

Preregistration of Survey and Planned Analysis Contributing to “Measuring Social Media Network Effects: Evidence from Online Choice Experiments”

Authors: Sinan Aral, Seth Benzell, Avi Collis, Christos Nicolaides

Registered: 10/30/2023

Narrative Summary of Preregistration

We have already conducted a series of surveys through Lucid on individuals' demographics, connections on a digital platform, and their valuations of these connections and digital platforms.

In this document we are pre-registering a follow-up survey and analysis plan. The survey will collect information on 2,500 users of each of four popular social media platforms (Instagram, Facebook, Twitter/X, and LinkedIn), with a focus on their number of contacts, valuation of the platform, social media usage, and other socio-economic and demographic information.

The first question we are interested in is the relationship between the number of contacts and platform valuation. Our main specification will be a linear regression with an intercept term. In secondary specifications we will replace the number of contacts with the log of the number of contacts, and include additional linear controls.

We are also interested in how platform valuations vary with socio-economic, demographic and platform usage characteristics, as well as question framing. Our main analysis will first select covariates (and their interactions) with explanatory power using LASSO. We will then run a regression of platform valuation on the LASSO selected coefficients and interactions. As secondary specifications, we will estimate how the average platform valuation varies by demographically defined sub-groups.

To ensure responses to questions about platform valuations are incentive compatible, we will follow up with a small randomly selected subset of users to actually enforce the giving up the platform in exchange for a randomly selected amount of money promised

Research Questions for the New Collection and Analysis

1. How does platform valuation vary with the number of contacts on the platform
2. How does platform valuation vary with other user characteristics

Survey Details:

The survey instrument for each of the 4 platforms studied is attached.

The study will be conducted by Lucid Technologies, which promises a representative sample of the users of each platform. The survey of users of each platform will have respondents with the same distribution of age, gender, incomes, and race as those found by a 2021 Pew study of

social media use.¹

We will reject surveys that don't meet the following criteria until we have 2,500 high quality responses for each platform. These criteria are: (1) Respondents who are inconsistent in their responses to the amount of money they would accept to stop using the platform for a month (2) Respondents who take less than one minute to complete the survey (3) Any respondents who report also using the platform "Mystro", a made up social media platform we include as a question response as an attention check and (4) Respondents must report using the focal platform of the survey.

Our primary outcome of interest is the platform valuation of the user for a month. Approximately $\frac{1}{3}$ of each platform's users will be randomly assigned to one of three versions of the main platform valuation question. $\frac{1}{3}$ will have a maximum platform valuation of \$100, $\frac{1}{3}$ will have a maximum platform valuation of \$150, and $\frac{1}{3}$ will have a maximum platform valuation of \$200.

To ensure that survey takers take the survey seriously, we will make responses incentive compatible following Brynjolfsson et. al. (2019).² We will randomly select $1/Z$ users, within each group, to follow up with the following game, where "Z" is the top amount of dollars offered in their version of the platform valuation question. In the follow-up game, we will randomly select a number from the set $[0, 5, 10, \dots, Z]$ ³ to implicitly offer survey takers to stop using the platform for a month. Survey takers who reported a maximum acceptable offer greater than or equal to our randomly selected offer will be offered the chance to stop using the platform for one month in return for the randomly selected compensation. We will require participating survey takers to report their social media information, and will check compliance. At the end of a month, study participants who comply with their offers will be compensated.

Planned Specifications:

1. Regarding Research Question 1

a. Main Specification:

i. $V_i = \beta * C_i + \text{intercept}$

1. There will be 6 separate regressions, 1 each for Facebook and LinkedIn, 2 each for Twitter/X and Instagram for followers and followings, separately

b. Secondary Specifications - the following linear regressions:

i. $V_i = \beta * C_i + \gamma_1 * X_i + \text{intercept}$

1. This is the primary specification, but with all demographic and socio-economic controls included

ii. $V_i = \beta * C_i + \gamma_1 * Z_i + \text{intercept}$

1. This is the primary specification, but with an additional control for what version of the valuation question was specified

¹ <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>

² <https://www.pnas.org/doi/10.1073/pnas.1815663116>

³ I.e. The random number will be selected from the discrete uniform distribution $[0, Z]$, with only multiples of 5 being possibilities, each with a $(1/(Z/5 + 1))$ chance of each outcome.

- iii. $V_i = \beta * C_i + \gamma_1 * X_i + \gamma_2 * W_i + \gamma_3 * Z_i + \text{intercept}$
 - 1. This is the primary specification, but with all additional demographic, valuation question specification, and platform usage covariates
- iv. $V_i = \beta * \log(C_i) + \text{intercept}$
 - 1. This is the primary specification, but with the number of connections logged

2. Regarding Research Question 2

a. Main Specification (a):

- i. A regression of V_i on the parameters selected by a LASSO regression of V_i on:
 - 1. C_i (with followers and followings included as separate explanatory variables for Instagram and Twitter/X), $\log(C_i)$, X_i , W_i , Z_i , P_i and all 2 and 3-way interactions of these covariates
- ii. Main Specification (b):
 - 1. A variation on the main LASSO specification (immediately above) but excluding the connection count and platform usage covariates, i.e.:
 - 2. All the covariates in: X_i , Z_i , P_i and all 2 and 3-way interactions of these covariates

b. Secondary Specifications:

- i. We will estimate $E[V_i | \text{Platform, Single Ego Variable}]$ and $E[V_i | \text{Single Ego Variable}]$, for belonging to different categories on a demographic question: i.e. we will report the average V_i for members of different groups both across all platforms as well as on specific platforms.
 - 1. Gender
 - 2. Decade of Birth
 - 3. US Census Region
 - 4. Urbanity
 - 5. Ethnicity
 - 6. Income
 - 7. Political Views
 - 8. Job Search
 - 9. Relationship Status
 - 10. Household Size
 - 11. Extroversion Likert 1
 - 12. Extroversion Likert 2
 - 13. How many moves in past 10 years
 - 14. Years of time using the platform
 - a. Binned into:
 - i. <2 years
 - ii. 2-4 years

- iii. 4-8 years
 - iv. 8+ years
15. The following categories will only be evaluated for specific platforms. For each of these, unlike the above, categories are not mutually exclusive:
- a. Which Facebook Features Do You Use
 - b. Which LinkedIn Features Do You Use
 - c. Which Instagram Features Do You Use
 - d. Which Twitter/X Features Do You Use
16. Time spent using the platform in the last week

We will also explore the data for additional specifications with interesting patterns.

Definition of Variables:

- i = lowercase i indexes survey takers. Each survey taker answers questions about a single platform.
- V_i = Platform Valuation for survey taker i . This is defined as the largest selection they answer yes to in survey question beginning “In the following question, we will ask you which amounts of money (in US \$) you would accept to deactivate Facebook for 1 month.”
- Z_i = Which variation on the platform valuation question the respondent faces. Respondents will be randomly assigned to experiments with \$100, \$150, and \$200 upper bounds in valuation.
 - We will treat the \$100 version as the omitted category in regressions
- C_i = Self reported number of contacts of the ego on the platform of a given type. For LinkedIn this is contacts, for Facebook this is friends. For Instagram and Twitter/X there are both followers and followings and we will evaluate both separately (i.e. as explanatory variables in separate regression specifications).
- X_i = List of demographic and socio-economic covariates solicited of all survey takers. These are:
 - Gender
 - Male
 - Female
 - Other
 - The omitted category in regressions will be the most common response
 - Year of birth
 - This will be treated as a continuous variable
 - US Census Region
 - Northeast
 - Midwest
 - South
 - West
 - As determined by reported Zipcode

- The omitted category in regressions will be the most common response
- Regional urbanity
 - A continuous variable, %urban of the Zipcode as measured in the most recent US census
- Ethnicity
 - White
 - Hispanic or Latino
 - Black or African American
 - Native American or American Indian
 - Asian/ Pacific Islander
 - Other
 - The omitted category in regressions will be the most common response
- Household income
 - <\$25,000
 - \$25,000-\$49,999
 - \$50,000-\$99,999
 - \$100,000-\$149,999
 - \$150,000 or more
 - The omitted category in regressions will be "<25,000"
- Political Views
 - Extremely liberal
 - Liberal
 - Slightly liberal
 - Moderate
 - Slightly conservative
 - Conservative
 - Extremely conservative
 - The omitted category in regressions will be "Extremely liberal"
- Are you currently looking for a job? Please select the situation that best applies to you
 - Yes
 - No, but I looked in the past year
 - No, and I have not looked in the past year
 - The omitted category is "No, and I have not looked"
- What is your relationship status?
 - Single, looking for a relationship
 - Single, not looking for a relationship
 - In a serious relationship
 - Married or domestic partnership
 - Prefer not to say
 - Omitted category is "single not looking for a relationship"
- How many people are in your household

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7+
 - Omitted category is 1
- Likert scales for two questions designed to solicit extroversion
 - Categories: Strongly disagree, disagree, neither agree nor disagree, somewhat agree, strongly agree
 - Statements about yourself:
 - “I keep in the background”
 - “I talk to a lot of different people at parties”
 - Omitted categories will be “strongly agree” for the first, and “strongly disagree” for the second. In reporting results a scale may be inverted so they both “point” in the same direction.
- How many different places have you lived in the past 10 years
 - 1
 - 2
 - 3
 - 4
 - 5 or more
 - 1 will be the omitted category
- W_i = List of platform usage covariates solicited. This list varies slightly from platform to platform.
 - When did you start using the platform?
 - This will be treated as a continuous variable
 - How often did you use the platform in the last week:
 - Less than 10 minutes per day
 - 10–30 minutes per day
 - 31–60 minutes per day
 - 1–2 hours per day
 - 2–3 hours per day
 - More than 3 hours per day
 - Less than 10 minutes per day will be the omitted category
 - A series of platform specific questions on whether users use the following features.
 - Which platform (besides the focal platform in the study) do you use:
 - Facebook
 - Instagram
 - Twitter/X
 - LinkedIn

- TikTok
- Snapchat
- Which of the following features do you use?
 - Facebook List:
 - Joining Facebook Groups
 - Posting content on your News Feed
 - Facebook Marketplace
 - Playing games
 - Engaging with content from your friends on your News Feed
 - Organize and attend Events
 - Facebook live video streaming
 - Messenger
 - Following pages
 - Instagram List:
 - Post photos
 - Like and comment on photos
 - Post reels
 - Like and comment on reels
 - Post stories
 - Like and comment on stories
 - Instagram live
 - Direct message
 - LinkedIn List:
 - Keep profile up to date
 - LinkedIn groups
 - Like, comment and share content
 - Search and apply for jobs
 - Twitter/X List:
 - Post tweets
 - Like/ reply/ retweet tweets
 - Direct message
 - Bookmark tweets
 - Twitter lists
 - Spaces