Codebook for Publicly Available Data on Social Capital

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel
Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin,
Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg,
Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang,
Mike Bailey, Pablo Barberá, Monica Bhole, Nils Wernerfelt

July 2022

Contents

1 Overview

This document presents details on the publicly available data released as part of the Social Capital Atlas, constructed in Chetty et al. (2022a) and Chetty et al. (2022b). This data release includes a number of social capital statistics aggregated to the level of U.S. counties, ZIP codes, high schools, and colleges. We use methods from the differential privacy literature to add noise to these aggregate statistics to protect privacy while maintaining a high level of statistical reliability (see Section 4 for details). In addition, the data release does not include measures for cells below a minimum size, as discussed in Section 4.4. Our measures of social capital fall into three groups:

1. **Connectedness**: The extent to which people with different characteristics (e.g., low vs. high socioeconomic status) are friends with each other.

2. **Cohesiveness**: The degree to which friendship networks are clustered into cliques and whether friendships tend to be supported by mutual friends.

3. **Civic Engagement**: Indices of trust or participation in civic organizations.

In addition to the social capital measures, in Chetty et al. (2022b) we decompose our measure of the degree to which people with different SES interact with each other – which we term *economic connectedness* – into two determinants: *exposure* (the extent to which people with low versus high socioeconomic status (SES) participate in the same groups) and friending bias conditional on exposure (the tendency for low-SES people to befriend high-SES people at lower rates even conditional on exposure). We release measures of exposure and connectedness as well.

As described in greater detail in Chetty et al. (2022a) and Chetty et al. (2022b), the primary analysis we use to construct these statistics consists of Facebook users aged between 25 and 44 who reside in the United States, were active on the Facebook platform at least once in the prior 30 days, have at least 100 U.S.-based Facebook friends, and have a non-missing residential ZIP code as of May 28, 2022.

For high school and college-level statistics, we focus on individuals in the the 1986-1996 birth cohorts for measures using own SES and individuals in the 1990-2000 birth cohorts for measures using parental SES.

Note that the estimates obtained from these publicly released data will not match those reported in the published papers because of the steps taken to protect privacy – namely the exclusion of small cells (e.g., those with fewer than 100 low-SES and 100 high-SES Facebook users for EC measures) and the addition of noise. In practice, because our estimates are population weighted, the estimates are very similar but the count of the number of observations and the exact point estimates may differ slightly.

2 Codebook

County Connectedness Statistics

Variable Name	Description
`ec_county`	Baseline definition of economic connectedness: two times the share of above-median-SES friends among below-median-SES individuals. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is mapped in Figure 2a of Chetty et al. (2022a).
`ec_se_county`	The standard error of economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy. The variance due to sampling error is calculated using a bootstrap approach described in Chetty et al. (2022a). We then add the noise variance that we apply to protect privacy to generate a combined standard error.
`child_ec_county`	Childhood economic connectedness: two times the share of above-median-SES friends among below-median-SES individuals, calculated using only an individual's high school friends and the individual's and friends' parental SES. This statistic is estimated on the subsample of individuals who can be linked to a parent with a valid SES prediction and matched to a high school. We link individuals to parents as described in Supplementary Information A.1 of Chetty et al. (2022a), and calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a).
`child_ec_se_county`	The standard error of childhood economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
`ec_grp_mem_county`	Two times the share of above-median-SES friends among below-median-SES individuals, restricting attention to friendships that we can allocate to the group in which they were formed as described in Supplementary Information B.1 and B.2 of Chetty et al. (2022b).
`ec_high_county`	Economic connectedness for high-SES individuals: two times the share of above-median-SES friends among above-median-SES individuals.
`ec_high_se_county`	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy.
`child_high_ec_county`	Childhood economic connectedness (calculated using only an individual's high school friends and the individual's and friends' parental SES) for above-median-SES individuals.

child_high_ec_se_county	The standard error of childhood economic connectedness for above-median-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
ec_grp_mem_high_county	Two times the share of above-median-SES friends among above-median-SES individuals, restricting attention to friendships that we can allocate to the group in which they were formed.
exposure_grp_mem_county	Mean exposure to high-SES individuals by county for low-SES people: two times the average share of high-SES individuals in individuals' groups, averaged over low-SES users. We assign Facebook users to groups within settings as described in Supplementary Information B.1 of Chetty et al. (2022b). We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). This variable is plotted in Figure 4A of Chetty et al. (2022b).
exposure_grp_mem_high_county	Mean exposure to high-SES individuals by county for high-SES people: two times the average share of high-SES individuals in individuals' groups, averaged over high-SES users.
child_exposure_county	Mean exposure to above-median-parental-SES peers in high school, averaged over below-median-parental-SES individuals. We assign Facebook users to high schools as described in Supplementary Information B.1 of Chetty et al. (2022b). This statistic is estimated on the subsample of individuals who can be linked to a parent with a valid SES prediction and matched to a high school. We link individuals to parents as described in Supplementary Information A.1 of Chetty et al. (2022a), and calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a).
child_high_exposure_county	Mean exposure to above-median-parental-SES peers in high school, averaged over above-median-parental-SES individuals.
bias_grp_mem_county	ec_grp_mem_county divided by exposure_grp_mem_county, all subtracted from one. Note that this estimate of friending bias is not identical to what one would obtain by calculating friending bias at the group level and then taking means by county because of covariances between exposure and friending bias across groups (see the Exposure, bias and upward income mobility section of Methods of Chetty et al. (2022b) for further discussion). Nevertheless, these approximate estimates of friending bias are correlated above 0.85 with estimates that are aggregated up from group-level statistics. This variable is plotted in Figure 4C of Chetty et al. (2022b).
bias_grp_mem_high_county	ec_grp_mem_high_county divided by exposure_grp_mem_high_county, all subtracted from one.
child_bias_county	child_ec_county divided by child_exposure_county, all subtracted from one.
child_high_bias_county	child_high_ec_county divided by child_high_exposure_county, all subtracted from one.

County Cohesiveness Statistics

Variable Name	Description
clustering_county	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a). We include links to people outside the county when calculating individual clustering (equation 4), but only average clustering over individuals in the relevant county to compute clustering at the county level (equation 5). We add noise to protect privacy, as described in Section 3 of this document.
support_ratio_county	The proportion of within-county friendships where the pair of friends share a third mutual friend within the same county. See equation (6) of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document.

County Civic Engagement Statistics

Variable Name	Description
volunteering_rate_county	The percentage of Facebook users who are members of a group which is predicted to be about volunteering or activism based on its title. We do not include groups that have the privacy setting 'secret' enabled. We additionally manually review the 50 largest such groups in the United States and the largest group in each state, and remove the very small number of groups that are clearly misclassified. We add noise to protect privacy, as described in Section 3.
civic_organizations_county	The number of "public good" Facebook Pages, classified based on page title and page category, per 1,000 users in the county. We remove pages that do not have a website linked, do not have a description on their Facebook page or do not have an address listed. We then assign the page to a county on the basis of its listed address. We add noise to protect privacy, as described in Section 3.

Zip Connectedness Statistics

Variable Name	Description
ec_zip	Baseline definition of economic connectedness: two times the share of above-median-SES friends among below-median-SES individuals. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is mapped for the Los Angeles area in Figure 2b of Chetty et al. (2022a).
ec_se_zip	The standard error of economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy. The variance due to sampling error is calculated using a bootstrap approach described in Chetty et al. (2022a). We then add the noise variance that we apply to protect privacy to generate a combined standard error.
nbhd_ec_zip	Economic connectedness using within-neighborhood friends. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the green neighborhood bar in Figure 2A of Chetty et al. (2022b).
ec_grp_mem_zip	Two times the share of above-median-SES friends among below-median-SES individuals, restricting attention to friendships that we can allocate to the group in which they were formed as described in Supplementary Information B.1 and B.2 of Chetty et al. (2022b). This variable is used in the first row of Table 2 of Chetty et al. (2022b).
ec_high_zip	Economic connectedness for high-SES individuals: two times the share of above-median-SES friends among above-median-SES individuals.
ec_high_se_zip	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy.
nbhd_ec_high_zip	High-type economic connectedness calculated using only an individual's neighborhood friends. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the orange neighborhood bar in Figure 2A of Chetty et al. (2022b).
ec_grp_mem_high_zip	Two times the share of above-median-SES friends among above-median-SES individuals, restricting attention to friendships that we can allocate to the group in which they were formed.

exposure_grp_mem_zip	Mean exposure to high-SES individuals by zip for low-SES people: two times the average share of high-SES individuals in individuals' groups, averaged over low-SES users. We assign Facebook users to groups within settings as described in Supplementary Information B.1 of Chetty et al. (2022b). We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). This variable is mapped for the Los Angeles area in Figure 4B of Chetty et al. (2022b).
exposure_grp_mem_high_zip	Mean exposure to high-SES individuals by zip for high-SES people: two times the average share of high-SES individuals in individuals' groups, averaged over high-SES users.
nbhd_exposure_zip	Exposure calculated using only users living in the relevant zip code. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the green and orange neighborhood bars in Figure 2B of Chetty et al. (2022b). Note that this is the same for high- and low-SES individuals who live in the same zip.
bias_grp_mem_zip	ec_grp_mem_zip divided by exposure_grp_mem_zip, all subtracted from one. This variable is mapped for the Los Angeles area in Figure 4D of Chetty et al. (2022b).
bias_grp_mem_high_zip	ec_grp_mem_high_zip divided by exposure_grp_mem_high_zip, all subtracted from one.
nbhd_bias_zip	nbhd_ec_zip divided by nbhd_exposure_zip, all subtracted from one. This variable is used to construct the green neighborhood bar in Figure 2C of Chetty et al. (2022b).
nbhd_bias_high_zip	nbhd_ec_high_zip divided by nbhd_exposure_zip, all subtracted from one. (Note again that, within the same neighborhood, exposure is the same for low-SES and high-SES individuals) This variable is used to construct the orange neighborhood bar in Figure 2C of Chetty et al. (2022b).

Zip Cohesiveness Statistics

Variable Name	Description
`clustering_zip`	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a). We include links to people outside the zip when calculating individual clustering (equation 4), but only average individual clustering over users in the relevant zip to compute clustering at the zip level (equation 5). We add noise to protect privacy, as described in Section 3 of this document.
`support_ratio_zip`	The proportion of within-zip friendships where the pair of friends share a third mutual friend within the same zip. See equation (6) of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document.

Zip Civic Engagement Statistics

Variable Name	Description
`volunteering_rate_zip`	The percentage of Facebook users who are members of a group which is predicted to be about volunteering or activism based on its title. We do not include groups that have the privacy setting 'secret' enabled. We additionally manually review the 50 largest such groups in the United States and the largest group in each state, and remove the very small number of groups that are clearly misclassified. We add noise to protect privacy, as described in Section 3 of this document.
`civic_organizations_zip`	The number of "public good" Facebook Pages, classified based on page title and page category, per 1,000 users in the zip. We remove pages that do not have a website linked, do not have a description on their Facebook page or do not have an address listed. We then assign the page to a zip on the basis of its listed address. We add noise to protect privacy, as described in Section 3 of this document.

High-School Connectedness Statistics

Variable Name	Description
`ec_own_ses_hs`	Baseline definition of economic connectedness: two times the share of above-median-SES friends within three birth cohorts among below-median-SES individuals. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We estimate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is used in Supplementary Information Figure 3A of Chetty et al. (2022b).
`ec_own_ses_se_hs`	The standard error of economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy. The variance due to sampling error is calculated using a bootstrap approach described in Chetty et al. (2022a). We then add the noise variance that we apply to protect privacy to generate a combined standard error.
`ec_parent_ses_hs`	Economic connectedness with parental SES: two times the share of high-parental-SES friends (who attended the same school within three birth cohorts of the individual) averaged over low-parental-SES individuals at the school. See equations (1), (2), and (3) of Chetty et al. (2022a) for more details on the calculation. We link individuals to parents as described in Supplementary Information A1 of Chetty et al. (2022a), and estimate parental SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is used in Figure 5A of Chetty et al. (2022b).
`ec_parent_ses_se_hs`	The standard error of economic connectedness with parental SES, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
`ec_high_own_ses_hs`	Economic connectedness for high-SES individuals: two times the share of above-median-SES friends within three birth cohorts among above-median-SES individuals.
`ec_high_own_ses_se_hs`	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
`ec_high_parent_ses_hs`	Economic connectedness for high-parental-SES individuals using parental SES: two times the share of above-median-parental-SES friends within three birth cohorts among above-median-parental-SES individuals.
`ec_high_parent_ses_se_hs`	The standard error of economic connectedness for high-parental-SES individuals using parental SES, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.

`exposure_own_ses_hs`	Mean exposure to high-SES individuals by high school for low-SES people: two times the average share of high-SES individuals within three birth cohorts, averaged over low-SES users. This variable is used in Supplementary Information Figure 3A of Chetty et al. (2022b).
`exposure_parent_ses_hs`	Mean exposure to above-median-parental-SES individuals by high school for below-median-parental-SES individuals: two times the average share of high-parental-SES individuals within three birth cohorts, averaged over low-parental-SES users. This variable is used in Figure 5A of Chetty et al. (2022b).
`bias_own_ses_hs`	`ec_own_ses_hs` divided by `exposure_own_ses_hs`, all subtracted from one. This variable is used in Supplementary Information Figure 3A of Chetty et al. (2022b).
`bias_parent_ses_hs`	`ec_parent_ses_hs` divided by `exposure_parent_ses_hs`, all subtracted from one. This variable is used in Figure 5A of Chetty et al. (2022b).
`bias_high_own_ses_hs`	`ec_high_own_ses_hs` divided by `exposure_own_ses_hs`, all subtracted from one.
`bias_high_parent_ses_hs`	`ec_high_parent_ses_hs` divided by `exposure_parent_ses_hs`, all subtracted from one.

High-School Cohesiveness Statistics

Variable Name	Description
clustering_hs	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a). We include only links to friends within the school when calculating individual clustering (equation 4). We add noise to protect privacy, as described in Section 3 of this document.
support_ratio_hs	The proportion of within-school friendships where the pair of friends share a third mutual friend within the same high school. See equation (6) of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. Note that there is very little variation in support ratio by school.

High-School Civic Engagement Statistics

Variable Name	Description
volunteering_rate_hs	The percentage of Facebook users who are members of a group which is predicted to be about volunteering or activism based on its title. We do not include groups that have the privacy setting 'secret' enabled. We additionally manually review the 50 largest such groups in the United States and the largest group in each state, and remove the very small number of groups that are clearly misclassified. We add noise to protect privacy, as described in Section 3.

College Connectedness Statistics

Variable Name	Description
ec_own_ses_college	Baseline definition of economic connectedness: two times the share of above-median-SES friends within three birth cohorts among below-median-SES individuals. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We estimate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is used in Supplementary Information Figure 3B of Chetty et al. (2022b).
ec_own_ses_se_college	The standard error of economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy. The variance due to sampling error is calculated using a bootstrap approach described in Chetty et al. (2022a). We then add the noise variance that we apply to protect privacy to generate a combined standard error.
ec_parent_ses_college	Economic connectedness with parental SES: two times the share of high-parental-SES friends (who attended the same school within three birth cohorts of the individual) averaged over low-parental-SES individuals at the college. See equations (1), (2), and (3) of Chetty et al. (2022a) for more details on the calculation. We link individuals to parents as described in Supplementary Information A1 of Chetty et al. (2022a), and estimate parental SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is used in Figure 5B of Chetty et al. (2022b).
ec_parent_ses_se_college	The standard error of economic connectedness with parental SES, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
ec_high_own_ses_college	Economic connectedness for high-SES individuals: two times the share of above-median-SES friends within three birth cohorts among above-median-SES individuals.
ec_high_own_ses_se_college	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.
ec_high_parent_ses_college	Economic connectedness for high-parental-SES individuals using parental SES: two times the share of above-median-parental-SES friends within three birth cohorts among above-median-parental-SES individuals.
ec_high_parent_ses_se_college	The standard error of economic connectedness for high-parental-SES individuals using parental SES, incorporating both sampling error and error from the addition of noise to protect privacy, calculated as described above.

exposure_own_ses_college	Mean exposure to high-SES individuals by college for low-SES people: two times the average share of high-SES individuals within three birth cohorts, averaged over low-SES users. This variable is used in Supplementary Information Figure 3B of Chetty et al. (2022b).
exposure_parent_ses_college	Mean exposure to above-median-parental-SES individuals by college for below-median-parental-SES individuals: two times the average share of high-parental-SES individuals within three birth cohorts, averaged over low-parental-SES users. This variable is used in Figure 5B of Chetty et al. (2022b).
bias_own_ses_college	ec_own_ses_college divided by exposure_own_ses_college, all subtracted from one. This variable is used in Supplementary Information Figure 3B of Chetty et al. (2022b).
bias_parent_ses_college	ec_parent_ses_college divided by exposure_parent_ses_college, all subtracted from one.
bias_high_own_ses_college	ec_high_own_ses_college divided by exposure_own_ses_college, all subtracted from one.
bias_high_parent_ses_college	ec_high_parent_ses_college divided by exposure_parent_ses_college, all subtracted from one.

College Cohesiveness Statistics

Variable Name	Description
clustering_college	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a). We include only links to friends within the college when calculating individual clustering (equation 4). We add noise to protect privacy, as described in Section 3 of this document.
support_ratio_college	The proportion of within-college friendships where the pair of friends share a third mutual friend within the same college. See equation (6) of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document.

College Civic Engagement Statistics

Variable Name	Description
volunteering_rate_college	The percentage of Facebook users who are members of a group which is predicted to be about volunteering or activism based on its title. We do not include groups that have the privacy setting 'secret' enabled. We additionally manually review the 50 largest such groups in the United States and the largest group in each state, and remove the very small number of groups that are clearly misclassified. We add noise to protect privacy, as described in Section 3.

3 Privacy Protection

We've protected this data set using tools from differential privacy, which adds enough noise to the data to provide precise guarantees that no significant additional information can be learned from the data about individuals (beyond what is already available from any external source). The objective of our approach to privacy protection is to add sufficient noise and to aggregate over sufficiently many individuals that it is not reasonably possible to learn about any one individual from the data. In particular, when releasing an aggregated statistic calculated over hundreds of individuals, it is possible to provide strong privacy protections in the above sense while applying only a small amount of noise, since each individual's measurement contributes a small amount to the overall statistic. Thus, the statistical usefulness of the data is maintained while the privacy of any one individual is protected. The key to determining how much noise to add requires determining how much any individual's measurement matters in the overall statistic. We describe below how we determine this for the key social capital measure, economic connectedness.

3.1 Privacy Protection for Economic Connectedness

3.1.1 Notation

- i and j denote a generic pair of individuals in the primary analysis sample.
- Let c be a cell: the set of Facebook users whom we are considering, such as students in a certain school or residents of a county; let N_c be the number of users in cell c.
- Let $g \in \{0,1\}^{n \times n}$ be a matrix representing the friendships we are considering for the users in cell c. In some cases (such as schools), we only consider friendships within the same school, so $n = N_c$. In other cases, such as county, we also consider friendships outside the county, so $n > N_c$.
- Let L denote the low-SES agents and H denote the high-SES agents.
- Let N_{Lc} denote the number of low-SES users in cell c.
- Let $\mathbf{g} - j$ denote the subgraph induced on the network $\mathbf{g}$ by removing node j and all related edges from the network.
- Let $\mathbf{g} + j$ denote a new network in which a node j has been added, together with some edges.
- Let $d_i(\mathbf{g}) = \sum_j g_{ij}$ denote the number of friends i has (i's degree), and $H_i(\mathbf{g}) = \sum_{j \in H} g_{ij}$ denote the number of high-SES friends i has.

3.1.2 Economic Connectedness

We calculate economic connectedness (EC_c) of c as follows:

$$\mathrm{IEC}_i(\mathbf{g}) \equiv \left\{ \frac{H_i(\mathbf{g})}{d_i(\mathbf{g})} \right\} / 0.5$$

$$\mathrm{EC}_c(\mathbf{g}) = \frac{\sum_{i \in L \cap c} \mathrm{IEC}_i(\mathbf{g})}{N_{Lc}}$$

3.1.3 Calculating Local Sensitivity

We characterize the local sensitivity of EC with respect to both deletions and additions of a node. Cells always have more than one user and their degrees are always more than one, so we never have to worry about division by 0 in what follows. The local sensitivity of $\mathrm{EC}_c(\mathbf{g})$ is defined as:

$$LS_c(\mathbf{g}) \equiv \max\left[\max_j |\mathrm{EC}_c(\mathbf{g}+j) - \mathrm{EC}_c(\mathbf{g})|, \max_j |\mathrm{EC}_c(\mathbf{g}) - \mathrm{EC}_c(\mathbf{g}-j)|\right].$$

In what follows, the terms H_i and d_i are always relative to the starting network $\mathbf{g}$ and so we omit that notation. The following result characterizes the local sensitivity of Economic Connectedness as a function of the network, and determines how much noise we apply to the raw connectedness statistic in each cell.

Theorem 1. *The local sensitivity of EC for cell c at a given network* $\mathbf{g}$ *is:*

$$LS_c(\mathbf{g}) = \max\left\{\frac{2}{N_{Lc}}\sum_{i\in L\cap c}\frac{d_i - H_i}{d_i(d_i-1)}, \frac{2}{N_{Lc}-1}\sum_{i\in L\cap c}\frac{H_i}{d_i(d_i-1)}, \frac{2}{N_{Lc}}\right\}$$

Proof. The proof proceeds by considering four scenarios.

1. The addition or deletion of a high-SES node that has negative influence.
2. The addition or deletion of a high-SES node that has positive influence.
3. The addition or deletion of a low-SES node that has negative influence.
4. The addition or deletion of a low-SES node that has positive influence.

In each case, we consider the effect of an arbitrary addition and then an arbitrary deletion, showing that an arbitrary deletion has a larger bound, and hence is the relevant term for computation of sensitivity.

Case 1: A high-SES node can only have (weakly) positive influence because it does not directly enter the EC sum and can only cause the IEC of terms in the sum to increase; hence, this case is not relevant.

Case 2 Additions: The most an *additional* high-SES node can move EC up by occurs when a high-SES node enters which befriends every low-SES node. Then the change in EC is:

$$\begin{aligned}&\frac{1}{N_{Lc}}\left(2\sum_{i\in L\cap c}\frac{H_i+1}{d_i+1} - 2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)\\ =&\frac{1}{N_{Lc}}\left(2\sum_{i\in L\cap c}\frac{d_i-H_i}{d_i(d_i+1)}\right)\end{aligned}$$

Case 2 Deletions: The most the *removal* of high-SES node can move EC down occurs when one removes a high-SES node who was friends with every low-SES node. Then the change in EC is:

$$\begin{aligned}&\frac{1}{N_{Lc}}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i} - 2\sum_{i\in L\cap c}\frac{H_i-1}{d_i-1}\right)\\ =&\frac{1}{N_{Lc}}\left(2\sum_{i\in L\cap c}\frac{d_i-H_i}{d_i(d_i-1)}\right)\end{aligned}$$

Note that this is larger than the case of additions (the denominator for each term in the sum is smaller for deletions), and is the first term the max operator of Theorem 1.

Case 3 Additions: The most an *additional* low-SES node can move EC down is when the IEC of the new node is 0 and it befriends every other low-SES node.

$$\begin{aligned}
&\frac{1}{N_{Lc}}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)-\frac{1}{N_{Lc}+1}\left(0+2\sum_{i\in L\cap c}\frac{H_i}{d_i+1}\right)\\
\leq&\frac{1}{N_{Lc}+1}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i}-0-2\sum_{i\in L\cap c}\frac{H_i}{d_i+1}\right)\\
=&\frac{1}{N_{Lc}+1}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i(d_i+1)}\right)
\end{aligned}$$

Case 3 Deletions: The most the *removal* of a low-SES node can move EC up is when the IEC of the deleted node is 0 and it was friends with every other low-SES node.

$$\begin{aligned}
&\frac{1}{N_{Lc}-1}\left(\sum_{i\in L\cap c,i\neq j}\frac{H_i}{d_i-1}-0\right)-\frac{1}{N_{Lc}}\sum_{i\in L\cap c}\frac{H_i}{d_i}\\
\leq&\frac{1}{N_{Lc}-1}\left(\sum_{i\in L\cap c,i\neq j}\frac{H_i}{d_i-1}-0-2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)\\
\leq&\frac{1}{N_{Lc}-1}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i-1}-2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)\\
=&\frac{1}{N_{Lc}-1}\left(2\sum_{i\in L\cap c}\frac{H_i}{d_i(d_i-1)}\right)
\end{aligned}$$

Note that this is larger than the case of additions, and is the second term of the max operator of Theorem 1.

Case 4 Additions: The most an *additional* low-SES node can move EC up is when the additional low-SES node only befriends high-SES nodes. Then it does not change the IEC of any other low-SES node down and its own IEC is maximized.

$$\begin{aligned}
&\frac{1}{N_{Lc}+1}\left(2+2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)-\frac{2}{N_{Lc}}\sum_{i\in L\cap c}\frac{H_i}{d_i}\\
\leq&\frac{1}{N_{Lc}}\left(2+2\sum_{i\in L\cap c}\frac{H_i}{d_i}-2\sum_{i\in L\cap c}\frac{H_i}{d_i}\right)\\
=&\frac{2}{N_{Lc}}
\end{aligned}$$

Case 4 Deletions: The most a *removal* of a low-SES node can move EC down is when the removed low-SES node was only friends with high-SES nodes. Then its own IEC was maximized and it did not bring down the IEC of other low-SES nodes.

$$\begin{aligned}
&\frac{2}{N_{Lc}}\sum_{i\in L\cap c}\frac{H_i}{d_i}-\frac{2}{N_{Lc}-1}\left(\sum_{i\in L\cap c, i\neq j}\frac{H_i}{d_i}\right)\\
\leq&\frac{2}{N_{Lc}}\left(1+\sum_{i\in L\cap c, i\neq j}\frac{H_i}{d_i}\right)-\frac{2}{N_{Lc}-1}\left(\sum_{i\in L\cap c, i\neq j}\frac{H_i}{d_i}\right)\\
\leq&\frac{2}{N_{Lc}}+\frac{2}{N_{Lc}-1}\left(\sum_{i\in L\cap c, i\neq j}\frac{H_i}{d_i}\right)-\frac{2}{N_{Lc}-1}\left(\sum_{i\in L\cap c, i\neq j}\frac{H_i}{d_i}\right)\\
=&\frac{2}{N_{Lc}}
\end{aligned}$$

This is the third term of the max operator.

Theorem 1 follows from combining the deletion cases for cases 2, 3, and 4. □

3.1.4 Constructing an Envelope from the Sensitivities

Using the local sensitivities S_c for each cell, we construct a smooth envelope based on one non-noisy parameter χ which we do not release to the public.

$$\chi = \max_c \left\{ \frac{S_c}{\frac{1}{N_{Lc}}\sum_{i\in L\cap c}\frac{1}{d_i}} \right\}$$

We then calculate smoothed noise $\tilde{S}_c$ as:

$$\tilde{S}_c = \chi \times \frac{1}{N_{Lc}}\sum_{i\in L\cap c}\frac{1}{d_i}$$

applying noise to EC calculated in each cell c from the distribution:

$$\text{Laplace}\left(0, \frac{\tilde{S}_c}{\varepsilon}\right)$$

using $\varepsilon = 8$, as in Chetty et al. (2018) .

3.2 Privacy Protection for Other Statistics

We protect privacy for the other statistics we release as follows.

Exposure and Volunteering Rate. These variables are simple means over independent values. Individual-level values of exposure lie between 0 and 2, so we follow standard results in the differential privacy literature for means of bounded variables and apply noise from the Laplace(0, $2/N\varepsilon$) distribution, where N is the number of users in the cell. Individual-level volunteering is a binary value equal to either zero or one, so we apply noise from the Laplace(0, $1/N\varepsilon$) distribution.

Friending Bias. We approximate friending bias as the ratio of two privacy-protected (EC and exposure) we release publicly; since it is simply a function of publicly available, privacy-protected statistics, no further noise is added.

Clustering and Support Ratio. For cohesiveness measures (clustering and support ratio), which do not use any information on individuals' characteristics, we follow the privacy procedures developed for the Social Connectedness Index, which was released in Fall 2020 by the Facebook Data for Good team (see Bailey et al., 2018, 2020, 2021). Specifically, we compute the statistic over the subgraph from a 99% random sample of users. We then apply additional noise from the Laplace(0, 0.001/8) distribution to the cell-level averages of the node-level network statistics.
Civic Organizations. Public good page density is a variable based on a count (the number of pages in an area) and only indirectly on users (through the density calculation), so we add noise from the Laplace(0, 0.001/8) distribution.

Finally, to further protect privacy, we only release statistics on economic connectedness, exposure, and friending bias for cells that contain at least 100 low-SES and at least 100 high-SES Facebook users. We only release statistics on volunteering rates, clustering, and support ratios for cells that contain at least 100 Facebook users.

4 Citing the Data

Please cite the following two publications as the source of the data:

- Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt (2022a). "Social Capital I: Measurement and Associations with Economic Mobility." *Nature*, XX(Y), AA–BB.

```
@article{
    chetty2022socialcapitalone,
    title = {Social Capital I: Measurement and Associations with Economic Mobility},
    author = {
        Chetty, Raj and Jackson, Matthew O. and Kuchler, Theresa and
        Stroebel, Johannes and Hendren, Nathaniel and Fluegge, Robert and
        Gong, Sara and Gonzalez, Federico and Grondin, Armelle and
        Jacob, Matthew and Johnston, Drew and Koenen, Martin and
        Laguna-Muggenberg, Eduardo and Mudekereza, Florian and Rutter, Tom and
        Thor, Nicolaj and Townsend, Wilbur and  Zhang, Ruby and
        Bailey, Mike and Barber\'{a}, Pablo and Bhole, Monica and Wernerfelt, Nils},
    journal = {Nature},
    volume = {608},
    number = {7921},
    pages = {AA--BB},
    year = {2022}
}
```

- Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt (2022b). "Social Capital II: Determinants of Economic Connectedness." *Nature*, XX(Y), CC–DD.

```
@article{
    chetty2022socialcapitaltwo,
    title = {Social Capital II: Determinants of Economic Connectedness},
```

```
    author = {
        Chetty, Raj and Jackson, Matthew O. and Kuchler, Theresa and
        Stroebel, Johannes and Hendren, Nathaniel and Fluegge, Robert and
        Gong, Sara and Gonzalez, Federico and Grondin, Armelle and
        Jacob, Matthew and Johnston, Drew and Koenen, Martin and
        Laguna-Muggenberg, Eduardo and Mudekereza, Florian and Rutter, Tom and
        Thor, Nicolaj and Townsend, Wilbur and  Zhang, Ruby and
        Bailey, Mike and Barber\'{a}, Pablo and Bhole, Monica and Wernerfelt, Nils},
    journal = {Nature},
    volume = {608},
    number = {7921},
    pages = {CC--DD},
    year = {2022}
}
```

References

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel, "International trade and social connectedness," *Journal of International Economics*, 2021, *129*, 103418.

_ **, Drew Johnston, Theresa Kuchler, Dominic Russel, Johannes Stroebel et al.**, "The determinants of social connectedness in Europe," in "International Conference on Social Informatics" Springer 2020, pp. 1–14.

_ **, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong**, "Social connectedness: Measurement, determinants, and effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–80.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," October 2018.

_ **, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenberg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt**, "Social Capital I: Measurement and Associations with Economic Mobility," *Nature*, 2022, *XX* (Y), AA–BB.

_ , _ **, and** _ , "Social Capital II: Measurement and Associations with Economic Mobility," *Nature*, 2022, *XX* (Y), CC–DD.