EMOTIONAL PEER EFFECTS IN ONLINE EATING DISORDERED COMMUNITY

A. IANNI [1], E. MENTZAKIS [2], T. WANG [3], AND M. BREDE [4]

Abstract. This paper studies emotional peer effects in online communities centered around eating disorders (ED), using a novel large-scale Twitter dataset constructed for this purpose. We develop a structural model that distinguishes two types of peer influence: conformity to the average emotional tone of peers and reinforcement through cumulative emotional exposure. Leveraging directed network data and sentiment analysis via LIWC, we estimate a linear-in-means model with instrumental variables derived from higher-order network characteristics to address endogeneity concerns. Our results show that among ED-identified users, positive emotions are amplified through aggregate peer exposure (emotional reinforcement), whereas general users exhibit conformity without reinforcement. Negative emotions are less influenced by peer dynamics among ED users but show strong conformity among general users. These findings suggest that platform structures facilitating emotional saturation may exacerbate vulnerability in ED communities. Policy implications emphasize the need for content moderation strategies that disrupt reinforcing emotional environments rather than solely shifting normative baselines.

1. Introduction

Recent media reports and academic studies have highlighted a resurgence of harmful content related to eating disorders on social media platforms. Trends such as ”SkinnyTok” on TikTok promote extreme thinness and unhealthy dieting practices, posing significant mental and physical health risks to adolescents[1]. Similarly, the ”Oatzempic” trend, involving a restrictive oat-based smoothie, has raised concerns among health professionals about its potential to encourage disordered eating behaviors among teens[2]. Investigative journalism has revealed that platforms like X (formerly Twitter) have become hubs for pro-eating-disorder communities, with lax content moderation allowing harmful content to proliferate[3]. Despite existing policies against promoting self-harm, enforcement remains inconsistent, leading to the normalization of dangerous behaviors. In response to these concerns, some

Date: April 2025.

[1,4] University of Southampton, [2] City, [3] King's College London.

[1]https://www.parents.com/what-is-skinnytok-trend-11716497

[2]https://www.parents.com/what-is-the-oatzempic-diet-trend-11718034

[3]https://www.theatlantic.com/technology/archive/2024/12/eating-disorder-content-x/681036/

policymakers are advocating for stricter regulations on social media platforms. For instance, legislation in New Jersey aims to prohibit social media platforms from using algorithms that promote harmful content related to eating disorders[4] and France is recently threatening to take action against some social media[5].

Academic research corroborates these findings, indicating that exposure to idealized body images and diet culture on social media correlates with increased body dissatisfaction and disordered eating behaviors among adolescents (Chu J (2024), Nawaz et al. (2024)). Studies have also shown that excessive screen time and social media use are linked to a higher risk of developing eating disorder symptoms in young adolescents (Chu J (2024)). Dane and Bhatia (2023) provide a large scoping review to investigate the association between social media, body image and eating disorders among young people.

These developments underscore the urgent need for comprehensive strategies that address both the content and structural aspects of social media platforms to mitigate their impact on vulnerable populations. This research contributes to this discourse by providing a structural estimation of emotional peer effects in online environments, offering insights into how social networks can influence disordered eating behaviors. While prior studies have explored emotional dynamics in online mental health communities (e.g., Birnbaum et al. (2020); Chancellor et al. (2016a)), these have largely focused on predictive modeling or descriptive correlations. In contrast, our study advances the literature by offering a structural econometric model that separately identifies normative conformity and emotional reinforcement mechanisms. We also leverage an original network-based dataset specifically constructed to preserve directed social relations and sentiment dynamics, enabling causal inference about peer effects. To our knowledge, this is among the first papers to structurally estimate emotional peer effects in a high-dimensional, vulnerable online population using instrumental variables.

To the aim of analysing these issues, this study uses an original, large-scale dataset drawn from Twitter. To the best of our knowledge, this dataset represents one of the first attempts to systematically capture both emotional expression and social network structures within online communities concerned with eating disorders. The data were collected explicitly for the purpose of this analysis, using a multi-stage strategy that combines keyword-based filtering, manual validation of self-reported ED identification, and snowball sampling of followees and followers. In total, the dataset comprises emotional and network information for over 3,000 ED-identified users and a comparable random reference sample, alongside their

[4]https://www.njspotlightnews.org/2024/12/nj-looks-to-blunt-social-media-impact-on-eating-disorders/

[5]https://www.theguardian.com/world/2025/apr/22/french-minister-reports-skinnytok-to-regulator-anorexia-concerns-tiktok-videos

extended social graphs. This bespoke dataset allows us to analyse eating disorders in a high-dimensional, real-world social environment while preserving rich information about network topology and user-level heterogeneity.

Our quantitative analysis distinguishes between two conceptual types of peer effects that operate through the social network. First, social conformity refers to the behavioural tendency of individuals to adjust their emotional expression to match the average level of emotion exhibited by their reference group. This aligns with classical models of social norms and identity, wherein deviation from a normative benchmark entails psychological costs. Second, social complementarity arises when the aggregate emotional intensity of peers enhances the marginal utility of expressing similar emotions, akin to a multiplier effect in network externalities. These two forces -conformity and reinforcement - are modelled separately and allow us to empirically test which type of social influence dominates in different reference populations.

Our approach is complementary to recent work on peer effects in mental health forums (Birnbaum et al. (2020); Sowles et al. (2018)), but goes further in combining a utility-based behavioral model, structural estimation, and policy-relevant implications for social media platform design. Our utility specification builds on the literature incorporating identity into economic modeling, especially the frameworks of Akerlof and Kranton (2000a), Bénabou and Tirole (2011), Bénabou and Tirole (2002), and Bernheim (1994a), by allowing emotional self-expression to be shaped by both conformity to peer norms and reinforcement from social interaction. Our model shares conceptual foundations with the literature on social image and peer pressure (Bursztyn and Jensen (2017)), which emphasizes how observable behavior and social norms shape individual decisions. We extend this framework to emotional expression in online settings, where peer visibility is both high and algorithmically mediated.

From a methodological perspective, our paper contributes to a small but growing literature that structurally estimates peer effects using network econometrics. Building on identification strategies developed by Bramoullé et al. (2009) and de Giorgi et al. (2010), we estimate a linear-in-means model that includes both direct and second-order peer exposure variables, while addressing endogeneity through instrumental variables derived from the network topology.

To our knowledge, this is one of the first papers to:

(1) structurally estimate *emotional* peer effects in an online environment;
(2) apply this framework to a *vulnerable population* (self-identified eating disorder users);
(3) leverage a *high-dimensional real-world social network* for both direct and second-order peer influence;

(4) empirically distinguish between *conformity (average)* and *reinforcement (aggregate)* mechanisms.

2. Modelling Emotional Peer Effects in Online ED Communities

This section introduces a formal model of peer influence tailored to online eating disorder (ED) communities. Motivated by empirical evidence from psychology and sociology—as well as identity-based utility models in economics—we seek to capture how individuals' emotional expression on social media platforms is shaped by both their own internal state and the social context provided by their digital network environment.

We begin by specifying a microeconomic utility function that integrates two central forces. The first is *self-image*, which represents the internalized perception individuals have of themselves, shaped by comparisons to others and aspirations toward particular social identities. The second is *peer effects*, through which the emotional content expressed by one's online connections (i.e., their followees) can influence one's own affective behavior. Together, these components reflect both introspective and interactive motivations for public emotional expression in digital spaces.

In the context of Twitter-based ED communities, where emotions are publicly revealed through linguistic content of tweets, we treat emotional expression as a choice variable that may be affected by normative expectations (i.e., conformity to the average sentiment of peers) as well as aggregate group dynamics (i.e., emotional amplification through dense, emotionally aligned networks).

We formalize this behavioral framework using a utility function defined over emotional intensity and self-image, with self-image itself being endogenously determined by the emotional landscape of one's peers. The model explicitly distinguishes between two types of endogenous peer effects:

- *Local average effects*, which reflect an individual's desire to conform to the mean sentiment expressed by their followees;
- *Local aggregate effects*, which capture the possibility of emotional reinforcement driven by the cumulative emotional intensity within the peer group (i.e., a social multiplier effect).

These distinctions map naturally onto different theoretical constructs in the literature on social interactions. Conformity mechanisms reflect disutility from deviation from group norms, while aggregate effects represent utility gains from coordination or contagion of emotional states.

Having developed this utility-based behavioral model, we then derive its best-response function, which forms the basis for our empirical strategy. This function links an individual's emotional expression to both their peers' emotional states and observable covariates, enabling us to estimate and compare the relative strengths of average versus aggregate peer effects across user types (ED vs. general users). The resulting equations motivate a structural econometric model, estimated using two-stage least squares (2SLS) to address endogeneity concerns inherent in social network data.

2.1. **Utility Function with Self-Image and Peer Influence.** We develop a utility-based framework in which individuals derive utility from expressing emotions in online environments, such as Twitter, while also being influenced by the social context provided by their network of connections. The model incorporates both an intrinsic component of emotional expression and an identity-based self-image component that is shaped by peer comparisons.

Let $i \in \{1, \ldots, n\}$ denote an individual in a directed social network G, represented by an adjacency matrix $G = [g_{ij}]$, where $g_{ij} = 1$ if individual i follows individual j, and $g_{ij} = 0$ otherwise (with $g_{ii} = 0$). The set of followees of individual i is denoted $N_i = \{j \neq i : g_{ij} = 1\}$. We also define the row-normalized matrix $G^* = [g^*_{ij}]$, where $g^*_{ij} = g_{ij} / \sum_j g_{ij}$, such that G^*Y gives the vector of peer group averages.

Let $y_i \in \mathbb{R}$ represent the intensity of emotional expression by user i, as inferred from their language in tweets. This serves as the individual's endogenous action. Emotional expression is costly but may generate utility through psychological expression, conformity, and social approval. We assume that each individual derives utility not only from their own emotional intensity but also from maintaining a self-image SI_i, which is shaped by both their own and their peers' emotional expressions. Comparisons between peers is an important factor of developing body image - a part of self-image - and appearance dissatisfaction is a prime contributor in mental disorders (e.g., ED and depression in Helfert and Warschburger (2013); Polivy (2017)).

Our modeling strategy builds on a self-identity model akin to Akerlof and Kranton (2000b). This is in line with extensive work emphasizing the importance of self-image (or identity) in explaining ED Hyatt et al. (1958); Costa-Font and Jofre-Bonet (2013). The utility function is specified as:

$$U_i = \lambda SI_i - \mu y_i^2,$$

where:

- $\lambda > 0$ captures the marginal utility of maintaining a desirable self-image;
- $\mu > 0$ captures the intrinsic cost of emotional expression;
- SI_i is the individual's self-image, defined below.

We model self-image SI_i as a function of both one's own emotional expression and peer-based comparisons:

$$SI_i = \alpha y_i - \frac{1}{2}\gamma \left(y_i - \sum_j g^*_{ij} y_j \right)^2 + \omega \left(\sum_j g_{ij} y_j \right) y_i. \tag{1}$$

This expression includes:

- *Intrinsic valuation of emotion* (αy_i): benefit from self-expression;
- *Conformity penalty* ($-\frac{1}{2}\gamma(y_i - \overline{y}_{N_i})^2$): cost for deviating from the average of peers;
- *Aggregate reinforcement* ($\omega \sum_j g_{ij} y_j \cdot y_i$): benefit from coordination with peer emotional intensity.

An individual i's self-image depends, first of all, on i's own actions y_i, i.e., emotions expressed in tweets. Coefficient α is i's *ex-ante* evaluation of an action, which is possibly heterogeneous and depends on i's attributes and environmental factors. Self-image is further influenced by the actions of peers with whom i interact. Since comparisons between peers is an important factor of developing body image and appearance dissatisfaction, as well as mental disorders like ED and depression (Helfert and Warschburger (2013); Polivy (2017)), we specify peer effects as the difference between an ego's emotions, y_i, and the average level of emotions expressed by the individual's peers, $\sum_{j=1}^n g^*_{ij} y_j$. As we assume that individuals wish to conform the social norm of their reference groups, i.e. the average level of emotions expressed by peers, we set the social-conformity coefficient $\gamma \geq 0$. Finally, to account for heterogeneity of peer effects (as in Ballester et al. (2006)), we assume that the marginal value of self-image of any given level of emotions is increasing in the aggregate level of emotions among peers, $\sum_{j=1}^n g_{ij} y_j$. ω is a social-multiplier coefficient which reflects the strength of spillover effects that high levels of one attribute among peers can have on an individual. The last term $\omega \sum_{j=1}^n g_{ij} y_j$ also allows us to explicitly formalize that individuals with different social statuses, particularly captured by different locations in the network here, may have different evaluations of self-images.

Substituting SI_i into the utility function, we obtain:

$$U_i = \left(\alpha + \omega \sum_j g_{ij} y_j \right) y_i - \frac{1}{2} \left[\gamma \left(y_i - \sum_j g^*_{ij} y_j \right)^2 + y_i^2 \right] \tag{2}$$

The two components of the utility function Eq. (2) can be interpreted as follows. The first component represents the benefits of expressing emotions y_i, increasing in the *ex-ante* utility value of emotions, α, and in the aggregate level of emotion encountered in the network, $\omega \sum_{j=1}^{n} g_{ij} y_j$. As individuals may have different locations in the network, $\sum_{j=1}^{n} g_{ij} y_j$ is heterogeneous in i even if every individual in the network has the same emotional level. The second component represents the cost of expressing emotions y_i, which includes two terms: one is the cost due to deviation from the social norm of the reference group, i.e. the distance from the local average level of emotions $\gamma \left(y_i - \sum_{j=1}^{n} g_{ij}^{*} y_j \right)^2$, and the other is the cost of maintaining a given emotion level, y_i^2.

The first-order condition for utility maximization yields the best-response function:

$$y_i = \phi \sum_j g_{ij} y_j + \psi \sum_j g_{ij}^{*} y_j + \eta_i, \tag{3}$$

where:

- $\phi = \frac{\omega}{1+\gamma}$ is the coefficient on the aggregate effect (social multiplier)
- $\psi = \frac{\gamma}{1+\gamma}$ is the coefficient on the average peer effect (social conformity), and
- $\eta_i = \frac{\alpha}{1+\gamma}$is the normalized intrinsic preference for emotional expression.

This linear structure links ego's emotional expression to both the aggregate and average behaviour of peers[6]. This best-reply formulation captures the dual influence of peers: the average behaviour of peers sets the social norm, while the aggregate behaviour may generate emotional reinforcement. The resulting system of equations forms the basis for the econometric model and estimation strategy that follows in Section (2.2).

Emotional reinforcement, modeled through cumulative peer exposure, captures the intuition that individuals may experience increasing returns to emotional engagement when surrounded by emotionally intense peers. This mechanism is particularly salient in the context of online eating disorder (ED) communities, where affective states such as anxiety, dissatisfaction, or idealization of thinness are often central to social identity. In these environments, positive emotional expressions (e.g., shared achievements toward disordered goals) may be perceived not only as individually rewarding but also as socially validated, creating a feedback loop of emotional amplification. By contrast, general online communities typically exhibit greater emotional heterogeneity, leading to weaker reinforcement effects and greater normative convergence rather than contagion.

[6]Given the linear structure of the best-reply functions and typical sparsity of social network matrices, the model satisfies standard contraction mapping conditions, ensuring the existence and uniqueness of equilibrium. Uniqueness ensures that our empirical estimates correspond to a single behavioral mapping from peer influences to individual emotional expressions, rather than to an average over multiple equilibria.

Interpreting Peer Effects in Digital Networks. In the context of online social platforms such as Twitter, users interact through directed, asynchronous relationships in which emotional content is publicly visible and temporally persistent.

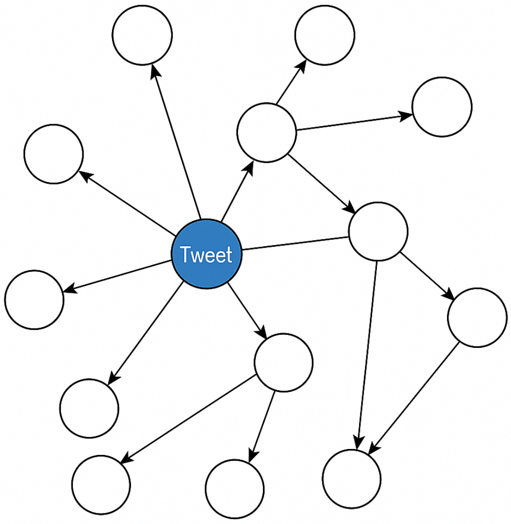


FIGURE 1. Stylized representation of a Twitter-based peer network. Nodes represent users, and directed edges denote follow relationships. Node color or label can represent emotional valence (e.g., positive vs. negative tone).

Within this environment, peer effects on behavior can be decomposed into two distinct channels: average effects and aggregate effects. These reflect fundamentally different mechanisms of social influence and carry important implications for emotional contagion and norm formation in online communities.

- Average Peer Effects (Conformity to Norms). The average peer effect, parameterized by ψ in our model, reflects the extent to which individuals align their behavior with the *mean emotional expression* of their network neighbors (i.e., followees). This type of influence corresponds to norm conformity: users compare their emotional expression to that of their peers and incur a disutility when deviating from the perceived social norm. In this way, emotional behavior is shaped by an internalized standard, fostering harmonization in affective tone. This mechanism is well-documented in the economics of identity and peer interactions (e.g., Bernheim (1994b); Brock and Durlauf (2001)).

  On Twitter, users encounter a curated timeline reflecting the average tone of their followees' tweets. For instance, in an ED recovery community, if most followees express moderate optimism, an individual may avoid posting highly negative content to preserve group membership or avoid alienation. This process may function as an implicit reputational constraint or as a psychological cost of norm deviation.

- Aggregate Peer Effects (Emotional Reinforcement). In contrast, the aggregate peer effect, captured by the coefficient ϕ, reflects the influence of the *total emotional input* received from peers—independent of its average. Rather than aligning with a benchmark, users are influenced by the *intensity and volume* of emotional signals in their environment. This effect is akin to strategic complementarities or social multipliers Glaeser et al. (2003); Ballester et al. (2006), where the marginal utility of an action increases with the overall prevalence of that action among peers.

  This mechanism captures emotional reinforcement: when a user is exposed to a high density of emotionally charged content (e.g., numerous expressions of despair or euphoria), they may be more inclined to echo similar sentiments, even if such behavior deviates from the group mean. The online setting facilitates this process by creating environments in which emotional saturation is common due to high followee activity or algorithmic amplification of expressive content.

These two peer effects may interact or oppose one another, depending on the structure and affective composition of the social network. In dense, homophilous networks (e.g., tightly knit pro-ED or pro-recovery clusters), average and aggregate exposures may be highly correlated, reinforcing one another. In sparser or emotionally diverse networks, however, the two may diverge, with average exposure encouraging moderation and aggregate exposure encouraging extremity.

Table 1. Comparison of Average vs. Aggregate Peer Effects in Online Settings

Aspect	Average Effect (Conformity)	Aggregate Effect (Reinforcement)
Target statistic	Mean emotion of peers	Total emotion from peers
Mechanism	Norm compliance	Emotional contagion
Behavioral logic	Fit in with norm	Echo emotional volume
Sensitive to	Consistency	Intensity and density
Online manifestation	Timeline tone	Emotional saturation
Economic analogy	Identity, peer norms	Strategic complements
Policy leverage	Shift norms	Disrupt feedback loops

Disentangling these channels is critical for understanding emotional propagation in vulnerable populations. For example, if ED users are more sensitive to aggregate than average effects, interventions should prioritize reducing the volume of negative content, rather than merely shifting norms. Conversely, if general users are more norm-sensitive, content-based nudges to reset perceived norms may be effective.

2.2. **Empirical Specification.** We derive an empirical specification based on the best-response function (3)). Following the peer effects literature (e.g., Liu et al. (2014)), we allow an individual's emotional expression to depend on their own characteristics, the emotional expression of their peers, and the characteristics of those peers.

Let $Y = (y_1, \ldots, y_n)'$ be the vector of observed emotional expressions (e.g., positivity or negativity scores) for n users in a directed social network. Let $X = (x_1, \ldots, x_n)'$ be a matrix of individual-level covariates (e.g., number of tweets, followers, and friends). The network is captured by two adjacency matrices:

- G: the binary directed adjacency matrix ($g_{ij} = 1$ if i follows j),
- G^*: the row-normalized version of G, such that G^*Y gives the average behavior of followees.

Our general model is:

$$Y = \phi GY + \psi G^*Y + \eta, \tag{4}$$

where:

- ϕ captures the effect of the aggregate emotional tone of followees (total exposure),
- ψ captures conformity to the average emotional tone of followees,
- η is a vector of error terms

Based on the best-reply function in Eq. 3, we specify econometric models to identify local aggregate and local average effects. Following the notation of Liu et al. (2014), we first define the *ex-ante* heterogeneity in the social network, η, as:

$$\eta = \beta X + \lambda G^*X + \epsilon. \tag{5}$$

where:

- X is a vector of exogenous variables,

- ϵ is a normally distributed error term with mean zero and variance σ^2 that is independent and identically distributed (i.i.d) across individuals,
- β and λ are parameters.

Then, from 3, the general econometric model for emotional peer effects is:

$$Y = \phi GY + \psi G^{*}Y + \beta X + \lambda G^{*}X + \epsilon. \tag{6}$$

The general model in Eq. 6 assumes that an ego's choice/outcome are influenced by:

(1) *endogenous peer effects*, captured by ϕ for the local aggregate effects and ψ for the local average effects;
(2) *exogenous* characteristics of the individual, captured by β,
(3) exogenous characteristics of the individual's followees, *contextual effects*, captured by λ.

A notable difference between the local aggregate and local average effects is the influence of positions in a network. In the local average effect, ψ, network positions would not influence an individual's behavior, and equilibrium utility would be the same for individuals who are *ex-ante* identical, e.g., with the same emotional state. However, in the local aggregate effect, ϕ, individuals in different positions of a social network can have different equilibrium utility and emotional states, even if these individuals are *ex-ante* identical (Liu et al. (2014); Ghiglino and Goyal (2010)).[7]

To account for cascading influences beyond immediate followees, we extend the model to include second-order network terms. Let G^2 denote the matrix capturing paths of length 2 (i.e., friends-of-friends), and G^{*2} its row-normalized version. The extended model is:

$$Y = \phi_1 GY + \psi_1 G^{*}Y + \phi_2 G^2 Y + \psi_2 G^{*2} Y + \beta X + \lambda G^{*}X + \varepsilon. \tag{7}$$

This allows us to separately identify whether ego's emotional expression is more influenced by direct peers or by the broader emotional environment formed by

[7]Note that we do not explicitly include *network fixed effects* which describe that individuals in different networks may behavior differently as they have different unobserved individual characteristics or they face a different social environment (Liu et al. (2014); Manski (1993)). This is because (i) our samples in both two datasets are connected within a single weakly connected network (as explained later), and (ii) we use separate models to model each group of user samples, instead of mixed models applied to both two groups. Thus, the network fixed effects are implicitly included in the error term ϵ.

their second-order connections. This nonlinear extension allows us to test for diminishing or increasing marginal effects of peer exposure. For instance, a negative coefficient on would suggest that aggregate peer exposure initially reinforces emotional expression but saturates or reverses beyond a certain threshold.

Our model of emotional peer effects is structurally similar to games with local interactions under strategic complements and substitutes (e.g., Ballester, Calvó-Armengol, and Zenou 2006). When the coefficient on average exposure is large, players align with their peers (norm conformity), while a high coefficient on aggregate exposure introduces complementarity in emotional expression, akin to positive spillovers. Depending on parameter values, the equilibrium emotional state may exhibit amplification or attenuation. Though we do not analyze equilibrium multiplicity formally, the empirics we produce allow us to test for the dominant channel of social influence in the data.

2.3. **Reverse Causality and the Reflection Problem.** A central challenge in estimating peer effects in network settings is the issue of *reverse causality*, as formalized in the seminal work of Manski (1993). When individuals are embedded in social networks, observed correlations between a person's behavior and that of their peers may arise from simultaneous influence, rather than unidirectional causality. This so-called *reflection problem* arises because an individual's outcome may affect and be affected by their peers' outcomes, leading to a mutual causality that is observationally indistinguishable in cross-sectional data. For instance, if both ego and their followees express high levels of emotional content, it is unclear whether ego is responding to peers, peers are responding to ego, or both are responding to a common shock. To overcome this identification problem, structural models typically impose additional assumptions or leverage instrumental variables—often constructed from exogenous characteristics of higher-order peers—to isolate truly exogenous variation in peer exposure.

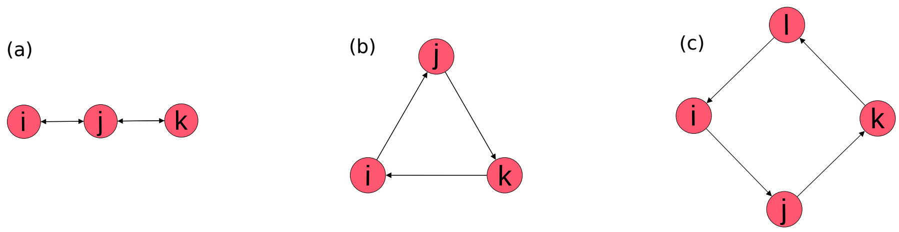


FIGURE 2. Examples of simultaneity between dependent variable and instruments caused by (a) pairwise mutual influence and (b-c) cycles of social networks.

In linear models such as the one we estimate, these effects enter additively, but are not separately identified without additional structure or instruments. Specifically, if both ego and their followees exhibit high emotional intensity, we cannot determine from correlation alone whether this is due to imitation (endogenous effect), demographic similarity (exogenous effect), or a shared environmental shock (correlated effect).

To identify endogenous peer effects while mitigating endogeneity from simultaneity and omitted variables, we employ two complementary strategies. First, we restrict attention to single-directional followees, i.e., users that ego follows but who do not follow ego back. This breaks potential feedback loops in mutual visibility and reduces the likelihood that the ego's emotional expressions influence those peers. Second, we implement instrumental variables (IV) estimation, where instruments are derived from the characteristics and actions of higher-order peers (as in Bramoullé et al. (2009) and de Giorgi et al. (2010)). Specifically, we use exogenous characteristics of second-degree peers, (i.e., followees-of-followees) such as the average number of tweets, friends, and followers, as instruments for endogenous peer outcomes. The validity of these instruments relies on the assumption that while second-degree peers influence ego's friends, they are not directly connected to the ego and thus less likely to correlate with ego's unobserved emotional state. These characteristics affect ego's exposure to peer emotion but are plausibly uncorrelated with ego's own unobserved determinants of emotional expression, thereby satisfying the exclusion restriction.

This identification strategy allows us to disentangle the influence of aggregate and average peer emotion while mitigating reverse causality bias that would otherwise compromise the validity of causal inference.

3. Twitter Data Set

We apply our empirical strategy to large-scale data collected from Twitter, a microblogging platform where users post and interact with short messages known as tweets. The platform supports directed social links (following relationships), enabling the construction of communication and exposure networks. All data used in this study were collected via the official Twitter APIs and consist entirely of publicly available information. We construct two complementary datasets: one focusing on users who self-identify with eating disorders (ED) and a second reference sample of general Twitter users, used for comparative analysis of behavioral patterns and peer influence.

3.1. **ED Sample.** To identify individuals likely to be affected by eating disorders on Twitter, we follow a method inspired by prior work on detecting ED-related communities in online platforms Wang et al. (2017). Our strategy combines keyword filtering with network-based snowball sampling to create a high-precision sample of ED-identified users.

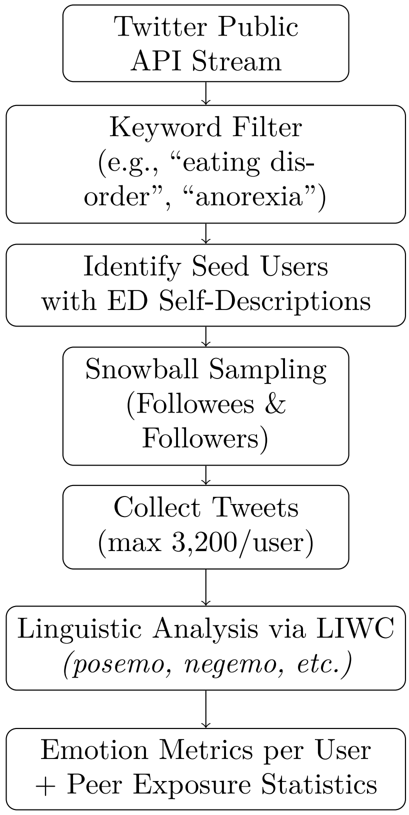


Figure 3. Data collection and linguistic analysis pipeline. Tweets are collected using keyword-filtered access to Twitter's public API. Users self-identifying as having an ED are used as seeds, whose networks are expanded via directed links. Sentiment analysis is performed using the LIWC dictionary to produce emotion variables used in the peer effects model.

We begin by collecting tweets containing ED-related keywords, including "eating disorder," "anorexia," "bulimia," and "EDNOS," from January 8 to January 15, 2016. This process yields 1,169 tweets mentioning common ED terminology. From the authors of these tweets, we identify a seed set of 33 users who self-report both (i) ED-related terms (e.g., "edprob," "proana") and (ii) personal bio-information (e.g., body weight) in their profile descriptions. These criteria ensure that the user is both aware of and publicly disclosing ED-related status, which improves signal precision.

We then expand the sample via snowball sampling over the directed social network of followees and followers. At each stage, we filter out non-English accounts and manually inspect ambiguous cases. The final sample consists of 3,380 unique ED users who explicitly self-identify with ED in their profile descriptions. To validate the quality of this sample, we implement a human annotation protocol. For a randomly selected subset of 1,000 users, we reviewed the content of users' timelines, including tweets, profile descriptions, and friend networks. Approximately 95.2% of users were judged to be highly likely to have ED, indicating a high degree of sample precision.

To construct the network structure and obtain behavioral measures, we collect metadata for all direct friends (both followees and followers) of the 3,380 ED users, resulting in 208,063 additional users. For each of these users, we retrieve up to 3,200 of their most recent tweets (the maximum allowed by the Twitter API at the time), resulting in a comprehensive dataset of 241,243,043 tweets. The data collection process for the ED sample concluded on March 2, 2016.

3.2. **Reference Sample.** To construct a representative baseline of general Twitter behavior, we leverage a pre-existing snapshot of the Twitter social graph captured in 2009 by Kwak et al. (2010). This dataset contains the complete "who-follows-whom" network as of September 24, 2009, encompassing approximately 1.47 billion directed edges among over 41 million users.

Due to platform restrictions, the original dataset includes only network topology. To augment it with behavioral and emotional data, we use the Twitter API to retrieve user profile and tweet information for a subset of users. In June 2018, we attempted to access all user accounts in the 2009 snapshot and successfully recovered 10,160,610 active or public profiles. We define a user as active if they had posted at least one tweet by June 2018; missing users were either deleted or made private in the intervening period.

We then collect up to 200 tweets (the maximum number allowed per user per request under the Twitter API at that time) posted by each user prior to September 30, 2009. This restriction ensures consistency with the original network snapshot and enables construction of temporally aligned behavioral measures. After filtering for users with more than one tweet, the final reference sample consists of 6,999,514 users and a total of 205,626,126 tweets.

Together, the ED and reference datasets allow us to examine peer influence mechanisms across two structurally and behaviorally distinct user populations—one characterized by self-disclosed vulnerability and dense emotional content, the other by a broad cross-section of typical Twitter usage. These datasets form the empirical foundation for our network-based analysis of emotional peer effects.

4. Measurements

4.1. **Emotion outcomes.** We measure individuals' emotions through their language use in tweets, as growing evidence has shown that different linguistic constructs capture a wide range of psychological phenomena, such as happiness and sadness (e.g., Oxman et al. (1982); Pennebaker et al. (2003)). While there are a variety of sentiment analysis algorithms to measure emotions in texts (e.g., Gonçalves et al. (2013); Ribeiro et al. (2016)), we use LIWC (Tausczik and Pennebaker (2010)) and SentiStrength (Thelwall et al. (2010)), because (i) these algorithms have been widely used to measure emotional content on social media (e.g., Oksanen et al. (2015); Ferrara and Yang (2015); Coviello et al. (2014); Chancellor et al. (2016b, 2017); De Choudhury (2015)); (ii) both two algorithms have shown good inter-rater reliability and consistently outperformed other algorithms across various sentiment analysis tasks (e.g., Gonçalves et al. (2013); Ribeiro et al. (2016); Oksanen et al. (2015)), and (iii) each algorithm measures emotions in a different way, which allows a cross validation between results of the two algorithms. LIWC reads a given text and counts the percentages of words that reflect different emotions, thinking styles, and social concerns. For a more reliable evaluation, we combine all historical tweets of each user as a document. All re-tweets are excluded, since they reflect cognitive attributes of their original authors rather than those of re-tweeters. After removing mention marks, hashtags and URLs, each document is split into tokens by white-space characters. Only documents containing more than 50 words are processed with LIWC for more trustworthy results. For each user i, we obtain the percentages of positive emotion words p_i and negative emotion words n_i in i's tweets.

4.2. **Estimation covariates.** Several control variables are included in our estimations to capture *ex-ante* heterogeneity of individuals. First, we notice that recent studies have shown that popular users who have a large pool of friends tend to post tweets with positive emotions like happiness (e.g., Bollen et al. (2017)). To control for this status effect, we consider sociometric metrics of an ego on Twitter, including the number of followees (i.e., people who are followed by an ego) which indicates the popularity of the ego on Twitter, and the number of followers (i.e., people who follow the ego) which indicates the degree of the ego's gregariousness (as in Snijders et al. (2013)). These metrics of ED users are obtained from user profile information, and those of random users are obtained by counting in-degrees and out-degrees of nodes in the full 2009 following networks[8]. Second, we consider the number of tweets that are used to measure emotions is controlled for heterogeneity in confidence of emotional outcomes, as a large number of tweets are expected to produce a more reliable measurement on users' emotions. Moreover, we measure the average numbers of followees, followers and tweets of an ego's followees as the contextual factors of the ego.

4.3. **Social networks.** We construct networks representing social contracts among users based on their who-follows-whom relations on Twitter, in which a directed link runs from node i representing user i to node j representing user j if i follows j.

[8]The profiles of random users are collected in 2018, and the numbers of followees/followers in profiles may change compared to those in 2009.

We only consider users who did not delete their Twitter accounts, whose privacy settings allow us to view their profiles, and who have a sufficiently large number of tweets to measure their emotion variables. Although this reduces the sizes of samples, it is reasonable to assume that those users who maintain public profiles and post tweets would be those users mostly likely to influence and be influenced by peers (because they emotions posted in tweets are visible to others and they actively receive content posted by other users). The social network built on ED data consists of 181,176 non-isolated nodes and 1,260,294 directed links, where 181,176 (accounting for 99%) nodes connected within the largest weakly connected component while all the rest nodes are isolated and have no connections to others. Due to the dominance, we focus on analyzing the largest connected component which contains 181,176 nodes and 1,260,294 directed links. The network in random samples contains 1,606,190 non-isolated nodes and 76,536,843 directed links, where 1,604,802 nodes are connected within the largest weakly connected component through 76,535,565 links, with only 5 nodes in the next-largest component. Similarly, we focus on analyzing the largest connected component for random samples, including 1,604,802 nodes and 76,535,565 directed links.

5. Results

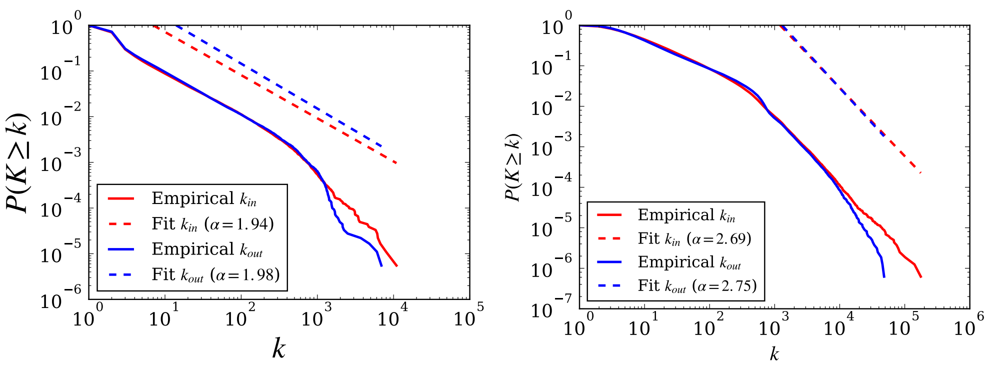
5.1. **Descriptive Statistics.** Table 2 gives descriptive statistics for variables that are used in our empirical estimations, where the ED users are those who self-identified as eating disordered in their Twitter profile descriptions. Since some users in our random samples deleted their accounts before our data collection, the full snapshots of social networks cannot be recovered for some random users. To avoid bias due to incomplete observation of users' social networks in estimating peer effects, we only consider random users who have a large proportion of friends and these friends have measurement outcomes in our data. Specifically, we select users who have more than 45% of their followees and more than 45% of their followers on Twitter. We choose 45% as a threshold since it gives us a similar size of samples as ED samples. Moreover, for both sets of samples, we only consider users who have more than 1 first-order followees ($|N_i| > 1$) and 1 second-order followees ($|N_i^2| > 1$) in the social networks, as they are the only users for whom we can distinguish the aggregate and average peer effects using IV models[9]. The differences between ED and random samples are measured by Mann-Whitney U-tests and the Bonferroni correction is used to counteract the problem of multiple comparisons.

[9]Users with $|N_i| = 0$ have no followees, making it hard to examine the effects of social networks on these users. For users with $|N_i| = 1$ have a single followee, the aggregate and average effects of the single followee on these users are the same, making it hard to distinguish these peer effects. Similarly, IVs of users with $|N_i^2| = 0$ are not available, while IVs for the aggregate and average peer effects are perfectly correlated for users with $|N_i^2| = 1$.

Compared to random users, ED users have less positive emotions and more negative emotions than random users in the results the sentiment analysis tools. This aligns with evidence that individuals with ED often have tendencies for depression, mental instability and irritability (e.g., De Choudhury (2015); Association et al. (2013)). Also, the typically negative tone of ED users reflects a lowered sense of self-esteem, likely due to normative dissatisfaction with one's body weight and shape (e.g., Wolf et al. (2007)).

Table 2. Descriptive statistics for ED and non-ED samples

	ED users (N = 3060)				Random users (N = 3482)			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
posemo	0.04	0.02	0.00	0.27	0.05	0.02	0.00	0.55
favg-posemo	0.05	0.00	0.01	0.08	0.05	0.01	0.00	0.10
fsum-posemo	5.97	7.66	0.05	70.71	0.97	2.42	0.01	41.34
ffavg-posemo	0.05	0.00	0.03	0.08	0.05	0.01	0.01	0.08
ffsum-posemo	63.54	71.87	0.08	530.10	74.65	185.21	0.04	2326.26
negemo	0.04	0.02	0.00	0.15	0.01	0.01	0.00	0.12
favg-negemo	0.04	0.01	0.01	0.06	0.01	0.00	0.00	0.04
fsum-negemo	4.32	5.57	0.04	60.02	0.25	0.63	0.00	8.72
ffavg-negemo	0.04	0.00	0.02	0.05	0.01	0.00	0.00	0.04
ffsum-negemo	48.89	54.69	0.06	392.98	19.39	47.08	0.01	603.68
followers-count	286.51	677.30	0.00	18128.00	173.79	294.45	1.00	4423.00
friends-count	296.64	474.08	3.00	10649.00	148.15	273.44	2.00	2674.00
tweets count	411.40	628.28	3.00	3234.00	85.21	69.64	6.00	200.00
favg-followers-count	611742.06	1463802.34	280.50	20180660.31	24895.82	82449.19	19.50	1031975.00
favg-friends-count	3290.39	7444.35	77.50	291463.87	1209.23	2923.21	6.50	54544.50
favg-tweets count	1585.88	326.24	334.27	3005.25	141.25	36.31	15.00	200.00
fnum	119.64	149.92	2.00	1471.00	19.19	45.37	2.00	706.00
ffnum	1317.79	1468.87	2.00	10595.00	1412.90	3468.84	2.00	44925.00



(A) ED samples (B) Random samples

Figure 4. Distributions of in-degree (k_{in}) and out-degree (k_{out}) in social networks of (a) ED data and (b) random data.

Table 3. Network attributes in ED data and Random data.

	#Node	#Link	Mean k	S.D. k_{in}	S.D. k_{out}	$\tau(k_{in}, k_{out})$	Reciprocity	Clustering
ED	181176	1260294	6.96	68.92	54.05	0.21	0.48	0.02
Random	1,604,802	76,535,565	47.69	375.74	258.11	0.71	0.67	0.03

Figure 4 shows the distributions of in- and out-degrees in social networks built on ED and random datasets. In both networks, in- and out-degrees have a heavy-tailed distribution, indicating that most users have a small number of connections to other users, while a few users have a very large number of connections. Such heterogeneity in degree distributions eliminates concerns that aggregate and average peer effects cannot be identified[10] due to a perfectly linear correlation between the local-aggregate endogenous variable GY and the local-average endogenous variable G^*Y. Fitting a power-law distribution for the degree distributions, the fitted values of exponent α are different in the two networks ($\alpha \approx 1.9$ in ED data and $\alpha \approx 2.7$ in random data, with comparable values of α between k_{in} and k_{out} in each network), indicating different network structures in the two groups of users. This is confirmed by the results of detailed network statistics shown in Table 3. We see that the average degree in random data is much larger than that in ED data, indicating relatively denser connections within random users on average though with larger variations in numbers of connections for individual users. Random users show a stronger tendency toward reciprocity, i.e., random users tend to follow their followers back[11]. This is confirmed by the larger value of Kendall rank correlation between in-degree and out-degree $\tau(k_{in}, k_{out})$ in random users than in ED users. Moreover, the results of global clustering coefficient show that both groups of users have a weak tendency toward transitivity, i.e., two neighbors of a node are less likely to be connected. Such intransitivity in social networks allows to use attributes of second-order followees as IVs for an ego's attributes.

5.2. **Estimation.** In this section, we present our empirical estimates of peer effects, focusing on emotional outcomes as proxied by linguistic sentiment in user tweets. For ease of interpretation, we begin with a core set of results using the primary specification and then summarize robustness checks (e.g., alternative thresholds for observed networks and second-order exposure).

Instruments. We are interested in estimating the values of peer effect factors ϕs and ψs in the model of equation (7):

[10] If all nodes had the same degree k, then $G = kG^*$ and hence GY would be perfectly correlated to G^*Y, see Proposition 3 in Liu et al. (2014) for more details.

[11] Our network statistics on the random data may differ from those reported in previous work Kwak et al. (2010) because (i) our network is a subset of the full 2009 network; (ii) inactive users, i.e., those who posted a small number of tweets, are excluded in building our networks.

$$Y = \phi_1 GY + \psi_1 G^* Y + \phi_2 G^2 Y + \psi_2 G^{*2} Y + \beta X + \lambda G^* X + \varepsilon.$$

Previous work (e.g., Liu et al. (2014)) has shows that these factors can be identified as long as nodes have different degrees in networks, while identification challenges exist due to endogenous issues. First, as already mentioned, *reverse causality* might arise from the inherent simultaneity in the process of peer effects (e.g., user i can affect her friend j's emotions and vice versa), but also from cycles in social networks (e.g., user i might affect user k's emotions and k then influence j's emotions). Second, like all social media studies, only a limited number of individuals' characteristics are available for the estimations and these are mostly observed through user-generated data online. This leads to *omitted variable* bias, as unobservable factors can be correlated with emotions of individuals and their followees. For example, both users and their friends might post content on positive emotions during holidays. Any of these issues can lead to biased estimates of ϕs and ψs.

We used two different approaches to address such endogeneity issues and obtain consistent estimates of peer effects. First, we only consider single-way followees of a user i, i.e., those who are followed by the user but do not follow the user back, as the friends of the user N_i. This is because posts of a user are less likely to be visible to their single-way followees, N_i, and hence emotions that the user expressed in the posts are less likely to influence friends' emotions, which breaks the reverse causality of a user's emotions on N_i's emotions. Second, we use instrumental variable (IV) estimation to address the omitted variable bias[12]. Following previous studies (e.g., Liu et al. (2014)), we use $Z = [GX, |N_i|]$ as instruments for $E = [GY, G^*Y]$. Based on these instruments, we then use two-stage least squares (2SLS) to estimate the model in Eq. (6). Similarly, instruments $Z = [GX, G^2X, |N_i|, |N_i^2|]$ are used for $E = [GY, G^*Y, G^2Y, G^{*2}Y]$ to estimate the extended medel in Eq. (7).

Empirical support for instrument relevance is established via strong first-stage F-statistics (all exceeding the standard threshold of 10). To further test instrument exogeneity, we conduct overidentification tests (Sargan tests). While we acknowledge that the exclusion restriction is ultimately untestable, our use of network structure for instrumentation is grounded in prior literature (e.g., Bramoullé, Djebbari, and Fortin 2009; De Giorgi, Pellizzari, and Redaelli 2010) and supported by the heavy-tailed degree distributions in both ED and random samples, which reduce concerns of perfect multicollinearity between aggregate and average exposures.

[12]An instrument for an endogenous variable E is a variable Z that (i) does not occur in the model equation, (ii) is correlated to E (i.e., $cov(Z, E) \neq 0$, conditional on all the exogenous explanatory variables) and (iii) is not correlated to the error term (i.e., $cov(Z, \epsilon) = 0$).

Estimation results. Our results are displayed in Table 4:

TABLE 4. IV Results, using original LIWC outcome, select random users whose fractions of observed followees are larger than 0.45.

	Dependent variable:							
	PosEmo		NegEmo		PosEmo		NegEmo	
	ED	Random	ED	Random	ED	Random	ED	Random
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
favg	0.394***	0.420***	0.100	0.488***	0.233	−0.065	0.092	0.768***
	(0.109)	(0.108)	(0.080)	(0.078)	(0.143)	(0.214)	(0.089)	(0.154)
fsum	0.114***	−0.111***	−0.033	−0.011	0.161**	−0.186***	−0.111*	−0.037
	(0.038)	(0.029)	(0.029)	(0.026)	(0.070)	(0.051)	(0.064)	(0.028)
ffavg					0.025	0.514***	0.090**	−0.138
					(0.053)	(0.185)	(0.038)	(0.156)
ffsum					−0.035	0.079*	0.100*	0.058**
					(0.049)	(0.041)	(0.053)	(0.026)
Xs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3060	3482	3060	3482	3060	3482	3060	3482
First-stage F-statistic[a]	22.102	26.972	29.761	33.808	3.228e+01	55.019	62.609	39.592
First-stage F-statistic[b]	3854.332	851.588	914.004	631.078	2.837e+03	795.845	1291.824	509.893
First-stage F-statistic[c]					1.129e+02	74.630	147.599	48.198
First-stage F-statistic[d]					1.346e+05	24962.532	24486.362	12455.708
P-value of Wu-Hausman test	0.152	7.72e-05	0.00112	0.000735	0.322	4.68e-06	0.0066	2.33e-05
P-value of Sargan test	0.481	0.387	0.08847	0.089711	0.137	0.966	0.1876	0.511

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4 shows estimation results for the peer effects on individuals' emotion outcomes in ED and random users using 2LSL. The coefficients on average exposure (favg) represent conformity effects: how strongly an individual's emotions align with the mean sentiment of their peers. The coefficients on aggregate exposure (fsum) reflect complementarity or multiplier effects, whereby the total volume of emotional input from peers modifies ego's emotional utility. All regressions control for ego and peer-level covariates.

The first two columns of Table 4 contain the results of positive emotions in ED and random samples when only first-degree friends are considered. We see that the average peer effect behaves similarly in both ED and random users on positive emotions, i.e., individual tend to follow and act more like the majority of a group. However, the aggregate peer effect (GY) shows different behaviors in ED and random users. Positive coefficients of GY among ED users (column 1) indicate that an ego who follows a larger number of positive friends tend to be more positive. This reflects a strategic complementarity (Bulow et al. (1985), also called social multiplier in Liu et al. (2014)), i.e., an increase in one friend's positive emotions increases the marginal utility of an ego being positive, motivating the ego to be more positive. To interpret magnitude, consider a one standard deviation increase in the aggregate positivity of peers. Among ED users, this is associated with a 0.11-point increase in ego positivity (Table 4, Column 1), indicating strategic complementarity. Among random users, the sign flips, suggesting strategic substitutability in emotional expression. In contrast, negative coefficients of GY among random users (column 2) show that an ego who follows a large number of positive friends tend be less positive. This reveals a strategic substitutability (Ballester et al. (2006)), i.e., an increase in one friend's positive emotions decreases the marginal utility of an ego being positive, giving the ego an incentive to be more negative.

The second two columns show the results of negative emotions in ED and random samples when only first-degree friends are considered. We see that both average and aggregated effects do not significantly influence an ego's negative emotions for ED users. In other words, changes of emotions among first-degree friends do not necessarily influence an ED user's emotions. This also aligns with evidence that people affected by ED often have normative dissatisfaction about their body image and tend to be depressed. In contrast, average peer effects significantly influence an ego's negative emotions in random users. This suggests that the endogenous peer effects on an ego's negative emotions are mostly captured by a social-conformity effect in random samples. In other words, the network position of an ego in an online community has no effect on the ego' emotions if all individuals in the network have the same emotional state. Moreover, coefficients of the average peer effects have positive values in all models, indicating that ED individuals attempt to conform a given social norm of their reference groups. These results align with previous evidence that an ego's friends posting positive/negative updates increases the positive/negative updates of the ego on social media (e.g., Coviello et al. (2014); Ferrara and Yang (2015); Kramer et al. (2014)).

The last four columns show the estimation results of peer effects when second-degree friends are considered. Although most results are consistent with those in the first four columns, several new patterns occur. First, the average effects of first-degree friends for positive emotions become insignificant, while the significantly aggregated effects of first-degree friends still hold. Moreover, in random users, the average effects are more derived from second-degree friends rather than the first-degree friends. Second, although the effects of the first-degree friends on negative emotions align with those in the first four columns, second-degree friends influence an ego's negative emotions in different ways in ED and random users. Specifically, the endogenous peer effects of second-degree friends on an ego's negative emotions are mostly captured by a social-conformity effect in ED users while a strategic complementarity effect of second-degree friends matters more in random users. Results highlight meaningful heterogeneity in emotional peer effects across network degrees. First-degree peers exert the strongest influence, particularly for conformity in emotional tone among both ED and random users. Aggregate reinforcement through first-degree exposure is a salient channel for ED users, consistent with emotional amplification mechanisms. In contrast, second-degree peers (friends of friends) exhibit weaker but detectable effects, especially for negative emotional expression among ED users. This suggests that emotional norms can cascade beyond direct connections, but the strength of influence attenuates with network distance. Such findings emphasize the localized nature of emotional contagion in vulnerable online communities, with diminishing peer effects at higher orders of separation.

TABLE 5. Summary of peer effects of positive and negative emotions in different groups.

		Without second-degree friends		With second-degree friends	
Effects	Group	Positive	Negative	Positive	Negative
Conform (first-degree friends)	ED	(+)			
	Random	(+)	(+)		(+)
Reinforce (first-degree friends)	ED	(+)		(+)	
	Random	(−)		(−)	
Conform (second-degree friends)	ED				(+)
	Random			(+)	
Reinforce (second-degree friends)	ED				
	Random				(+)

6. Policy Implications

Our findings shed light on how peer dynamics in online networks shape emotional expression, particularly within vulnerable communities such as those affected by eating disorders (ED). By distinguishing between average and aggregate peer effects, we reveal that social influence in digital settings operates through multiple behavioral channels—conformity to group norms and reinforcement through cumulative emotional exposure. These mechanisms have differential effects across populations and carry important implications for both clinical understanding and content moderation policies.

Among ED-identified users, we find strong evidence for aggregate peer effects in emotional expression. In particular, positive emotional exposure from peers is associated with a significant amplification in ego's own positivity. This result suggests that emotional contagion or reinforcement mechanisms are active in ED communities: the total volume of emotional content from followees increases the utility of emotional expression for the individual. By contrast, non-ED users exhibit negative aggregate peer effects for the same emotional domain, consistent with a crowding-out or emotional substitutability process. These differential responses highlight that platform structures can interact with group-level vulnerabilities to either amplify or dampen emotional behavior.

Average peer effects—reflecting norm conformity—are also present in both groups but play a more modest and symmetric role. This suggests that while users adjust their emotional tone to align with the perceived norm, the intensity and direction of affective engagement are more strongly influenced by the total emotional climate of their peer environment.

These findings are particularly relevant in the context of growing public concern over the role of social media in exacerbating mental health risks among adolescents and vulnerable groups. Recent investigative reports highlight that ED-related content continues to circulate widely on platforms such as Instagram and TikTok, often circumventing content moderation policies through euphemisms and coded imagery. Advocacy groups and regulators have called for structural reforms to algorithmic recommendation systems and visibility constraints on high-risk content.

Our results provide a behavioral micro-foundation for these concerns. The presence of strong aggregate peer effects implies that platforms which algorithmically amplify emotionally intense content may inadvertently reinforce the affective states associated with ED vulnerability. The structure of exposure—not just the content itself—matters: users embedded in dense, emotionally homogeneous networks are more likely to experience contagion-like amplification of sentiment.

This has several policy implications. First, interventions aimed at reducing the prevalence of ED-related harm online should not focus solely on removing individual items of content. Instead, platforms should develop tools to detect and disrupt structurally reinforcing emotional environments—e.g., through limiting exposure to high-density emotional clusters or introducing network-level diversity constraints. Second, public health strategies should consider emotional reinforcement as a critical channel through which digital ecosystems influence mental health. Exposure diversity, emotional balance in content feeds, and emotional education programs may all serve as preventative measures.

In a fully fledged policy study, the benefits of these interventions are, however, to be balanced against potential unintended costs. For example, Figure 5 summarizes both the intended benefits and the possible unintended consequences of interventions targeting emotional reinforcement mechanisms.

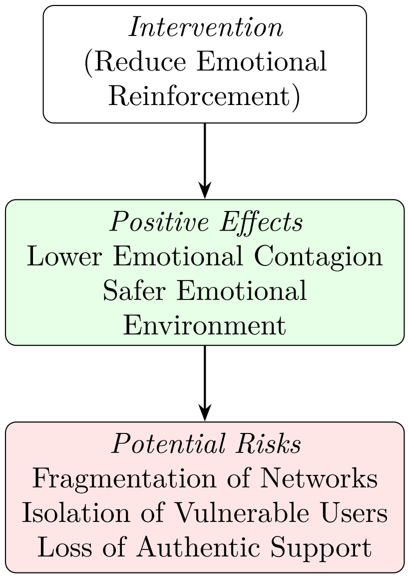


FIGURE 5. Potential Effects of Reducing Emotional Reinforcement in Online Communities.

DISCUSSION

6.1. **Methodology.** The distinction between average and aggregate peer effects has implications for policy design in online health communities. If conformity dominates (as in the case of negative emotions among random users), interventions may be more effective when targeted at reshaping perceived norms, for instance, by promoting emotionally positive content from a broad peer base. Conversely, if aggregate effects dominate (as for positive emotions among ED users), targeting structurally influential individuals—such as those with high out-degree or centrality—may amplify the reach of interventions via emotional contagion. Future work could operationalize these strategies using platform-level tools or community-based peer support programs.

On the Monotonicity of Peer Effects. While our model assumes a linear and additive relationship between an individual's emotional expression and peer exposure, the empirical findings suggest a more nuanced and potentially non-monotonic structure. Specifically, although the best-response formulation implies that increases in peer emotional expression (whether measured as an average or aggregate) should have proportionate and directionally consistent effects, our estimates reveal heterogeneity in both sign and magnitude across user groups, sentiment types, and network layers.

For instance, ED users exhibit positive aggregate peer effects (strategic complementarity) for positive sentiment, while random users display negative aggregate effects (strategic substitutability). Moreover, we observe sign reversals when moving from first-order to second-order peer exposure, particularly for negative sentiment. These patterns are difficult to reconcile with a strictly monotonic peer influence mechanism.

This heterogeneity points to possible threshold effects, saturation, or asymmetries in emotional responsiveness that are not captured by our linear specification. As such, future work could enrich the modelling framework by incorporating nonlinear peer exposure terms (e.g., quadratic, spline-based, or log transformations), interaction effects (e.g., between average and aggregate exposure), or even regime-switching dynamics where the nature of peer influence depends on the density or emotional valence of the network. Such extensions would allow us to more fully characterize the conditions under which emotional contagion intensifies, dampens, or reverses, and could provide more targeted insights for intervention design in vulnerable digital communities. In the next Section we highlight some possible extensions.

6.2. **Possible Extensions.** Our theoretical model captures emotional expression as a strategic response to peer influence through both conformity to norms and reinforcement from emotional exposure. While the baseline formulation offers a tractable structure for empirical estimation, it can be naturally extended to yield policy-relevant insights about how emotional environments in online networks shape behavior—particularly in vulnerable populations.

Several theoretical extensions and associated econometric strategies for estimating emotional peer effects in online social networks, especially within vulnerable populations, are listed below.

6.2.1. *Nonlinear Emotional Reinforcement and Thresholds.* The reinforcement mechanism in our model assumes linear amplification: greater exposure to peer emotion linearly increases the marginal utility of emotional expression. However, in psychologically charged contexts such as eating disorder (ED) communities, exposure may exhibit non-monotonic effects. Users may initially be drawn into affective amplification, but beyond a certain threshold of emotional saturation, they may disengage, suppress, or withdraw from expression.

To capture this, we generalize the reinforcement term:

$$\omega \left(\sum_j g_{ij} y_j \right) y_i \quad \longrightarrow \quad \omega \cdot f \left(\sum_j g_{ij} y_j \right) \cdot y_i,$$

where $f(\cdot)$ is a nonlinear function with a saturation point $\tau > 0$. For instance,

$$f(z) = \begin{cases} z, & \text{if } z < \tau, \\ 2\tau - z, & \text{if } z \geq \tau, \end{cases}$$

models reinforcement followed by emotional overload beyond the critical threshold τ. Such nonlinearity introduces a new behavioral dynamic where exposure may initially amplify emotion, but excessive density triggers avoidance or emotional burnout.

Hypothesis: Emotional reinforcement is not monotonic; high peer emotional exposure leads to diminishing or reversing effects on ego emotion. To capture potential saturation effects in reinforcement, we include a squared term:

$$y_i = \phi_1 \cdot \texttt{fsum}_i + \phi_2 \cdot \texttt{fsum}_i^2 + \psi \cdot \texttt{favg}_i + X_i'\beta + \varepsilon_i. \tag{8}$$

A significant negative ϕ_2 would indicate diminishing returns or emotional overload.

6.2.2. *Network-Specific Susceptibility to Influence.* The model may also be extended to allow heterogeneity in peer effect sensitivity. Specifically, we let γ_i (norm conformity) and ω_i (reinforcement) vary with ego's network characteristics:

$$\gamma_i = \bar{\gamma} + \delta_1 \cdot \text{degree}_i, \quad \omega_i = \bar{\omega} + \delta_2 \cdot \text{exposure heterogeneity}_i.$$

This formulation recognizes that more central users or those embedded in emotionally homogeneous subnetworks may experience stronger conformity pressures or contagion effects. Such heterogeneity is consistent with evidence on structural vulnerability in digital mental health research.

Hypothesis: The strength of peer effects varies depending on ego's network position or exposure structure. To test for heterogeneous reinforcement based on ego's network position, we include interaction terms:

$$y_i = \phi_1 \cdot \texttt{fsum}_i + \phi_2 \cdot (\texttt{fsum}_i \cdot \texttt{degree}_i) + \psi \cdot \texttt{favg}_i + X_i'\beta + \varepsilon_i. \tag{9}$$

A significant ϕ_2 implies variation in sensitivity to peer influence depending on network centrality.

6.2.3. *Policy Levers: Reinforcement Disruption.* To model intervention strategies, we introduce a policy parameter $\pi \in [0, 1]$ that attenuates the reinforcement channel:

$$U_i = \alpha y_i - \frac{\gamma}{2}\left(y_i - \sum_j g_{ij}^* y_j\right)^2 + (1-\pi)\omega\left(\sum_j g_{ij} y_j\right) y_i - \mu y_i^2.$$

Here, $\pi = 0$ corresponds to no intervention (full reinforcement), while $\pi = 1$ represents complete elimination of reinforcement dynamics—achievable, for instance, through algorithmic disruption of affectively concentrated content exposure.

Hypothesis: Platform-level interventions that reduce reinforcement (π) flatten emotional expression. To evaluate the effects of content moderation policies, we simulate reduced reinforcement:

$$y_i^{(\pi)} = (1-\pi) \cdot \phi \cdot \texttt{fsum}_i + \psi \cdot \texttt{favg}_i + X_i'\beta + \varepsilon_i, \tag{10}$$

where $\pi \in [0, 1]$ represents the intensity of intervention (e.g., $\pi = 1$ fully removes reinforcement).

We could then scale ϕ by $(1-\pi)$ for $\pi \in [0, 1]$ and use estimated coefficients to predict new y_i values under different policy regimes. Graph distributions of y_i under different π levels would illustrate effects of limiting reinforcement.

7. Conclusions

This paper makes several contributions to the empirical study of peer effects and emotional dynamics in online environments.

First, we develop a utility-based model that distinguishes between *average* and *aggregate* peer effects, corresponding respectively to social conformity and emotional reinforcement. While the linear-in-means framework is well established in the peer effects literature, our model incorporates both mechanisms in a unified specification and provides a behavioral interpretation grounded in affective identity and psychological utility.

Second, we operationalize this framework using large-scale, real-world network data, derived from observed Twitter relationships and linguistic sentiment. To our knowledge, this is one of the first studies to structurally estimate emotional peer effects in a high-dimensional online environment with vulnerable populations, leveraging both direct and second-order peer measures. This approach, grounded in the econometric network literature allows us to isolate endogenous peer effects from contextual and correlated effects, while accounting for the layered topology of digital social networks.

Third, our findings contribute to a growing literature on peer effects and mental health in digital environments. Unlike prior work that has primarily documented associative patterns or predictive relationships, we develop a structural framework capable of distinguishing conformity from reinforcement mechanisms. By combining directed network data with a causal identification strategy, we provide sharper insights into how emotional environments propagate in vulnerable online communities. These methodological innovations open new avenues for research on digital public health interventions, but also yield several policy-relevant contributions to the current debate on how to intervene and regulate social media platforms to mitigate their potential impact on eating disordered communities:

- *Targeting amplification, not just norms:* While shifting platform norms (e.g., via visible moderation or counter-messaging) may influence conformity behavior, regulating aggregate exposure—by capping the emotional volume shown to users—directly disrupts the contagion mechanism.
- *Threshold-aware content design:* Algorithms that detect emotionally saturated timelines can be used to intervene before overload effects occur, mitigating emotional reinforcement at critical points.
- *Personalized safeguards based on network position:* Users with high centrality or affective clustering may warrant greater protection or content diversity nudges, given their heightened susceptibility to contagion.
- *Behavioral heterogeneity as a design input:* Recognizing that peer effects are structurally asymmetric across users enables adaptive policy targeting—particularly relevant in ED-related or mental health contexts.

This framework can support platform-level interventions, such as throttling emotionally extreme content, diversifying timelines, or recommending followees that reduce affective homogeneity. It is, however, worth conjecturing that while such interventions hold promises, they may also generate unintended consequences. Reducing aggregate emotional exposure could fragment vulnerable users into smaller, more homogeneous sub-communities, where emotional feedback loops may become even more intense. Additionally, algorithmic diversification strategies, if applied too aggressively, might suppress authentic community support or inadvertently isolate users seeking emotional connection (e.g., Figure 5). These risks underscore the importance of carefully balancing exposure management with preservation of social integration and emotional authenticity. Future research should explore

Table 6. Mapping Policy Recommendations to Empirical Findings

Policy Recommendation	**Supporting Empirical Findings**
Targeting amplification, not just norms	Significant *aggregate peer exposure* (fsum) effects in ED users, especially for positive emotions, reflecting *strategic complementarity* (emotional amplification).
Threshold-aware content design	Influence from second-degree friends attenuates but remains detectable; emotional impact depends on *intensity* of network exposure, suggesting threshold and nonlinear effects.
Personalized safeguards based on network position	Emotional influence varies with *network distance*: users more embedded in the network (higher centrality) are more exposed to reinforcement effects, requiring position-sensitive protection.
Behavioral heterogeneity as a design input	Peer effects differ sharply between ED and Random users: ED users show stronger amplification and less negative emotion conformity, highlighting *group-specific emotional sensitivity.*

adaptive content moderation strategies that dynamically monitor emotional density and network fragmentation.

Finally, our framework is extensible: it accommodates nonlinear peer exposure terms, enabling tests of monotonicity, thresholds, and saturation effects in emotional reinforcement. This flexibility makes it suitable for application in other emotionally salient digital contexts, such as anxiety or self-harm forums, and offers a basis for modeling policy interventions that target either normative alignment or structural amplification.

Appendix A. Robustness checks

Table 7. IV Results, using original LIWC outcome, select random users whose fractions of observed followees are larger than 0.48.

	Dependent variable:							
	PosEmo		NegEmo		PosEmo		NegEmo	
	Random	ED	Random	ED	Random	ED	Random	ED
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
favg	0.547***	0.394***	0.502***	0.100	0.308*	0.233	0.398**	0.092
	(0.077)	(0.109)	(0.065)	(0.080)	(0.187)	(0.143)	(0.201)	(0.089)
fsum	−0.022	0.114***	0.023	−0.033	−0.042	0.161**	0.024	−0.111*
	(0.020)	(0.038)	(0.015)	(0.029)	(0.028)	(0.070)	(0.016)	(0.064)
ffavg					0.138	0.025	0.124	0.090**
					(0.140)	(0.053)	(0.171)	(0.038)
ffsum					0.019	−0.035	−0.006	0.100*
					(0.022)	(0.049)	(0.018)	(0.053)
Xs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7653	3060	7653	3060	7653	3060	7653	3060
First-stage F-statistic[a]	50.979	22.102	46.736	29.761	44.068	3.228e+01	32.653	62.609
First-stage F-statistic[b]	521.673	3854.332	319.086	914.004	637.600	2.837e+03	344.585	1291.824
First-stage F-statistic[c]					103.904	1.129e+02	64.696	147.599
First-stage F-statistic[d]					17837.821	1.346e+05	6437.116	24486.362
P-value of Wu-Hausman test	4.69e-06	0.152	0.000743	0.00112	0.000742	0.322	0.00508	0.0066
P-value of Sargan test	0.37	0.481	0.899503	0.08847	0.065659	0.137	0.33559	0.1876

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8. Summary of peer effects of positive and negative emotions in different groups (subset of random users with >48% observed followees).

		Without second-degree friends		With second-degree friends	
Effects	Group	Positive	Negative	Positive	Negative
Conform (first-degree friends)	ED	(+)			
	Random	(+)	(+)	(+)	(+)
Reinforce (first-degree friends)	ED	(+)		(+)	(−)
	Random				
Conform (second-degree friends)	ED				(+)
	Random				
Reinforce (second-degree friends)	ED				
	Random				

Table 9. IV Results, using original LIWC outcome, select random users whose fractions of observed followees are larger than 0.52.

	Dependent variable:							
	PosEmo		NegEmo		PosEmo		NegEmo	
	Random	ED	Random	ED	Random	ED	Random	ED
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
favg	0.722***	0.394***	0.457***	0.100	0.852**	0.233	0.191	0.092
	(0.150)	(0.109)	(0.112)	(0.080)	(0.366)	(0.143)	(0.185)	(0.089)
fsum	−0.036	0.114***	0.033	−0.033	−0.030	0.161**	0.041*	−0.111*
	(0.028)	(0.038)	(0.024)	(0.029)	(0.038)	(0.070)	(0.023)	(0.064)
ffavg					−0.210	0.025	0.236	0.090**
					(0.189)	(0.053)	(0.167)	(0.038)
ffsum					−0.003	−0.035	−0.018	0.100*
					(0.043)	(0.049)	(0.022)	(0.053)
Xs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contextual Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3847	3060	3847	3060	3847	3060	3847	3060
First-stage F-statistic[a]	12.514	22.102	23.582	29.761	13.246	3.228e+01	13.643	62.609
First-stage F-statistic[b]	1551.069	3854.332	1159.409	914.004	856.150	2.837e+03	619.877	1291.824
First-stage F-statistic[c]					54.199	1.129e+02	37.738	147.599
First-stage F-statistic[d]					10370.765	1.346e+05	5523.597	24486.362
P-value of Wu-Hausman test	0.0214	0.152	0.283	0.00112	0.123	0.322	0.562	0.0066
P-value of Sargan test	0.5152	0.481	0.732	0.08847	0.622	0.137	0.831	0.1876

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10. Summary of peer effects of positive and negative emotions in different groups (subset of random users with >52% observed followees).

		Without second-degree friends		With second-degree friends	
Effects	Group	Positive	Negative	Positive	Negative
Conform (first-degree friends)	ED	(+)			
	Random	(+)	(+)	(+)	(+)
Reinforce (first-degree friends)	ED	(+)		(+)	(−)
	Random				(+)
Conform (second-degree friends)	ED				(+)
	Random				
Reinforce (second-degree friends)	ED				
	Random				(+)

References

Akerlof, G. A. and R. E. Kranton (2000a): "Economics and identity," *Quarterly Journal of Economics*, 115, 715–753.

——— (2000b): "Economics and identity," *The Quarterly Journal of Economics*, 115, 715–753.

Association, A. P. et al. (2013): *Diagnostic and statistical manual of mental disorders (DSM-5®)*, American Psychiatric Pub.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: the key player," *Econometrica*, 74, 1403–1417.

Bénabou, R. and J. Tirole (2002): "Self-confidence and personal motivation," *Quarterly Journal of Economics*, 117, 871–915.

——— (2011): "Identity, morals, and taboos: Beliefs as assets," *Quarterly Journal of Economics*, 126, 805–855.

Bernheim, B. D. (1994a): "A theory of conformity," *Journal of Political Economy*, 102, 841–877.

——— (1994b): "A theory of conformity," *Journal of Political Economy*, 102, 841–877.

Birnbaum, M. L., S. K. Ernala, S. L. Rizvi, M. De Choudhury, and J. M. Kane (2020): "Modeling peer influence on disclosure of mental illness on social media," *Psychiatric Services*, 71, 365–371.

Bollen, J., B. Gonçalves, I. van de Leemput, and G. Ruan (2017): "The happiness paradox: your friends are happier than you," *EPJ Data Science*, 6, 4.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

Brock, W. A. and S. N. Durlauf (2001): "Interactions-based models of social influences," *Handbook of Econometrics*, 5, 3297–3380.

Bulow, J. I., J. D. Geanakoplos, and P. D. Klemperer (1985): "Multimarket oligopoly: Strategic substitutes and complements," *Journal of Political economy*, 93, 488–511.

Bursztyn, L. and R. Jensen (2017): "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure," *Annual Review of Economics*, 9, 131–153.

Chancellor, S., Y. Kalantidis, J. A. Pater, M. De Choudhury, and D. A. Shamma (2017): "Multimodal Classification of Moderated Online Pro-Eating Disorder Content," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 3213–3226.

Chancellor, S., Z. Lin, E. Goodman, S. Zerwas, and M. De Choudhury (2016a): "Quantifying and predicting mental illness severity in online pro-eating disorder communities," *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1171–1184.

Chancellor, S., T. Mitra, and M. De Choudhury (2016b): "Recovery Amid Pro-Anorexia: Analysis of Recovery in Social Media," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 2111–2123.

Chu J, Ganson KT, T. A. A.-S. A. J. D. R. R. H. J. B. F. N. J. (2024): "Screen time, problematic screen use, and eating disorder symptoms among early adolescents: findings from the Adolescent Brain Cognitive Development (ABCD) Study," .

Costa-Font, J. and M. Jofre-Bonet (2013): "Anorexia, body image and peer effects: evidence from a sample of European women," *Economica*, 80, 44–64.

Coviello, L., Y. Sohn, A. D. Kramer, C. Marlow, M. Franceschetti, N. A. Christakis, and J. H. Fowler (2014): "Detecting emotional contagion in massive social networks," *PloS one*, 9, e90315, [PMID: 24621792] [DOI: 10.1371/journal.pone.0090315].

Dane, A. and K. Bhatia (2023): "The social media diet: A scoping review to investigate the association between social media, body image and eating disorders amongst young people." *PLOS global public health*, 3.

De Choudhury, M. (2015): "Anorexia on tumblr: A characterization study," in *Proceedings of the 5th International Conference on Digital Health 2015*, ACM, 43–50.

de Giorgi, G., M. Pellizzari, and S. Redaelli (2010): "Identifying peer effects on health in social networks: A comparison of econometric techniques," *Journal of Health Economics*, 29, 659–673.

Ferrara, E. and Z. Yang (2015): "Measuring emotional contagion in social media," *PloS one*, 10, e0142390.

Ghiglino, C. and S. Goyal (2010): "Keeping up with the neighbors: social interaction in a market economy," *Journal of the European Economic Association*, 8, 90–119.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (2003): "Social interactions and the efficiency of crime prevention," *American Economic Review*, 90, 247–251.

Gonçalves, P., M. Araújo, F. Benevenuto, and M. Cha (2013): "Comparing and combining sentiment analysis methods," in *Proceedings of the first ACM conference on Online social networks*, ACM, 27–38, [DOI: 10.1145/2512938.2512951].

Helfert, S. and P. Warschburger (2013): "The face of appearance-related social pressure: gender, age and body mass variations in peer and parental pressure during adolescence," *Child and Adolescent Psychiatry and Mental Health*, 7, 16.

Hyatt, R. E., D. P. Schilder, and D. L. Fry (1958): "Relationship between maximum expiratory flow and degree of lung inflation," *Journal of Applied Physiology*, 13, 331–336.

Kramer, A. D. I., J. E. Guillory, and J. T. Hancock (2014): "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences*, 111, 8788–8790.

Kwak, H., C. Lee, H. Park, and S. Moon (2010): "What is Twitter, a social network or a news media?" in *Proceedings of the 19th international conference on World wide web*, ACM, 591–600.

Liu, X., E. Patacchini, and Y. Zenou (2014): "Endogenous peer effects: local aggregate or local average?" *Journal of Economic Behavior & Organization*, 103, 39–59.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542.

Nawaz, F., M. M. A. Riaz, N. Banday, A. Singh, Z. Arshad, H. Derby, and M. Sultan (2024): "Social media use among adolescents with eating disorders: a double-edged sword," *Frontiers in Psychiatry*, 15.

Oksanen, A., D. Garcia, A. Sirola, M. Näsi, M. Kaakinen, T. Keipi, and P. Räsänen (2015): "Pro-anorexia and anti-pro-anorexia videos on YouTube: Sentiment analysis of user responses," *Journal of medical Internet research*, 17.

Oxman, T. E., S. D. Rosenberg, and G. J. Tucker (1982): "The language of paranoia." *The American journal of psychiatry.*

Pennebaker, J. W., M. R. Mehl, and K. G. Niederhoffer (2003): "Psychological aspects of natural language use: Our words, our selves," *Annual review of psychology*, 54, 547–577.

Polivy, J. (2017): "What's that you're eating? Social comparison and eating behavior," *Journal of eating disorders*, 5, 18.

Ribeiro, F. N., M. Araújo, P. Gonçalves, M. A. Gonçalves, and F. Benevenuto (2016): "Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods," *EPJ Data Science*, 5, 1–29.

Snijders, T. A., A. Lomi, and V. J. Torló (2013): "A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice," *Social networks*, 35, 265–276.

Sowles, S. J., M. McLeary, A. Optican, E. Cahn, M. J. Krauss, E. E. Fitzsimmons-Craft, D. E. Wilfley, and P. A. Cavazos-Rehg (2018): "A content analysis of an online pro-eating disorder community on Reddit," *Body image*, 24, 137–144.

Tausczik, Y. R. and J. W. Pennebaker (2010): "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of language and social psychology*, 29, 24–54.

Thelwall, M., K. Buckley, G. Paltoglou, D. Cai, and A. Kappas (2010): "Sentiment strength detection in short informal text," *Journal of the Association for Information Science and Technology*, 61, 2544–2558.

Wang, T., M. Brede, A. Ianni, and E. Mentzakis (2017): "Detecting and characterizing eating-disorder communities on social media," in *Proceedings of the Tenth International Conference on Web Search and Data Mining (WSDM) 2017*, ACM, 91–100.

Wolf, M., J. Sedway, C. M. Bulik, and H. Kordy (2007): "Linguistic analyses of natural written language: Unobtrusive assessment of cognitive style in eating disorders," *International journal of eating disorders*, 40, 711–717.