame

issues.

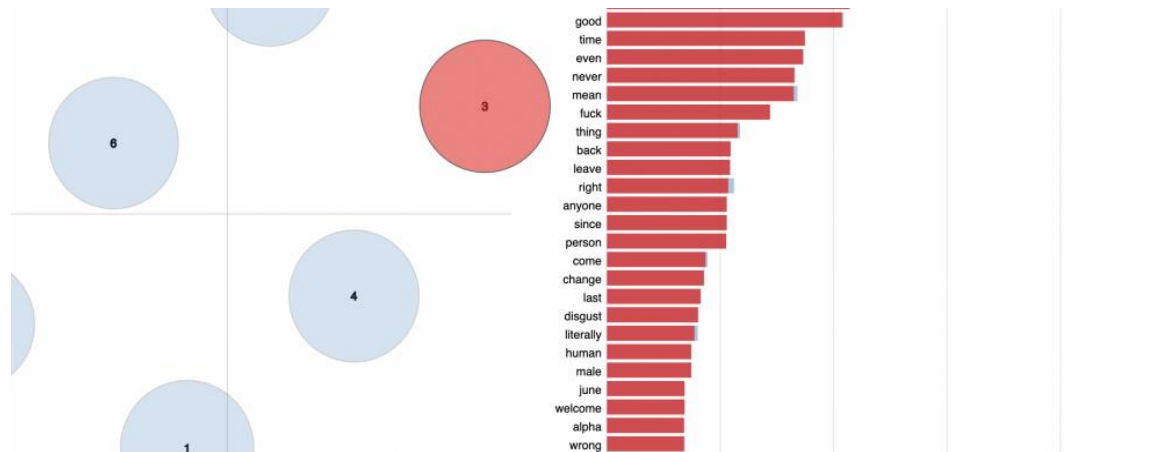


Figure 7: Summary statistics of latent code 3 for #bodyshame in 2022.

Additionally, the most frequent word appeared in code 2 is “woman”, which shows that the topic generates concerns and public discussions on Twitter mainly around women. Words such as “girl” and “sweet” further justify that tendency.

4. Discussion

4.1. Bodyshaming1

Research show that the anti-body shame movement and the third wave of body positivity was began in 2012. But as seen from the number of tweets collected, the #bodyshame social media movement was not yet popular among Twitter users at least not at the beginning of 2012.

4.2. Bodyshaming2

The body shame social media movement entered a middle period in 2017. The middle period is reflected in two aspects. Firstly, people began to realize that factors other than the body itself can also contribute to body shaming, such as low income. Secondly, it is concluded that women are often the main victims of body shame. In this middle period of the movement, people began to oppose and resist body shame through the hashtag. Previously, many people who suffered body shame were reluctant to show the real themselves, as they were worried and anxious about their bodies. However, with this hashtag, people were encouraged and thus willing to show the “real” and natural sides of their appearance. They were trying to spread their dissenting opinions towards body shame and were dedicated to promoting body positivity.

Facts were revealed, stories were told. Words such as “bully”, “slay” and “joke” indicate that people began to know more facts-many people suffered from bullies and humiliates due to body shame. Body shame are negatively viewed during this period. Encouragements were spread: When one receives a negative judgment about their body from others, they are supposed to think of it as “nothing”.

The higher number of hits also indicates that people were giving more attention to the movement. More people began to realize the severity and harmfulness of body shame as compared to before.

It is worth mentioning that “Adele” and “Lady Gaga” also appear to be in the most frequent word list. This shows a phenomenon that celebrities and influencers were engaged in this movement in this time period. They were either advocate of anti-body shaming, dedicated to spreading protests against body shaming using their influence. Or had aroused controversy within this movement. For example,

Adele's being blamed for body shaming others as an anti-body shame advocate herself (mentioned in the introduction).

4.3. Bodyshaming³

Data analysis of the 2022 dataset displays a vast difference from previous years' results.

Firstly, there is a general trend of increasing attention on this body shame movement, which can be seen from many tweets (number of "documents" in the framework). Such an increase in discussion on #bodyshame represents an increasing awareness of body identity in society. And the change is, to a certain extent, related to the growing level of Internet penetration, since the diverse public opinions on this movement generate a number of debates and discussions.

Secondly, concerns and discussions on the body shame topic are mainly around gender issues, with a strong tendency towards females. According to summary statistics of different latent codes, the body shame issue seems more related to women than men. A possible reason is that social norms make women feel dissatisfied with their bodies and thus depressed. Besides, stereotypes toward women's aesthetics have caused intensive judgments on their appearances. Moreover, the image-heavy era of the media generates countless "perfect" bodies, which led to higher attention paid to women's bodies.

Additionally, the increasing awareness of body shame topics has led to emotional responses from many tweets. Emotions (Figure 7) vary from positive to highly negative, reflecting the issue's complexity and proving this study's significance. Social problems are complex due to the controversial responses they generate, thus leading to debates between different groups. It is difficult to conclude a "correct" opinion. At the same time, the increasing attention paid to the body shame issue is an aspiring signal of people's recognition and concerns about social events, ultimately contributing to societal progress.

5. Conclusion

The study initially hypothesized that the body shame movement has changed from positive to toxic in the past decade since Adele's case had several controversies. While a series of positive phenomena and trends acquired by applying data analyses demonstrated that # bodyshame is still much of a positive movement than a toxic one. The growing popularity of this movement and the rising public awareness of this topic are shown. People are less anxious and worried about their physical drawbacks. They are instead spontaneous to show the real and natural side of their appearance. They are willing to share their experience of being body shamed and how they have overcome it by changing their mindsets.

6. Limitations

This study attempts to study the alternation of Twitter users' attitudes toward the body shame issue in the past decade. But the study did not fully illustrate the growth of the movement from its beginning to its "middle period" (4.2) since the period during which the movement significantly grew in its popularity does not meet with the time frame (2012.1.1-2012.3.31) of the study.

As the study has removed non-English tweets from the dataset, the study is mainly focused on the tweets in English. Thus, the results are not considered universal.

The study does analyze the influence of celebrities and influencers through word frequency. But the study does not examine the significance of the number of retweets, number of replies, number of likes, and number of quotes. Little attention is given to the influence of active users in this hashtag and the power of some heated-discussed tweets.

7. Future Inquiry

Future studies could look into how sub-movements of #bodyshame affect the overall trend. For example, #TheySaid (2017) is a Twitter movement initiated by Sally Bergesen. It encouraged women to share their experiences of being body shamed with this hashtag. Another sub-movement worth discussing is #BigThighTwitter, a body positivity movement on Twitter to highlight the ridiculous and unrealistic beauty standards. People shared images of so-called “big thighs” rather than hiding them with unflattering clothing.

As suggested in the Limitation, future studies could devote more effort to analyzing the meaning of retweet numbers, reply to numbers, like numbers and quote numbers, and the significance of active users and popular tweets.

Studies that analyze the universal trend of the #bodyshame movement, i.e., keeping the non-English, texts, could be conducted.

Future inquiries could also be conducted to seek alternate points within this movement. For example, from which point did the users start to post photos of their own to show their endorsement of anti-body shaming? Or when did the users begin to analyze the causation of body shame?

References

- [1] Ariane Resnick, C. N. C. (2022), *What is body shaming?* very well mind. <http://www.verywellmind.com/what-is-body-shaming-5202216>
- [2] Erin Nolen(2019), *Is the body positivity Social Movement Toxic?*, UT News. <https://news.utexas.edu/2019/12/18/is-the-body-positivity-social-movement-toxic/>
- [3] A.M. Chiesi(2001), *Network Analysis*, ScienceDirect. <https://www.sciencedirect.com/topics/social-sciences/network-analysis>
- [4] Freeman, Linton C (2004), *The development of social network analysis*, Empirical Press.
- [5] Wikipedia org(2022), *Social Network Analysis*, Wikipedia. https://en.wikipedia.org/wiki/Social_network_analysis#History
- [6] Manuel S. González Canché(2020), *Machine Driven Classification of Open-ended Responses (MDCOR)*.