Measuring Social Media Network Effects Across Platforms

Preliminary Extended Abstract

DO NOT CITE OR DISTRIBUTE

Sinan Aral, Seth G Benzell, Avinash Collis

July 12, 2021

1 Introduction

In this research, we seek to measure the value of connections of different types created on four of the most common social media platforms. The social media platforms we study are Facebook, Twitter, Instragram, and LinkedIn. To do so, we are conducting surveys on users of social media. In one type of survey we ask social media users to identify a set of connections on the platfrom, their demographic characteristics, the nature of their relationship, and to rank these connections in terms of importance. In a second set of surveys we conduct a 'willingness to accept' experiment, where we ask platform users how much they would need to be paid to give up connections to users of various types.

There are several reasons why it is important to understand how different sorts of connections create value on different social media platforms. One important reason is that it may play a large role in determining why some social media platforms succeed, while others are flops. Myspace had 88 million more users than Facebook in September 2006, but MySpace was built around music interest groups, while Facebook was built around college friendships. It is theorized this might have contributed to their divergent outcomes (Aral, 2020).

More broadly, measuring the network effects among different groups is essential for modelling participation on social medial platforms. It is therefore key to both platform managers and aspiring regulators. Participation by users on a social media platform can be modelled as a recursive function, where the quality of a platform, and therefore a user's desire to participate on it, is a function of the participation (in previous periods) by users of different types.

Let P_t be a vector of participation rates for members of I demographic groups, and ϕ a vector of fees for each group. Then,

$$P_{t+1} = f(P_t, \phi) \tag{1}$$

Taking the partial derivative of participation with respect to participation in previous periods, for each different type of user yields a matrix B

$$\frac{\partial P_{t+1}}{\partial P_t} = B \tag{2}$$

In the 'cannonical' model of network effects (e.g. Rohlfs (1974), Weyl (2010), Weyl and White (2014)) where i's choice to use the platform P_i is a function of others' use and monetization, B is the product of the matrix of network effects and each group's elasticity of demand

$$B_{i,j} = \underbrace{\frac{\partial \mu_i}{\partial P_j}}_{\text{Network Effect of } j \text{ on } i} \times \underbrace{\frac{\partial P_i}{\partial \mu_i}}_{\text{Demand Elasticity of Group } i} \tag{3}$$

where μ_i is the value that a user of type i gets from participating on the platform.

2 Research Plan

We are partway through conducting two series of surveys.

First, have launched, but do not have results yet from, a large N series of willingness to accept surveys implemented through Google Surveys. The surveys ask participants whether they would be willing to give up all connections to users of a particular type or demographic, for one month, in exchange for a payment. This approach to soliciting the value of network effects has precedents in Allcott et. al. (2019), Brynjolfsson, Collis and Eggers (2019), and Benzell and Collis (2020). We plan to collect 320,000 responses, with 80,000 responses per platform.

Second, we have preliminary results ready from a medium N series of surveys conducted through the Lucid survey platform. More similar to a focus group, these surveys take much longer to complete than the single-response Google Survey questionnaires. After filtering responses for quality, we have 1516 responses to our Lucid survey of Facebook users. We are currently in the process of soliciting a similar amount of high quality responses for the other three platforms.

In the Lucid surveys, we ask the users of the platforms to identify eight connections on the platform – their four highest valued connections, and four additional connections. We then ask several questions about he characteristics of those friends, and ask them to rank those friends in terms of connection value. Finally, we ask users their total value for being able to use the platform for a month.

As part of the survey, we require survey takers to link their social media platforms. This allows

	(1) friendranking_	(2) friendranking_
friend_num=1	0.000	0.000
	(.)	(.)
friend_num=2	0.929***	0.972***
	(0.074)	(0.092)
friend_num=3	1.423***	1.464***
	(0.074)	(0.095)
friend_num=4	2.066***	2.208***
	(0.075)	(0.096)
friend_num=5	2.554***	2.734***
	(0.076)	(0.097)
friend_num=6	2.971***	3.182***
	(0.074)	(0.093)
friend_num=7	3.063***	3.269***
	(0.074)	(0.094)
friend_num=8	3.208***	3.534***
	(0.075)	(0.094)
Constant	2.336***	2.212***
	(0.053)	(0.067)
Observations	11666	6801

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Regression of friend value rank, as a function of the order in which friends were entered in the survey. Unsurprisingly, friends listed in the first four (highest valued friends) are ranked higher. The second column restricts attention to survey takers with Facebook pages with visible friend lists of more than 10 friends.

us to collect additional verified information about the user's mix of friends, and total friend count. It also allows us to verify the authenticity of the participant's responses.

At some point in the future, either as part of this paper or a subsequent one, we plan a small scale in-lab study where incentive compatability for the willingness to accept experiments can be enforced (i.e. by actually compensating users for giving up connections, and verifying compliance).

3 Preliminary Results

These are the preliminary results of our Lucid survey of Facebook users.

3.1 OLS Results

Each of the following tables has four columns.

	(1) friendranking_	(2) friendranking_	(3) friendranking_	(4) friendranking_
College or University	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Family member or family friend	-0.268**	-0.176	-0.460***	-0.296*
	(0.095)	(0.090)	(0.135)	(0.124)
I only know them online	0.838***	0.374**	0.769***	0.301
	(0.121)	(0.117)	(0.169)	(0.159)
None of these describe how I know this person	0.147	-0.032	0.100	0.005
	(0.136)	(0.121)	(0.193)	(0.163)
School (K-12)	0.355***	0.205*	0.272	0.129
	(0.107)	(0.101)	(0.149)	(0.137)
Through a shared interest in real life	0.252*	0.134	0.190	0.099
	(0.108)	(0.101)	(0.153)	(0.139)
Work	0.542***	0.338**	0.404**	0.228
	(0.109)	(0.104)	(0.153)	(0.141)
Constant	4.336***	4.382***	4.503***	4.496***
	(0.090)	(0.086)	(0.129)	(0.120)
Order FE		X		X
N	1.2e+04	1.2e+04	6801.000	6801.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Regression of friend value rank, as a function of how the friend is known. Family members have the highest value, followed by College friends. Columns 2 and 4 employ order fixed effects (i.e. they only identify off of 'random' friends). Columns 3 and 4 restrict attention to the subset of survey takers with public friend lists and more than 10 friends.

	(1) friendranking_	(2) friendranking_	(3) friendranking_	(4) friendranking_
Female	0.047	1.046***	-0.974*	0.155
	(0.236)	(0.238)	(0.435)	(0.429)
Male	0.190	1.140***	-0.848	0.241
	(0.235)	(0.237)	(0.434)	(0.428)
Other/ Don't know	0.705*	1.053***	-0.095	0.266
	(0.286)	(0.288)	(0.495)	(0.492)
same_gender_as_alter=0	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
same_gender_as_alter=1	0.291***	0.223***	0.365***	0.278***
	(0.046)	(0.041)	(0.059)	(0.052)
Constant	4.100***	3.178***	5.095***	4.050***
	(0.231)	(0.234)	(0.431)	(0.426)
Order FE		X		X
N	1.2e+04	1.2e+04	6801.000	6801.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regression of friend value rank, as a function of how the gender of the friend (alter) and the subject's gender (ego). Female connections are valued more highly, as well as opposite gender connections, meaning female-to-male connections are the highest valued. Columns 2 and 4 employ order fixed effects (i.e. they only identify off of 'random' friends). Columns 3 and 4 restrict attention to the subset of survey takers with public friend lists and more than 10 friends.

	(1) friendranking_	(2) friendranking_	(3) friendranking_	(4) friendranking_
Connected on INST	-0.488***	-0.274***	-0.565***	-0.365***
	(0.046)	(0.041)	(0.060)	(0.051)
Connected on TWIT	-0.140*	-0.075	-0.163	-0.050
	(0.069)	(0.062)	(0.094)	(0.080)
Connected on LINK	0.008	0.044	-0.074	0.029
	(0.076)	(0.068)	(0.105)	(0.089)
Constant	4.605***	4.508***	4.655***	4.555***
	(0.027)	(0.024)	(0.035)	(0.030)
Order FE		X		X
N	1.2e+04	1.2e+04	6801.000	6801.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression of friend value rank, as a function of whether the ego is connected to the alter on other platforms. Alters also connected to on Instagram are more highly valued than ones which are not, suggesting connections on these platforms are not perfect substitutes. Columns 2 and 4 employ order fixed effects (i.e. they only identify off of 'random' friends). Columns 3 and 4 restrict attention to the subset of survey takers with public friend lists and more than 10 friends.

	(1) friendranking_	(2) friendranking_	(3) friendranking_	(4) friendranking_
Asian / Pacific Islander	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Black or African American	0.203	0.183	0.107	0.106
	(0.124)	(0.110)	(0.186)	(0.152)
Hispanic or Latino	0.079	0.114	-0.081	-0.015
	(0.128)	(0.115)	(0.194)	(0.161)
Native American or American Indian	0.162	0.196	-0.096	-0.072
	(0.211)	(0.189)	(0.344)	(0.294)
Other/ Don't know	0.326	0.014	0.569*	0.267
	(0.167)	(0.152)	(0.245)	(0.203)
White	0.270*	0.249*	0.261	0.243
	(0.114)	(0.101)	(0.175)	(0.140)
same_ethnic_as_alter=1	-0.159**	-0.015	-0.222**	-0.062
	(0.059)	(0.053)	(0.081)	(0.073)
Constant	4.287***	4.197***	4.378***	4.266***
	(0.115)	(0.102)	(0.176)	(0.140)
Order FE		X		X
N	1.2e+04	1.2e+04	6801.000	6801.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Regression of friend value rank, as a function of alter race and whether the ego and alter share the same race. Alter race has no strong direct effect (white alters are perhaps slightly less valued), but alters of the same race as the ego are more likely to be ranked among the top four friends. Columns 2 and 4 employ order fixed effects (i.e. they only identify off of 'random' friends). Columns 3 and 4 restrict attention to the subset of survey takers with public friend lists and more than 10 friends.

	(1) friendranking_	(2) friendranking_	(3) friendranking_	(4) friendranking_
About once a month	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
About once a week	-0.297***	-0.159*	-0.212*	-0.037
	(0.077)	(0.072)	(0.103)	(0.093)
I live with this person	-1.737***	-0.815***	-1.757***	-0.724***
	(0.083)	(0.081)	(0.106)	(0.104)
Less than 4 times last year	0.167**	0.113	0.222**	0.143
	(0.065)	(0.059)	(0.084)	(0.076)
Multiple times per week	-0.495***	-0.190**	-0.515***	-0.181*
	(0.070)	(0.064)	(0.092)	(0.083)
Never	0.570***	0.308***	0.796***	0.476***
	(0.067)	(0.062)	(0.087)	(0.079)
Constant	4.515***	4.429***	4.470***	4.381***
	(0.050)	(0.047)	(0.066)	(0.061)
Order FE		X		X
N	1.2e+04	1.2e+04	6801.000	6801.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Regression of friend value rank, as a function of how often the ego sees the alter. Connections seen more often are more highly valued, suggesting that in-person contact is a complement to social media connections rather than a substitute. Columns 2 and 4 employ order fixed effects (i.e. they only identify off of 'random' friends). Columns 3 and 4 restrict attention to the subset of survey takers with public friend lists and more than 10 friends.

3.2 OLS Results - Value of FB as Outcome

3.3 LASSO Results

In principle, the marginal value that one group gets from another could be an extremely complex function of ego characteristic, alter characteristic, and time (as preferences, may shift over time as well).

An OLS regression of friend importance on all of these characteristics and their interactions would tend to be overfit however, and OLS regressions on each of these characteristics individually would face multiple hypothesis testing problems. Therefore, we proceed to estimate LASSO specifications on a wide array of our ego characteristics (left of the $\times$ interaction) and alter characteristics (right of the $\times$ interaction) and their interactions. Using cross fold variation, we select two OLS models. The best model without including friend order fixed effects (and interactions) has 172 terms (with an R^2 of .1272), and the best model including them has 188 terms (with an an R^2 of .2956).

The columns below report the OLS results of a regression of friend rank on the terms selected by OLS. Only selected terms are reported. The column on the left is from the model without hot-coded order fixed effects and interactions included, and the one on the right reports selected coefficients when they are included.

	(1) total_FB_monthly_val
FB_friend_count_tot	0.004***
	(0.001)
FB_friend_pctblack_tot	-5.365***
	(1.414)
FB_friend_pctwhite_tot	-5.167***
	(1.431)
FB_friend_pct_hispanic_tot	-5.061***
	(1.441)
FB_friend_pct_asian_tot	-5.450***
	(1.406)
FB_friend_pct_native_am_tot	-4.766**
	(1.680)
FB_friend_pct_2race_tot	3.744
	(3.246)
FB_friend_pct_female_tot	-7.008**
	(2.502)
FB_lname_commonness_tot	0.033*
	(0.016)
Constant	584.634***
	(144.587)
Observations	6968

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Regression of ego reported total monthly value from Facebook (top-coded at $100) as a function of their total number of friends, the share of friends by race and gender, and the commonness of friends' last names. Race and gender of friends are estimated by the first (gender) and last (race) names of friends, so sample is restricted to respondents with public friend lists of at least 10 friends. With these controls for friend type, the average Facebook value per connection is .4 cents per month. Individuals with a higher share of Native American, mixed race and male friends also have a higher value from Facebook in this specification.

	(1) friendranking_	(2) friendranking_
Less than $25,000 × Other/ Don't know	0.826**	
	(0.313)	
Less than $25,000 × 55-64 years old	-0.395*	-0.352*
	(0.183)	(0.158)
$50,000 - $99,999 × Hispanic or Latino	-0.566**	-0.475**
	(0.174)	(0.181)
$150,000 or more × College or University	-1.185***	-1.111***
	(0.282)	(0.290)
Extremely conservative × 55-64 years old	-0.695**	
	(0.240)	
Liberal × I only know them online	-0.717*	
	(0.283)	
Slightly liberal × About once a week	-0.867**	-0.443
	(0.288)	(0.234)
Asian / Pacific Islander × Under 18 years old	-3.037***	-3.022*
	(0.654)	(1.252)
Black or African American × 35-44 years old	0.441*	0.355*
	(0.189)	(0.178)
Black or African American × 45-54 years old	-0.558*	-0.894***
	(0.240)	(0.216)
Native American or American Indian × 45-54 years old	2.632***	
	(0.379)	
Other × Under 18 years old	2.144*	1.835**
	(1.038)	(0.703)
Black or African American × College or University	0.633*	0.606
	(0.304)	(0.326)
Other × College or University	-1.768**	-1.940***
	(0.555)	(0.443)
Black or African American × About once a month	-0.585**	
	(0.224)	
Other × College or University	-1.760**	
	(0.620)	
Female × Multiple times per week	-0.294*	-0.097
	(0.143)	(0.116)
Other × I live with this person	-1.771***	-1.600**
	(0.264)	(0.539)
65 years or older × Female	-0.385*	
	(0.181)	
25-34 years old × Under 18 years old	0.782*	0.637
	(0.357)	(0.387)
45-54 years old × Under 18 years old	-1.302**	-0.402
	(0.399)	(0.404)
65 years or older × Asian / Pacific Islander	2.066***	
	(0.486)	
25-34 years old × Less than 4 times last year	0.342**	0.137
	(0.126)	(0.102)
Hispanic or Latino × older_than_alter=1	0.669**	0.699**
	(0.233)	(0.223)
Extremely liberal × same_age_as_alter=1	-0.314*	
	(0.137)	
Constant	4.174***	4.574***
	(0.279)	(0.220)
N	6747.000	6747.000

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: OLS regression of friend rank on ego and alter characteristics, with terms selected by a CFV LASSO procedure. Only some terms reported. The term on the left of the interaction is an ego characteristic, and the term on the right is an alter characteristic. Sample is restricted to respondents with public friend lists of at least 10 friends.