

C A G E

Working Paper

824/2026
September 2026

Social Comparison and Young Adult Mental Health with Endogenous Network Formation *

Junwei Fan,
Santiago Montoya-Blandon,
Eugenio Proto

ISSN: 2978-0276
Grant number: ES/Z504701/1

Social Comparison and Young Adult Mental Health with Endogenous Network Formation ^{*}

Junwei Fan[†] Santiago Montoya-Blandon[‡] Eugenio Proto[§]

September 25, 2026

Abstract

We examine how social comparison within endogenously formed peer networks affects different measures of young adult mental health. Using over 2 million smartphone-based communication records from the NetHealth Study, we construct objective social networks among university students and estimate a structural model in which individuals choose their peers and mental health depends on the income of chosen peers. When networks are treated as exogenous, higher average peer income predicts better mental health; once network formation is modelled endogenously, the relationship reverses: exposure to higher-income peers significantly worsens depressive symptoms. Income gaps harm mental health in both directions, consistent with inequality aversion rather than relative deprivation. The reversal survives extending binary links to message-weighted networks, and counterfactual exercises show that policies reducing the volume of messages exchanged with peers improve mental health, increasingly so with the size of the reduction, though the effect of targeted reductions depends on which ties are attenuated. Ignoring endogenous peer selection can thus yield qualitatively misleading conclusions about peer effects on mental health.

Keywords: social networks; relative status; mental health; network formation.

JEL: D63, D85, D91, I14, I31.

^{*}We thank Eleonora Brandimarti, Maria Nareklshvili, and Áureo de Paula, as well as participants at the Adam Smith Business School (ASBS) Economics PhD Reading Group (October 2025), the ASBS Applied Economics Brownbag Seminar (February 2026), the ASBS PhD Workshop (April 2026), the 2026 Scottish Behavioural Science Conference in Stirling, the Social Network Analysis in Scotland (SNAS) seminar, the 2026 International Association for Applied Econometrics (IAAE) Annual Conference in Lisbon, and the Tenth International Meeting on Experimental and Behavioural Social Sciences (IMEBESS 2026) in Prague. We also thank the High Performance Computing Cluster (HPCC) at the University of Glasgow.

[†]University of Glasgow, Adam Smith Business School

[‡]University of Glasgow, Adam Smith Business School

[§]University of Glasgow, Adam Smith Business School; CEPR; CESifo; IZA

1 Introduction

Concerns about young adult mental health have intensified in recent years, as indicators of anxiety, depression, and loneliness have deteriorated sharply among adolescents and young adults (e.g. [Currie, 2025](#); [Zhang et al., 2023](#)). A growing body of work argues that this decline is closely related to the rise of digital social platforms. [Twenge \(2024\)](#) and [Haidt \(2024\)](#) document that the widespread adoption of smartphones and social media in the early 2010s coincided with an abrupt and unprecedented deterioration in youth psychological well-being. Their evidence suggests this new environment in which social comparisons increasingly take place may amplify isolation and vulnerabilities among young people.

This perspective is consistent with a large interdisciplinary literature showing that individuals evaluate their well-being not only in absolute terms but also relative to the outcomes of others. Evidence from psychology indicates that social comparisons can affect stress and psychological well-being in complex ways, particularly when the reference group itself is endogenously determined (e.g. [Diener and Fujita, 2013](#); [Rosenquist et al., 2011](#); [Bárcena-Martín et al., 2017](#); [Buunk and Dijkstra, 2017](#)). In parallel, contributions in economics link relative income to psychological well-being (e.g. [Ferrer-i Carbonell, 2005](#); [Luttmer, 2005](#); [Clark and Senik, 2010](#)), emphasizing the importance of the specific peers or reference groups individuals compare themselves to.

Because the effects of social comparison depend crucially on whom individuals select as reference points, properly defining peer groups is central to empirical analysis of the determinants of subjective well-being and mental health ([Falk and Knell, 2004](#)). Yet much of the existing literature relies on administratively or geographically defined reference groups, even though extensive work in network science demonstrates that individuals choose peers in ways systematically correlated with both observed and unobserved traits ([McPherson et al., 2001](#); [Jackson, 2008](#); [Feiler and Kleinbaum, 2015](#)). The networks chosen by individuals are also known to be sparser and less structured than the complete groupings used in much applied research, with strategic behaviour and noise playing large roles in network formation processes. When these unobservables influence both network formation and mental health, conventional peer-effect estimates risk substantial endogenous-selection bias, requiring corrections in the spirit of Heckman selection models ([Heckman, 1979](#); [Bramoullé et al., 2009](#); [Goldsmith-Pinkham and Imbens, 2013](#); [Hsieh and Lee, 2016](#); [Hsieh et al., 2022](#)).

This paper examines how social comparison within endogenously formed social networks affects mental health among young adults. We develop a structural framework in which individuals form communication ties through strategic linking based on homophily, popularity, personality traits, and relative-income considerations. Mental health is modeled as a function of an individual’s own income and the income of his or her endogenously chosen peers, enabling identification of income peer effects on mental health while correct-

ing for reference group selection. Our empirical analysis draws on the NetHealth Study, a longitudinal dataset of university students with over 2.19 million smartphone-based communication records and containing information on mental-health assessments, personality tests, parental income, and other demographic characteristics for each responder. These high-frequency communication records allow us to construct objectively measured social networks of varying intensity, addressing a well-documented concern that self-reported networks and censored friendship surveys incompletely capture individuals’ true interaction sets (Marmaros and Sacerdote, 2006; Eagle et al., 2009; Griffith, 2022). We infer social structure directly from communication activity, following approaches developed by Kossinets and Watts (2006) and Marmaros and Sacerdote (2006) to identify social ties from observed interaction patterns. This avoids relying on administratively assigned or self-reported peer groups, as done in large-scale studies such as Add Health, which have been central to research on peer effects and network formation (Bramoullé et al., 2009; Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Jackson et al., 2017).

The paper makes at least four contributions. First, and most strikingly, we show that not accounting for endogenous peer selection leads to substantively biased peer-effect estimates, such that the sign can be reversed. Under exogenous network assumptions, exposure to higher-income peers appears to improve mental health, consistent with positive spillovers from a wealthier social environment. Once we allow networks to form endogenously, individuals surrounded by higher-income peers exhibit significantly worse mental health outcomes (conditional on selection), consistent with adverse upward social comparison (Fehr and Schmidt, 1999; Clark et al., 2018). This reversal demonstrates that endogenous peer selection is a first-order econometric concern and a substantive force that shapes the relationship between peer income and mental health.

Second, decomposing average peer income into upward and downward comparisons reveals that income gaps are harmful in both directions: both being poorer and being richer than one’s peers increases depression. This pattern is inconsistent with standard relative deprivation theory — which assigns welfare costs only to unfavourable upward comparisons — and points instead toward inequality aversion, a generalised discomfort with economic distance from one’s reference group. This finding provides novel empirical support for homophily as a factor impacting welfare: individuals prefer economically similar peers, and deviations from similarity in either direction carry psychological costs.

Third, we integrate personality traits into the network formation process. Personality is known to shape sociability, behavioural responses, and economic decision-making (Borghans et al., 2008; Almlund et al., 2011; Becker et al., 2012), yet personality-driven heterogeneity is typically omitted from empirical network formation models (Fang et al., 2015; Landis, 2016). Embedding personality into link formation decomposes standard degree heterogeneity effects (Graham, 2017), providing a largely unexplored behavioural foundation, and helping the econometric methodology purge peer-income estimates of a source of confounding that would otherwise remain unaddressed.

Finally, exploiting our rich data on individual communication intensity, we expand upon standard models of binary link formation and introduce a model that can accommodate weighted networks, where the weight reflects link strength across peers. We show that personality similarity becomes a major factor driving stronger communication, while the intensity likelihood identifies the response of communication to income distance only as a whole, and not its upward and downward components separately. The intensity specification produces similar qualitative conclusions for peer effects: exogenous networks point to improvements in mental health, whereas accounting for endogenous communication intensity shows a sign-reversal; income gaps produce negative impacts on young adult mental health.

Because the model describes jointly how income shapes mental health and how it shapes the network, the estimated parameters support counterfactual exercises that a reduced-form specification cannot deliver. We conduct two, corresponding to the two margins of peer influence in our framework. The first perturbs incomes and traces the response of depressive symptoms once the network is allowed to re-form. Raising the incomes of all others worsens young adults’ depressive symptoms by more than one BDI level, whereas the fixed-network benchmark predicts an improvement; raising only the incomes of an individual’s existing peers has a markedly smaller, though still adverse, effect. The asymmetry is informative about the mechanism: when only current peers become richer, individuals partially insulate themselves by severing those ties and forming new ones with relatively poorer individuals, so that relative-status concerns in link formation act as a self-protective force. When incomes rise across the board, that escape route is unavailable and the full cost of comparison is borne — a distinction between individual and aggregate income growth that is invisible to models treating reference groups as fixed. The second exercise uses the intensity model to simulate policies that reduce the volume of messages exchanged with peers, the margin on which school phone bans and platform time caps operate. A uniform reduction improves mental health, increasingly so with the size of the cut, but targeted reductions work in opposite directions: cutting communication with poorer peers worsens mental health, while cutting it with richer peers improves it, since each reweights the peer income distribution the other way. Policies restricting digital communication among young adults therefore have effects that depend on which ties they attenuate, not merely on how much contact they remove.

The remainder of the paper is organised as follows. Section 2 presents the econometric framework, developing in turn the mental health model under exogenous reference groups and its extension to endogenous network formation. Section 3 describes the NetHealth dataset used to estimate both models, discussing its key features and the variables of interest. Section 4 presents the main empirical results, comparing estimates from the exogenous and endogenous specifications and documenting the central role of peer selection on the mental health effect of peer’s income. Section 5 refines the analysis by decomposing average peer income into its upward and downward comparison components, examining the

distinct mental health consequences of being poorer or richer than one’s peers. Section 6 extends the framework from binary links to weighted networks, modelling the endogenous intensity of communication between peers and re-estimating the mental health equation with message-weighted peer income. Section 7 exploits the estimated structural parameters to conduct two counterfactual exercises, one for each margin of peer influence: the first perturbs incomes and traces the response of depressive symptoms when the network is allowed to re-form, benchmarked against a model in which it is held fixed; the second simulates policies that reduce the volume of messages exchanged with peers, applied uniformly and targeted at richer or poorer peers. Section 8 summarises the main findings and concludes. Technical details of the estimation procedure are collected in Appendix A, data definitions and descriptive statistics in Appendix B, and additional results and counterfactual exercises in Appendix C.

2 Mental Health Peer Effects with Social Networks

Consider a set of n agents $\mathcal{N} = \{1, \dots, n\}$ connected through an undirected network $G \in \mathcal{G}_n$, where $\mathcal{G}_n$ is the set of all potential undirected networks with n players. G is represented by its $n \times n$ binary adjacency matrix, with common element $g_{ij} \in \{0, 1\}$; $g_{ij} = 1$ if there is a connection between agent i and j , $g_{ij} = 0$ otherwise. We rule out self-connections by setting $g_{ii} = 0 \ \forall i \in \mathcal{N}$, and the undirected nature of the network implies $g_{ij} = g_{ji} \ \forall i, j \in \mathcal{N}$, so that G is symmetric and a link is formed only if both i and j agree (mutual consent).

The reference group for agent i is denoted as $\mathcal{N}_i = \{j \in \mathcal{N} : g_{ij} = 1\}$ and $d_i = \sum_{j \in \mathcal{N}} g_{ij}$ denotes the degree, or number of friends, for agent i . We additionally denote an $n \times n$ real-valued adjacency matrix G^* as the row-normalized version of G ; i.e., weighting the rows by the number of friends for each agent i , resulting in $g_{ij}^* = g_{ij} / \sum_{k \in \mathcal{N}} g_{ik} \in [0, 1]$ for each off-diagonal entry $i, j \in \mathcal{N}$ such that $i \neq j$.

2.1 Mental Health and Peers’ Income in Exogenous Networks

Consider the following specification to model individual mental health for $i \in \mathcal{N}$, including own- and peer-effects of income:

$$MH_i = \beta_0 + \beta_1 \underbrace{y_i}_{\text{own income}} + \beta_2 \underbrace{y_i^*}_{\text{average peers' income}} + \beta_X^\top \mathbf{X}_i + \eta_i, \quad (1)$$

where MH_i denotes a measure of mental health for individual i , y_i represents parental income (an exogenous proxy for income), $y_i^* = \sum_{j=1}^n g_{ij}^* y_j = \sum_{j \in \mathcal{N}_i} g_{ij}^* y_j$ is the average

income of peers over G expressed in terms of individual i 's reference group $\mathcal{N}_i$, $\mathbf{X}_i$ is a vector of individual characteristics, and η_i are unobserved factors that impact mental health, assumed to be independent and identically distributed among individuals. In this specification β_1 represents the direct marginal effect of income and β_2 represents the peer effect of marginally higher income within the reference group (which can follow from either larger peer income or increased connectivity). In the model, MH_i is a stock of mental health — a form of human capital that enters positively, so that higher values denote better psychological well-being. The measures available to us, however, are depression inventories, on which higher scores denote more severe symptoms. Estimated coefficients therefore carry the opposite sign to their theoretical counterparts: a negative estimate corresponds to an improvement in mental health and a positive estimate to a deterioration. We return to this when describing the data in Section 3.

The assumption that the network or reference group G is exogenous imposes a restriction on its statistical relationship with the unobservables η_i that affect well-being. In particular, this exogeneity assumption can be expressed as

$$\mathbb{E}[\eta_i \mid G, \mathbf{y}, \mathbf{X}] = 0,$$

which requires the unobserved determinants of mental health to be mean-independent of the network structure conditional on observables and the income distribution. Similar assumptions appear in the literature on linear-in-means and social interaction models, which rely on comparable orthogonality conditions to identify peer effects (see, for example, [Bramoullé et al., 2009](#); [Lewbel et al., 2023](#)).

In the next section, we relax this assumption by endogenizing network formation. Once the network is allowed to respond to individual characteristics and unobserved heterogeneity, the condition above no longer holds, and the unobservables are generally correlated with G .

2.2 Network Formation

Because we construct the network from observed interpersonal communication rather than from administrative or geographic groupings, the reference groups $\mathcal{N}_i$ in Equation (1) cannot be treated as exogenously given. Individuals choose whom they interact with and, by extension, whom they compare themselves to ([Falk and Knell, 2004](#); [Ghiglini and Goyal, 2010](#); [Immorlica et al., 2017](#)). Among the attributes they weigh when forming those ties is the income of potential peers — precisely the variable whose effect on mental health Equation (1) seeks to identify. Network formation is therefore shaped by the same economic and psychological factors that determine well-being, and any unobservable entering both processes confounds the estimated peer effect. Modelling link formation explicitly is thus necessary to account for the endogeneity induced by self-selection into

peer groups.

We specify that process on its own here, and integrate it with the mental health equation in Section 2.3. The utility an individual derives from a link depends on the characteristics of both parties, and a link is formed only by mutual consent: i and j must each find the connection worthwhile. This bilateral-consent requirement underlies the notion of pairwise stability introduced by Jackson and Wolinsky (1996).

Formally, individual i 's utility from the network configuration is determined by,

$$U_i = \sum_{j \neq i} g_{ij} \cdot u_i(j). \quad (2)$$

That is, total utility is additively separable across bilateral links, and depends on the utility derived from each connection. This marginal utility of individual i forming a link with j is captured by $u_i(j)$. Individuals i and j will agree to be linked (i.e., $g_{ij} = g_{ji} = 1$) if and only if the marginal utility of i connecting with j is positive for both, i.e.

$$g_{ij} = g_{ji} = \mathbb{1}\{u_i(j) > 0\} \cdot \mathbb{1}\{u_j(i) > 0\}.$$

The structure we adopt is standard in the econometric literature on network formation: linking utility is additively separable across dyads, the idiosyncratic linking shocks are logistic, and observable covariates enter through homophily and degree heterogeneity, with unobserved heterogeneity entering as a latent individual type once the network is endogenized in Section 2.3 (Goldsmith-Pinkham and Imbens, 2013; Graham, 2017; Hsieh and Lee, 2016). We add to it the two terms discussed below: the personality distance $d(P_i, P_j)$, which these models typically omit, and the upward and downward income ratios through which social comparison enters the linking decision. The linking utility i receives connecting to j , conditional on income $\mathbf{y}$, exogenous characteristics $\mathbf{X}$, personality traits $\mathbf{P}$, and the baseline network G^0 , is given by

$$\begin{aligned} u_i(j \mid \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = & \delta_0 + \delta_1 \underbrace{d(P_i, P_j)}_{\text{Personality distance}} + \delta_2 y_j + \\ & \delta_3 \underbrace{\max\{\frac{y_j}{y_i}, 1\}}_{\text{Upward comparison}} + \delta_4 \underbrace{\max\{\frac{y_i}{y_j}, 1\}}_{\text{Downward comparison}} + \\ & \delta_5^\top \underbrace{|X_i - X_j|}_{\text{Homophily}} + \delta_6 \underbrace{d_i^0}_{\text{Own popularity}} + \delta_7 \underbrace{d_j^0}_{\text{Other's popularity}} + \underbrace{\epsilon_{ij}}_{\text{Linking shocks}}, \end{aligned} \quad (3)$$

where δ_0 is a constant and linking shocks ϵ_{ij} are assumed to be independently and identically distributed across all dyads, following a standard logistic distribution. The parameter δ_1 measures how an individual forms a link with another based on the Euclidean distance between their respective Big Five personality traits: Openness, Conscientiousness, Ex-

traversion, Agreeableness, and Neuroticism, i.e. $P_i = (O_i, C_i, E_i, A_i, N_i)$. Formally, the Euclidean distance of personality traits is defined by $d(P_i, P_j) = \sqrt{\sum_{k=1}^5 (P_{i,k} - P_{j,k})^2}$. Therefore, the parameter δ_1 captures how differences in personality between individuals influence their propensity to form social connections.¹

Using information on baseline communication in Year 1 of the NetHealth data, denoted as G^0 , we can include d_i^0 as the number of friends for i in this network. A positive δ_6 implies that the more popular an individual is in the initial network, the higher the likelihood of forming new links in the subsequent period. A larger existing network signals sociability and provides greater access to information and resources, which increases the expected benefits of forming a new link.²

We assume that when individuals decide whether to form a link, they consider both the income of the potential peer (i.e., y_j) and the ratio between themselves and the peer (i.e. relative income). We incorporate the concept of upward and downward comparison from [Fehr and Schmidt \(1999\)](#), and use the ratio of between own and others' income instead of the absolute difference to distinguish the effects of income and the relative income. The upward comparison term, $\max\{\frac{y_j}{y_i}, 1\}$, captures the additional utility (or disutility) that individual i derives from connecting with someone wealthier than themselves. Conversely, the downward comparison term, $\max\{\frac{y_i}{y_j}, 1\}$, reflects the utility (or disutility) associated with linking to someone less wealthy.

Given the above setting, defining $\Delta_{ij} = u_i(j) - \epsilon_{ij}$ and letting p_{ij} denote the probability that individual i agrees to connect with j for each $i \neq j$, we have³

$$p_{ij} = \frac{\exp(\Delta_{ij})}{1 + \exp(\Delta_{ij})}.$$

Thus, given the network G^0 , income distribution y , and relevant individual characteristics $\mathbf{X}, \mathbf{P}$, the probability that i and j are connected is,

$$Pr(g_{ij} = 1 \mid \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = p_{ij} \cdot p_{ji}.$$

The probability that agent i forms a link with agent j is not necessarily equal to the probability that agent j forms a link with agent i , i.e., $p_{ij} \neq p_{ji}$. This differs from the model in [Goldsmith-Pinkham and Imbens \(2013\)](#) that assumes perfect symmetry, i.e. $p_{ij} = p_{ji}$. In our specification, the asymmetry arises because we account for individual attributes (e.g., personality traits and popularity), as well as upward and downward comparisons.

Finally, given our assumptions, each potential link between agents i and j can be

¹We also consider alternative specifications where we either include $\sum_{k=1}^5 \delta_{1,k}^{\text{own}} P_{i,k} + \delta_{1,k}^{\text{other}} P_{j,k}$ or $\sum_{k=1}^5 \delta_{1,k} |P_{i,k} - P_{j,k}|$ instead of $\delta_1 d(P_i, P_j)$; i.e., a coefficient per trait or per trait distance. Results of these higher-dimensional alternative specifications are provided in Appendix C.

²This reflects the well-documented preferential attachment or “the rich get richer” phenomenon in social networks ([Jackson, 2008](#)).

³As ϵ_{ij} is assumed to be i.i.d., following a standard logistic distribution F , the probability that individual i agrees to connect with j is $p_{ij} = \Pr(u_i(j) > 0) = \Pr(\epsilon_{ij} > -\Delta_{ij}) = 1 - F(-\Delta_{ij})$.

treated as a Bernoulli random variable with success probability $p_{ij} \cdot p_{ji}$, such that the likelihood function of the network formation model can be specified as:

$$\mathcal{L}_{network}(\boldsymbol{\delta} \mid G, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = \prod_{i < j} [(p_{ij} \cdot p_{ji})^{g_{ij}} (1 - p_{ij} \cdot p_{ji})^{1-g_{ij}}].$$

This distribution is obtained by multiplying the Bernoulli contributions over all unordered pairs of agents. This product yields the likelihood function, representing the joint probability of observing a network configuration given parameters $\boldsymbol{\delta}$.

2.3 Well-being Peer Effects with Endogenous Social Networks

As discussed above, we endogenize the network for two main reasons. First, individuals actively choose their reference groups rather than inheriting them exogenously (Falk and Knell, 2004; Ghiglini and Goyal, 2010; Immorlica et al., 2017). Second, link formation may itself depend on strategic incentives and comparisons: when deciding whom to befriend or communicate with, individuals may take into account the income level and relative income position of potential peers. These features imply that the realized network can be systematically related to unobserved determinants of mental health. Friends may share unobservable traits—such as values, beliefs, self-esteem, lifestyle, or habits—that directly affect mental health; if so, unobservables are correlated across linked individuals, confounding estimates of income peer effects and making it difficult to distinguish true social influence from sorting on unobserved characteristics.

To mitigate this potential bias, following Goldsmith-Pinkham and Imbens (2013), we introduce unobserved individual-specific heterogeneity modelled as a binary latent type, $\xi_i \in \{0, 1\}$, with equal a priori probability $\pi = \Pr(\xi_i = 1)$ assigned to each type.

$$\pi = \Pr(\xi_i = 1) = 1 - \Pr(\xi_i = 0) = \frac{1}{2}. \quad (4)$$

As a stylized example, suppose ξ_i captures an unobserved characteristic such as holding a paid job during term time. Students who work tend to meet and befriend one another, through shared shifts and the limited leisure time their schedules leave them, so there is homophily in ξ_i . Term-time work also bears on mental health directly, through longer hours, less sleep and greater financial strain, so that ξ_i enters the well-being equation. Because students work largely out of financial necessity, ξ_i is in addition correlated with family income; homophilous sorting then implies that the peer-income measure over the reference group y_i^* will be correlated with ξ_i , with working students facing poorer reference groups. In this case, regressions that omit ξ_i may attribute its effect to peer income, biasing estimated peer effects.

This latent type is assumed to influence both the well-being and the network-formation

processes in the following way:

$$MH_i = \beta_0 + \beta_1 y_i + \beta_2 y_i^* + \beta_X^\top \mathbf{X}_i + \beta_\xi \underbrace{\xi_i}_{\text{unobserved heterogeneity}} + \nu_i, \quad (5)$$

$$\begin{aligned} u_i(j \mid \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = & \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\left\{\frac{y_j}{y_i}, 1\right\} + \delta_4 \max\left\{\frac{y_i}{y_j}, 1\right\} \\ & + \delta_5^\top |\mathbf{X}_i - \mathbf{X}_j| + \delta_6 d_i^0 + \delta_7 d_j^0 + \delta_\xi \underbrace{|\xi_i - \xi_j|}_{\text{unobserved homophily}} + \varepsilon_{ij}. \end{aligned} \quad (6)$$

We assume that ξ_i is a priori independent of the observed covariates in Equation (5), though they are correlated a posteriori. Conditional on $(G, \boldsymbol{\xi}, \mathbf{y}, \mathbf{X})$, the well-being shocks are independent across individuals and normally distributed:

$$\boldsymbol{\nu} \mid G, \boldsymbol{\xi}, \mathbf{y}, \mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n). \quad (7)$$

The link-specific shocks ε_{ij} follow a standard logistic distribution and are independent of $(\boldsymbol{\xi}, \boldsymbol{\nu})$; moreover, ε_{ij} are independent across pairs (i, j) . Note that equations (5) and (6) reflect the assumption that, conditional on observables, individuals do not take into account their own mental health when forming the network other than through the unobserved homophily caused by ξ_i .

To see how considering an endogenous model of network formation affects the sign of the estimated peer-income effect, consider the Ordinary Least Squares (OLS) regression of MH_i on y_i^* (and controls): if ξ_i is omitted from the correctly specified well-being equation (5), the probability limit of the peer-income coefficient satisfies the usual omitted-variable bias decomposition:

$$\tilde{\beta}_2 = \beta_2 + \beta_\xi \cdot \frac{\text{Cov}(y_i^*, \xi_i)}{\text{Var}(y_i^*)}.$$

Under the sign pattern of the example above —working students report more depressive symptoms ($\beta_\xi > 0$) but face poorer reference groups ($\text{Cov}(y_i^*, \xi_i) < 0$)— the bias term is negative. Hence $\tilde{\beta}_2 < \beta_2$, meaning the peer-income effect is biased downward in the exogenous model and can appear spuriously negative, suggesting that richer peers protect mental health, even when $\beta_2 \geq 0$. This is precisely the pattern we document in Section 4. Conversely, when the unobserved type moves depressive symptoms and peer income in the same direction, the bias term is positive and the estimated peer-income effect is biased upward.

As in the exogenous model, letting $\Lambda_{ij} = u_i(j) - \varepsilon_{ij}$, the probability that individual i agrees to connect with j , for each $i \neq j$, is defined as

$$q_{ij} = \frac{\exp(\Lambda_{ij})}{1 + \exp(\Lambda_{ij})}.$$

Conditional on $(G, \boldsymbol{\xi}, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0)$, the likelihood function for the network process is

$$\mathcal{L}_{\text{network}}(\boldsymbol{\delta} \mid G; \boldsymbol{\xi}, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}],$$

where the product runs over unordered pairs, since G is undirected and the mutual-consent probability $q_{ij}q_{ji}$ is symmetric. Conditional on observed and first-stage quantities, the likelihood function for the mental health equation is

$$\mathcal{L}_{MH}(\boldsymbol{\beta}, \sigma^2 \mid MH; G, \boldsymbol{\xi}, \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right).$$

where μ_i denotes the systematic component of Equation (5), so that $MH_i = \mu_i + \nu_i$. Finally, the conditional likelihood function for the joint structural model (where $\boldsymbol{\xi}$ is added to the parameters in order to update its prior) is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \boldsymbol{\xi} \mid MH, G; \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = \\ \mathcal{L}_{\text{network}}(\boldsymbol{\delta} \mid G; \boldsymbol{\xi}, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \times \mathcal{L}_{MH}(\boldsymbol{\beta}, \sigma^2 \mid MH; G, \boldsymbol{\xi}, \mathbf{y}, \mathbf{X}) \times \prod_{i=1}^n \pi^{\xi_i} (1 - \pi)^{1-\xi_i}. \end{aligned}$$

Given the independent Bernoulli priors for ξ_i , once we set the prior for $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and σ^2 , we can use a mix of Gibbs and Metropolis-Hastings steps to sample them sequentially from their posterior distribution (see Appendix A for details).

3 Data

We estimated the models described in the previous sections using the data from NetHealth (Purta et al., 2016), a longitudinal study that investigates social networks and health outcomes among undergraduate students at the University of Notre Dame (focusing on the 2015 to 2019 cohort). Initially, 700 students were recruited for a two-year study, with 300 participants continuing for an additional two years. The project collected extensive data on participants' social networks, physical activity, mental health, demographics and other relevant factors.

The basic survey data in NetHealth includes individual characteristics such as gender, family income, academic grade, and personality traits. It also collected various mental health measures including depression scores. Participants are fairly similar as they are all first-year undergraduate students at the same university.

We utilize four waves of survey data from NetHealth, collected in summer 2015 (wave 1), winter 2016 (wave 2), summer 2016 (wave 3), and fall 2016 (wave 4). Since the original dataset does not collect most key variables in every wave, we pooled the data into two academic years: the first running from 2015-08-03 to 2016-06-19 and the second from 2016-06-20 to 2017-04-09. The first academic year includes data from waves 1 and 2, the

second from waves 3 and 4 (see Tables B.8 and B.9 of the Appendix). All references to the first and second academic years below follow this definition.

3.1 Mental Health

Our two main outcomes for measuring mental health are depression scores. The first is the Center for Epidemiologic Studies Depression scale (CES-D; Radloff, 1977), a widely used 20-item self-report questionnaire designed to measure the frequency and severity of depressive symptoms, particularly in the general population. The second is the Beck Depression Inventory (BDI; Beck et al., 1961), another widely used 21-item self-report questionnaire that measures the severity of depression in individuals over the age of 13. It rates symptoms like sadness and pessimism over the past two weeks. Both are severity scales: higher scores indicate more frequent or more severe depressive symptoms, and therefore worse mental health. They consequently run in the opposite direction to MH_i as specified in Section 2. All estimates below are reported on the depression scale, so that a positive coefficient signals a deterioration in mental health and a negative coefficient an improvement. The items and scaling defining these variables are in Tables B.4 and B.5 in Appendix B, respectively. The descriptive statistics of the main variables are in Table B.9 of the Appendix.

3.2 Network Structure

All participants voluntarily signed up to take part in the NetHealth survey. As part of their enrollment, NetHealth obtained access to full communication data (e.g., call and message logs), collected from participants' smartphones.⁴

There are over 60 million communication events in the raw data. As suggested by NetHealth, we filter out the following records: (1) calls of zero duration, (2) messages of zero length or bytes, (3) WhatsApp Group chats. Excluding these records, as well as those belonging to non-participants or participants who could not be identified, approximately 3.28 million events were observed in the first academic year and 2.45 million in the second academic year, yielding an approximate total of 5.73 million communication records across the two-year period. After merging with the available survey data and removing single-direction communication records, the final dataset contains approximately 2.19 million records across the two years (1.14 million in Year 1 and 1.05 million in Year 2).

For each message, we have the following information: (1) the date and time of message, (2) identifiers for both sender and receiver, (3) message direction and (4) message types. We use this data (from the first two years of study) to map individual social networks from four sources: phone call (1.05%), MMS (Multimedia Messaging Service, 3.33%), SMS

⁴For Android users, an app transmitted daily logs to NetHealth's servers. iPhone users instead provided periodic backups, which were processed by a computer app to extract and upload communication events. Over 95% of events in the dataset came from iPhone users.

(Short Message Service, 93.86%), and WhatsApp (1.76%). Note that while the receiver can potentially refer to a non-participant, we use the communication records only between participants to construct the social networks as we can then observe survey data for both, which would be unavailable for non-participants.

Figure 1 presents the distribution of communication record counts over time during the first two academic years. On average, individuals engaged in 8 communication events per day across the two academic years. These communication records enable us to objectively identify social interactions between participants, as they capture daily patterns of communication and contact.

From these raw communication records, we construct social network ties by mapping actual exchanges between participants: a link is established between two individuals whenever the volume of two-way communication between them exceeds a threshold, the precise choice of which we set out below.

Inferring social ties from observed interaction rather than from surveys follows [Kossinets and Watts \(2006\)](#) and [Marmaros and Sacerdote \(2006\)](#). What is distinctive here is the coverage of the record across pairs and over time: within the smartphone call and message channels available to us, every exchange between every pair of participants is recorded continuously across two academic years, so reference groups are neither capped at a fixed number of nominations nor dependent on recall. The record is not exhaustive across modes, since it does not capture email or messaging through social media platforms, and ties sustained primarily through those channels will be understated. A key feature of this dataset is that while the researchers only see anonymised identifiers, the participants themselves were communicating with real acquaintances whom they know in daily life (friends, peers, colleagues, etc.). Thus, the networks we analyze reflect genuine social interactions, not artificially imposed or survey-reported links.

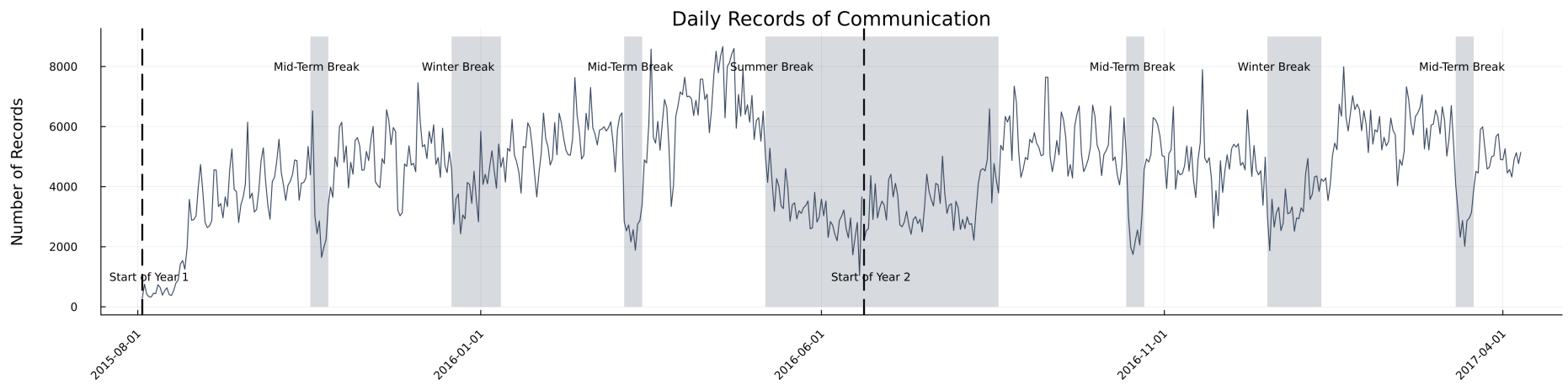


Figure 1: Distribution of Communication Records in NetHealth

Consistent with the model, as we define links from two-way communication, the adjacency matrix for communication networks, denoted as G , is symmetric. Each element g_{ij} represents whether there is at least a certain amount of recorded communication between participants i and j . We distinguish between weak and strong ties, based on the distribution of the number of exchanged messages in the communication records. Specifically, we define a strong tie as one where the number of exchanged messages falls above the median during the second academic year. Conversely, a weak tie is defined as at least one two-way communication. As a robustness check, we also consider the first quartile (Q1) and the third quartile (Q3) as alternative thresholds for defining communication networks. Finally, in a methodological contribution and extension, we also consider the communication intensity network M , whose elements $m_{ij} = m_{ji}$ directly represent the number of messages exchanged between i and j (formally defined in Section 6).

In our empirical analysis, we remove isolated individuals (i.e. individuals who do not have at least two-way communication with any other participants) from networks defined above. Specifically, there are 2 isolated individuals in the network with weak ties, 3 in Q1, 16 in the network with strong ties, and 66 in Q3. Figure 2 presents the distribution of the natural logarithm of communication records between participants in our dataset. On average, around 565 messages were exchanged between pairs of participants who had any contact in Year 2. The minimum number of messages exchanged is 2 by definition, the maximum number of messages exchanged is 69,863, which is similar in scale to existing studies on individual communication (e.g., Battestini et al., 2010).

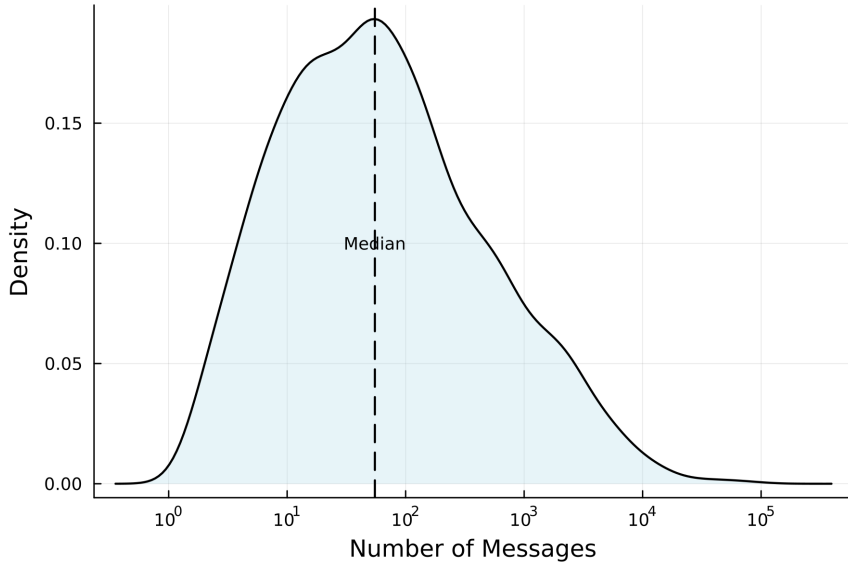


Figure 2: Distribution of Communication Records Between Participants

Note: This figure shows the distribution of communication records between participants at Year 2 in the NetHealth study. The sample includes only pairs with at least one two-way communication record. The x-axis shows the number of communication records on a logarithmic scale, and the y-axis represents the density estimated via kernel density estimation. First quartile is 14, median is 55, and third quartile is 228.

The number of social ties both weak and strong (denoted as edges following the lit-

erature) and the average number of friends (denoted as alters) in the network are shown in Tables B.1 and B.2 of the Appendix. The network graph along with the mental health distribution is presented in Figure B.1 of the Appendix.

3.3 Income, Demographics and Other Characteristics

The factor y in our analysis is measured by the parents' income before the first academic year. In NetHealth, parental income is elicited through the question: "What is the best estimate of your parents' total income last year? Consider income from all sources before taxes." Responses are recorded on an eight-point ordinal scale, ranging from 1 (less than 25,000 USD) to 8 (250,000 USD or more) (see Table B.3 and Figure B.2 of the Appendix).

For the analysis, these ordinal categories are mapped to the midpoint of each interval, also to facilitate the computation of average income differences across peers. Specifically, level 1 is assigned a value of 12,500 and level 2 a value of 37,500; because the intervals widen in the upper half of the scale, the assigned values are not equally spaced, and the full mapping is reported in Table B.3. For the open-ended top category (level 8), we estimate the midpoint by extending the interval width of the preceding category (level 7), resulting in an assigned value of 275,000. In the estimation, income is expressed in units of ten thousand US dollars, so the reported coefficients measure the effect of a 10,000 USD change (see Table B.9).

Figure 3 shows the percentage distribution of income levels among peers. The heatmap reveals strong assortative mixing at the top of the income distribution: the largest share of links involving the wealthiest individuals is concentrated in the $\{8, 8\}$ cell, indicating that individuals in income category 8 are especially likely to connect with others with similar income level. At the same time, mixing patterns are more diffuse in the middle of the distribution, with middle-income individuals forming ties across a wider range of income levels. In addition to income, we control for gender (1 if male and 0 if female) in the mental health equation. The homophily term of the linking utility uses the absolute difference between i and j in gender and an indicator equal to one when the two individuals differ in degree aspiration⁵, the latter elicited in wave 3 (see Table B.8).

⁵In NetHealth, participants are asked the question: "What is the highest academic degree that you intend to obtain?". The response categories, reported in Table B.9, are not ordered, so the homophily term enters as a same-versus-different indicator rather than as a distance on a scale.

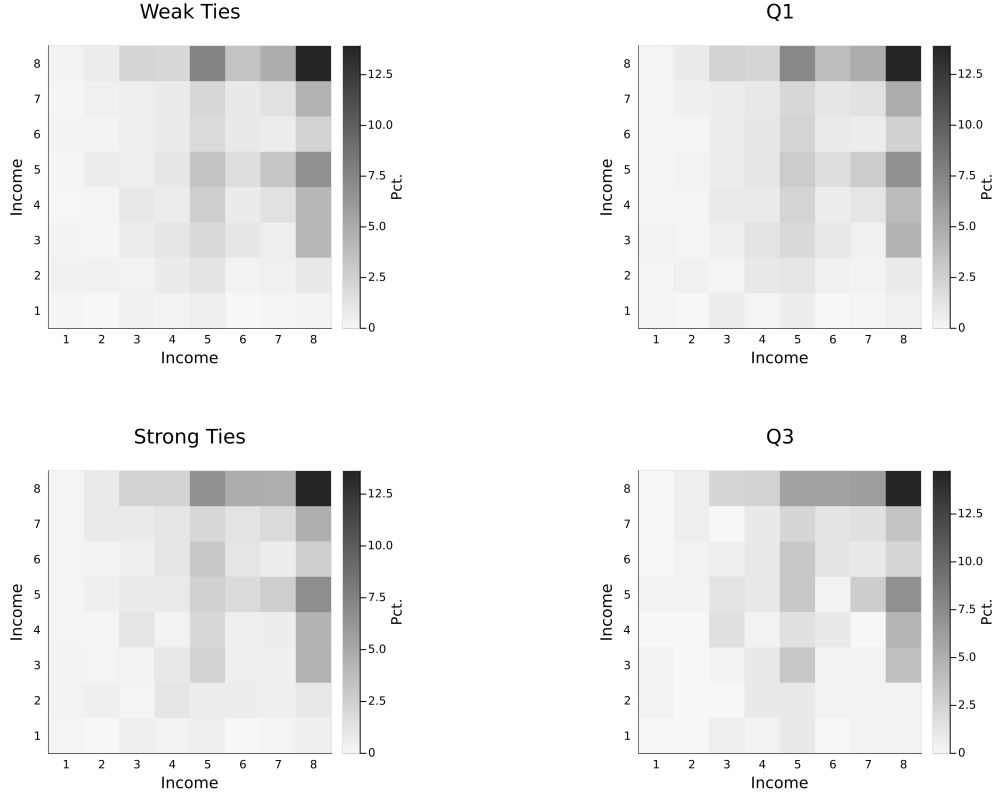


Figure 3: Distribution of Income Levels among Peers

Note: The heatmaps display the joint distribution of income levels between individuals (x-axis) and their peers (y-axis). The color gradient represents the density of peer connections with such income level combinations, with darker shades indicating higher densities. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (the median threshold). Q1 and Q3 represent the first quartile (14 messages) and third quartile (228 messages) of the message exchange distribution, respectively.

Personality traits are measured using the Big Five Inventory (BFI; [John et al., 1991](#)), which assesses the five dimensions of personality: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The inventory comprises 44 items, each rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).⁶ In the main specifications, we use the Euclidean distance of personality traits, exploring alternative personality specifications in Appendix C.⁷

⁶The full list of survey questions for personality is available in Table B.6.

⁷NetHealth measures personality in four waves: W1 (Summer 2015), W2 (Winter 2016), W6 (Fall 2017) and W8 (Spring 2019). Because personality is stable over time ([Borghans et al., 2008](#)) and the study spans only four years, we treat traits as time-invariant and measure them as the average of the W1 and W2 responses, retaining the available wave when one of the two is missing (see Table B.8).

4 Results

We first present the result from the estimation of the model of mental health (1) given the exogenous model of network formation (3), which allows us to determine individuals' peers. We then present the result of the model of mental health with endogenous network formation (5).

4.1 Peer's Income and Well-Being with Exogenous Networks

Tables 1 and 2 report estimated effects on mental health under the exogenous network specification, across networks defined by varying communication intensity thresholds. Two results are worth noting. First, own income is negatively associated with depression, indicating that higher personal income is associated with better mental health. This effect is consistently significant when depression is measured by the CESD scale, though somewhat less precisely estimated for BDI. This is consistent with a large literature documenting a positive relationship between absolute income and psychological wellbeing (e.g. [Ferrer-i Carbonell, 2005](#); [Blanchflower and Oswald, 2004](#)), and provides a useful benchmark against which to interpret the peer income estimates.

Second, and more substantively for our purposes, average peer income is negatively and significantly associated with depression – that is, higher peer income predicts better mental health – across all network definitions and both depression scales. The most natural interpretation in the exogenous setting is that peer income proxies for the broader socioeconomic environment in which an individual is embedded. Individuals who communicate primarily with higher-income peers are likely to inhabit wealthier social milieus, with greater access to material resources, social capital, and emotional support. In this setting, peer income captures positive spillovers from the social environment rather than the effect of income-driven comparison processes, since network composition is held fixed and treated as independent of individual characteristics. As we show in Section 4.3, this interpretation changes fundamentally once network formation is modelled endogenously.⁸

⁸Female respondents report systematically higher depression levels across both scales and all network specifications, consistent with well-documented gender differences in the prevalence and self-reporting of depression (e.g. [Nolen-Hoeksema, 2001](#)).

Table 1: Exogenous Network: CESD and Peer's Income

Dependent: CESD	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	20.9526*** (2.2454)	21.4783*** (2.1041)	21.2999*** (2.0426)	21.1945*** (2.1784)
income	-0.1690*** (0.0641)	-0.1671*** (0.0634)	-0.1926*** (0.0639)	-0.1780** (0.0741)
average peers' income	-0.1961 (0.1193)	-0.2284** (0.1068)	-0.1895* (0.0971)	-0.2020** (0.0963)
gender	-2.3175** (1.1033)	-2.3149** (1.1036)	-2.4197** (1.1238)	-2.6270** (1.2679)
Number of connections	1221	906	602	305
Average connections	8	6	4	2
Observations	297	296	283	233

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: OLS regression results examining the relationship between CESD and peers' income. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 2: Exogenous Network: BDI and Peer's Income

Dependent: BDI	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	11.8898*** (1.5602)	12.0115*** (1.4634)	11.3698*** (1.4229)	11.4345*** (1.4101)
income	-0.0626 (0.0446)	-0.0640 (0.0441)	-0.0736* (0.0445)	-0.0176 (0.0480)
average peers' income	-0.1652** (0.0829)	-0.1711** (0.0742)	-0.1284* (0.0676)	-0.1889*** (0.0623)
gender	-1.8840** (0.7666)	-1.8662** (0.7675)	-1.8876** (0.7829)	-1.9263** (0.8207)
Number of connections	1221	906	602	305
Average connections	8	6	4	2
Observations	297	296	283	233

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents OLS regression results examining the relationship between BDI and peers' income. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

4.2 Exogenous Network Formation

Table 3 shows the estimates of the exogenous network formation models. Others' income (y_j) is not a robust predictor of the connection; on the other hand, the ratios between income of the matching individuals are significant predictors of a connection in the strong-tie network. In particular an individual is more likely to have strong ties the lower is the upward ratio and the higher is the downward ratio of their income with their peer's in-

come. In other words, individuals tend to connect with richer peers when the income ratio between them is not too large (i.e., when peers are only slightly richer than themselves). In contrast, when peers are poorer, individuals are more likely to connect with those whose incomes differ substantially rather than with those whose incomes are very similar.⁹ The effect of the other determinants in Table 3 is quite intuitive: The Euclidean distance in the Big Five personality traits affects network formation, greater similarity increases the likelihood of forming connections.¹⁰ Likewise, individuals are more likely to form connections with others of the same gender and with those who share similar degree aspirations. In the strong-tie network, individuals with a larger number of friends in the previous period are more likely to establish new connections (own popularity), whereas people are less likely to form connections with others who had a large number of friends in the previous period; neither effect is statistically significant at the remaining thresholds. In other words, popularity increases one’s own ability to form links (or provides greater incentives to form connections regardless of the characteristics of others), but reduces one’s attractiveness as a potential connection. This asymmetric effect of degree is consistent with models of network formation with heterogeneous linking capacity and congestion: while more connected individuals may face lower marginal costs of forming additional links, highly connected partners may offer lower marginal benefits due to limited attention or increased competition for their connections.

⁹We also tested upward and downward comparisons defined by income difference (i.e. upward: $\max\{y_j - y_i, 0\}$, downward: $\max\{y_i - y_j, 0\}$) in Appendix B. Similarly, others’ income doesn’t affect network formation, but the income difference does. The effects of upward comparison on network formation remain negative across most networks (weak ties, Q1, and Q3). (see column (1) and (2) in Table C.1, C.2, C.3). The effects of downward comparison are positive (as with the ratio measure) when both upward and downward comparisons are included. (see column (1) in Table C.1, C.2, C.3). However, when only downward comparison is included, the effects are negative (see column (3) in Table C.1, C.2, C.3).

¹⁰Alternative model specification with personality similarity (PS) is also tested. Similarly, results from the PS model with strong ties support this pattern (see Table C.4). Besides, similarity in extraversion levels also influences network formation in network with stronger ties (Q3): the more similar individuals are in extraversion, the more likely they are to form a connection.

Table 3: Exogenous Network Formation

Linking utility: $u_i(j)$	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-2.1821*** (0.1173)	-2.4169*** (0.1335)	-2.5082*** (0.1618)	-2.4742*** (0.2238)
$d(P_i, P_j)$	-0.0419** (0.0190)	-0.0368** (0.0202)	-0.0468* (0.0256)	-0.0296 (0.0354)
y_j	0.0052 (0.0033)	0.0040 (0.0038)	0.0073* (0.0045)	0.0033 (0.0066)
upward comparison	-0.0347 (0.1040)	-0.0368 (0.1247)	-0.1486*** (0.0291)	-0.1415 (0.1858)
downward comparison	0.0159 (0.1036)	0.0416 (0.1292)	0.1990*** (0.0616)	0.1927 (0.2015)
gender difference	-0.3906*** (0.0381)	-0.4240*** (0.0438)	-0.4671*** (0.0528)	-0.5149*** (0.0729)
degree aspiration difference	-0.1326*** (0.0401)	-0.1150*** (0.0474)	-0.1442** (0.0555)	-0.1411* (0.0750)
own popularity	0.0885 (0.1008)	0.1184 (0.1440)	0.2815*** (0.0279)	0.3575 (0.2820)
peer's popularity	0.0165 (0.1007)	0.0173 (0.1437)	-0.1151*** (0.0243)	-0.1313 (0.2820)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

*HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.*

Note: This table presents the posterior means and standard errors of the parameters in the exogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

4.3 Joint Peer Effects and Endogenous Reference Groups

The central result of the paper is in Panel A of Tables 4 and 5 (and Tables C.5 and C.6 in the appendix). Conditional on the endogenously determined reference group, the coefficient on average peers' income is positive for both depression scales and at every intensity threshold: individuals whose chosen peers are richer report more severe depressive symptoms. In the exogenous specification of Tables 1 and 2 the same coefficient was negative. Modelling how reference groups are formed does not merely sharpen the estimate; it reverses its sign.

The magnitudes are economically meaningful. A \$10,000 increase in peers' average income is associated with an increase in the BDI score of between 0.13 and 0.20 points, and an increase in the CESD score of between 0.33 and 0.56. The BDI estimate is not statistically significant at the Q3 threshold, where ties are strongest, whereas the CESD estimate is significant at every threshold.

The effect also weakens monotonically as ties become stronger, on both scales. A plausible reading is that income comparisons lose salience as relationships become closer: emotional closeness supplies support that buffers the individual against the comparison, so the same income gap costs less among close friends than among acquaintances.

The reversal follows from what the endogenous model conditions on. Individuals sort strongly on the latent type: the coefficient on $|\xi_i - \xi_j|$ in the linking utility ranges from -0.87 to -1.09 across the weak-tie, Q1 and strong-tie networks and is significant at the 99% level in each, for both outcomes, so that individuals sharing the same unobserved type are considerably more likely to be linked (the estimate is small and insignificant at the Q3 threshold). The type in turn enters the mental health equation with a positive sign throughout, though it is precisely estimated only in the strong-tie network. Unobserved homophily of this kind is exactly the channel identified in Section 2.3: because individuals sort on a characteristic that also bears on depression, a specification that omits it attributes part of that sorting to peer income. The exogenous estimate is contaminated by selection, while the endogenous estimate, purged of it, measures comparison.

The two specifications therefore support different readings of the same variable. When networks are taken as given, peer income captures average spillover effects that operate independently of how peers are chosen. Variation in average peer income reflects income differences within a fixed set of connections, holding network composition constant, so that higher peer income primarily captures the benefits of belonging to a wealthier social environment – improved access to material resources, information, and emotional support – which is associated with lower depression levels.

In contrast, when network formation is modelled endogenously, exposure to peer income is no longer treated as random but as the outcome of individuals' linking incentives. Two features of the linking estimates explain why this changes the sign. First, income position is salient in the choice of peers: in the strong-tie network, individuals are markedly

less likely to connect with peers far above them in the income distribution (Table 3), so relative income enters the linking decision and not only the mental health equation. Second, this avoidance is necessarily incomplete. Links are also formed on gender, degree aspiration, personality and popularity, and the pool of available partners is fixed by the cohort, so individuals end up in reference groups richer than they would choose on income grounds alone. It is among those left with richer reference groups that depressive symptoms are more severe. The specification is linear in y_i^* , so a poorer reference group maps mechanically into fewer symptoms; whether what matters is the direction of the income gap or its size is the question we take up in Section 5.

The remaining determinants of link formation, reported in Panel B of the same tables, are largely consistent with those of the exogenous model in Table 3. Estimating the network and mental health processes jointly therefore does not disturb the picture of how ties are formed; what it changes is the interpretation of the reference group those ties define.

Table 6 sets out the related distinction in what a change in y_i^* represents under each model: income growth within a fixed peer group when the network is held fixed, and income growth together with a change in peer composition once the network is allowed to respond. These are the two margins that the counterfactual exercises of Section 7 separate, and there we show that they work in opposite directions. The broader lesson is that failing to account for endogenous network formation leads not merely to attenuated estimates, but to qualitatively misleading conclusions about the direction of peer effects on mental health.

Table 4: Endogenous Network: Effects of Peer's Income on BDI

(A) Mental Health and Peer's Income				
Dependent:BDI	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.7255*** (0.9383)	3.0504*** (0.9113)	3.1060*** (0.8912)	3.3055*** (0.8787)
income	0.0054 (0.0461)	0.0172 (0.0425)	0.0266 (0.0438)	0.1079** (0.0477)
average peers' income	0.2031*** (0.0659)	0.1477** (0.0591)	0.1294** (0.0564)	0.0267 (0.0554)
gender	-0.6399 (0.6155)	-0.5336 (0.6172)	-0.4954 (0.6280)	-0.4091 (0.6514)
ξ_i	1.0153 (1.9722)	1.6831 (1.2524)	1.6648* (0.9644)	1.5659 (0.9022)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.9572*** (0.1549)	-2.1753*** (0.1748)	-2.2217*** (0.1747)	-2.4922*** (0.2636)
$d(P_i, P_j)$	-0.0349 (0.0206)	-0.0280 (0.0238)	-0.0420 (0.0269)	-0.0313 (0.0428)
y_j	0.0060* (0.0033)	0.0049 (0.0045)	0.0078* (0.0048)	0.0045 (0.0072)
upward comparison	0.0172 (0.1015)	-0.1073** (0.0356)	-0.1428*** (0.0273)	-0.2295*** (0.0604)
downward comparison	-0.0284 (0.1023)	0.1150** (0.0598)	0.2015*** (0.0647)	0.2910** (0.1097)
gender difference	-0.3982*** (0.0387)	-0.4302*** (0.0480)	-0.4847*** (0.0574)	-0.5210*** (0.0971)
degree aspiration difference	-0.1278*** (0.0441)	-0.1122** (0.0531)	-0.1507** (0.0586)	-0.1478 (0.0973)
own popularity	0.0200 (0.0989)	0.2107*** (0.0173)	0.2696*** (0.0315)	0.4760*** (0.0828)
peer's popularity	0.0881 (0.0978)	-0.0726*** (0.0144)	-0.0990*** (0.0274)	-0.2510*** (0.0746)
$ \xi_i - \xi_j $	-0.8709*** (0.0537)	-0.9670*** (0.0684)	-1.0949*** (0.1014)	-0.0206 (0.0975)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 5: Endogenous Network: Effects of Peer's Income on CESD

(A) Mental Health and Peer's Income				
Dependent:CESD	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.7558*** (0.9587)	3.0715*** (0.9430)	3.2559*** (0.9449)	2.9516*** (0.9564)
income	-0.0297 (0.0689)	0.0113 (0.0653)	0.0379 (0.0641)	0.1227* (0.0760)
average peers' income	0.5605*** (0.0895)	0.4678*** (0.0811)	0.4274*** (0.0778)	0.3296*** (0.0848)
gender	-0.1632 (0.7423)	-0.0312 (0.7625)	0.0197 (0.7706)	0.0793 (0.8298)
ξ_i	1.1515 (2.0535)	1.8760 (1.5399)	1.6832** (0.8330)	1.5069 (0.9905)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.9874*** (0.1371)	-2.1332*** (0.1410)	-2.2227*** (0.1672)	-2.4257*** (0.2473)
$d(P_i, P_j)$	-0.0297* (0.0192)	-0.0306 (0.0219)	-0.0443 (0.0280)	-0.0339 (0.0394)
y_j	0.0059* (0.0035)	0.0044 (0.0040)	0.0079 (0.0047)	0.0035 (0.0069)
upward comparison	-0.0263 (0.1011)	-0.1096*** (0.0306)	-0.1467*** (0.0261)	-0.1983** (0.1134)
downward comparison	0.0199 (0.1004)	0.1169** (0.0566)	0.2075*** (0.0596)	0.2455* (0.1434)
gender difference	-0.4024*** (0.0363)	-0.4416*** (0.0442)	-0.4796*** (0.0564)	-0.5043*** (0.0842)
degree aspiration difference	-0.1230*** (0.0431)	-0.1155** (0.0524)	-0.1542*** (0.0575)	-0.1554* (0.0917)
own popularity	0.0869 (0.0990)	0.2068*** (0.0177)	0.2693*** (0.0287)	0.4420** (0.1682)
peer's popularity	0.0221 (0.0997)	-0.0708*** (0.0149)	-0.0985*** (0.0256)	-0.2215** (0.1635)
$ \xi_i - \xi_j $	-0.8788*** (0.0488)	-0.9740*** (0.0685)	-1.0845*** (0.0988)	-0.0203 (0.0893)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 6: Effects of Changes in Peer’s Income on Depression

$\sum_{j \in \mathcal{N}} g_{ij}^* y_j$	Model	Source	Mechanism	Effect on depression
Increase	Exogenous	y_j	$\uparrow y_j$ within existing network G	Decrease
Increase	Endogenous	y_j and G	$\uparrow y_j$ + More slightly richer peers or fewer slightly poorer peers	Increase
Decrease	Exogenous	y_j	$\downarrow y_j$ within existing network G	Increase
Decrease	Endogenous	y_j and G	$\downarrow y_j$ + More much poorer peers or less much richer peers	Decrease

5 Social Comparison and Well-Being

The previous section established that average peer income has a significant effect on mental health once network formation is modelled endogenously. We now probe the mechanism further by decomposing this effect into its upward and downward components. Specifically, we ask whether the mental health cost of peer income exposure is symmetric – that is, whether being poorer than one’s peers is as harmful as being richer than them – or whether one direction dominates.

Following the network formation specification, upward relative income is defined as the ratio of peers’ income to own income when peers are richer on average, and downward relative income as the ratio of own income to peers’ income when peers are poorer. We replace average peer income in the mental health equation with these two terms, and estimate it jointly with the linking utility (6), in which the same comparison terms enter at the pair level:

$$MH_i = \beta_0 + \beta_1 y_i + \beta_2 \underbrace{\max\left\{\frac{y_i^*}{y_i}, 1\right\}}_{\text{upward relative income}} + \beta_3 \underbrace{\max\left\{\frac{y_i}{y_i^*}, 1\right\}}_{\text{downward relative income}} + \beta_X^\top \mathbf{X}_i + \beta_\xi \xi_i + \nu_i \quad (8)$$

When networks are taken as exogenously given, the two comparison terms are almost never significant: of the sixteen estimates in Tables 7 and 8, only the downward term for BDI at the Q3 threshold differs from zero. Once again, the exogenous model largely fails to detect effects that emerge clearly once peer selection is accounted for.

Tables 9 and 10 report results from the endogenous model. Both upward and downward relative income have significant positive effects on depression — that is, both increase CESD and BDI scores — indicating that any income gap between an individual and their peers, regardless of its direction, is harmful to mental health.¹¹

This pattern is consistent with inequality aversion rather than pure relative depriva-

¹¹Own income enters positively and significantly in the CESD specification, and is insignificant for BDI. Because y_i appears in both comparison terms, β_1 no longer isolates the effect of absolute income once the two ratios are included, and we do not interpret it as such. The same caveat applies to the exogenous columns, where own income enters negatively, so the change of sign across the two specifications is not informative either.

tion. Under pure relative deprivation, only upward comparisons – being worse off than one’s peers – would matter; the fact that downward gaps also raise depression suggests that individuals are uncomfortable with economic distance from their reference group in either direction. The relative size of the two effects, however, is not stable across instruments. For CESD the two coefficients are close and their ordering varies across thresholds, whereas for BDI the downward coefficient is roughly three times the upward one at every threshold. We therefore interpret only the common sign, which is estimated consistently, and not the relative magnitudes. This is worth stating plainly, because the inequality-aversion formulation of [Fehr and Schmidt \(1999\)](#) weighs disadvantageous inequality more heavily than advantageous inequality, an ordering the BDI estimates invert; our estimates are not precise enough to adjudicate between the two. Read alongside the network estimates, the results also reveal a wedge between choice and welfare. Individuals form ties with peers substantially poorer than themselves — the downward comparison term enters the linking utility positively and significantly (Table 3) — yet the mental health equation attaches a cost to economic distance in that direction as well. This is consistent with the assumption maintained in Section 2.3, that individuals do not weigh their own mental health when choosing whom to connect with, beyond the unobserved homophily captured by ξ_i . What carries the mental health cost is distance from the reference group rather than its sign, and that distance is not what individuals appear to be optimising when they choose their peers.

Table 7: Exogenous Network: BDI and Income Comparison

Dependent: BDI	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	10.3019*** (1.5671)	9.9479*** (1.4177)	9.5915*** (1.3076)	8.8213*** (1.5625)
income	-0.1231** (0.0595)	-0.1405** (0.0572)	-0.1303** (0.0543)	-0.1144* (0.0588)
$\max\{y_i^*/y_i, 1\}$	-0.2718 (0.2353)	-0.2871 (0.2241)	-0.1934 (0.2070)	-0.3816 (0.2992)
$\max\{y_i/y_i^*, 1\}$	0.1802 (1.0950)	0.7116 (0.8208)	0.6202 (0.5095)	1.1658** (0.4711)
gender	-1.9287** (0.7719)	-1.9274** (0.7738)	-1.8705** (0.7870)	-2.0071** (0.8245)
Number of connections	1221	906	602	305
Average connections	8	6	4	2
Observations	297	296	283	233

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the results of OLS regressions examining the relationship between BDI (depression) and income comparison factors with peers. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 8: Exogenous Network: CESD and Income Comparison

Dependent: CESD	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	14.9103*** (2.2439)	15.3827*** (2.0344)	16.4267*** (1.8759)	14.4523*** (2.4081)
income	-0.1555* (0.0852)	-0.1771** (0.0821)	-0.2112*** (0.0780)	-0.1521* (0.0906)
$\max\{y_i^*/y_i, 1\}$	0.5184 (0.3370)	0.3939 (0.3216)	0.2508 (0.2970)	0.7398 (0.4612)
$\max\{y_i/y_i^*, 1\}$	1.1621 (1.5679)	1.2470 (1.1778)	1.0580 (0.7310)	1.0239 (0.7260)
gender	-2.3751** (1.1053)	-2.4552** (1.1105)	-2.4925** (1.1290)	-2.6293** (1.2707)
Number of connections	1221	906	602	305
Average connections	8	6	4	2
Observations	297	296	283	233

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the results of OLS regressions examining the relationship between CESD (depression) and income comparison factors with peers. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 9: Endogenous Network: Effects of Income Comparison on BDI

(A) Mental Health and Income Comparison				
Dependent:BDI	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.8781*** (0.8572)	3.0695*** (0.8783)	3.2313*** (0.8126)	2.3346*** (0.8667)
income	0.0568 (0.0513)	0.0555 (0.0479)	0.0492 (0.0452)	0.0686 (0.0453)
$\max\{y_i^*/y_i, 1\}$	0.3921** (0.2021)	0.3923* (0.2013)	0.3808** (0.1866)	0.4079* (0.2444)
$\max\{y_i/y_i^*, 1\}$	1.1455* (0.7503)	1.1554* (0.6307)	1.0487** (0.4488)	1.3879*** (0.4220)
gender	-0.6176 (0.6165)	-0.6043 (0.6201)	-0.5244 (0.6038)	-0.6485 (0.6510)
ξ_i	1.7205* (0.8603)	1.2525 (1.4424)	0.9903 (0.7084)	1.1651 (0.9122)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.9321*** (0.1057)	-2.1443*** (0.1561)	-2.2329*** (0.1748)	-2.4843*** (0.3125)
$d(P_i, P_j)$	-0.0380** (0.0186)	-0.0284 (0.0225)	-0.0390 (0.0269)	-0.0325 (0.0456)
y_j	0.0060* (0.0032)	0.0045 (0.0040)	0.0077 (0.0048)	0.0050 (0.0081)
$\max\{y_j/y_i, 1\}$	-0.0941** (0.0371)	0.0486 (0.1248)	-0.1489*** (0.0293)	-0.1512 (0.1734)
$\max\{y_i/y_j, 1\}$	0.0786 (0.0564)	-0.0384 (0.1243)	0.2107*** (0.0647)	0.2002 (0.2004)
gender difference	-0.4028*** (0.0386)	-0.4400*** (0.0445)	-0.4764*** (0.0570)	-0.5012*** (0.0922)
degree aspiration difference	-0.1337*** (0.0403)	-0.1216** (0.0493)	-0.1546** (0.0599)	-0.1467* (0.0930)
own popularity	0.1579*** (0.0122)	0.0213 (0.1309)	0.2746*** (0.0292)	0.3642 (0.2413)
peer's popularity	-0.0499*** (0.0098)	0.1150 (0.1313)	-0.1045*** (0.0249)	-0.1427 (0.2439)
$ \xi_i - \xi_j $	-0.8676*** (0.0574)	-0.9783*** (0.0611)	-1.0666*** (0.0958)	-0.0185 (0.0993)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table 10: Endogenous Network: Effects of Income Comparison on CESD

(A) Mental Health and Income Comparison				
Dependent:CESD	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.9064*** (0.8914)	3.0793*** (0.9129)	3.3942*** (0.8925)	2.1954** (0.9487)
income	0.2149*** (0.0645)	0.2120*** (0.0633)	0.1996*** (0.0607)	0.2282*** (0.0652)
$\max\{y_i^*/y_i, 1\}$	1.7619*** (0.2699)	1.6552*** (0.2640)	1.4749*** (0.2517)	2.1261*** (0.3516)
$\max\{y_i/y_i^*, 1\}$	1.8506** (0.8190)	1.7499** (0.7616)	1.6841*** (0.6060)	1.4296** (0.6015)
gender	-0.1112 (0.7207)	-0.1486 (0.7481)	-0.0292 (0.7475)	-0.1975 (0.8101)
ξ_i	1.2931 (1.3294)	1.3388 (1.1776)	1.1003 (0.8335)	1.1172 (1.0058)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.8903*** (0.1161)	-2.1501*** (0.1526)	-2.2259*** (0.1740)	-2.4619*** (0.2814)
$d(P_i, P_j)$	-0.0359* (0.0187)	-0.0298 (0.0226)	-0.0396 (0.0272)	-0.0309 (0.0397)
y_j	0.0042 (0.0034)	0.0044 (0.0039)	0.0078 (0.0050)	0.0044 (0.0077)
$\max\{y_j/y_i, 1\}$	-0.0968*** (0.0278)	0.0506 (0.1292)	-0.1487*** (0.0276)	-0.2240*** (0.0574)
$\max\{y_i/y_j, 1\}$	0.0796* (0.0464)	-0.0380 (0.1255)	0.2089*** (0.0616)	0.2777** (0.1071)
gender difference	-0.4029*** (0.0373)	-0.4331*** (0.0459)	-0.4767*** (0.0566)	-0.5213*** (0.0896)
degree aspiration difference	-0.1331*** (0.0389)	-0.1176** (0.0466)	-0.1541*** (0.0606)	-0.1512* (0.0931)
own popularity	0.1583*** (0.0118)	0.0211 (0.1306)	0.2739*** (0.0280)	0.4667*** (0.0844)
peer's popularity	-0.0511*** (0.0096)	0.1151 (0.1312)	-0.1046*** (0.0251)	-0.2456*** (0.0757)
$ \xi_i - \xi_j $	-0.9017*** (0.0532)	-0.9650*** (0.0612)	-1.0670*** (0.0988)	-0.0178 (0.0919)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

6 Endogenous Communication Intensity Networks

The previous sections focus on the binary representation of social networks, where a link is either present or absent between two individuals. However, in many real-world scenarios, the strength or intensity of connections can vary significantly. For instance, some friendships may involve frequent interactions, while others may be more casual with infrequent contact. In that case, the peer effects on individual well-being may vary according to the strength of the connections. To capture this variation, we extend our model to incorporate the intensity of connections using communication data, such as the number of messages exchanged between individuals (see Section 3).

Similarly, let M be an $n \times n$ network representing intensity of messaging connections, in addition to the binary connectivity representation offered by G . Each element m_{ij} takes values on $\mathbb{N} = \{0, 1, \dots\}$, where it is still maintained that $m_{ij} = m_{ji}$. The $n \times n$ real-valued adjacency matrix M^* is the graph M weighted by the number of messages transferred, with element $m_{ij}^* = \frac{m_{ij}}{\sum_{k \in \mathcal{N}} m_{ik}}$. Namely, M^* is the row-normalized version of M , and the entry $0 \leq m_{ij}^* \leq 1$.

Likewise, the well-being equation is modified to incorporate the intensity of connections as follows,

$$MH_i = \beta_0 + \beta_1 \underbrace{y_i}_{\text{own income}} + \beta_2 \underbrace{y_i^m}_{\text{weighted average peers' income}} + \beta_X^\top \mathbf{X}_i + \eta_i \quad (9)$$

where $y_i^m = \sum_{j \in \mathcal{N}} m_{ij}^* y_j$ measures the average peers' income weighted by the number of messages exchanged with peers.

Alternatively, we also consider the income comparison with the weighted average peer's income,

$$MH_i = \beta_0 + \beta_1 y_i + \beta_2 \max\left\{\frac{y_i^m}{y_i}, 1\right\} + \beta_3 \max\left\{\frac{y_i}{y_i^m}, 1\right\} + \beta_X^\top \mathbf{X}_i + \eta_i \quad (10)$$

Network formation is modelled as before, but for message counts rather than for the presence of a link. The linking utility of Section 2 is replaced by a latent index Λ_{ij} carrying the same covariates, and the Bernoulli likelihood over g_{ij} by a count likelihood over m_{ij} . The bilateral structure is preserved, since the indices of the two parties continue to enter multiplicatively, in the same way that mutual consent made the link probability

the product $p_{ij}p_{ji}$. Extending the original framework in the simplest way, we assume

$$\begin{aligned} m_{ij} \mid M^0, \mathbf{y}, \mathbf{P} &\sim \text{Poisson}(\psi_{ij} \cdot \psi_{ji}) \\ \psi_{ij} &\equiv \exp(\Lambda_{ij}) \\ \Lambda_{ij} &= \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\{\frac{y_j}{y_i}, 1\} \\ &\quad + \delta_4^\top |X_i - X_j| + \delta_5 D_i^0 \end{aligned}$$

where $D_i^0 \equiv \sum_{j \in \mathcal{N}} m_{ij}^0$ denotes the total number of messages i exchanged with peers in the baseline network M^0 . The downward comparison term $\max\{y_i/y_j, 1\}$ is excluded because, given the bilateral sum structure $\Lambda_{ij} + \Lambda_{ji}$, upward and downward comparison effects enter the likelihood only through their sum and cannot be separately identified in the intensity model.

As $\mathbb{E}[m_{ij} \mid M^0, \mathbf{y}, \mathbf{P}] = \psi_{ij} \cdot \psi_{ji}$, when Λ_{ij} grows while keeping Λ_{ji} fixed (e.g., through a popularity increase via D_i^0), the mean number of messages sent between dyad (i, j) grows. The same holds when Λ_{ij} is held fixed and Λ_{ji} varies.

Since the two indices enter through an exponential link, their product collapses to the exponential of their sum,

$$\mathbb{E}[m_{ij} \mid M^0, \mathbf{y}, \mathbf{P}] = \psi_{ij} \cdot \psi_{ji} = \exp(\Lambda_{ij}) \exp(\Lambda_{ji}) = \exp(\Lambda_{ij} + \Lambda_{ji}),$$

as the log-likelihood below makes explicit.

In that case, under a Poisson distribution with the expectation of $\psi_{ij} \cdot \psi_{ji}$ messages transferred, the probability that individuals i and j have m_{ij} messages transferred is,

$$\frac{(\psi_{ij} \cdot \psi_{ji})^{m_{ij}} \exp(-\psi_{ij} \cdot \psi_{ji})}{m_{ij}!}$$

The likelihood function of the exogenous network formation model is,

$$\mathcal{L}_{\text{network}}(\boldsymbol{\delta} \mid M^0, \mathbf{y}, \mathbf{P}) = \prod_{i < j} \frac{(\psi_{ij} \cdot \psi_{ji})^{m_{ij}} \exp(-\psi_{ij} \cdot \psi_{ji})}{m_{ij}!}$$

The log-likelihood function is,¹²

$$\log \mathcal{L}_{\text{network}}(\boldsymbol{\delta} \mid M^0, \mathbf{y}, \mathbf{P}) = \sum_{i < j} (m_{ij}(\Lambda_{ij} + \Lambda_{ji}) - \exp(\Lambda_{ij} + \Lambda_{ji}) - \log(m_{ij}!))$$

The specification set out above treats the intensity network as exogenous. Its endogenous counterpart introduces the latent type into both processes exactly as in Section 2.3: the term $\beta_\xi \xi_i$ is added to the mental health equations (9) and (10), the term $\delta_\xi |\xi_i - \xi_j|$ is added to Λ_{ij} , and the two processes are estimated jointly using the sampler described in

¹²Because M contains large integers, we use $\Gamma(m_{ij} + 1)$ in place of $m_{ij}!$ to avoid overflow, where $\Gamma(\cdot)$ is the gamma function.

Appendix A. We report both versions below.

6.1 Results: Communication Intensity Model

As in the binary network model, peer income is negatively associated with depressive symptoms when the intensity network is treated as exogenous (Table 11), whereas the association is positive once the intensity network is modelled endogenously (Table 15), though it is statistically significant only for CESD. The specification based on income comparison delivers the same conclusion: young adults display inequality aversion when their own income is compared with message-weighted average peer income (Table 16).

The determinants of message volume differ from those of binary link formation in some respects but not others. Peers' income is positively associated with message volume, and personality distance is a strong correlate of communication intensity (Table 15). Decomposing the index by trait (Table 14), individuals exchange more messages with those closer to them in Openness, Extraversion, and Neuroticism, but more messages with those more distant in Conscientiousness and Agreeableness. Income distance reduces message volume — recall from the previous section that the comparison term (reported as upward comparison in the tables) is identified only as the response of communication to income distance as a whole — and homophily in gender and degree aspiration remains significant; however, we find no significant homophily in unobserved characteristics. ¹³

Table 11: Empirical Results: MH and Peer's Income (Intensity)

Dependent:	BDI	CESD
constant	11.6586*** (1.2781)	20.0345*** (1.8475)
income	-0.0607 (0.0440)	-0.1734*** (0.0636)
weighted average peers' income	-0.1544** (0.0607)	-0.1401 (0.0878)
gender	-1.8097** (0.7644)	-2.2591** (1.1050)
Number of connections		1221
Average connections		8
Observations		297

*Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Note: This table presents the results of OLS regressions examining the relationship between mental health and weighted average peers' income.

¹³The network formation parameters in Panel B are nearly identical across BDI and CESD specifications. This is because the Poisson likelihood exploits the full distribution of message counts, providing sufficient information to identify the network formation parameters, including the unobserved type term $|\xi_i - \xi_j|$, from the communication network alone. As a result, the mental health specification contributes little additional information for identifying ξ , leaving the network formation estimates robust to the choice of outcome variable. This interpretation is consistent with the insignificant estimates of both ξ_i in Panel A and $|\xi_i - \xi_j|$ in Panel B, suggesting that unobserved heterogeneity plays a limited role in either mental health outcomes or communication sorting in this sample.

Table 12: Empirical Results: MH and Income Comparison (Intensity)

Dependent:	BDI	CESD
constant	9.4367*** (1.3837)	15.0654*** (1.9854)
income	-0.1412*** (0.0543)	-0.1786** (0.0779)
$\max\{y_i^m/y_i, 1\}$	-0.2406 (0.2138)	0.4198 (0.3067)
$\max\{y_i/y_i^m, 1\}$	1.0179 (0.7075)	1.4368 (1.0152)
gender	-1.9124** (0.7692)	-2.4382** (1.1036)
Number of connections		1221
Average connections		8
Observations		297

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Note: This table presents the results of OLS regressions examining the relationship between mental health and income comparison.

Table 13: Exogenous Network Formation: Intensity Model

Number of message: m_{ij}	
constant	1.2856*** (0.0032)
$d(P_i, P_j)$	-0.0461*** (0.0006)
y_j	0.0067*** (0.0001)
upward comparison	-0.0710*** (0.0006)
gender difference	-0.3682*** (0.0013)
degree aspiration difference	-0.0848*** (0.0014)
messages last year	0.0001*** (0.0000)
Number of connections	1221
Average number of connections	8
Observations	297

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the exogenous network formation model with intensity. Communication networks are exogenous defined by the number of two-way communication records. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the number of messages, m_{ij} . Upward comparison is defined by $\max\{y_j/y_i, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Messages last year is the total number of messages exchanged of individual i at year 1.

Table 14: Exogenous Network Formation: Intensity Model with Personality-Trait Decomposition

Number of message: m_{ij}	
constant	1.2601*** (0.0033)
$ O_i - O_j $	-0.1283*** (0.0008)
$ C_i - C_j $	0.0467*** (0.0007)
$ E_i - E_j $	-0.0783*** (0.0008)
$ A_i - A_j $	0.0648*** (0.0007)
$ N_i - N_j $	-0.0048*** (0.0007)
y_j	0.0065*** (0.0001)
upward comparison	-0.0725*** (0.0006)
gender difference	-0.3698*** (0.0013)
degree aspiration difference	-0.0852*** (0.0014)
messages last year	0.0001*** (0.0000)
Number of connections	1221
Average number of connections	8
Observations	297

Posterior standard errors in parentheses

*HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.*

Note: This table presents the posterior means and standard errors of the parameters in the exogenous network formation model with intensity. Communication networks are exogenous defined by the number of two-way communication records. The marginal effects of personality similarities on the number of messages (m_{ij}) are shown in $|O_i - O_j|$, $|C_i - C_j|$, $|E_i - E_j|$, $|A_i - A_j|$, $|N_i - N_j|$ for a pair $\{i, j\}$. Upward comparison is defined by $\max\{y_j/y_i, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Messages last year is the total number of messages exchanged of individual i at year 1.

Table 15: Endogenous Network (Intensity): Effects of Peer's Income on MH

(A) MH and Peer's Income		
Dependent:	BDI	CESD
constant	3.8980*** (0.8805)	3.7401*** (0.9920)
income	0.0443 (0.0466)	0.0521 (0.0712)
weighted average peers' income	0.0864 (0.0566)	0.3946*** (0.0802)
gender	-0.3383 (0.6494)	0.1705 (0.7983)
ξ_i	0.8290 (0.7494)	1.0177 (0.9629)
(B) Communication Intensity Network		
Number of message: m_{ij}		
constant	1.3186*** (0.0264)	1.3186*** (0.0264)
$d(P_i, P_j)$	-0.0451*** (0.0018)	-0.0451*** (0.0018)
y_j	0.0071*** (0.0008)	0.0071*** (0.0008)
upward comparison	-0.0699*** (0.0017)	-0.0699*** (0.0017)
gender difference	-0.3686*** (0.0024)	-0.3686*** (0.0024)
degree aspiration difference	-0.0872*** (0.0024)	-0.0872*** (0.0024)
messages last year	0.0001*** (0.0000)	0.0001*** (0.0000)
$ \xi_i - \xi_j $	-0.1176 (0.1064)	-0.1176 (0.1064)
Number of connections		1221
Average number of connections		8
Observations		297

Posterior standard errors in parentheses

*HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.*

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model with intensity. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the number of messages, m_{ij} . Upward comparison is defined by $\max\{y_j/y_i, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Messages last year is the total number of messages exchanged of individual i at year 1.

Table 16: Endogenous Network (Intensity): Effects of Income Comparison on MH

(A) MH and Income Comparison		
Dependent:	BDI	CESD
constant	3.2254*** (0.8555)	3.2196*** (0.9432)
income	0.0289 (0.0473)	0.1871*** (0.0659)
$\max\{y_i^m/y_i, 1\}$	0.3572* (0.1956)	1.5971*** (0.2742)
$\max\{y_i/y_i^m, 1\}$	1.6291*** (0.5936)	2.1789*** (0.7392)
gender	-0.5650 (0.6515)	-0.1016 (0.7973)
ξ_i	0.6667 (0.7245)	0.7468 (0.8843)
(B) Communication Intensity Network		
Number of message: m_{ij}		
constant	1.3205*** (0.0281)	1.3205*** (0.0281)
$d(P_i, P_j)$	-0.0450*** (0.0020)	-0.0450*** (0.0020)
y_j	0.0071*** (0.0007)	0.0071*** (0.0007)
$\max\{y_j/y_i, 1\}$	-0.0700*** (0.0016)	-0.0700*** (0.0016)
gender difference	-0.3686*** (0.0025)	-0.3686*** (0.0025)
degree aspiration difference	-0.0871*** (0.0021)	-0.0871*** (0.0021)
messages last year	0.0001*** (0.0000)	0.0001*** (0.0000)
$ \xi_i - \xi_j $	-0.1224 (0.1095)	-0.1225 (0.1095)
Number of connections		1221
Average number of connections		8
Observations		297

Posterior standard errors in parentheses

*HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.*

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model with intensity. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j . y_j denotes the marginal effects of peer's income on the number of messages, m_{ij} . Upward comparison is defined by $\max\{y_j/y_i, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Messages last year is the total number of messages exchanged of individual i at year 1.

7 Counterfactual Analysis

In this section, we conduct two counterfactual exercises corresponding to our two models of the network: the binary network and the communication-intensity network. The first perturbs incomes and traces the effect on depressive symptoms when the network is allowed to re-form, benchmarked against a model in which it is not; the second reduces the intensity of peer communication, which in that model constitutes the network itself. The exercises thus correspond to the two margins of peer influence in our framework — the composition of the peer group and the intensity of contact within it.

7.1 Income Shocks and Network Formation

The estimated model delivers a joint description of how income shapes both mental health and the structure of the social network. We exploit this structure to ask a policy-relevant question that a standard regression cannot answer: what happens to an individual’s mental health when incomes change, once the social network is allowed to re-form in response? Because links are endogenous to income, an income shock does not merely move own income and the incomes of a fixed set of peers; it also reshapes who an individual is connected to, and hence the composition of the peer group that enters the mental health equation. In this section, we implement various counterfactual exercises and report how mental health on average responds to different income shocks. We compare two cases: one where the network is allowed to re-form (endogenous model) and one where it is not (exogenous model).

The benchmark shock is one standard deviation of income, $y_i^{\text{new}} = y_i + \sigma_y$. Throughout, $\{\theta_s\}_{s=1}^S$ are the posterior draws $\theta = (\beta, \delta, \xi)$, y_i^* is i ’s average peer income, and $Q_{ij} = q_{ij} q_{ji}$ is the mutual-consent link probability, with $q_{ij} = \exp(\Lambda_{ij}) / (1 + \exp(\Lambda_{ij}))$ as in Section 2.3; the probability responds to the shock through the income channels entering Λ .

Network sampling. Fix a posterior draw θ_s and an individual i . Using the observed and the perturbed incomes we recompute the link-probability row,

$$Q_{ij}^{\text{base}} = Q(y_i, y_j; \theta_s), \quad Q_{ij}^{\text{new}} = Q(y_i^{\text{new}}, y_j^{\text{new}}; \theta_s), \quad j \neq i. \quad (11)$$

Equation (11) states the case in which all incomes are perturbed; for the shocks to own income alone, or to the incomes of i ’s peers alone, the same construction applies with only the relevant incomes replaced by their perturbed values. Since the network is stochastic, we integrate over it by Monte Carlo: for each of R replications we draw an independent

network row and average the resulting peer income,¹⁴

$$g_{ij}^{(r)} \sim \text{Bernoulli}(Q_{ij}), \quad g_{ii}^{(r)} = 0, \quad \hat{y}_i^* = \frac{1}{R} \sum_{r=1}^R \frac{\sum_{j \neq i} g_{ij}^{(r)} y_j}{\sum_{j \neq i} g_{ij}^{(r)}}. \quad (12)$$

This yields $\hat{y}_i^{*,\text{base}}$ from Q^{base} with observed incomes and $\hat{y}_i^{*,\text{new}}$ from Q^{new} with the perturbed incomes y^{new} . The averaging in (12) captures the genuine re-formation of the network: the shock changes which peers are likely to be linked, not only the income they contribute.

Change in outcomes. For mental health, the individual change is the difference of the MH equation f evaluated at the counterfactual and the baseline,

$$\begin{aligned} \Delta \text{MH}_{i,s} &= \text{MH}_{i,s}^{\text{new}} - \text{MH}_{i,s}^{\text{base}} \\ &= f(y_i^{\text{new}}, \hat{y}_i^{*,\text{new}}, x_i, \xi_{i,s}; \beta_s) - f(y_i, \hat{y}_i^{*,\text{base}}, x_i, \xi_{i,s}; \beta_s), \end{aligned} \quad (13)$$

which we average over the n individuals and the S posterior draws,

$$\overline{\Delta \text{MH}} = \frac{1}{S} \sum_{s=1}^S \frac{1}{n} \sum_{i=1}^n \Delta \text{MH}_{i,s}. \quad (14)$$

For the linking utility, we first compute the change for each pair, $\Delta \Lambda_{ij,s} = \Lambda(y_i^{\text{new}}, y_j^{\text{new}}; \delta_s) - \Lambda(y_i, y_j; \delta_s)$, average over i 's potential partners, $\Delta u_{i,s} = \frac{1}{n-1} \sum_{j \neq i} \Delta \Lambda_{ij,s}$, and then aggregate in the same way to obtain $\overline{\Delta u}$. Both quantities integrate over posterior uncertainty in (β, δ, ξ) (the S draws) and the empirical income distribution (the n individuals); the mental health change additionally integrates over the stochastic network (the R inner replications in (12)).

To illustrate the mechanisms at work, we use the posterior draws from the BDI specifications estimated on the strong-ties network to conduct a series of counterfactual exercises. We consider two classes of counterfactuals, corresponding to the two BDI specifications: the specification with average peers' income (5) and the specification with income comparison (8). For the former, we examine the average effect on BDI of counterfactual changes in own income, in the incomes of all other individuals, and in the incomes of peers. For the latter, we examine the average effect of counterfactual changes in the incomes of richer and poorer individuals, distinguishing between others and peers. In the exercises involving others, we change, for each individual i , the incomes of all individuals other than i , and then average the resulting effects on BDI across individuals. In the exercises involving peers, we instead change only the incomes of i 's peers before averaging.

Figure 4 shows the counterfactual results for the specification with average peer's income. Consistent with the estimation results, the exogenous and endogenous models yield

¹⁴A replication in which i draws no links contributes zero to the average peer income, so that $\hat{y}_i^*$ is an average over all R replications with isolated draws treated as having no peer income effect.

markedly different predictions, particularly for the shock that raises others' incomes. In the exogenous model, an increase in others' incomes reduces young adults' BDI scores. By contrast, once individuals are allowed to re-form their networks, a one-standard-deviation increase in others' incomes raises young adults' BDI scores by more than one point. Panel B lends additional support to the role of relative-status concerns in link formation: an increase in own income raises average linking utility, whereas an increase in others' incomes lowers it.

The effect of peers' income on depressive symptoms is smaller but remains positive. Because this shock applies only to an individual's existing peers, individuals can reshape their networks in response, severing ties with peers who have become substantially richer and forming new ties with relatively poorer individuals. The change in average peers' income y_i^* therefore reflects two opposing forces: the income gains of current friends mechanically raise y_i^* , while network re-formation (replacing much richer peers with less rich ones) lowers it. As a result, average peers' income need not rise substantially, since individuals can attenuate the adverse mental health effect by adjusting their social circles.

The economic intuition is as follows: when individuals are placed in a social environment in which incomes are higher across the board, their depressive symptoms worsen: they cannot rearrange their networks to find relatively poorer peers. When only their existing peers' incomes rise, however, relative-status concerns in the link-formation process enable individuals to protect themselves by re-forming their networks.

Figure 5 presents the counterfactual results for the specification with income comparison. The results are consistent with the estimates in Tables 9 and 10, which indicate that young adults are inequality averse: income gaps in either direction are associated with more severe depressive symptoms. Accordingly, an increase in richer others' incomes (which widens the income gap) raises BDI scores, whereas an increase in poorer others' incomes (which narrows the gap) lowers them. Moreover, because individuals benefit from connecting with relatively poorer others, increases in the incomes of both richer and poorer others reduce linking utility. Similarly, increases in peers' incomes have smaller effects on mental health, as young adults can adjust their networks to attenuate the impact.

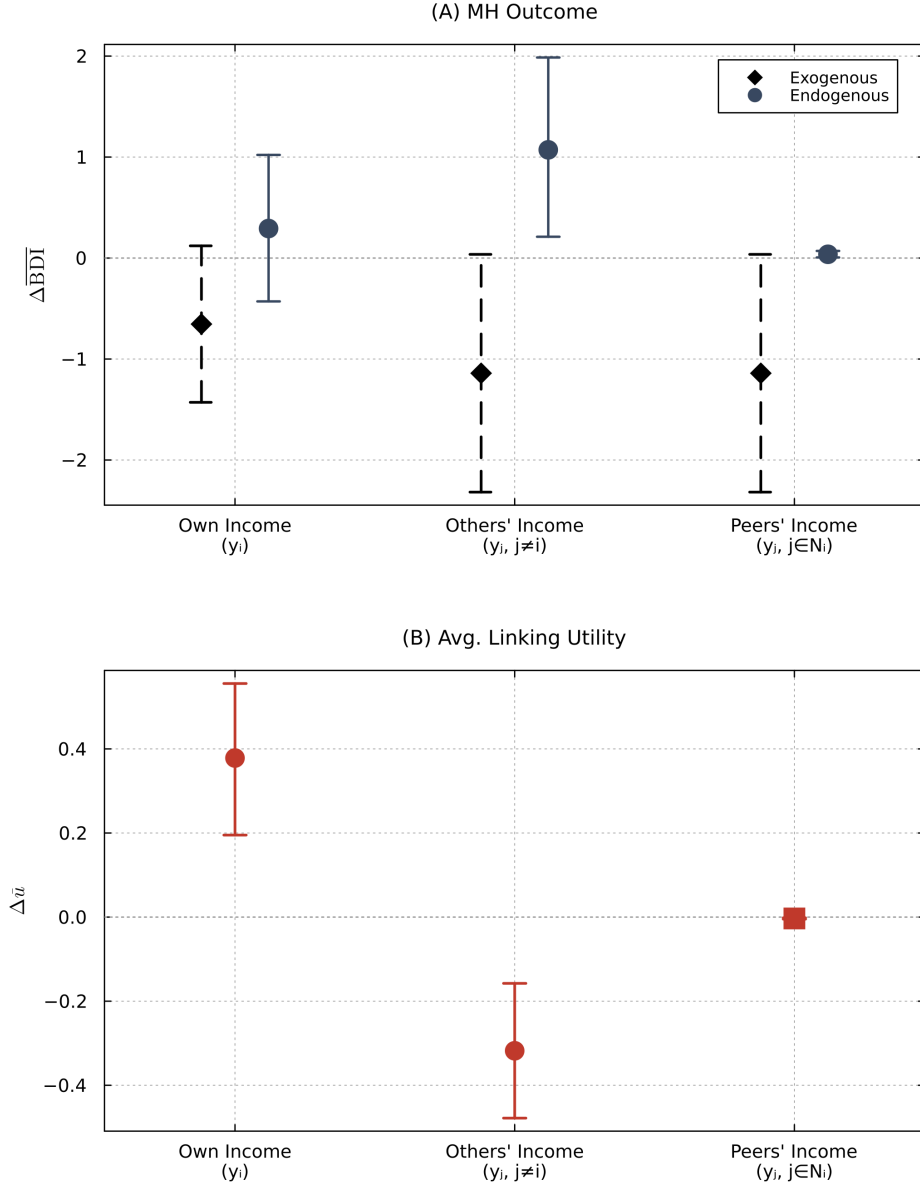


Figure 4: Income Increase and BDI (Own, Others', and Peers')

Notes: **Panel A** reports the average change in the BDI depression score under the indicated income shock. **Panel B** reports the corresponding change in individual i 's average linking utility, $\bar{u}_i = \frac{1}{n-1} \sum_{j \neq i} \Lambda_{ij}$, where Λ_{ij} is the pre-sigmoid linear index of i 's utility of linking to j . The three exercises apply a one-standard-deviation shock to own income (y_i), all others' income ($y_j, j \neq i$), and peers' income ($y_j, j \in N_i$), respectively. Black diamonds in Panel A denote the exogenous OLS estimates. The 95% confidence intervals are $\bar{\Delta}^{\text{exo}} \pm 1.96 \cdot \widehat{\text{SE}}$. For own income, $\widehat{\text{SE}} = \text{SE}(\hat{\beta}_2)$ directly. For peer effects, $\widehat{\text{SE}} = W \cdot \text{SE}(\hat{\beta}_3)$, where W is the sample-average marginal change in average peer income with respect to the income shock. Navy circles (Panel A) and red circles (Panel B) denote the endogenous model estimates for Own Income and Others exercises; red squares (Panel B) denote the endogenous model estimates for Peers exercises. The 95% HPD intervals are constructed from the posterior distribution of the cross-individual average change $\frac{1}{n} \sum_{i=1}^n \Delta_{i,s}$, integrating over all S posterior draws. Red squares in Panel B indicate that the average change in linking utility for peer income exercises is empirically small, as the income change affects only a subset of potential pairs while the change in linking utility is averaged over all potential pairs. All estimates use the strong-ties network and the average peer income (API) specification for BDI (Table 1, column 3 and Table 4, column 3).

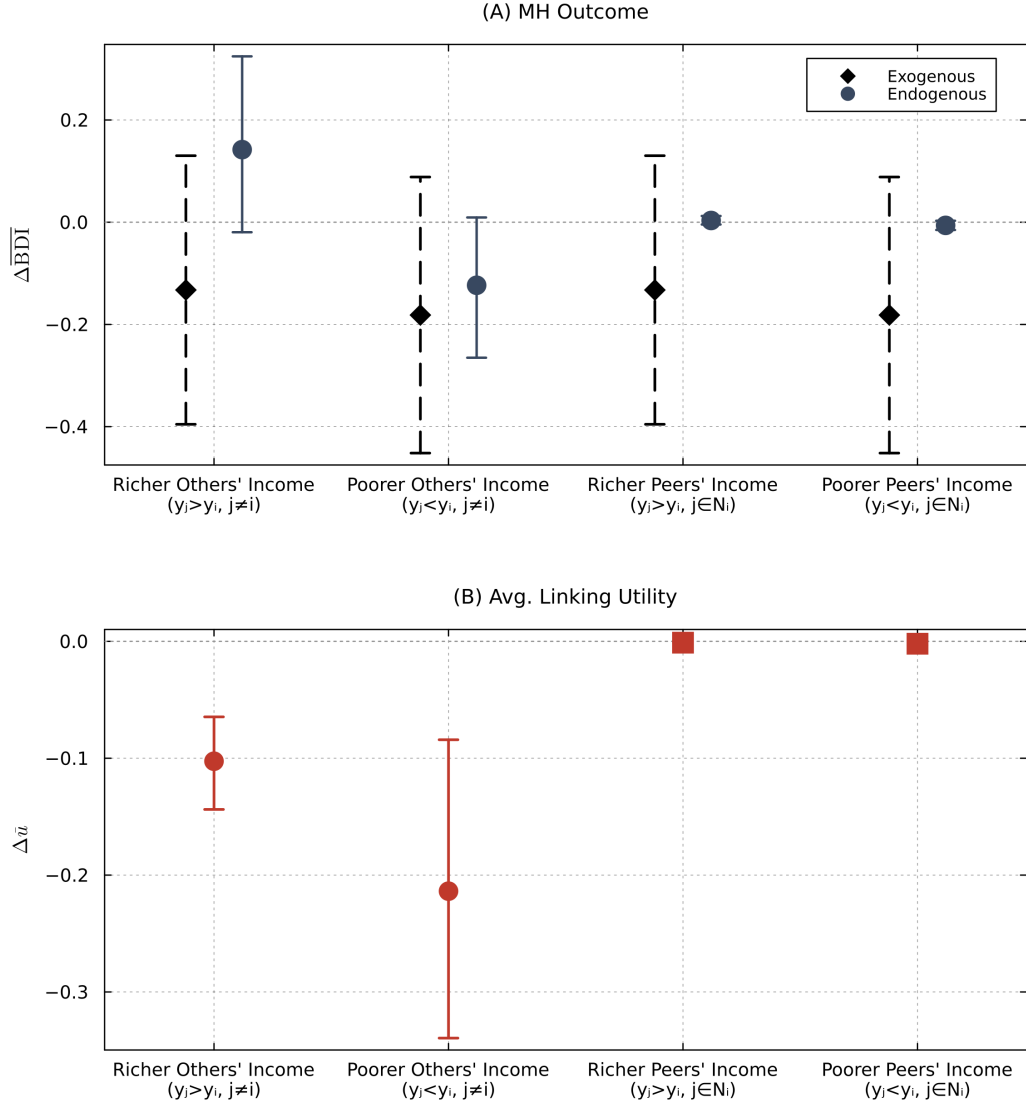


Figure 5: Income Increase and BDI (Richer vs. Poorer Others and Peers)

Notes: **Panel A** reports the average change in the BDI depression score under the indicated income shock. **Panel B** reports the corresponding change in individual i 's average linking utility, $\bar{u}_i = \frac{1}{n-1} \sum_{j \neq i} \Lambda_{ij}$, where Λ_{ij} is the pre-sigmoid linear index of i 's utility of linking to j . The four exercises apply a one-standard-deviation shock to richer others' income ($y_j > y_i, j \neq i$), poorer others' income ($y_j < y_i, j \neq i$), richer peers' income ($y_j > y_i, j \in \mathcal{N}_i$), and poorer peers' income ($y_j < y_i, j \in \mathcal{N}_i$), respectively. Black diamonds in Panel A denote the exogenous OLS estimates. The 95% confidence intervals are $\bar{\Delta}^{\text{exo}} \pm 1.96 \cdot \widehat{\text{SE}}$, where $\widehat{\text{SE}} = \sqrt{W_C^2 \cdot \text{SE}(\hat{\beta}_3)^2 + W_D^2 \cdot \text{SE}(\hat{\beta}_4)^2}$ is obtained by the delta method, where W_C and W_D are the sample-average marginal changes in the above- and below-income comparison terms with respect to the income shock, respectively. Navy circles (Panel A) denote the endogenous model estimates across all exercises. Red circles (Panel B) denote the endogenous estimates for Others exercises; red squares (Panel B) denote the endogenous estimates for Peers exercises. The 95% HPD intervals are constructed from the posterior distribution of the cross-individual average change $\frac{1}{n} \sum_{i=1}^n \Delta_{i,s}$, integrating over all S posterior draws. Red squares in Panel B indicate that the average change in linking utility for peer income exercises is empirically small, as the income change affects only a subset of potential pairs while the change in linking utility is averaged over all potential pairs. All estimates use the strong-ties network and the income comparison (IC) specification for BDI (Table 7, column 3 and Table 9, column 3).

7.2 Reduction in Messages

Policies restricting young people’s digital communication — school phone bans, platform-imposed time caps — act on the intensity of peer contact rather than its composition, a margin whose welfare effects are *ex ante* ambiguous. We therefore ask whether young adults’ mental health would respond to policies that alter the volume of messages they exchange with peers. Using the estimates of the intensity model in Section 6, we simulate individual CESD scores under counterfactual reductions in message volume, implemented separately for all peers, for richer peers only, and for poorer peers only.¹⁵ Unlike in the previous exercise on income, where the network can be reformed, we directly manipulate the communication network through its intensity.

First, we consider a uniform reduction in the number of messages exchanged with all peers.¹⁶ As Figure 6 shows, reducing message volume lowers young adults’ CESD scores, that is, it improves their mental health — and the magnitude of the improvement is increasing in the size of the reduction. Both specifications (Average Peer Income and Income Comparison) deliver the same qualitative pattern.

Second, we consider proportional reductions targeted at particular groups of peers (see Figure 7). Reducing the messages exchanged with richer peers lowers CESD scores, whereas reducing those exchanged with poorer peers raises them. This asymmetry follows from the way targeted reductions reweight the peer income distribution: assigning less weight to richer peers lowers weighted peer income y_i^m , while assigning less weight to poorer peers raises it — and weighted peer income enters the mental health equation with a positive sign in the endogenous model (Table 15). As in the uniform case, the effect grows with the size of the reduction. Under the income-comparison specification, we find weak evidence that reductions in message volume are associated with lower depression; however, the estimated effects are not statistically distinguishable from zero based on the 95% HPD interval.

¹⁵We focus on CESD because, in the endogenous intensity model, the weighted peer-income effect is precisely estimated only for CESD; the BDI coefficient has the same sign but its HPD interval includes zero (Table 15).

¹⁶As a uniform proportional cut leaves the weights $m_{ij}^* = m_{ij} / \sum_k m_{ik}$ unchanged and hence y_i^m unchanged, we only consider the uniform reduction in the number of messages rather than a proportion.

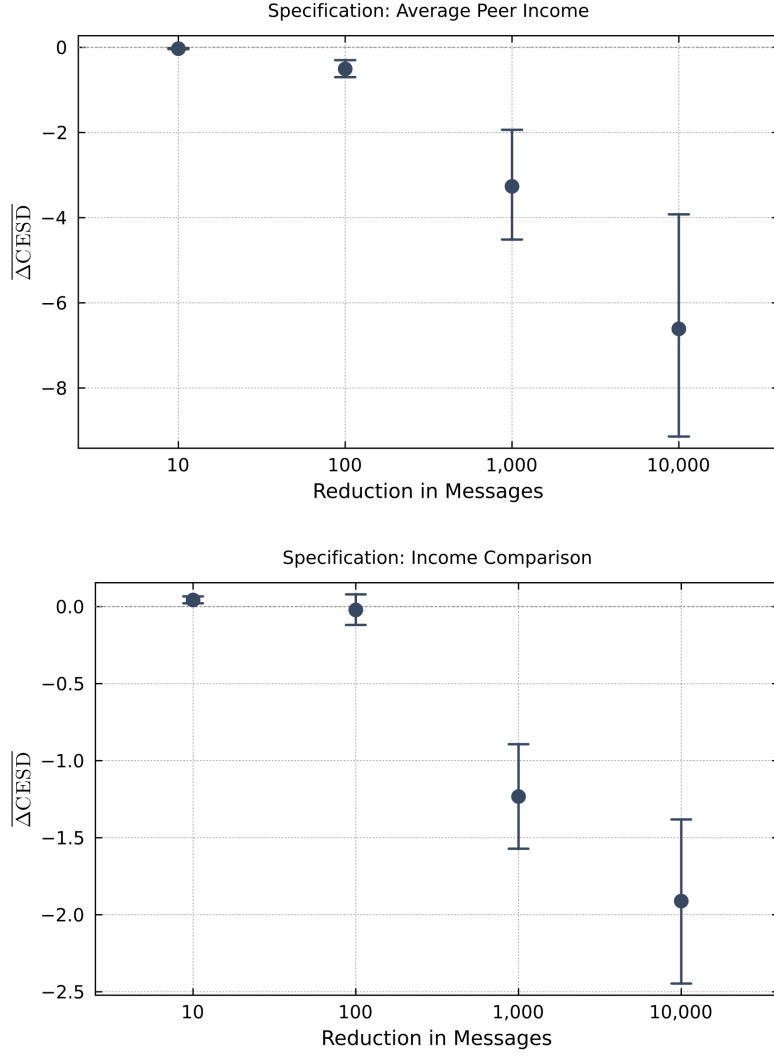


Figure 6: Communication Intensity Reduction and CESD

Notes: The figure reports the average change in the CESD depression score under reductions in the number of messages. The upper panel corresponds to the average peer income (API) specification, and the lower panel to the income comparison (IC) specification. Each panel displays the posterior mean (circle) and 95% HPD interval for the cross-individual average change $\overline{\Delta\text{CESD}} = \frac{1}{n} \sum_{i=1}^n \Delta\text{CESD}_{i,s}$, integrating over all S posterior draws. The x -axis indicates the number of messages subtracted from every pairwise link; connections falling below 2 messages after the reduction are treated as severed. In the upper panel, the channel is the resulting change in weighted-average peer income y_i^m . All estimates use the posterior from the API specification (Table 15, column 2). In the lower panel, the channels are the corresponding changes in the income comparison terms $\max\{y_i^m/y_i, 1\}$ and $\max\{y_i/y_i^m, 1\}$; individuals who become completely isolated after the reduction are assigned neutral comparison values of unity. All estimates use the posterior from the IC specification (Table 16, column 2).

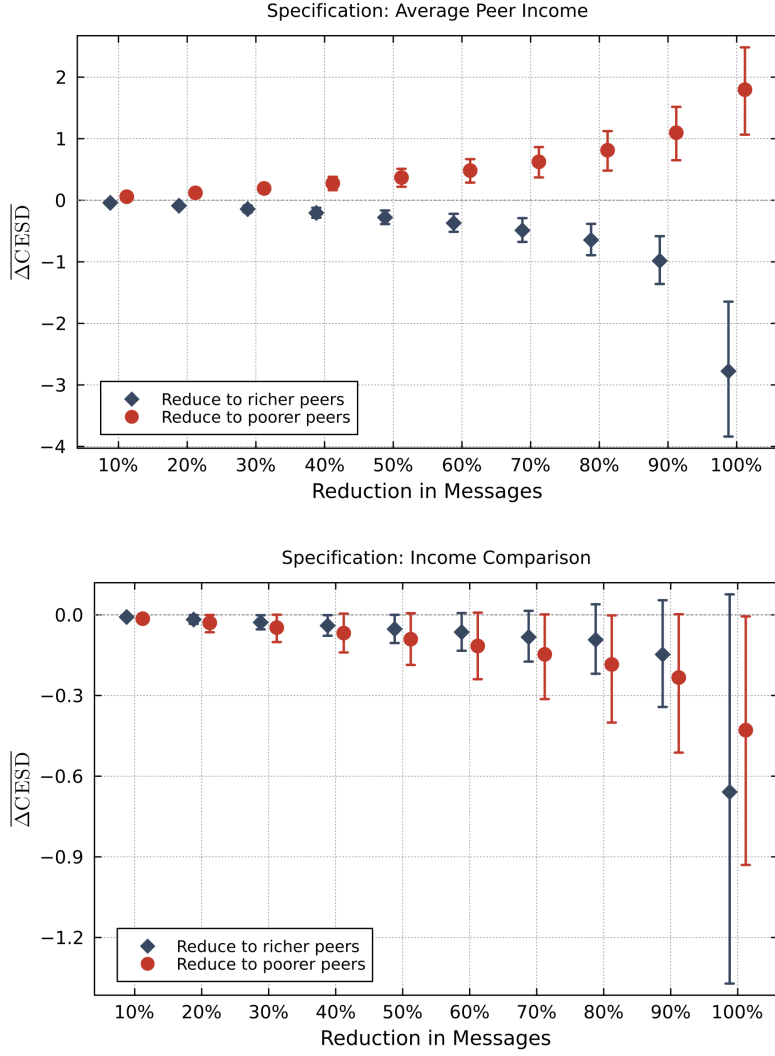


Figure 7: Communication Intensity Reduction and CESD (Richer vs. Poorer Peers)

Notes: The figure reports the average change in the CESD depression score under selective reductions in message intensity directed at richer or poorer peers. The upper panel corresponds to the average peer income (API) specification and the lower panel to the income comparison (IC) specification. Navy diamonds denote reductions applied only to messages sent to peers with strictly higher income ($y_j > y_i$); red circles denote reductions applied only to messages sent to peers with strictly lower income ($y_j < y_i$). Each series reports the posterior mean and 95% HPD interval for the cross-individual average change $\overline{\Delta CESD} = \frac{1}{n} \sum_{i=1}^n \Delta CESD_{i,s}$, integrating over all S posterior draws. The x -axis indicates the percentage by which the affected messages are reduced; connections falling below 2 messages after the reduction are treated as severed. In the upper panel, the channel is the change in weighted-average peer income. All estimates use the posterior from the API specification (Table 15, column 2). In the lower panel, the channels are the changes in the income comparison terms, with isolated individuals assigned neutral comparison values of unity. All estimates use the posterior from the IC specification (Table 16, column 2).

8 Conclusion

This paper has examined how social comparison within endogenously formed peer networks affects mental health among young adults. Using high-frequency smartphone-based communication records from the NetHealth Study, we constructed objective measures of social interaction intensity and estimated a structural framework in which individuals form ties strategically and mental health is shaped by the income of endogenously chosen peers.

Our main finding is that network endogeneity is not a second-order econometric con-

cern but a substantively important feature of the data that fundamentally shapes the empirical relationship between peer income and mental health. Under the exogenous model, higher average peer income appears to improve mental health — a result consistent with straightforward positive spillovers from a wealthier social environment, such as access to resources, information, and emotional support. Once network formation is modelled endogenously, this relationship reverses entirely. Conditional on peer selection, exposure to higher-income peers is associated with significantly worse mental health outcomes. This sign reversal indicates that the compositional effect of peer selection — the tendency to form ties with individuals who intensify social comparison pressure — dominates the direct income spillover effect of a fixed peer group.

Decomposing the peer income effect into its upward and downward components reveals a further result that goes beyond the standard relative deprivation framework. Both upward and downward income gaps are harmful to mental health: individuals are worse off when their peers are richer than them, but also when their peers are poorer. This bidirectional pattern is inconsistent with pure relative deprivation, which assigns welfare costs only to unfavourable upward comparisons, and points instead toward inequality aversion — a generalised discomfort with economic distance from one’s peers in either direction. This finding connects naturally to the homophily-driven network formation process we document: individuals prefer economically similar peers, and deviations from similarity, regardless of direction, carry psychological costs.

Extending the framework from binary links to weighted networks, in which tie strength reflects the volume of messages exchanged, leaves these conclusions intact while changing what drives link formation itself. Social comparison plays little role in determining how much individuals communicate with one another; personality distance instead emerges as the dominant correlate of communication intensity, and homophily in unobserved characteristics disappears. The peer income effect, however, exhibits the same sign reversal as in the binary model, indicating that the selection problem is not an artefact of how ties are discretised.

Because the estimated model describes jointly how income shapes mental health and how it shapes the network, it can be used to answer counterfactual questions that a reduced-form specification cannot address. Two exercises are instructive. Raising incomes across the board worsens depressive symptoms by more than one BDI level, whereas the fixed-network benchmark predicts an improvement; raising only the incomes of an individual’s existing peers has a markedly smaller, though still adverse, effect. The asymmetry is revealing: when a shock is confined to current peers, individuals partially insulate themselves by severing those ties and forming new ones with relatively poorer individuals, so that relative-status concerns in link formation operate as a form of self-protection. When incomes rise throughout the distribution, that adjustment margin is unavailable and the full cost of comparison is borne — a distinction between individual and aggregate income growth that models with fixed reference groups cannot represent.

The second exercise speaks more directly to policy. Reducing the volume of messages young adults exchange with peers improves their mental health on average, and increasingly so with the size of the reduction. Targeted reductions, however, work in opposite directions: attenuating contact with richer peers improves mental health, while attenuating contact with poorer peers worsens it, since each reweights the peer income distribution the other way. The effects are estimated less precisely under the income-comparison specification, and we regard the targeted results as suggestive rather than definitive. They nonetheless indicate that measures such as school phone bans or platform-imposed time caps act on a margin whose consequences depend on which ties are attenuated, and not on the volume of contact alone.

Taken together, these results carry three broader implications. First, they suggest that the deterioration in youth mental health documented in recent years may be partly driven not by absolute economic conditions but by the composition of peer networks and the comparison dynamics they generate — a mechanism that is invisible to models that treat reference groups as fixed. Second, they highlight a general methodological lesson for the peer effects literature: failing to account for endogenous network formation risks not merely attenuating estimates but reaching qualitatively wrong conclusions about the direction of peer influence. Third, they caution against evaluating interventions in young people’s social environments without accounting for the endogenous response of the network: because individuals adjust whom they interact with, and how intensively, policies that appear beneficial when the peer group is held fixed may prove ineffective or perverse once that adjustment is allowed for.

Future work could extend the framework to more diverse populations, explore dynamic network formation over longer time horizons, and investigate the specific behavioral mechanisms — social media exposure, consumption signaling, direct interaction — through which peer income translates into mental health costs. Understanding these mechanisms in greater detail remains an important priority for both research and policy.

References

- Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education*, volume 4, pages 1–181. Elsevier.
- Bárcena-Martín, E., Cortés-Aguilar, A., and Moro-Egido, A. I. (2017). Social comparisons on subjective well-being: The role of social and cultural capital. *Journal of Happiness Studies*, 18(4):1121–1145.
- Battestini, A., Setlur, V., and Sohn, T. (2010). A large scale study of text-messaging use. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 229–238.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561–571.
- Becker, A., Deckers, T., Dohmen, T., Falk, A., and Kosse, F. (2012). The relationship between economic preferences and psychological personality measures. *Annual Review of Economics*, 4(1):453–478.
- Blanchflower, D. G. and Oswald, A. J. (2004). Well-being over time in britain and the usa. *Journal of Public Economics*, 88(7-8):1359–1386.
- Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059.
- Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55.
- Buunk, A. P. and Dijkstra, P. (2017). Social comparisons and well-being. In *The happy mind: Cognitive contributions to well-being*, pages 311–330. Springer.
- Clark, A. et al. (2018). The origins of happiness. *Princeton University Press*.
- Clark, A. E. and Senik, C. (2010). Who compares to whom? the anatomy of income comparisons in europe. *The Economic Journal*, 120(544):573–594.
- Currie, J. (2025). Presidential address: Investing in children to address the child mental health crisis. *American Economic Review*, 115(5):1369–1399.
- Diener, E. and Fujita, F. (2013). Social comparisons and subjective well-being. In *Health, coping, and well-being*, pages 329–357. Psychology Press.
- Eagle, N., Pentland, A., and Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278.

- Falk, A. and Knell, M. (2004). Choosing the joneses: Endogenous goals and reference standards. *The Scandinavian Journal of Economics*, 106(3):417–435.
- Fang, R., Landis, B., Zhang, Z., Anderson, M. H., Shaw, J. D., and Kilduff, M. (2015). Integrating personality and social networks: A meta-analysis of personality, network position, and work outcomes in organizations. *Organization science*, 26(4):1243–1260.
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.
- Feiler, D. C. and Kleinbaum, A. M. (2015). Popularity, similarity, and the network extraversion bias. *Psychological science*, 26(5):593–603.
- Ferrer-i Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of public economics*, 89(5-6):997–1019.
- Ghiglini, C. and Goyal, S. (2010). Keeping up with the neighbors: social interaction in a market economy. *Journal of the European Economic Association*, 8(1):90–119.
- Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.
- Griffith, A. (2022). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. *Journal of Labor Economics*, 40(4):779–805.
- Haidt, J. (2024). *The Anxious Generation*. Random House.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161.
- Hsieh, C.-S., König, M. D., and Liu, X. (2022). A structural model for the coevolution of networks and behavior. *Review of Economics and Statistics*, 104(2):355–367.
- Hsieh, C.-S. and Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.
- Immorlica, N., Kranton, R., Manea, M., and Stoddard, G. (2017). Social status in networks. *American Economic Journal: Microeconomics*, 9(1):1–30.
- Jackson, M. (2008). Social and economic networks.
- Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of economic literature*, 55(1):49–95.

- Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of economic theory*, 71(1):44–74.
- John, O. P., Donahue, E. M., and Kentle, R. L. (1991). Big five inventory. *Journal of personality and social psychology*.
- Kossinets, G. and Watts, D. J. (2006). Empirical analysis of an evolving social network. *science*, 311(5757):88–90.
- Landis, B. (2016). Personality and social networks in organizations: A review and future directions. *Journal of organizational behavior*, 37:S107–S121.
- Lewbel, A., Qu, X., and Tang, X. (2023). Social networks with unobserved links. *Journal of Political Economy*, 131(4):898–946.
- Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly journal of economics*, 120(3):963–1002.
- Marmaros, D. and Sacerdote, B. (2006). How do friendships form? *The Quarterly Journal of Economics*, 121(1):79–119.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.
- Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science*, 10(5):173–176.
- Purta, R., Mattingly, S., Song, L., Lizardo, O., Hachen, D., Poellabauer, C., and Striegel, A. (2016). Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits. In *Proceedings of the 2016 ACM international symposium on wearable computers*, pages 28–35.
- Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.
- Rosenquist, J. N., Fowler, J. H., and Christakis, N. A. (2011). Social network determinants of depression. *Molecular psychiatry*, 16(3):273–281.
- Twenge, J. M. (2024). *Generations*. Atria Books.
- Zhang, A., Gagné, T., Walsh, D., Ciancio, A., Proto, E., and McCartney, G. (2023). Trends in psychological distress in great britain, 1991–2019: evidence from three representative surveys. *J Epidemiol Community Health*, 77(7):468–473.

A Endogenous Network Model

Priors

We are interested in β, δ, σ^2 , and ξ is unobservable variable, and $\theta = (\beta, \delta, \sigma^2, \xi)$.

- β ($n_\beta \times 1$): normal distribution with mean zero and variance equal to one.

$$P(\beta) = \frac{1}{(2\pi)^{\frac{n_\beta}{2}}} \exp\left(-\frac{1}{2}\beta^\top \beta\right)$$

- δ ($n_\delta \times 1$): normal distribution with mean zero and variance equal to one.

$$P(\delta) = \frac{1}{(2\pi)^{\frac{n_\delta}{2}}} \exp\left(-\frac{1}{2}\delta^\top \delta\right)$$

- σ^2 : inverse χ^2 , $Inv - \chi^2(\nu_0, \tau_0^2)$

$$P(\sigma^2) \propto (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp\left(-\frac{\nu_0 \cdot \tau_0^2}{2\sigma^2}\right)$$

where ν_0 is the degrees of freedom and τ_0^2 is the scale parameter.

- $\xi_i \stackrel{iid}{\sim} \text{Bernoulli}(\pi)$: Bernoulli heterogeneity with prior weights π on types

$$P(\xi_i = 1) = \pi$$

We set $\pi = 1/2$ in all our results to allow for equal prior type weights.

Conditional Distributions

Full posterior distributions,

$$P(\theta \mid MH, G, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \propto P(MH \mid \beta, \sigma^2, \xi; G, \mathbf{y}, \mathbf{X}) \times P(G \mid \delta, \xi; \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \\ \times P(\beta) \times P(\sigma^2) \times P(\delta) \times P(\xi)$$

And the likelihood function for the outcome,

$$P(MH \mid \beta, \sigma^2, \xi; G, \mathbf{y}, \mathbf{X}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right)$$

where

$$\mu_i = \beta_0 + \beta_1 y_i + \beta_2 y_i^* + \beta_X^\top \mathbf{X}_i + \beta_\xi \xi_i$$

The likelihood function for the network,

$$P(G \mid \delta, \xi; \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) = \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}]$$

where

$$q_{ij} = \frac{\exp(\Lambda_{ij})}{1 + \exp(\Lambda_{ij})}$$

$$\Lambda_{ij} = \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\{y_j/y_i, 1\} + \delta_4 \max\{y_i/y_j, 1\} + \delta_5^\top |X_i - X_j| + \delta_6 d_i^0 + \delta_7 d_j^0 + \delta_\xi |\xi_i - \xi_j|$$

$$q_{ji} = \frac{\exp(\Lambda_{ji})}{1 + \exp(\Lambda_{ji})}$$

$$\Lambda_{ji} = \delta_0 + \delta_1 d(P_j, P_i) + \delta_2 y_i + \delta_3 \max\{y_i/y_j, 1\} + \delta_4 \max\{y_j/y_i, 1\} + \delta_5^\top |X_j - X_i| + \delta_6 d_j^0 + \delta_7 d_i^0 + \delta_\xi |\xi_j - \xi_i|$$

Conditional Distribution for β

$$\begin{aligned} & P(\beta \mid \sigma^2, \xi; MH, G, \mathbf{y}, \mathbf{X}) \\ & \propto \underbrace{P(MH \mid \beta, \sigma^2, \xi; G, \mathbf{y}, \mathbf{X})}_{\text{Likelihood function for the MH}} \times P(\beta) \\ & = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right) \times \frac{1}{(2\pi)^{\frac{n_\beta}{2}}} \exp\left(-\frac{1}{2}\beta^\top \beta\right) \\ & = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}(MH - \mu)^T(MH - \mu)\right) \times \frac{1}{(2\pi)^{\frac{n_\beta}{2}}} \exp\left(-\frac{1}{2}\beta^\top \beta\right) \end{aligned}$$

Given σ^2, ξ , we can first update β using a Metropolis-Hasting algorithm, the proposal is the candidate normal distribution with the current value as mean.

Conditional Distribution for σ^2

$$\begin{aligned} & P(\sigma^2 \mid \beta, \xi; MH, G, \mathbf{y}, \mathbf{X}) \\ & \propto \underbrace{P(MH \mid \beta, \sigma^2, \xi; G, \mathbf{y}, \mathbf{X})}_{\text{Likelihood function for the outcome}} \times P(\sigma^2) \\ & \propto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right) \times (\sigma^2)^{-(\frac{\nu_0}{2}+1)} \exp\left(-\frac{\nu_0 \cdot \tau_0^2}{2\sigma^2}\right) \\ & \propto (\sigma^2)^{-(\frac{n}{2}+\frac{\nu_0}{2}+1)} \exp\left(-\frac{\sum_{i=1}^n (MH_i - \mu_i)^2 + \nu_0 \cdot \tau_0^2}{2\sigma^2}\right) \end{aligned}$$

Thus it's a form of an inverse χ^2 distribution, and the conditional posterior of σ^2 given other parameters and the data is,

$$\begin{aligned} \sigma^2 \mid \beta, \xi; MH, G, \mathbf{y}, \mathbf{X} & \sim Inv - \chi^2(\nu_0 + n, \frac{\sum_{i=1}^n (MH_i - \mu_i)^2 + \nu_0 \cdot \tau_0^2}{\nu_0 + n}) \\ & \sim Inv - Gamma(\frac{\nu_0 + n}{2}, \frac{\sum_{i=1}^n (MH_i - \mu_i)^2 + \nu_0 \cdot \tau_0^2}{2}) \end{aligned}$$

Since we have the exact conditional posterior distribution, the next step is to update σ^2 using Gibbs, given the updated values of β in this iteration and ξ from last iteration.

Conditional Distribution for δ

$$\begin{aligned}
& P(\delta \mid \xi; G, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \\
& \propto \underbrace{P(G \mid \delta, \xi; \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0)}_{\text{Likelihood function for the network}} \times P(\delta) \\
& = \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}] \times \frac{1}{(2\pi)^{\frac{n_\delta}{2}}} \exp\left(-\frac{1}{2} \delta^\top \delta\right)
\end{aligned}$$

where

$$q_{ij} = \frac{\exp(\Lambda_{ij})}{1 + \exp(\Lambda_{ij})}$$

$$\Lambda_{ij} = \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\{y_j/y_i, 1\} + \delta_4 \max\{y_i/y_j, 1\} + \delta_5^\top |X_i - X_j| + \delta_6 d_i^0 + \delta_7 d_j^0 + \delta_\xi |\xi_i - \xi_j|$$

$$q_{ji} = \frac{\exp(\Lambda_{ji})}{1 + \exp(\Lambda_{ji})}$$

$$\Lambda_{ji} = \delta_0 + \delta_1 d(P_j, P_i) + \delta_2 y_i + \delta_3 \max\{y_i/y_j, 1\} + \delta_4 \max\{y_j/y_i, 1\} + \delta_5^\top |X_j - X_i| + \delta_6 d_j^0 + \delta_7 d_i^0 + \delta_\xi |\xi_j - \xi_i|$$

Update δ given ξ , so use Metropolis-Hasting with the current value as mean and covariance matrix equal to the covariance matrix estimated from the exogenous model.

Conditional Distribution for ξ

$$\begin{aligned}
& P(\xi_i = k \mid \beta, \sigma^2, \delta, \xi_{-i}; MH, G, \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \\
& \propto P(MH_i \mid \beta, \sigma^2, \xi_i = k; G, \mathbf{y}, \mathbf{X}) \times P(G \mid \delta, \xi_i = k, \xi_{-i}; \mathbf{y}, \mathbf{X}, \mathbf{P}, G^0) \times P(\xi_i = k) \\
& = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right) \times \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}] \times P(\xi_i = k)
\end{aligned}$$

where k takes values 0 or 1.

Thus, given parameters $\beta, \sigma^2, \delta, \xi_{-i}$ and the data,

$$P(\xi_i = 1 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right) \times \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}] \times \frac{1}{2}$$

where

$$q_{ij} = \frac{\exp(\Lambda_{ij})}{1 + \exp(\Lambda_{ij})}$$

$$\Lambda_{ij} = \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\{y_j/y_i, 1\} + \delta_4 \max\{y_i/y_j, 1\} + \delta_5^\top |X_i - X_j| + \delta_6 d_i^0 + \delta_7 d_j^0 + \delta_\xi |1 - \xi_j|$$

$$\mu_i = \beta_0 + \beta_1 y_i + \beta_2 y_i^* + \beta_X^\top \mathbf{X}_i + \beta_\xi \cdot 1$$

Similarly,

$$P(\xi_i = 0 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(MH_i - \mu_i)^2}{2\sigma^2}\right) \times \prod_{i < j} [(q_{ij} \cdot q_{ji})^{g_{ij}} (1 - q_{ij} \cdot q_{ji})^{1-g_{ij}}] \times \frac{1}{2}$$

where

$$q_{ij} = \frac{\exp(\Lambda_{ij})}{1 + \exp(\Lambda_{ij})}$$

$$\Lambda_{ij} = \delta_0 + \delta_1 d(P_i, P_j) + \delta_2 y_j + \delta_3 \max\{y_j/y_i, 1\} + \delta_4 \max\{y_i/y_j, 1\} + \delta_5^\top |X_i - X_j| + \delta_6 d_i^0 + \delta_7 d_j^0 + \delta_\xi |0 - \xi_j|$$

$$\mu_i = \beta_0 + \beta_1 y_i + \beta_2 y_i^* + \beta_X^\top \mathbf{X}_i + \beta_\xi \cdot 0$$

Normalize it,

$$\bar{P}_i = \frac{P(\xi_i = 1 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data)}{P(\xi_i = 1 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data) + P(\xi_i = 0 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data)}$$

However, it's challenging to get $\bar{P}_i$ directly due to the likelihood of the network formation model approaching 0 when n increases. Let $P(\xi_i = 0 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data) = P(\xi_i = 0 \mid \dots)$ and $P(\xi_i = 1 \mid \beta, \sigma^2, \delta, \xi_{-i}; Data) = P(\xi_i = 1 \mid \dots)$. Then we can rewrite the probability as,

$$\bar{P}_i = \frac{1}{1 + \exp(\log(P(\xi_i = 0 \mid \dots)) - \log(P(\xi_i = 1 \mid \dots)))}$$

Then we can sample ξ_i from a Bernoulli distribution ¹⁷,

$$\xi_i \mid \beta, \sigma^2, \delta; Data \sim \text{Bernoulli}(\bar{P}_i)$$

Update ξ_i using Gibbs since we know exactly the posterior, and update each ξ_i (from $i = 1$ to N) one at a time independently.

¹⁷So each ξ_i , $i = 1, \dots, n$, follows a Bernoulli distribution with probability given by $\bar{P}_i$.

B Data Definitions and Statistics

Table B.1: Communication Networks: Descriptive Statistics

Year 1	Weak Ties	Q1	Strong Ties	Q3
Number of connections	1265	1009	669	374
Average number of connections	8	7	4	3
Observations	298	291	285	249

Year 2	Weak Ties	Q1	Strong Ties	Q3
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Note: Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table B.2: Degree of Separation in Communication Networks, Year 2

Degree of separation	Weak Ties	Q1	Strong Ties	Q3
1	1221	906	602	305
2	7707	4414	2053	543
3	20520	13038	5559	909
4	11915	15352	10081	1403
5	2261	6275	10553	1927
6	308	1369	6334	2337
7	23	239	2183	2399
8	1	32	507	2264
9	0	2	76	1822
10	0	0	7	1266
11	0	0	0	806
12	0	0	0	457
13	0	0	0	250
14	0	0	0	140
15	0	0	0	61
16	0	0	0	25
17	0	0	0	4
18	0	0	0	0
∞	595	2924	6596	27633

Note: This table shows the number of pairs of individuals with a given degree of separation in networks with different thresholds. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively. The degree of separation is defined as the number of edges in the shortest path between two individuals. For example, in network with weak ties, there are 1221 pairs of individuals with a degree of separation of 1, meaning they are directly connected.

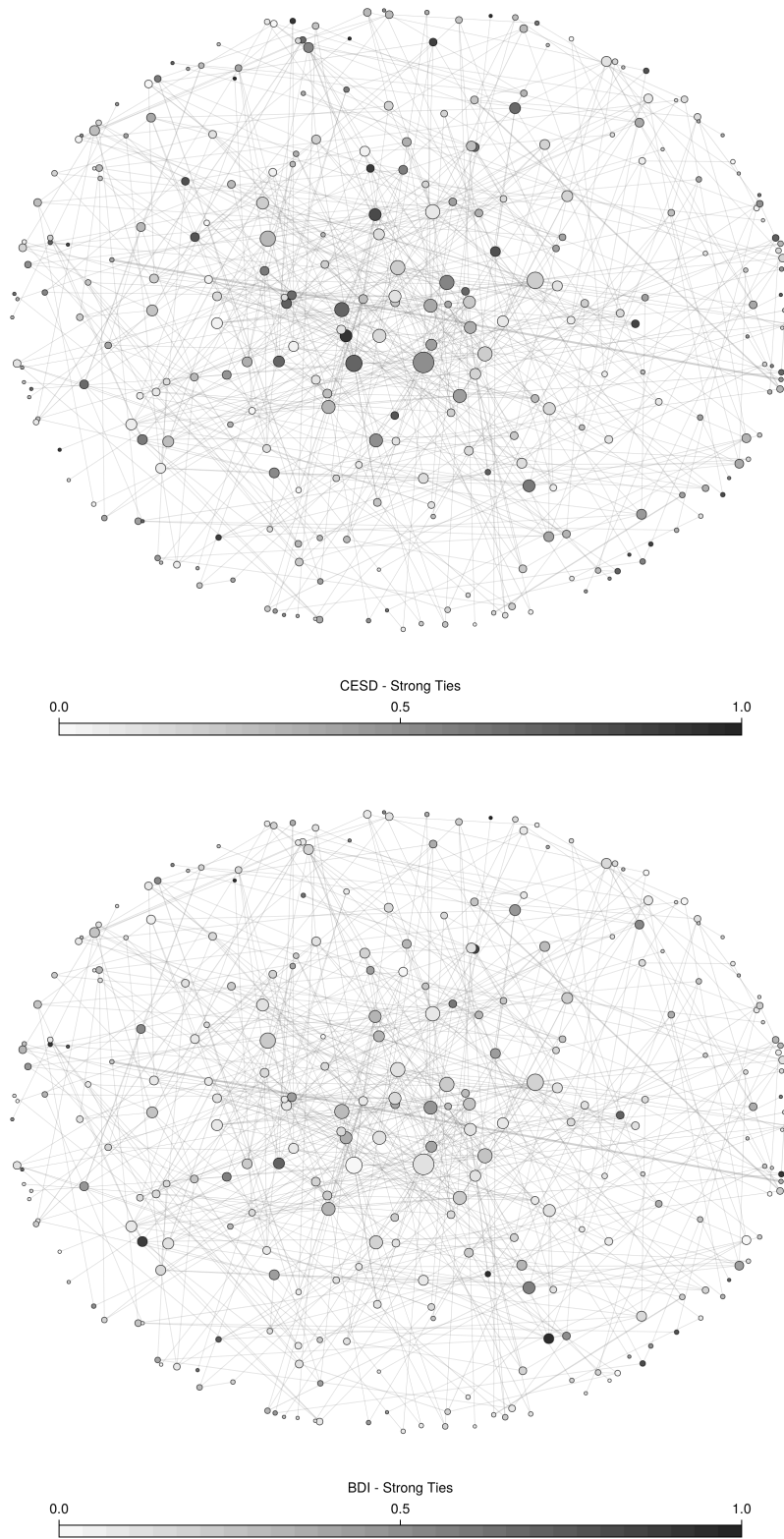


Figure B.1: Mental Health Distribution in Networks with Strong Ties

Note: This figure presents the distribution of various mental health metrics across social networks at Year 2, where darker colors represent higher values. The size of each node represents the number of connections, and the thickness of the lines between nodes represents the strength of communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Isolated individuals are removed from the figure.

Table B.3: Income Level Categories in NetHealth

Income level	Nominal income	Parents' income range per year
1	\$12,500	less than \$25,000
2	\$37,500	\$25,000 – \$49,999
3	\$62,500	\$50,000 – \$74,999
4	\$87,500	\$75,000 – \$99,999
5	\$125,000	\$100,000 – \$149,999
6	\$175,000	\$150,000 – \$199,999
7	\$225,000	\$200,000 – \$249,999
8	\$275,000	\$250,000 or more

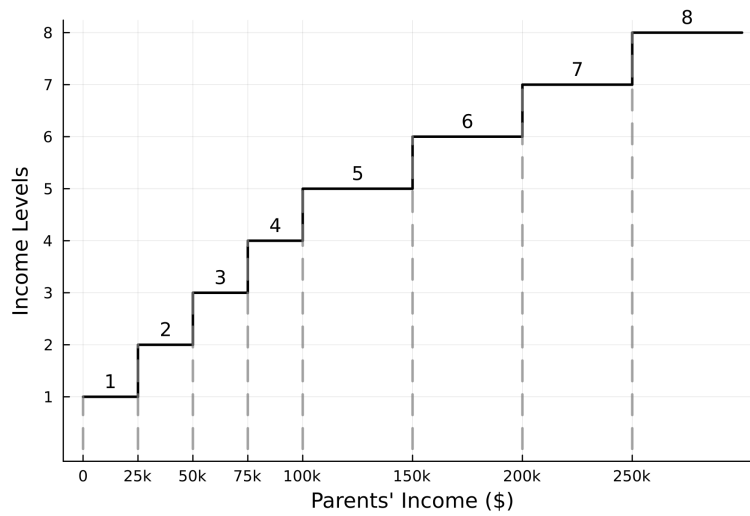


Figure B.2: Income Level in NetHealth

Table B.4: Center for Epidemiologic Studies Depression Scale (CES-D) Items

Item	Statement
1	I was bothered by things that usually don't bother me
2	I did not feel like eating; my appetite was poor
3	I felt that I could not shake off the blues even with help
4	I felt that I was just as good as other people
5	I had trouble keeping my mind on what I was doing
6	I felt depressed
7	I felt that everything I did was an effort
8	I felt hopeful about the future
9	I thought my life had been a failure
10	I felt fearful
11	My sleep was restless
12	I was happy
13	I talked less than usual
14	I felt lonely
15	People were unfriendly
16	I enjoyed life
17	I had crying spells
18	I felt sad
19	I felt that people disliked me
20	I could not "get going"

Note: Response options were scored based on frequency during the past week: 0 = Rarely or none of the time (less than 1 day); 1 = Some or a little of the time (1–2 days); 2 = Occasionally or a moderate amount of time (3–4 days); 3 = Most or all of the time (5–7 days).

Table B.5: Beck Depression Inventory (BDI) Items

Item	Symptom/Feeling
1	Sadness
2	Pessimism
3	Sense of failure
4	Dissatisfaction
5	Guilt
6	Sense of punishment
7	Self-dislike
8	Self-criticism
9	Suicidal thoughts/wishes
10	Crying
11	Irritability
12	Social withdrawal/loss of interest
13	Indecisiveness
14	Body image change
15	Work difficulties
16	Insomnia
17	Fatigability
18	Loss of appetite
19	Weight loss
20	Somatic preoccupation
21	Loss of libido

Note: Each item consists of 4–6 statements rated from 0–3, reflecting increasing symptom severity. Respondents select the statement(s) that best describe how they felt during the past two weeks.

Table B.6: Survey Question for Big Five Personality Traits in NetHealth

Big Five	I see myself as someone who...
Openness	Is original, comes up with new ideas. Is curious about many different things. Is ingenious, a deep thinker. Has an active imagination. Is inventive. Values artistic, aesthetic experiences. Prefers work that is routine. (R) Likes to reflect, play with ideas. Has few artistic interests. Is sophisticated in art, music, or literature.
Conscientiousness	Does a thorough job. Can be somewhat careless. (R) Is a reliable worker. Tends to be disorganized. (R) Tends to be lazy. (R) Perseveres until the task is finished. Does things efficiently. Makes plans and follows through with them. Is easily distracted. (R)
Extraversion	Is talkative. Is reserved. (R) Is full of energy. Generates a lot of enthusiasm. Tends to be quiet. (R) Has an assertive personality. Is sometimes shy, inhibited. (R) Is outgoing, sociable.
Agreeableness	Tends to find fault with others. (R) Is helpful and unselfish with others. Starts quarrels with others. (R) Has a forgiving nature. Is generally trusting. Can be cold and aloof. (R) Is considerate and kind to almost everyone. Is sometimes rude to others. (R) Likes to cooperate with others.
Neuroticism	Is depressed, blue. Is relaxed, handles stress well. (R) Can be tense. Worries a lot. Is emotionally stable, not easily upset. (R) Can be moody. Remains calm in tense situations. (R) Gets nervous easily.

Note: Reversed items are marked with (R).

Table B.7: Connection Dynamics in Net Health

		Year 2	
Year 1	n = 299	Connected	Not connected
		710 (1.59%)	555 (1.25%)
	Connected	710 (1.59%)	555 (1.25%)
	Not connected	511 (1.15%)	42775 (96.01%)

Note: This table reports pairwise connection status across two years in the Net Health dataset. Each cell shows the number of dyads and their share of all possible pairs. Rows indicate Year 1 status and columns indicate Year 2 status (n = 299 individuals).

Table B.8: Key Variable Constructions in NetHealth

Variables	Year 1		Year 2	
	Wave 1	Wave 2	Wave 3	Wave 4
CES-D				✓
BDI			✓	
Income	✓			
Big Five Personality	✓	✓		
Gender	✓			
Degree Aspiration			✓	

Note: This table reports the timing of data collection for the variables employed in this study. For the Big 5 personality traits, the average across the two waves is employed; if one wave is missing, the value from the other is retained.

Table B.9: Descriptive Statistics

	Mean	Std. Dev.	Min	Max
<i>Mental Health (MH)</i>				
CESD	13.645	9.692	0	49
BDI	7.100	6.753	0	46
<i>Key Variables</i>				
Parental income	16.626	8.886	1.250	27.500
Openness	0.000	1.000	-3.117	2.429
Conscientiousness	0.000	1.000	-2.905	2.489
Extraversion	0.000	1.000	-2.326	2.204
Agreeableness	0.000	1.000	-3.596	2.460
Neuroticism	0.000	1.000	-2.401	2.364
Gender	0.472	0.500	0	1
Degree Aspiration: Bachelor	0.328	0.470	0	1
Degree Aspiration: Master	0.288	0.453	0	1
Degree Aspiration: PhD/EdD	0.134	0.341	0	1
Degree Aspiration: MD/DO/DDS/DVM	0.177	0.383	0	1
Degree Aspiration: JD	0.070	0.256	0	1
Degree Aspiration: Other	0.003	0.058	0	1

Note: Income is expressed in units of ten thousand US dollars. Personality traits are standardized. Gender takes 1 if participants are male and 0 if female.

C Additional Results and Counterfactual Exercises

Table C.1: Empirical Results: Individual Personality Model

Linking utility: $u_i(j)$	Weak Ties			Q1			Strong Ties			Q3		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
constant	-2.158*** (0.137)	-2.203*** (0.083)	-2.170*** (0.119)	-2.354*** (0.112)	-2.408*** (0.094)	-2.402*** (0.096)	-2.504*** (0.123)	-2.429*** (0.110)	-2.489*** (0.113)	-2.429*** (0.156)	-2.326*** (0.157)	-2.379*** (0.158)
O_i	-0.097 (0.143)	-0.088 (0.185)	-0.065 (0.198)	-0.020 (0.198)	-0.118 (0.083)	-0.081 (0.162)	-0.008 (0.206)	-0.283** (0.112)	-0.114 (0.283)	-0.011 (0.151)	-0.100 (0.181)	-0.001 (0.220)
C_i	0.126 (0.172)	0.087 (0.074)	-0.065 (0.116)	0.032 (0.163)	-0.017 (0.080)	-0.025 (0.087)	-0.063 (0.194)	0.243** (0.114)	0.093 (0.258)	0.112 (0.168)	0.211 (0.226)	-0.034 (0.331)
E_i	0.081 (0.167)	-0.003 (0.077)	0.032 (0.082)	0.086 (0.282)	0.013 (0.101)	0.073 (0.112)	-0.085 (0.165)	0.081 (0.142)	0.021 (0.145)	0.293 (0.187)	0.346 (0.322)	-0.068 (0.458)
A_i	0.032 (0.202)	-0.053 (0.098)	-0.097 (0.103)	-0.019 (0.129)	0.051 (0.107)	0.008 (0.097)	0.020 (0.123)	0.117 (0.112)	0.050 (0.142)	0.021 (0.132)	0.020 (0.150)	0.032 (0.155)
N_i	-0.150 (0.169)	-0.131** (0.068)	-0.074 (0.133)	0.038 (0.097)	-0.067 (0.090)	-0.036 (0.136)	-0.003 (0.199)	-0.108 (0.126)	-0.033 (0.162)	-0.113 (0.139)	-0.033 (0.165)	0.069 (0.179)
O_j	0.085 (0.134)	0.075 (0.179)	0.047 (0.193)	0.011 (0.194)	0.104 (0.080)	0.060 (0.161)	-0.017 (0.198)	0.262** (0.110)	0.092 (0.280)	-0.011 (0.151)	0.075 (0.182)	-0.018 (0.217)
C_j	-0.078 (0.185)	-0.043 (0.075)	0.107 (0.113)	-0.022 (0.167)	0.029 (0.081)	0.039 (0.088)	0.076 (0.197)	-0.234** (0.118)	-0.087 (0.261)	-0.129 (0.172)	-0.231 (0.231)	0.010 (0.332)
E_j	-0.011 (0.166)	0.081 (0.079)	0.044 (0.081)	-0.018 (0.278)	0.050 (0.103)	-0.007 (0.111)	0.145 (0.161)	-0.024 (0.145)	0.035 (0.147)	-0.277 (0.194)	-0.333 (0.328)	0.083 (0.458)
A_j	0.003 (0.213)	0.079 (0.092)	0.130 (0.101)	0.059 (0.125)	-0.014 (0.107)	0.025 (0.100)	0.017 (0.123)	-0.080 (0.111)	-0.015 (0.142)	0.036 (0.130)	0.042 (0.147)	0.029 (0.158)
N_j	0.157 (0.157)	0.138** (0.065)	0.076 (0.130)	-0.034 (0.100)	0.062 (0.090)	0.032 (0.138)	0.028 (0.196)	0.128 (0.123)	0.052 (0.162)	0.203 (0.139)	0.124 (0.167)	0.023 (0.180)
y_j	0.005 (0.003)	0.005 (0.003)	0.005 (0.004)	0.003 (0.004)	0.004 (0.004)	0.004 (0.004)	0.007 (0.004)	0.006 (0.004)	0.008 (0.004)	0.002 (0.006)	0.001 (0.006)	0.003 (0.006)
upward comparison (difference)	-0.030* (0.016)	-0.015*** (0.005)		-0.015 (0.042)	-0.013** (0.005)		0.021 (0.050)	-0.010 (0.007)		-0.069*** (0.018)	-0.021** (0.010)	
downward comparison (difference)	0.021 (0.022)		-0.012** (0.005)	0.012 (0.042)		-0.009 (0.006)	-0.013 (0.047)		-0.003 (0.008)	0.067** (0.026)		-0.016 (0.011)
gender difference	-0.423*** (0.035)	-0.396*** (0.041)	-0.396*** (0.041)	-0.420*** (0.044)	-0.418*** (0.043)	-0.440*** (0.045)	-0.483*** (0.052)	-0.468*** (0.052)	-0.469*** (0.053)	-0.520*** (0.074)	-0.512*** (0.075)	-0.509*** (0.074)
degree aspiration difference	-0.138*** (0.052)	-0.131*** (0.044)	-0.148*** (0.043)	-0.148*** (0.048)	-0.120*** (0.047)	-0.130*** (0.047)	-0.151*** (0.055)	-0.141** (0.056)	-0.144** (0.056)	-0.145* (0.079)	-0.145* (0.079)	-0.139* (0.079)
own popularity	0.146*** (0.028)	0.172*** (0.015)	0.095 (0.116)	0.099 (0.095)	0.227*** (0.019)	0.116 (0.154)	0.032 (0.155)	0.274*** (0.040)	0.146 (0.189)	0.418*** (0.070)	0.348* (0.178)	0.057 (0.287)
peer's popularity	-0.051* (0.028)	-0.074*** (0.014)	0.000 (0.118)	0.022 (0.095)	-0.099*** (0.017)	0.010 (0.154)	0.123 (0.158)	-0.120*** (0.037)	0.008 (0.189)	-0.190*** (0.065)	-0.125 (0.175)	0.163 (0.287)
Number of connections		1221			906			602			305	
Average connections		8			6			4			2	
Number of isolated nodes		2			3			16			66	
Observations		297			296			283			233	

Standard errors in parentheses; HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the exogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. The marginal effects of own personalities and peer's personalities on the linking utility are shown in O_i , C_i , E_i , A_i , N_i for individual i, and O_j , C_j , E_j , A_j , N_j for peer j. y_j denotes the marginal effects of peer's income on the linking utility. We calculate the difference between peer's gender and Euclidean distance of degree intentions. Popularity is the number of connections of individual i. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table C.2: Empirical Results: Personality Similarity Model

Linking utility: $u_i(j)$	Weak Ties			Q1			Strong Ties			Q3		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
constant	-2.228*** (0.098)	-2.217*** (0.102)	-2.227*** (0.094)	-2.427*** (0.107)	-2.399*** (0.111)	-2.398*** (0.106)	-2.450*** (0.128)	-2.391*** (0.134)	-2.406*** (0.126)	-2.338*** (0.174)	-2.234*** (0.176)	-2.293*** (0.172)
$ O_i - O_j $	-0.031 (0.026)	-0.024 (0.022)	-0.029 (0.021)	-0.024 (0.024)	-0.026 (0.025)	-0.025 (0.025)	-0.027 (0.030)	-0.022 (0.029)	-0.022 (0.030)	0.004 (0.042)	-0.003 (0.041)	0.000 (0.041)
$ C_i - C_j $	0.006 (0.022)	-0.004 (0.021)	-0.004 (0.021)	-0.001 (0.023)	0.001 (0.024)	-0.005 (0.024)	-0.032 (0.031)	-0.029 (0.030)	-0.030 (0.032)	-0.028 (0.043)	-0.032 (0.042)	-0.038 (0.040)
$ E_i - E_j $	-0.017 (0.022)	-0.015 (0.022)	-0.019 (0.023)	-0.015 (0.025)	-0.017 (0.027)	-0.016 (0.028)	-0.030 (0.033)	-0.033 (0.031)	-0.032 (0.032)	-0.114*** (0.046)	-0.113** (0.044)	-0.112** (0.044)
$ A_i - A_j $	-0.037 (0.021)	-0.032 (0.023)	-0.032 (0.022)	-0.024 (0.027)	-0.029 (0.026)	-0.026 (0.026)	-0.020 (0.032)	-0.022 (0.032)	-0.024 (0.031)	0.016 (0.043)	0.016 (0.041)	0.019 (0.041)
$ N_i - N_j $	-0.035 (0.024)	-0.035 (0.023)	-0.032 (0.022)	-0.035 (0.024)	-0.035 (0.025)	-0.035 (0.028)	-0.027 (0.031)	-0.032 (0.032)	-0.032 (0.031)	0.024 (0.044)	0.017 (0.042)	0.019 (0.043)
y_j	0.006** (0.003)	0.007** (0.003)	0.007*** (0.003)	0.005 (0.003)	0.005 (0.004)	0.005* (0.003)	0.007 (0.004)	0.007 (0.005)	0.007 (0.004)	0.002 (0.006)	0.000 (0.006)	0.004 (0.006)
upward comparison (difference)	-0.033*** (0.012)	-0.013** (0.006)		-0.035*** (0.011)	-0.010* (0.006)		-0.011 (0.054)	-0.008 (0.008)		-0.080*** (0.016)	-0.021** (0.010)	
downward comparison (difference)	0.029** (0.016)		-0.011* (0.005)	0.035** (0.015)		-0.010* (0.006)	0.019 (0.055)		-0.005 (0.008)	0.087*** (0.026)		-0.020** (0.009)
gender difference	-0.385*** (0.042)	-0.387*** (0.037)	-0.383*** (0.037)	-0.421*** (0.044)	-0.411*** (0.042)	-0.421*** (0.043)	-0.470*** (0.056)	-0.459*** (0.052)	-0.461*** (0.052)	-0.523*** (0.076)	-0.500*** (0.073)	-0.502*** (0.072)
degree aspiration difference	-0.137*** (0.043)	-0.130*** (0.043)	-0.130*** (0.038)	-0.130*** (0.046)	-0.119*** (0.047)	-0.121*** (0.050)	-0.141** (0.058)	-0.145*** (0.058)	-0.144** (0.054)	-0.135* (0.078)	-0.132* (0.076)	-0.137* (0.072)
own popularity	0.154*** (0.014)	0.090 (0.106)	0.018 (0.106)	0.215*** (0.018)	0.122 (0.150)	0.015 (0.151)	0.148 (0.180)	0.149 (0.191)	0.146 (0.191)	0.448*** (0.069)	0.382* (0.197)	-0.178* (0.178)
peer's popularity	-0.047*** (0.013)	0.017 (0.106)	0.089 (0.106)	-0.077*** (0.015)	0.014 (0.149)	0.121 (0.151)	0.020 (0.181)	0.015 (0.189)	0.019 (0.190)	-0.225*** (0.063)	-0.167 (0.193)	0.390* (0.182)
Number of connections		1221			906			602			305	
Average connections		8			6			4			2	
Number of isolated nodes		2			3			16			66	
Observations		297			296			283			233	

Standard errors in parentheses; HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. The marginal effects of personality similarities on the linking utility are shown in $|O_i - O_j|$, $|C_i - C_j|$, $|E_i - E_j|$, $|A_i - A_j|$, $|N_i - N_j|$ for a pair $\{i, j\}$. y_j denotes the marginal effects of peer's income on the linking utility. We calculate the difference between peer's gender and Euclidean distance of degree intentions. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table C.3: Empirical Results: Euclidean Distance Model

Linking utility: $u_i(j)$	Weak Ties			Q1			Strong Ties			Q3		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
constant	-2.229*** (0.108)	-2.207*** (0.102)	-2.223*** (0.091)	-2.426*** (0.112)	-2.392*** (0.120)	-2.405*** (0.113)	-2.458*** (0.133)	-2.438*** (0.141)	-2.421*** (0.130)	-2.361*** (0.183)	-2.250*** (0.174)	-2.314*** (0.178)
$d(P_i, P_j)$	-0.049*** (0.019)	-0.040** (0.018)	-0.043** (0.017)	-0.036* (0.021)	-0.035 (0.022)	-0.038* (0.022)	-0.045* (0.026)	-0.048* (0.025)	-0.048* (0.026)	-0.027 (0.036)	-0.032 (0.035)	-0.032 (0.034)
y_j	0.007** (0.003)	0.007* (0.003)	0.008** (0.003)	0.005 (0.004)	0.005 (0.004)	0.005 (0.004)	0.007 (0.004)	0.008* (0.004)	0.007* (0.004)	0.001 (0.006)	-0.000 (0.006)	0.003 (0.006)
upward comparison (difference)	-0.006 (0.034)	-0.013** (0.005)		-0.034*** (0.011)	-0.011* (0.007)		0.027 (0.056)	-0.004 (0.009)		0.031 (0.080)	-0.022** (0.009)	
downward comparison (difference)	0.001 (0.033)		-0.012*** (0.005)	0.033*** (0.015)		-0.008 (0.006)	-0.018 (0.055)		-0.005 (0.008)	-0.026 (0.080)		-0.019** (0.010)
gender difference	-0.395*** (0.041)	-0.392*** (0.036)	-0.388*** (0.036)	-0.431*** (0.043)	-0.422*** (0.042)	-0.424*** (0.044)	-0.471*** (0.053)	-0.459*** (0.052)	-0.458*** (0.052)	-0.517*** (0.075)	-0.502*** (0.072)	-0.505*** (0.074)
degree aspiration difference	-0.118*** (0.042)	-0.130*** (0.040)	-0.128*** (0.042)	-0.117*** (0.048)	-0.118*** (0.049)	-0.117*** (0.048)	-0.143** (0.056)	-0.147*** (0.057)	-0.142*** (0.055)	-0.144* (0.080)	-0.137* (0.076)	-0.136* (0.076)
own popularity	0.089 (0.098)	0.090 (0.108)	0.018 (0.104)	0.217*** (0.016)	0.119 (0.151)	0.122 (0.151)	0.021 (0.181)	0.015 (0.190)	0.148 (0.193)	-0.002 (0.331)	0.431** (0.133)	-0.176* (0.195)
peer's popularity	0.018 (0.097)	0.016 (0.107)	0.088 (0.105)	-0.080*** (0.015)	0.016 (0.151)	0.014 (0.150)	0.147 (0.181)	0.150 (0.191)	0.017 (0.193)	0.229 (0.331)	-0.212** (0.128)	0.393* (0.196)
Number of connections		1221			906			602			305	
Average connections		8			6			4			2	
Number of isolated nodes		2			3			16			66	
Observations		297			296			283			233	

Standard errors in parentheses; HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the exogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. $d(P_i, P_j)$ denotes the euclidean distance of personality traits between individual i and peer j. y_j denotes the marginal effects of peer's income on the linking utility. We calculate the difference between peer's gender and Euclidean distance of degree intentions. Popularity is the number of connections of individual i. Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table C.4: Empirical Results: Exogenous Network Formation

Linking utility: $u_i(j)$	Intensity Threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-2.1854*** (0.1059)	-2.4114*** (0.1361)	-2.5052*** (0.1590)	-2.4873*** (0.2201)
$ O_i - O_j $	-0.0265 (0.0203)	-0.0258 (0.0257)	-0.0226 (0.0291)	0.0071 (0.0413)
$ C_i - C_j $	-0.0068 (0.0199)	-0.0003 (0.0245)	-0.0297 (0.0305)	-0.0313 (0.0435)
$ E_i - E_j $	-0.0152 (0.0242)	-0.0123 (0.0273)	-0.0288 (0.0303)	-0.1129*** (0.0443)
$ A_i - A_j $	-0.0321 (0.0238)	-0.0246 (0.0260)	-0.0207 (0.0300)	0.0189 (0.0419)
$ N_i - N_j $	-0.0316 (0.0218)	-0.0362 (0.0266)	-0.0310 (0.0313)	0.0201 (0.0426)
y_j	0.0049 (0.0031)	0.0040 (0.0037)	0.0075 (0.0045)	0.0050 (0.0062)
upward comparison	0.0293 (0.1020)	-0.0353 (0.1254)	-0.1483*** (0.0282)	-0.2206*** (0.0454)
downward comparison	-0.0476 (0.1008)	0.0394 (0.1270)	0.1973*** (0.0599)	0.2776*** (0.0831)
gender difference	-0.3941*** (0.0392)	-0.4235*** (0.0441)	-0.4666*** (0.0515)	-0.5150*** (0.0728)
degree aspiration difference	-0.1246*** (0.0398)	-0.1193*** (0.0477)	-0.1435*** (0.0576)	-0.1393* (0.0761)
own popularity	0.0179 (0.1012)	0.1173 (0.1425)	0.2809*** (0.0304)	0.4712*** (0.0641)
peer's popularity	0.0875 (0.1009)	0.0180 (0.1425)	-0.1144*** (0.0266)	-0.2503*** (0.0581)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Standard errors in parentheses

*HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.*

Note: Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. The marginal effects of personality similarities on the linking utility are shown in $|O_i - O_j|$, $|C_i - C_j|$, $|E_i - E_j|$, $|A_i - A_j|$, $|N_i - N_j|$ for a pair $\{i, j\}$. y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the number of connections of individual i . Weak ties are defined as connections where peers have exchanged at least one instance of two-way communication. Strong ties are defined as connections where peers have exchanged at least 55 messages (median). Q1 and Q3 refer to the first and third quartiles of the message exchange distribution, 14 and 228 messages respectively.

Table C.5: Endogenous Network: Effects of Peer's Income on BDI

(A) Mental Health and Peer's Income				
Dependent:BDI	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.6860*** (0.9591)	3.1191*** (0.9061)	3.0361*** (0.8921)	3.3103*** (0.8419)
income	0.0083 (0.0446)	0.0239 (0.0437)	0.0250 (0.0433)	0.1064** (0.0458)
average peers' income	0.2004*** (0.0656)	0.1591*** (0.0607)	0.1302** (0.0549)	0.0284 (0.0543)
gender	-0.6450 (0.6384)	-0.5639 (0.6404)	-0.4992 (0.5877)	-0.4368 (0.6295)
ξ_i	1.2090 (1.7708)	0.9840 (1.5142)	1.9164* (0.9551)	1.5980* (0.9281)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.9627*** (0.1293)	-2.1961*** (0.1775)	-2.2442*** (0.1821)	-2.4773*** (0.2556)
$ O_i - O_j $	-0.0194 (0.0223)	-0.0202 (0.0270)	-0.0186 (0.0332)	0.0045 (0.0529)
$ C_i - C_j $	-0.0005 (0.0227)	0.0054 (0.0247)	-0.0240 (0.0327)	-0.0298 (0.0536)
$ E_i - E_j $	0.0105 (0.0242)	-0.0051 (0.0264)	-0.0224 (0.0335)	-0.1114** (0.0541)
$ A_i - A_j $	-0.0285 (0.0227)	-0.0204 (0.0265)	-0.0220 (0.0319)	0.0221 (0.0527)
$ N_i - N_j $	-0.0383 (0.0249)	-0.0391 (0.0287)	-0.0296 (0.0344)	0.0133 (0.0510)
y_j	0.0048 (0.0035)	0.0055 (0.0042)	0.0080* (0.0047)	0.0052 (0.0072)
upward comparison	0.1029** (0.0510)	-0.0228 (0.1349)	-0.1440*** (0.0267)	-0.2332*** (0.0554)
downward comparison	-0.1066*** (0.0265)	0.0480 (0.1353)	0.2099*** (0.0645)	0.2978*** (0.0962)
gender difference	-0.4014*** (0.0373)	-0.4216*** (0.0479)	-0.4706*** (0.0512)	-0.5161*** (0.0873)
degree aspiration difference	-0.1282*** (0.0404)	-0.1287*** (0.0466)	-0.1545*** (0.0575)	-0.1436 (0.0954)
own popularity	-0.0493*** (0.0092)	0.1157 (0.1312)	0.2674*** (0.0297)	0.4759*** (0.0907)
peer's popularity	0.1575*** (0.0115)	0.0222 (0.1310)	-0.0980*** (0.0248)	-0.2540** (0.0829)
$ \xi_i - \xi_j $	-0.8967*** (0.0500)	-0.9770*** (0.0575)	-1.0244*** (0.0968)	-0.0544 (0.1858)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. The marginal effects of personality similarities on the linking utility are shown in $|O_i - O_j|$, $|C_i - C_j|$, $|E_i - E_j|$, $|A_i - A_j|$, $|N_i - N_j|$ for a pair $\{i, j\}$. y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the total number of messages of individual i at year 1.

Table C.6: Endogenous Network: Effects of Peer's Income on CESD

(A) Mental Health and Peer's Income				
Dependent:CESD	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	2.7336*** (0.9631)	3.0708*** (0.9508)	3.2341*** (0.9406)	2.9365*** (0.9332)
income	-0.0321 (0.0676)	0.0156 (0.0675)	0.0379 (0.0645)	0.1254* (0.0723)
average peers' income	0.5581*** (0.0889)	0.4710*** (0.0819)	0.4260*** (0.0781)	0.3308*** (0.0805)
gender	-0.1523 (0.7714)	-0.0306 (0.7819)	0.0027 (0.7610)	0.0405 (0.7862)
ξ_i	1.3803 (1.7870)	1.5375 (1.6322)	1.8313* (1.0051)	1.4617 (0.9976)
(B) Network Formation				
Linking utility: $u_i(j)$	Intensity threshold			
	Weak Ties	Q1	Strong Ties	Q3
constant	-1.9802*** (0.1246)	-2.1964*** (0.1680)	-2.2574*** (0.1709)	-2.4249*** (0.2720)
$ O_i - O_j $	-0.0188 (0.0218)	-0.0170 (0.0248)	-0.0194 (0.0319)	0.0026 (0.0496)
$ C_i - C_j $	0.0028 (0.0231)	0.0052 (0.0249)	-0.0287 (0.0304)	-0.0348 (0.0527)
$ E_i - E_j $	0.0084 (0.0240)	-0.0052 (0.0278)	-0.0221 (0.0332)	-0.1093* (0.0562)
$ A_i - A_j $	-0.0260 (0.0221)	-0.0263 (0.0271)	-0.0188 (0.0325)	0.0256 (0.0537)
$ N_i - N_j $	-0.0373 (0.0240)	-0.0372 (0.0266)	-0.0280 (0.0331)	0.0153 (0.0508)
y_j	0.0049 (0.0035)	0.0056 (0.0041)	0.0084* (0.0046)	0.0039 (0.0077)
upward comparison	0.1140** (0.0419)	-0.0312 (0.1171)	-0.1450*** (0.0284)	-0.2282*** (0.0612)
downward comparison	-0.1148*** (0.0231)	0.0392 (0.1190)	0.2178*** (0.0684)	0.2818*** (0.1050)
gender difference	-0.3940*** (0.0386)	-0.4313*** (0.0438)	-0.4797*** (0.0587)	-0.5262*** (0.0996)
degree aspiration difference	-0.1307*** (0.0429)	-0.1059** (0.0493)	-0.1571** (0.0587)	-0.1439 (0.0916)
own popularity	-0.0503*** (0.0096)	0.1157 (0.1313)	0.2670*** (0.0293)	0.4683*** (0.0825)
peer's popularity	0.1593*** (0.0118)	0.0233 (0.1304)	-0.0975*** (0.0256)	-0.2514*** (0.0714)
$ \xi_i - \xi_j $	-0.9079*** (0.0449)	-0.9647*** (0.0634)	-1.0315*** (0.1045)	-0.0375 (0.1001)
Number of connections	1221	906	602	305
Average number of connections	8	6	4	2
Observations	297	296	283	233

Posterior standard errors in parentheses

HPD (highest posterior density) interval excludes zero: * 90%, ** 95%, *** 99%.

Note: This table presents the posterior means and standard errors of the parameters in the endogenous network formation model. Communication networks are exogenous defined by the number of two-way communication records, intensity thresholds define the intensity of communication in the network. The marginal effects of personality similarities on the linking utility are shown in $|O_i - O_j|$, $|C_i - C_j|$, $|E_i - E_j|$, $|A_i - A_j|$, $|N_i - N_j|$ for a pair $\{i, j\}$. y_j denotes the marginal effects of peer's income on the linking utility. Upward comparison is defined by $\max\{y_j/y_i, 1\}$, downward comparison is defined by $\max\{y_i/y_j, 1\}$. Gender difference is calculated as the absolute difference between an individual's gender and their peer's gender. The degree aspiration difference dummy equals 1 if the two individuals have different degree aspirations and 0 if they have the same aspiration. Popularity is the total number of messages of individual i at year 1.

Table C.7: Counterfactual Results on Average Peer Income

Exercise	Endogenous			Exogenous (OLS)	
	$\overline{\Delta BDI}$	95% HPD	$\text{std}(\Delta BDI_i)$	ΔBDI	95% CI
Own Income	0.2924	$[-0.4294, 1.0213]$	0.1533	-0.6539	$[-1.4287, 0.1210]$
Others	1.0715	$[0.2109, 1.9859]$	0.1725	-1.1407	$[-2.3178, 0.0364]$
Peers	0.0373	$[0.0059, 0.0705]$	0.0315	-1.1407	$[-2.3178, 0.0364]$

Table C.8: Counterfactual Results on Income Comparison

Exercise	Endogenous			Exogenous (OLS)	
	$\overline{\Delta BDI}$	95% HPD	$\text{std}(\Delta BDI_i)$	ΔBDI	95% CI
Richer Others	0.1420	$[-0.0196, 0.3245]$	0.2399	-0.1326	$[-0.3954, 0.1301]$
Poorer Others	-0.1234	$[-0.2651, 0.0093]$	0.1996	-0.1817	$[-0.4517, 0.0884]$
Richer Peers	0.0034	$[-0.0045, 0.0122]$	0.0115	-0.1326	$[-0.3954, 0.1301]$
Poorer Peers	-0.0062	$[-0.0152, 0.0029]$	0.0112	-0.1817	$[-0.4517, 0.0884]$

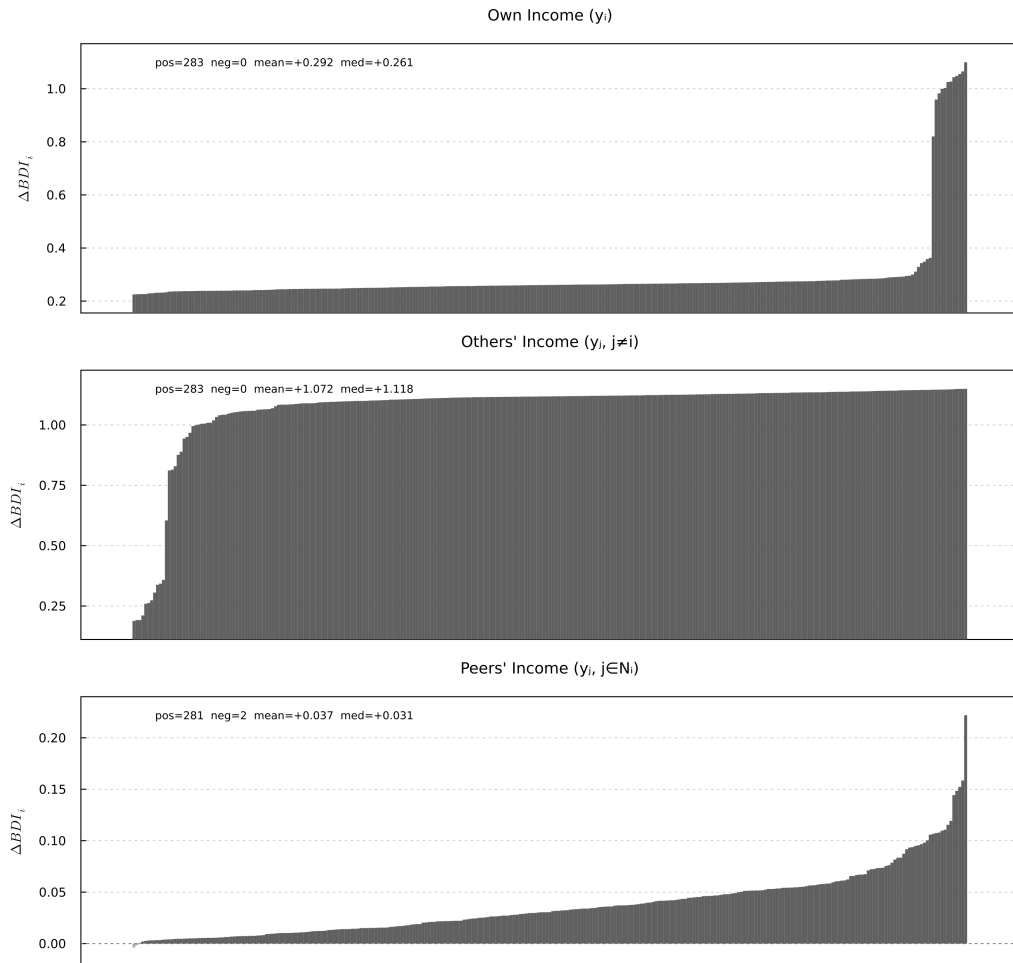


Figure C.1: Counterfactual Exercises on Average Peer Income at Individual Level

Notes: Each point reports the posterior mean change in BDI (ΔBDI_i) for a single individual under a one-standard-deviation income shock, using the average peer income specification.

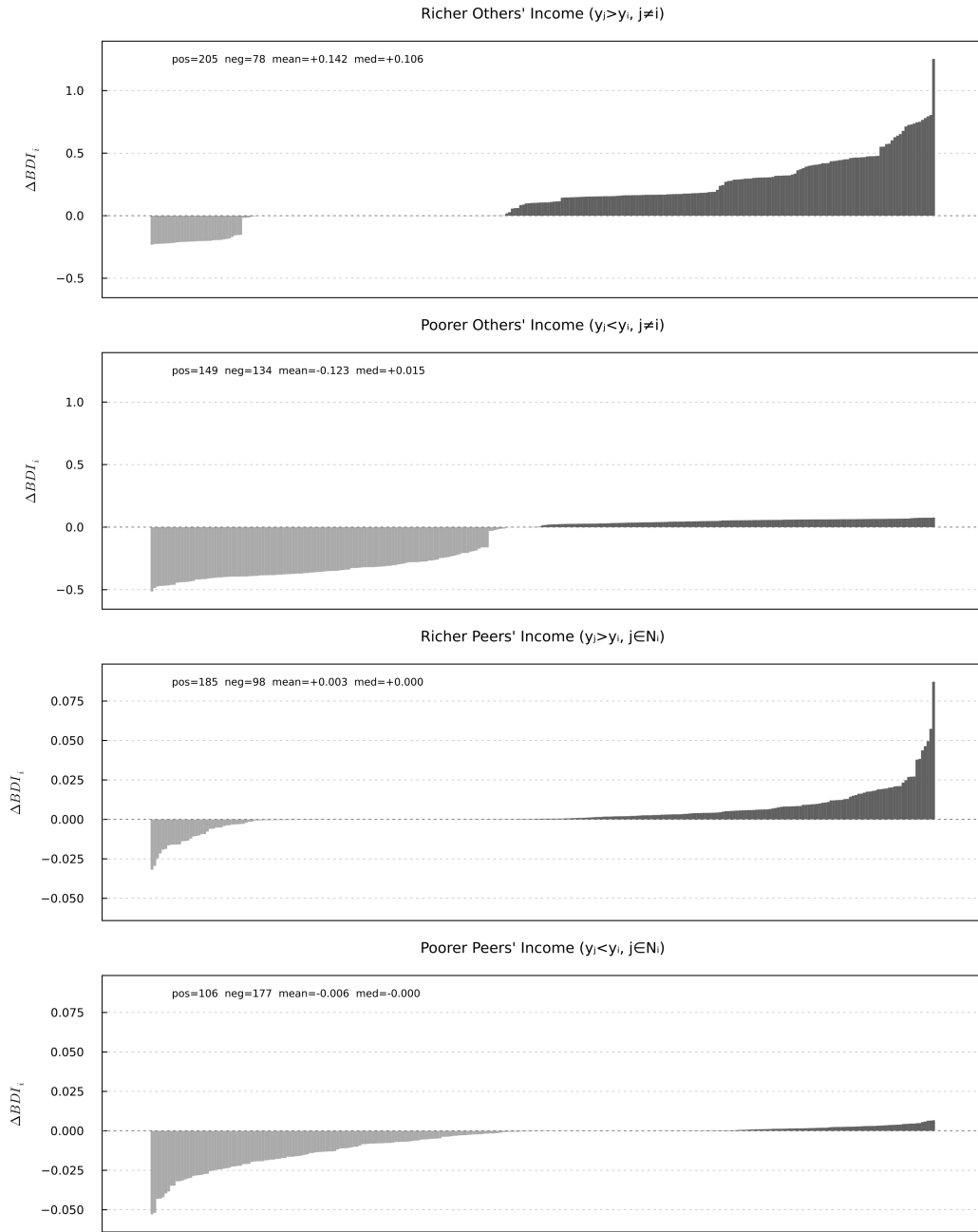


Figure C.2: Counterfactual Exercises on Income Comparison at Individual Level

Notes: Each point reports the posterior mean change in BDI (ΔBDI_i) for a single individual under a one-standard-deviation income shock, using the income comparison specification.