Computational Social Science Lab 2024/2025

Talking Bodies: Gendered Discourse on Body Image in Instagram Comments
A Computational Social Science Analysis

Francesca Michieletto (ID 255371)

Keywords
Body, Computational Social Science, Gender differences, Instagram, Comments, Text Analysis, Sentiment Analysis

1 Introduction

In the past few decades, media landscapes have changed dramatically. Communication, once dominated by broadcast systems operating on a "one-to-many" model, has shifted toward a participatory model shaped by the rapid spread of social media. Platforms such as Facebook, Instagram, YouTube, and TikTok have turned audiences into active participants and simultaneously producers and consumers of content. This shift has transformed how information circulates, how communities take shape, and how cultural meaning is negotiated [6].

Social networking sites (SNSs) have therefore become a cornerstone of digital culture. Unlike traditional media, they are characterized by two key features: content is largely peer-generated, and interaction is built into the design. Users can create profiles, share material, and engage with others through likes, comments, and other forms of feedback [10]. This "many-to-many" model has opened up new opportunities for communication and self-expression, while also contributing to what has been described as a demotic turn, where ordinary individuals can achieve visibility and recognition through memes, viral videos, or blogs [6].

Yet these opportunities are often fleeting and rarely supported by the financial or social capital of traditional fame. The spread of broadband and mobile devices has further reinforced the "always-on" nature of social media, encouraging both celebrities and aspiring public figures to share details of their daily lives on a near-constant basis [6].

Within this participatory environment, the body has become a central medium of communication and self-presentation. Social media not only circulates ideas and narratives but also invites users to manage, display, and negotiate their bodies in highly visible spaces. This visibility can foster empowerment and self-expression, but it also exposes individuals to judgment, comparison, and forms of regulation deeply embedded in digital culture.

For these reasons, this paper focuses on Instagram, where visual self-presentation and body management are especially prominent. Instagram is one of the most widely used social networking sites, particularly among younger generations [10]. According to the Digital 2025 Global Overview Report, 63.9% of the global population has a social media account (5.24 billion users), a 4.1% increase from the previous year. Internet users aged 16 and above spend an average of 2 hours and 21 minutes per day on social platforms, with Instagram being the preferred choice for 16.6% of active users, especially those between 16 and 34 years old [11].

Instagram's origins as a photo-sharing platform make it especially relevant for discussions about the body. Practices such as body shaming and body positivity frequently emerge, showing how digital cultures simultaneously reproduce and challenge normative ideals of appearance. With the growing use of online platforms, particularly among young people, body shaming has been amplified. This form of online aggression involves unsolicited and often negative comments about a person's body—whether related to size, shape, weight, or other features [3]. The anonymity of online interaction can further amplify these behaviours.

Against this backdrop, the present study explores how discourses about the body unfold in Instagram comment sections. Specifically, we analyse comments posted under images where the body is prominently displayed, such as selfies or full-body photographs. We place particular emphasis on comparing reactions to male and female creators. By investigating differences in tone, vocabulary, and emotional orientation, this study aims to show how gender shapes online talk about the body. We also consider whether engagement with such discourse might affect users' well-being.

In Western societies, both women and men are influenced by cultural ideals of physical appearance. Women are often confronted with the "thin ideal," which creates a gap between their actual bodies and media-portrayed ideals, frequently leading to body dissatisfaction (Mills, Musto, Williams, & Tiggemann, 2018). In recent years, however, the notion of the "ideal" male body has also gained strength. Media representations increasingly promote exaggerated muscular ideals for men, spread through advertising, television, and online content (Singh, Parsekar, & Bhumika, 2016) [7].

2 Literature Review

Research on Instagram and body image has increasingly embraced computational social science (CSS) approaches. These methods make it possible to combine large-scale data collection, advanced statistical modelling, and natural language processing (NLP). Compared to small-scale experiments or survey-based studies, CSS approaches allow for a more systematic analysis of how audiences react to body-related content in online environments and how such reactions contribute to broader cultural discourses.

One strand of research has focused on detecting harmful discourse. Grasso et al. (2023), for example, introduced the first dataset of Italian Instagram comments annotated for body shaming. They trained BERT-based language models to classify different forms of stigmatization, including fatphobia, skinny-shaming, and ableism. Their design combined large-scale scraping (almost 40,000 comments, reduced to 11,000 annotated) with supervised machine learning. The strength of this work lies in its scalability and its effort to capture nuanced, intersectional expressions of body-related discourse. At the same time, model performance was uneven due to category imbalance, and the restriction to Italian-language content limits its relevance for global debates on body image. This gap points to the need for similar approaches applied to English-language corpora, which dominate most international platforms [3].

Another important contribution comes from Pedalino and Camerini (2022), who studied how different Instagram practices—browsing, posting, and commenting—predict body dissatisfaction among young women, mediated by upward social comparison with peers and influencers. Their cross-sectional survey (n = 291) was analysed with structural equation modelling in R, which allowed them to test indirect pathways and assess model fit. The strength of the study lies in its integration of the Tripartite Influence Model with social comparison theory, offering fine-grained evidence that influencers, more than peers, play a key role in shaping dissatisfaction. However, reliance on self-report measures, the use of single-item indicators of Instagram use, and the cross-sectional design limit the ability to draw causal conclusions [8].

Survey-based CSS research has also contributed to understanding gender and cultural variation. Nimiya et al. (2024) conducted a balanced-sample survey of 212 Indian Instagram users (equal numbers of men and women) to examine sex differences in body image, exercise motivation, and social comparison. Using validated psychometric scales and inferential statistics (t-tests and ANOVA), they found that women report greater body concerns, while men show stronger exercise motivation. Both genders, however, engage in social comparison online. This study stands out for its methodological rigor, balanced sampling, and the development of an Instagram-specific social comparison scale. At the same time, its limitations lie in its cross-sectional design, possible self-report biases, and the cultural specificity of the Indian context, which may constrain generalization [7].

Finally, meta-analytic work has provided a broader perspective. Jiaqing et al. (2023) conducted a systematic review and bibliometric meta-analysis of 84 studies published between 2013 and 2023, following the PRISMA framework and using R-based bibliometric tools. Their methods included keyword co-occurrence mapping, centrality and density measures, and thematic clustering. The results confirmed consistent links between social media use, body dissatisfaction, and negative well-being outcomes, while also highlighting positive aspects such as greater social connection and opportunities for creative self-expression. The main strengths of this review are its comprehensive scope and methodological transparency [4]. Yet, reliance on Scopus-indexed publications and bibliometric mapping means that pooled effect sizes typical of traditional meta-analyses are missing, and the strong focus on female samples reduces inclusivity.

2.1 Research Question

Building on the reviewed literature, this study approaches the Instagram comment space as a setting where bodies are not simply displayed but actively negotiated. Prior research has shown that comments can both amplify and mitigate body image concerns, but most studies have focused primarily on women, relied on cross-sectional designs with limited ecological validity, and rarely analysed comments at scale. Moreover, systematic comparisons of responses to male and female creators remain scarce, and few works have examined whether linguistic, thematic, and emotional differences emerge in these interactions.

Against this background, the central research question guiding the project is: How do audiences react in the comment sections of Instagram posts that make the body visible, and how do these reactions differ depending on the creator's gender?

To answer this question, the study pursues three interconnected objectives. First, it maps lexical and structural patterns in body-related comments through document-feature matrices, keyness analysis, and semantic network modelling. Second, it identifies thematic differences using topic modelling (LDA and STM) to assess whether particular themes are more strongly linked to male or female creators. Third, it analyses the emotional dimension of comments with sentiment and emotion lexicons, in order to capture differences in tone and affective framing across genders.

Through this multi-method design, the study aims to uncover systematic gendered differences in audience reactions—highlighting not only the words and topics that dominate body-related discussions, but also the emotional orientations that shape how men's and women's bodies are evaluated on Instagram. The project thus offers a comparative analysis of discourse aimed at male and female creators, focusing on comments posted under images in which the creator's body or face is clearly visible.

The analytical strategy combines lexical frequency analysis, keyness comparison, semantic network analysis, topic modelling, and sentiment/emotion profiling. These methods are well suited to extract meaningful patterns from the short, noisy, and informal style typical of Instagram comments. To maintain methodological rigor and capture gender-specific linguistic and emotional dynamics, the study processes male and female datasets separately before comparing the results.

3. Project Design

3.1 Data Collection Strategy

The dataset for this project consists of Instagram comments collected from posts in which the creator's body is clearly visible, for example, selfies, full-body photographs, sports-related content, or modelling posts. The collection strategy was designed to balance ecological validity with methodological rigor, ensuring that the corpus reflects authentic user interactions while remaining suitable for systematic computational analysis.

To make male–female comparisons possible, we adopted a purposive sampling approach. In total, 40 public profiles were selected: 20 female creators (5 activists, 5 models, 5 athletes, 5 actresses) and 20 male creators (5 activists, 5 models, 5 athletes, 5 actors). From each profile, five posts were chosen, giving a dataset of 200 posts. Selection criteria required that the creator's body be clearly visible, that the account was public and accessible, and that each post contained enough comments to capture real audience interaction. To prevent over-representing individual accounts, only a limited number of comments were kept per post.

Data scraping was carried out using the IG Comments Export Tool together with Python-based solutions such as the Apify Instagram Scraper [12]. These tools retrieved both the text of the comments and minimal metadata (comment IDs, timestamps, usernames). All data came exclusively from public accounts, were anonymized at the user-identifier level, and analysed only in aggregated form.

All comments were pre-processed in R with the quanteda package. The following steps were applied to both male and female datasets:

1. Removal of non-linguistic elements (punctuation, numbers, symbols, URLs)
2. Tokenization adjustments (e.g., splitting hyphenated words and separating HTML tags)
3. Alphabetic filtering (keeping only alphabetic tokens)
4. Case normalization (conversion to lowercase)
5. Stop word removal (using a standard English stop word list)
6. Stemming (Porter algorithm for English)
7. Exclusion of social media mentions (e.g., @username)
8. Language detection to exclude non-English comments, ensuring comparability with NLP tools
9. Computation of co-occurrences between keywords for semantic network analysis, applying a minimum frequency threshold

Following preprocessing, comments were structured into a document-feature matrix (DFM), which formed the basis for subsequent analyses such as lexical frequency comparison, topic modelling, and semantic network construction. After trimming, the final corpora consisted of

45,322 comments in the male subset and 39,017 in the female subset. The male vocabulary was reduced from 12,030 to 1,214 terms, and the female vocabulary from 8,756 to 1,234. This ensured representativeness while avoiding sparsity. More conventional thresholds (e.g., excluding terms present in fewer than 5% or more than 95% of documents) proved too restrictive, leaving only a very small set of frequent words. By contrast, the adopted threshold (0.0001) maintained vocabulary diversity while remaining computationally feasible.

Ethical considerations were central throughout. Only comments from publicly accessible posts were included, and no attempt was made to access private content. The analysis focused exclusively on aggregated textual patterns, not individual behaviours. The study follows established standards in computational social science, aiming to minimize risks of harm while making responsible use of public digital data to investigate gendered patterns in body-related discourse.

3.2 Methodology

This methodological design addresses the gaps identified in previous research. We apply a large-scale, observational "text-as-data" approach to public Instagram comments, comparing the discourse directed at male and female creators.

The analysis combines multiple computational techniques to provide a multi-layered view of gendered, body-related discourse across lexical, structural, thematic, and affective dimensions. Document-Feature Matrices (DFMs) and keyness analysis are used to identify words that are statistically distinctive across genders, while semantic network analysis models word co-occurrences to explore how terms cluster and which ones act as central connectors. To uncover latent themes and compare their prevalence, we employ probabilistic topic models, including Latent Dirichlet Allocation (LDA) and a Structural Topic Model (STM) with creator gender as a covariate. Finally, the NRC multi-emotion lexicon is applied to assess the polarity and emotional orientation of comments.

3.2.1 Lexical and Structural Analysis

To address the first objective, we constructed separate document-feature matrices (DFMs) for the male and female datasets, which provided the foundation for the lexical and structural analyses. Building on these lexical results, we created semantic co-occurrence networks for each dataset, with nodes representing words and edges indicating co-occurrence within the same comment. To limit noise and enhance interpretability, low-frequency terms were excluded through thresholding, and only the giant component of each network was retained. We then calculated degree, betweenness, and closeness centrality to identify structurally important terms.

In the female network, central nodes included appearance- and emotion-laden words such as *beautiful, love,* and *sexy*, suggesting that discourse clusters around aesthetic and affective evaluations. The male network, by contrast, was shaped by words linked to strength, performance, and camaraderie (e.g., *strong, win, bro*).

The two graphs contained 1,054 nodes for the male corpus and 871 for the female corpus, with network densities of 0.041 and 0.057 and average degrees of 0.020 and 0.029 respectively. These values confirm the sparse structure of the networks, which reflects the fragmented and short nature of Instagram comments. Single words, emojis, and short interjections reduce the likelihood of repeated co-occurrences, preventing the formation of dense clusters. In the male dataset, the presence of multilingual tokens (e.g., *hai, la, ya, para*) further dispersed the vocabulary. As a result, no single term dominates the network; instead, many low-frequency connections create graphs with numerous peripheral nodes and only a few weakly connected hubs. Centrality scores confirm this pattern: normalized degree values for the most connected nodes remain below 0.25 (max ≈ 0.23 for *can*), betweenness peaks at 0.05 (*can*), and closeness does not exceed 0.47 (*can*). Rather than being a methodological artifact, these relatively low values capture the heterogeneous and interactional character of male-directed discourse, which is spread across multiple small clusters rather than organized around a central lexicon.

In the female semantic network, both degree and betweenness centrality values remain very low. This indicates that no single word dominates the discourse or acts as a strong connector between clusters. Instead, the network is highly fragmented, with many terms occurring only in localized contexts and few serving as bridges across communities. Such sparsity reflects the short and often repetitive nature of Instagram comments, where expressions of admiration or affect are dispersed across multiple small clusters rather than concentrated around a cohesive set of central terms.

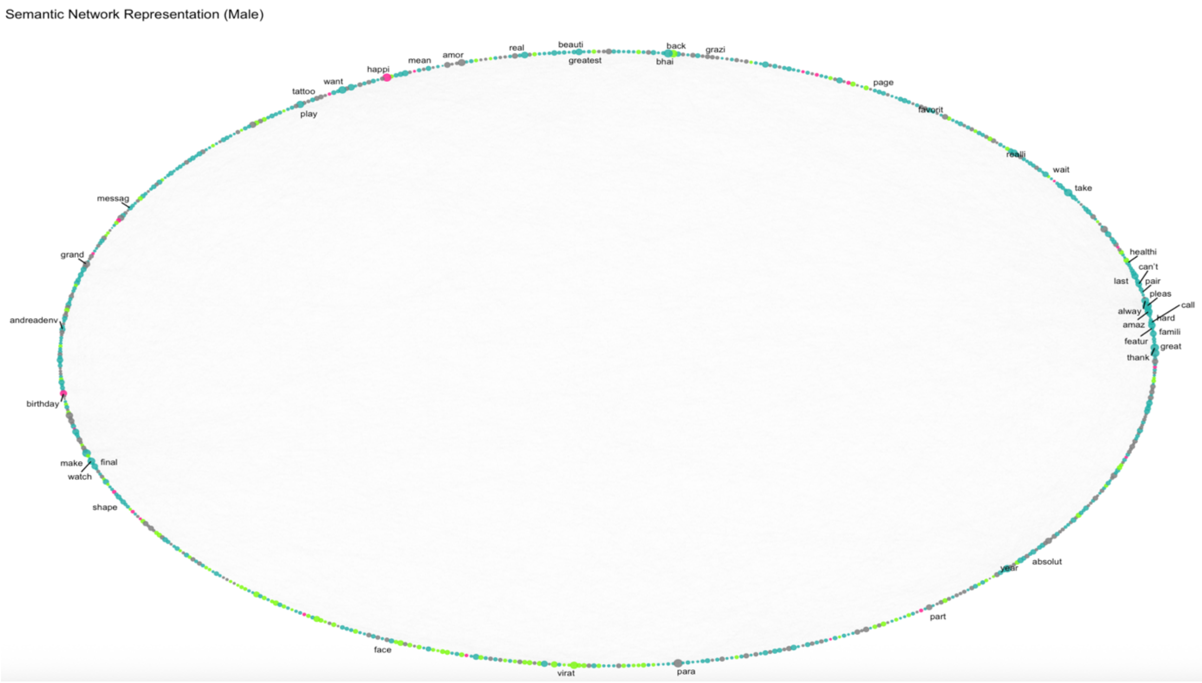

Figure 1. Semantic network Representation of male corpus

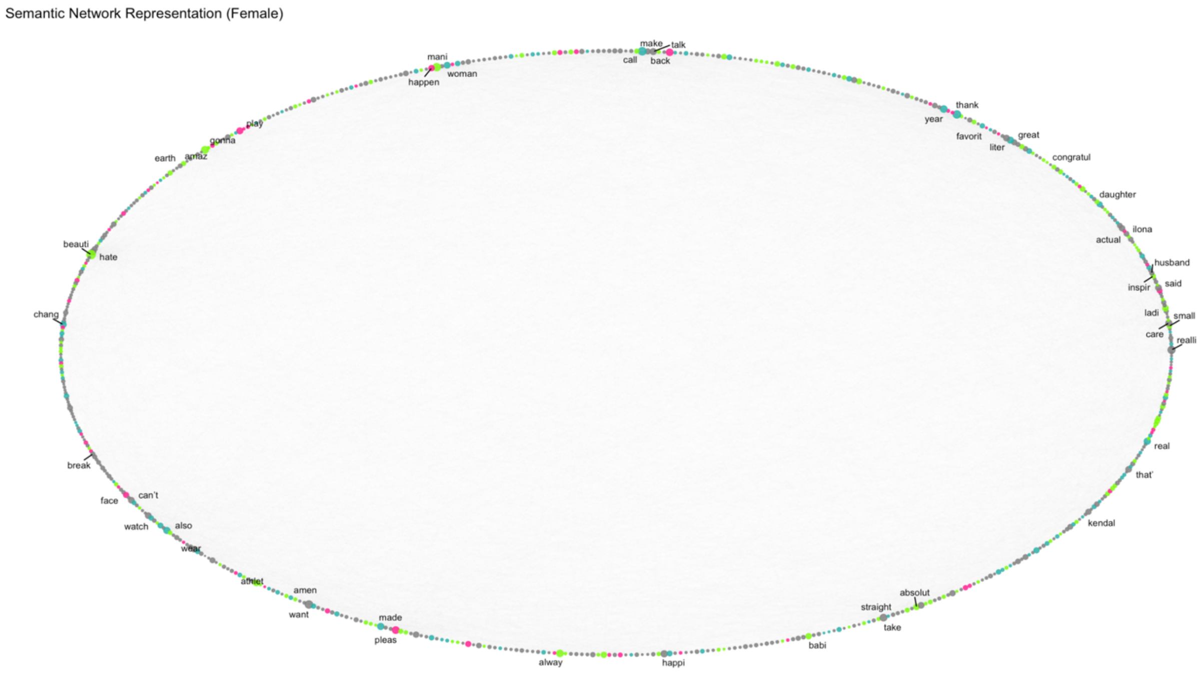

Figure 2. Semantic Network Representation of female corpus

The spatial arrangement of the networks also provides useful insight. Circular layouts (Figures 1 and 2) highlight overall density and balanced word distribution but compress peripheral nodes into uniform positions.

3.2.2 Thematic Analysis

To explore latent discourses and compare them across genders, we applied Latent Dirichlet Allocation (LDA) separately to the male and female corpora, and a Structural Topic Model (STM) to the combined corpus with gender included as a prevalence covariate. Unlike surface-level co-occurrence networks, topic models uncover hidden thematic structures by grouping words that tend to appear together into coherent semantic bundles.

Model selection was guided by semantic coherence, which measures the degree to which the most representative words of a topic co-occur across documents. Higher coherence values signal topics that are easier to interpret and more meaningful, whereas lower values point to noisier or less consistent clusters.

For comments directed at male creators, coherence values fluctuated widely, with a peak around K=10 topics. This pattern suggests that male-related discourse is fragmented and heterogeneous, characterized by a dispersed lexicon and multiple micro-themes, such as sports references, fandom, expressions of thanks, and multilingual slang. Capturing this diversity requires a relatively large number of topics.

The female corpus, in contrast, showed a more compact structure. Coherence peaked between K=6 and K=8 topics before dropping sharply, indicating that female-related discourse is concentrated around a small number of semantic cores, particularly those related to appearance and affective appreciation. Adding more topics in this case artificially fragmented clusters that were already cohesive, reducing interpretability.

These differences reflect the nature of Instagram comments themselves: short, informal, and often limited to a single word, emoji, or brief interjection. Within this format, male-directed comments tend to generate more heterogeneous discourses, where fandom, sports, and multilingual interactions coexist, while female-directed comments are more homogeneous, dominated by aesthetic evaluations and repeated formulas of appreciation.

On the basis of these diagnostics, we retained two main topics per gender, both semantically coherent and thematically interpretable. This ensured comparability while avoiding the excessive fragmentation that comes with larger K. For the female corpus, topics centred strongly on appearance and aesthetic admiration. One cluster was dominated by terms such as *beautiful, woman, gorgeous, amazing, pretty,* and *lady*, while another emphasized positive evaluation and emotional engagement, with words like *thank, happi, amaz, congrats,* and *birthday* (Figure 3). Together, these topics suggest that comments directed at women revolve around admiration for physical appearance, often intertwined with emotional support.

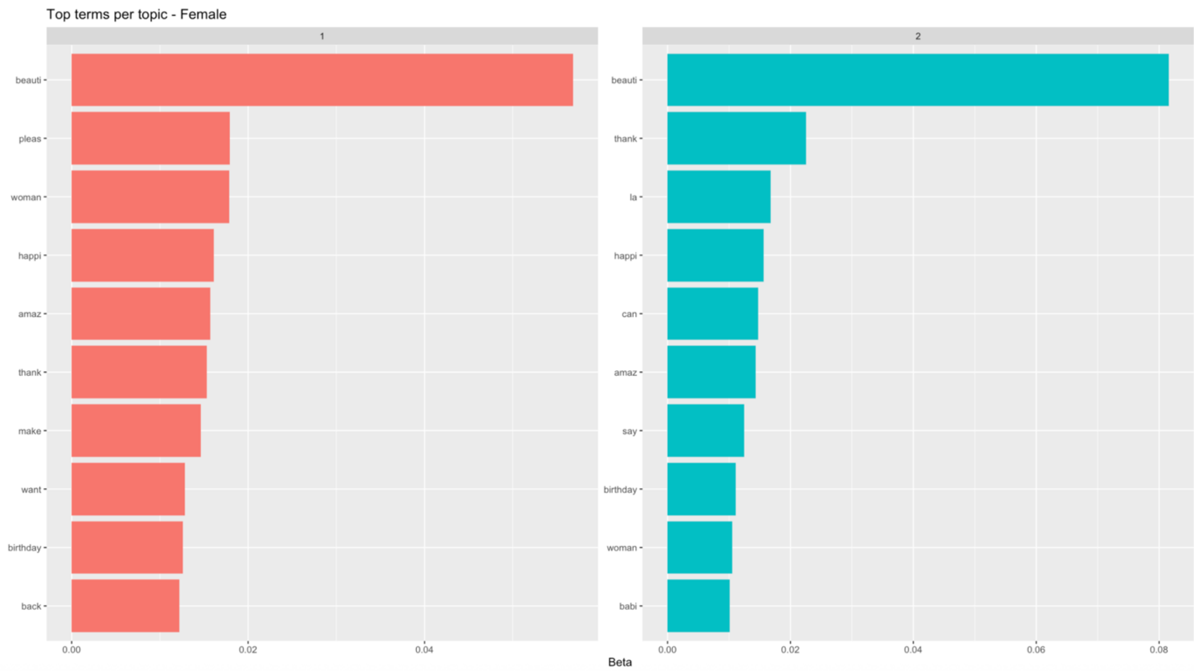

Figure 3. Top terms per topic plot (female)

Male-related topics were less tightly focused on appearance and more dispersed across identity and celebratory discourse. One topic revolved around terms such as *man, happi, thank, bro, handsome,* and *day*, reflecting camaraderie and informal recognition. Another emphasized fandom and celebratory talk, with words such as *birthday, goat, virat, fan, great,* and *thank* (Figure 4). These topics suggest that male creators are addressed more through identity markers, sporting admiration, and collective celebration than through appearance-based judgments.

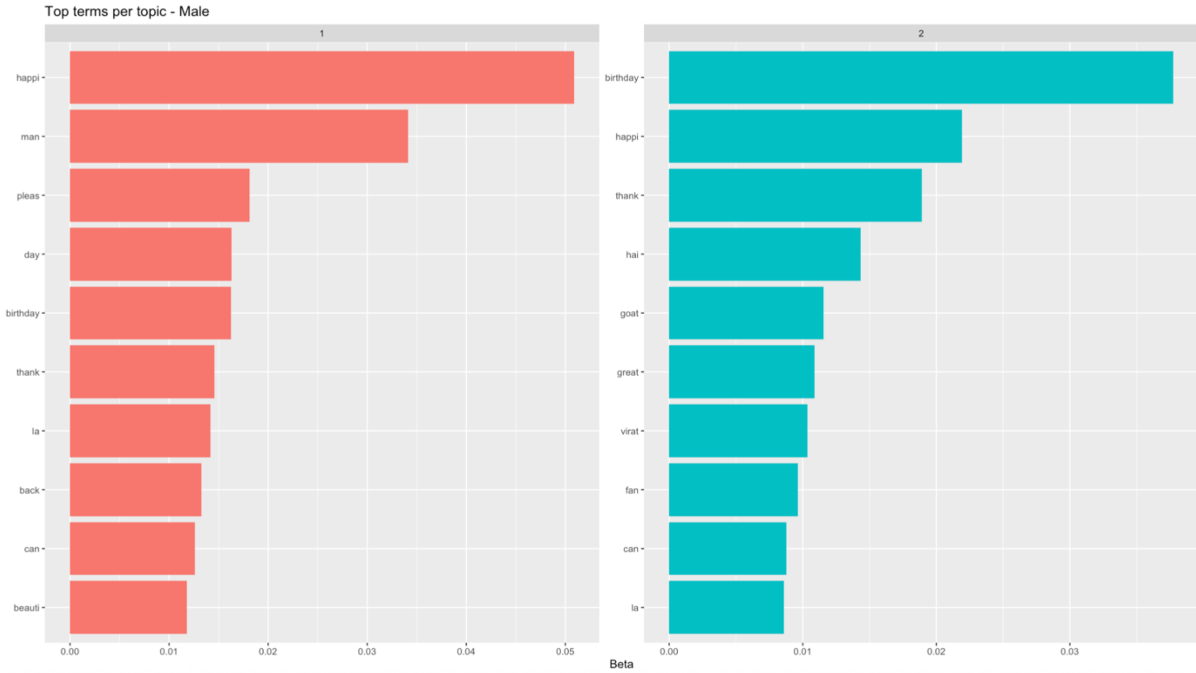

Figure 4. Top terms per topic plot (male)

STM analysis confirmed these tendencies. By including gender as a covariate, it showed that appearance and body-related topics were far more common in comments to female creators, while performance, fandom, and celebratory themes were more strongly associated with male creators. Cosine similarity scores between β-weighted topic vectors revealed only moderate overlap across genders, pointing to shared lexical bases but clear divergences in central themes: aesthetic admiration for women versus identity and fandom based engagement for men (Figure 5).

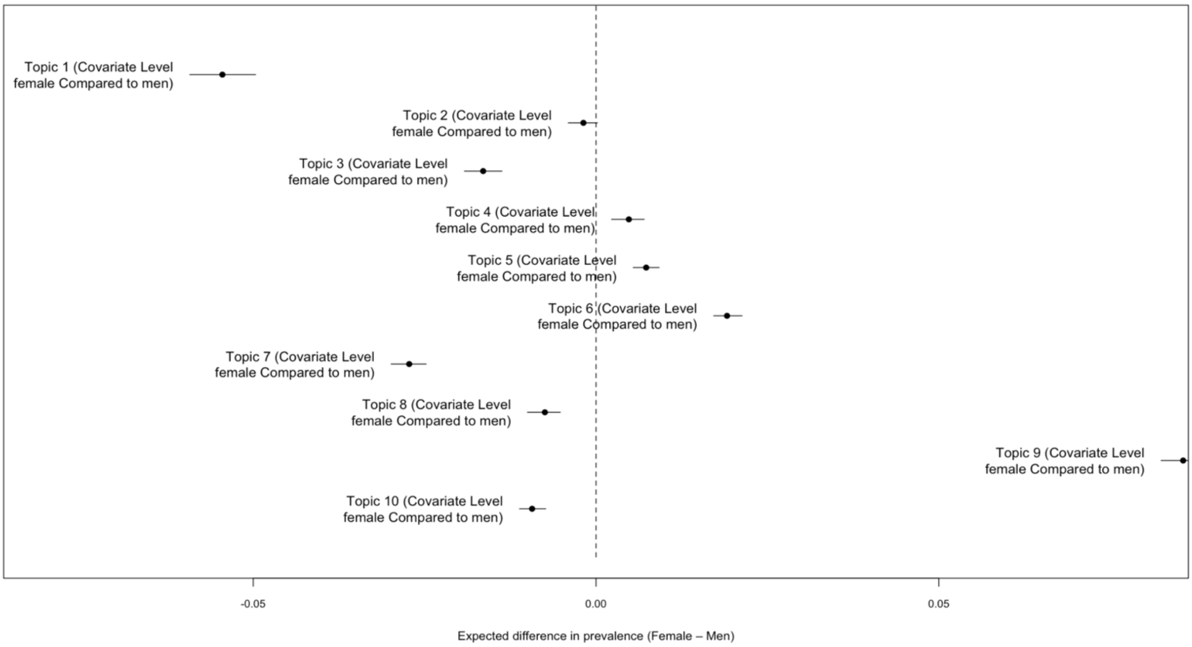

Figure 5. Cosine similarity plot

```
Topic 1 Top Words:
     Highest Prob: happi, birthday, mani, bad, hard, human, team, anyth, bath, may
     FREX: birthday, mani, bad, hard, human, team, anyth, bath, may, dad
     Lift: act, ain’t, amoooo, anyth, background, bath, beau, buena, date, generat
     Score: happi, birthday, mani, bad, hard, bday, human, dad, team, bath
Topic 2 Top Words:
     Highest Prob: pleas, say, mean, call, una, beach, chang, ha, mama, wanna
     FREX: pleas, say, mean, call, una, beach, chang, ha, mama, wanna
     Lift: amigo, así, bang, channel, facebook, hace, hahaha, hang, oscar, persona
     Score: pleas, say, una, mean, beach, call, mama, chang, ha, wanna
Topic 3 Top Words:
     Highest Prob: fan, goat, bhai, that’, ladi, stay, actual, vida, athlet, natur
     FREX: fan, goat, bhai, that’, ladi, stay, actual, vida, athlet, natur
     Lift: aren’t, aw, ay, bonita, exampl, gotta, pack, pass, sale, athlet
     Score: fan, goat, bhai, that’, ladi, stay, natur, actual, recogn, vida
Topic 4 Top Words:
     Highest Prob: amaz, great, day, want, make, real, para, wait, black, hair
     FREX: amaz, great, day, want, make, real, wait, black, hair, babe
     Lift: day, make, stand, ador, asf, babe, bar, bella, black, calvin
     Score: amaz, great, want, day, make, real, para, wait, hair, black
Topic 5 Top Words:
     Highest Prob: la, babi, year, mai, said, ma, liter, care, heart, famili
     FREX: la, babi, year, said, ma, liter, care, heart, famili, mayb
     Lift: disappoint, stomach, straight, babi, famili, heart, la, mayb, saw, ago
     Score: la, babi, year, ma, said, mai, liter, care, heart, dear
Topic 6 Top Words:
     Highest Prob: alway, woman, take, dream, congratul, share, yeah, anyon, ask, name
     FREX: alway, woman, take, dream, congratul, share, yeah, anyon, ask, name
     Lift: action, admir, angl, ask, ciao, comeback, dam, earth, easi, favourit
     Score: woman, alway, take, congratul, dream, yeah, share, ask, anyon, name
Topic 7 Top Words:
     Highest Prob: thank, damn, hai, ya, talk, ka, wear, esta, kya, head
     FREX: thank, damn, hai, ya, ka, wear, esta, head, read, daddi
     Lift: ahí, appreci, away, awww, cap, compar, cuando, daddi, dalla, das
     Score: thank, hai, damn, ya, ka, esta, talk, daddi, wear, read
Topic 8 Top Words:
     Highest Prob: back, way, da, can’t, made, favorit, ass, happen, okay, fabul
     FREX: back, way, da, can’t, made, favorit, ass, happen, okay, fabul
     Lift: hahahaha, back, actor, ah, alreadi, american, anybodi, appear, ass, bag
     Score: back, way, da, can’t, favorit, made, ass, okay, happen, fabul
Topic 9 Top Words:
     Highest Prob: beauti, man, amo, absolut, play, linda, amor, grand, awesom, face
     FREX: beauti, amo, absolut, linda, awesom, face, also, inspir, eat, ate
     Lift: ahh, ate, woah, also, linda, beauti, absolut, af, al, amiga
     Score: beauti, man, absolut, amo, linda, ate, awesom, also, face, slay
Topic 10 Top Words:
     Highest Prob: can, realli, watch, last, na, speak, agre, lmao, player, vai
     FREX: can, realli, watch, last, na, speak, agre, lmao, player, vai
     Lift: ain't, aliv, baddest, brain, breath, catch, clean, crack, creat, dia
     Score: can, realli, watch, last, na, vai, speak, agre, player, fat
```

Figure 6. List of Top Words for each topic

This reflects findings already present in the literature, which highlight the promotion of typical ideals such as the thin ideal for women and the muscular ideal for men [1][5][7]. Comments directed at female creators contained a higher share of appearance-focused terms such as *gorgeous, pretty, body,* and *dress*, whereas male creators attracted more sport- and performance-related vocabulary including *game, strong,* and *champion*.

3.2.3 Emotional and Sentiment Analysis

To examine the affective framing of comments, we used the NRC Emotion Lexicon, which classifies words into eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) as well as positive and negative polarity. Tokens were aligned through stemming and dictionary lookup, and corpus-level proportions were calculated separately for comments directed at male and female creators. This approach is particularly appropriate for Instagram, where comments are often short, informal, and rich in evaluative expressions. It allows the detection of affect even when it is conveyed through minimal textual units such as a single word or emoji.

The results show that both corpora are dominated by positive affect, though with noticeable gendered nuances. Comments addressed to female creators contained a higher share of positive words (29.7% vs. 26.2%) and more frequent expressions of joy (22.5% vs. 20.4%), suggesting that discourse aimed at women is framed primarily through admiration and affective reinforcement (Figure 7). In contrast, comments directed at men showed relatively higher levels of anticipation (12.7% vs. 7.8%), trust (13.5% vs. 10.9%), and surprise (8.1% vs. 6.3%). This pattern indicates that audiences engage with male creators more in terms of expectation, reliability, and collective enthusiasm, often linked to fandom and performance contexts.

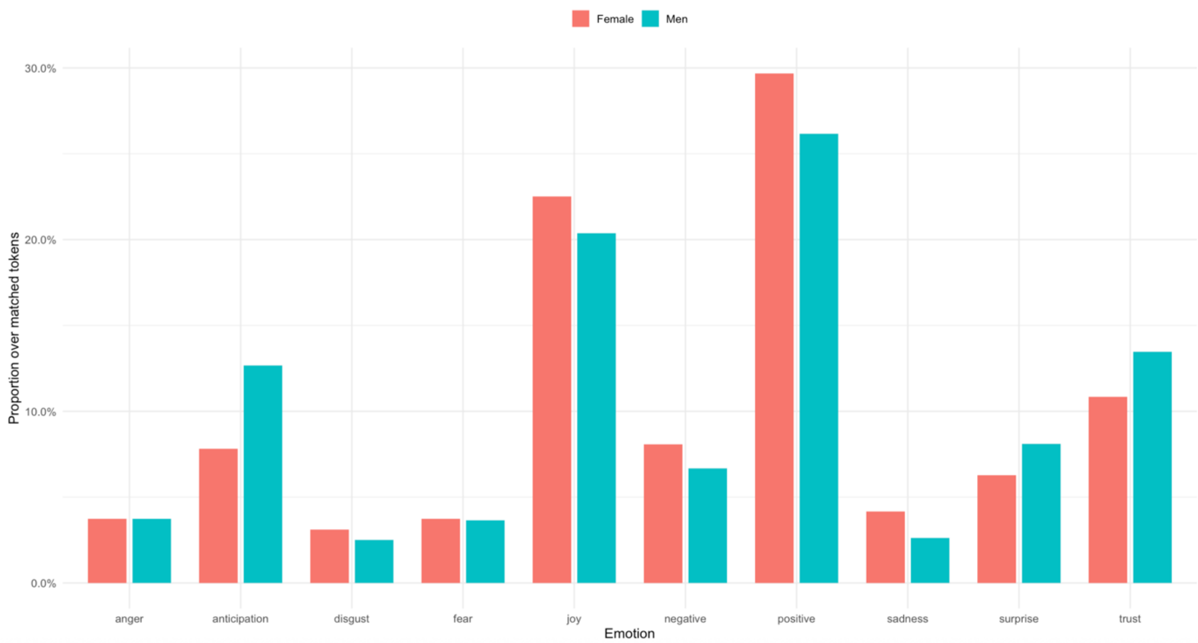

Figure 7. NRC emotional profile Men vs Female

Negative emotions were much less common overall but still revealed subtle gender differences. Female-directed comments contained more sadness (4.2% vs. 2.6%) and disgust (3.1% vs. 2.5%), pointing to occasional critical or body-related negative judgments that coexist with the generally positive tone. Male-directed comments, by contrast, were not significantly higher in any negative category, with anger and fear distributed almost equally across genders.

When aggregating emotions into overall polarity, the imbalance becomes clear: female creators receive a slightly higher proportion of positive comments (≈80% vs. 78%), whereas male creators show a marginally higher share of neutral-to-negative tones. These findings align with the thematic analysis: women are more often framed through appearance and emotional expression, while men are more frequently framed in terms of performance, reliability, and celebratory engagement, producing a distinct emotional signature for each gender.

Taken together, the emotion analysis shows that gender shapes not only the vocabulary and themes of Instagram comments but also their affective orientation. Even in micro-texts such as Instagram comments, often just a word, phrase, or emoji, cultural expectations emerge clearly: women are predominantly addressed through admiration of beauty, while men are framed in terms of trust, anticipation, and performance.

4 Results

The results are presented as a narrative sequence, moving from lexical and structural features of the corpora, through thematic modelling, and finally to the emotional profiles expressed in the comments.

At the lexical level, the document-feature matrices showed both overlap and divergence. The male dataset contained 45,322 comments and the female dataset 39,017, with a shared vocabulary of 581 terms. At the same time, each corpus included a substantial number of unique words—633 for men and 653 for women—indicating that while both groups of commenters draw on a common linguistic base, they also rely on gender-specific repertoires. Keyness analysis confirmed these distinctions: comments directed at women were more strongly associated with terms linked to beauty, aesthetics, and relational evaluation, while those directed at men more often included references to identity, fandom, and celebratory expressions. Semantic network analysis reinforced this picture. Both corpora produced sparse networks with low density and average degree, reflecting the fragmented and short nature of Instagram comments. Still, community detection revealed clusters of semantically related words, and centrality measures identified a handful of bridging terms that function as hubs in otherwise dispersed vocabularies. In male-directed comments, these hubs often referenced athletes and public figures, while in female-directed comments they tended to be evaluative or affective.

Building on these findings, topic modelling offered a more explicit view of thematic patterns. Latent Dirichlet Allocation (LDA) models retaining two main topics per gender highlighted clear gendered differences: male-directed comments mixed affective language with domain-specific references to sports and fandom, while female-directed comments were more consistently centred on beauty, admiration, and social evaluation. Structural Topic Modelling (STM) with gender as a covariate confirmed this divergence, showing that appearance and aesthetic related topics were significantly more prevalent in comments to women, whereas celebratory or performance-oriented discourse was more typical in comments to men. The cosine similarity between topics across genders was moderate (0.41–0.44), pointing to some lexical overlap but also clear thematic separation. These results suggest that while both men and women receive supportive comments, the way this support is framed reflects broader cultural expectations of masculinity and femininity.

The emotional analysis, based on the NRC Emotion Lexicon, revealed a broadly positive tone across both corpora. Around three-quarters of all mapped tokens expressed positive sentiment, with joy, trust, and anticipation as the most prominent emotions. Subtle gender differences emerged, however. Comments directed at women contained more positive language overall (29.7% vs. 26.2%) and more joy (22.5% vs. 20.4%), alongside slightly higher levels of sadness (4.2% vs. 2.6%) and disgust (3.1% vs. 2.5%). Male-directed comments, by contrast, showed relatively higher levels of anticipation (12.7% vs. 7.8%), trust (13.5% vs. 10.9%), and surprise (8.1% vs. 6.3%). Negative emotions such as anger and fear were rare in both datasets, each contributing less than 5% of the total. Overall polarity remained highly positive, with only minor differences in effect size.

Taken together, these findings show that gender shapes Instagram discourse about the body not through sharp differences in tone, but by subtly steering the vocabulary, themes, and emotional framings through which support and admiration are expressed.

5 Conclusions Based on Empirical Analysis

The conclusions of this study rest on multiple layers of evidence. Vocabulary comparison and keyness analysis highlight lexical differences, semantic networks reveal structural divergences, topic models quantify contrasts in thematic prevalence, and sentiment and emotion analysis shows not only the overall positivity of the comments but also the more subtle emotional distinctions between those directed at male and female creators. The convergence of these methods reinforces the reliability of the results.

Our analysis shows that, although comments directed at male and female creators on Instagram share a broadly positive and supportive tone, they differ in how this positivity is framed. Female-directed comments emphasize beauty, aesthetics, and emotional vulnerability, whereas male-directed comments are more often oriented toward identity, celebration, and anticipation. These findings suggest that gender plays a subtle but significant role in shaping online discourse about body representation.

The central research question guiding this project asked whether a creator's gender influences the tone and emotional profile of comments in online discussions of body representation. The results confirm that while the general sentiment is overwhelmingly positive for both genders, the framing of this positivity diverges. Women are more often addressed through appearance-related and affective language, while men are more frequently recognized in terms of excitement, admiration, and collective recognition.

Taken together, these conclusions highlight that gender shapes Instagram discourse about the body not by creating stark differences in tone, but by subtly steering the vocabulary, thematic structures, and emotional framings through which support and admiration are expressed.

5.1 Critical Analysis of the Adopted Strategy

A key strength of this project lies in the integration of complementary computational methods. By combining descriptive statistics, network analysis, topic modelling, and sentiment analysis, the study offers both breadth and depth in the exploration of gendered discourse. The use of the Structural Topic Model (STM) with gender as a covariate is especially valuable, as it provides a direct statistical test of differences in topic prevalence. The relatively large size of the datasets further increases the robustness of the findings.

At the same time, the project also faces several limitations. All data were collected from a single platform (Instagram), focusing on comments retrieved directly from selected posts where the creator's body was prominently displayed. While this strategy ensured ecological validity, it may still introduce selection bias, as the sample cannot capture the full diversity of Instagram interactions. Moreover, the short and noisy nature of Instagram comments—often reduced to emojis, single words, or informal expressions—requires extensive preprocessing before they can be used for systematic analysis. Even after cleaning, this fragmented structure limits the reliability of certain network-based measures, such as co-occurrence and centrality. Lexicon-based sentiment analysis, although informative, is similarly constrained, as it cannot fully capture context, irony, or multi-word expressions common in social media discourse. Another challenge concerned the collection of comments from some categories of male creators, such as activists or actors, who tend to attract fewer body-related interactions than athletes or models. This imbalance complicated the sampling process and may have shaped the thematic patterns observed in the male corpus.

Taken together, these limitations highlight the need for larger and more diverse datasets to enhance the reliability of future analyses. This is particularly relevant in the current cultural moment, where issues such as body positivity and body representation occupy a central place in public debate. Broader data would not only improve methodological robustness but also allow for a more accurate picture of how gendered discourses around the body are negotiated in digital environments.

Future work could address these limitations in several ways. One important step would be to expand the dataset across platforms and include a broader range of users, thereby improving representativeness. Applying more advanced NLP techniques, such as transformer-based sentiment classifiers or contextual embeddings, would also allow for a more nuanced analysis of meaning beyond lexicon matches. Finally, complementing computational findings with a qualitative examination of selected comment threads could provide interpretive depth and enable a more comprehensive interpretation of results.

6 Reference

[1] Boursier, V., & Gioia, F. (2022). Which are the effects of body-objectification and Instagram-related practices on male body esteem? *Clinical Neuropsychiatry, 19*(1), 8–19. https://doi.org/10.36131/cnfioritieditore20220103

[2] Febrianti, Y., & Fitria, K. (2020). The interpretation and attitude of body shaming behaviour on social media: A digital ethnography study on Instagram. *Diakom: Jurnal Media dan Komunikasi, 3*(1), 12–25. https://doi.org/10.17933/diakom.v3i1.78

[3] Grasso, F., Valese, A., & Micheli, M. (2023). Body-shaming detection and classification in Italian social media. In *Proceedings of the Conference on Natural Language Processing (CLiC-it 2023)*. University of Turin. https://github.com/ValeseA/BS-Detect

[4] Jiaqing, X., Alivi, M. A., Mustafa, S. E., & Dharejo, N. (2023). The impact of social media on women's body image perception: A meta-analysis of well-being outcomes. *International Journal of Interactive Mobile Technologies, 17*(20), 148–170. https://doi.org/10.3991/ijim.v17i20.44665

[5] Kim, H. M. (2021). What do others' reactions to body posting on Instagram tell us? The effects of social media comments on viewers' body image perception. *New Media & Society, 23*(12), 3448–3465. https://doi.org/10.1177/1461444820956368

[6] Marwick, A. E. (2015). Instafame: Luxury selfies in the attention economy. *Public Culture, 27*(1), 137–160. https://doi.org/10.1215/08992363-2798379

[7] Nimiya, A., Vasudha, K. G., Shetty, S. B., Pai, K., Reshma, N. S., Radhika, K., D'Souza, M., & D'Souza, P. (2024). Sex difference in body image, exercise motivation and social comparison among Instagram users: A cross-sectional study. *F1000Research, 12*, 1506. https://doi.org/10.12688/f1000research.134799.2

[8] Pedalino, F., & Camerini, A.-L. (2022). Instagram use and body dissatisfaction: The mediating role of upward social comparison with peers and influencers among young females. *International Journal of Environmental Research and Public Health, 19*(3), 1543. https://doi.org/10.3390/ijerph19031543

[9] Tiggemann, M., & Barbato, I. (2018). "You look great!": The effect of viewing appearance-related Instagram comments on women's body image. *Body Image, 27*, 61–66. https://doi.org/10.1016/j.bodyim.2018.08.009.

[10] Tiggemann, M., & Velissaris, V. G. (2020). The effect of viewing challenging "reality check" Instagram comments on women's body image. *Body Image, 33*, 257–263. https://doi.org/10.1016/j.bodyim.2020.04.004.

[11] Digital 2025: Global Overview Report. https://datareportal.com/reports/digital-2025-global-overview-reportApify

[12] IG comments Expert Tool.

chrome-extension://nmgmcehdhckaehgfokcomaboclhbdpkb/options.html

[13] For full access to the scripts, preprocessing procedures, and supplementary materials underlying the analysis, readers are referred to the project's GitHub repository: https://github.com/FraMichi/Computational-Social-Science