Peers and Persuasion Across Collegiate Social Networks

Jakina Debnam[1]

Abstract

This Draft: 12/31/2017

Using a unique set of text and network data from a social network, this paper measures the persuasiveness of peers' communications among college undergraduates' course selection at Cornell University. I use idiosyncratic shocks to students' information sets to create an instrumental variable and find that while in general, the effect of receiving an additional piece of information about a course is a decrease in the likelihood that a student enrolls in the course, if the message-giver is a peer, the effect of this additional message is up to a 7.4% increase in the likelihood that a student enrolls in that course. This finding is consistent with theories of information aggregation where individuals 'tag' information with sources as they incorporate these sources into their final decisions. I support key assumptions using exponential random graph models and in-person survey data which I collected from 112 undergraduate students. To the best of my knowledge, this work is the first in economics to empirically investigate theories of social influence using non-experimental field data.

JEL Codes: I2, D83

[1]Charles H. Dyson School of Applied Economics and Management, Cornell University, 410 Warren Hall, Ithaca, NY 14853 (e-mail: jrd294@cornell.edu). I am grateful to the Behavioral Economics Research Group and Social Dynamics Laboratory at Cornell University for helpful comments, as well as to the National Science Foundation Graduate Research Fellowship Program grant no. DGE-1144153 for research support.

I. Introduction

The importance of peer effects for educational and labor market outcomes is well documented. For post-secondary education, the relevant literature finds robust effects for academic outcomes such as grades and large effects for social behaviors such as smoking[2]. Given the importance of such peer effects, there exists relatively little work exploring the mechanisms by which these peer effects arise. In this paper, I use a novel set of de-identified social network and course choice data to investigate a particular mechanism through which undergraduate peer effects may arise – information transmission.

To investigate this mechanism, I exploit a feature of *Chatter*, the proprietary undergraduate social network which all Cornell University undergraduates within the College of Agriculture and Life Sciences are mandated to use, which gives rise to exogenous variation in students' information sets: students lack complete control over the sets of information they receive from their peers on the platform. Students' information receipt within the platform is driven by the set of groups to which they belong; students' group membership is partially determined by administrators. Further, due to a bug within the platform at the time of this writing, students receive messages from peers an idiosyncratically determined number of times. This random variation allows me to adopt an instrumental variable approach to measure the effect of peers' information on own course choice. Using this approach, I find that receiving a statement from a peer about a class increases the likelihood that a student enrolls in a course by up to 7.4%.

Well-known pitfalls of identifying peer effects include concerns of peers' simultaneous effects on one another, common shocks to the individual and her peers which drive outcomes, and disentangling contextual from endogenous peer effects (Manski (1993)). Others have achieved identification in this setting appealing to the random assignment of roommates (Sacerdote (2001); Zimmerman (2003)), or appealing to instrumental variables approaches (Case and Katz (1991)). In this particular setting, an

[2] For a somewhat recent review of the literature concerning peer effects in education, see Sacerdote (2011).

additional concern is that of motivated information seeking, or information receipt which depends individuals' preferences over outcomes. To address these concerns, I adopt an instrumental variable approach which provides unbiased estimates of the effects of peers' messages on own outcomes.

In addition to the peer effects literature, this paper also speaks to another, largely theoretical, literature which studies the way individuals within networks process information signals from diverse sources into a single decision. This literature originates with DeGroot's (1974) seminal model which features an environment in which individuals form decisions by updating their previously held beliefs by taking weighted averages of their own beliefs and others' past beliefs. Importantly, this model features an updating mechanism which fails to account for information sources, leading to double-counting of information which arrives to the agent from a single source but through multiple channels – an outcome known as persuasion bias. Refinements of this model, wherein individuals account for information sources in their decision-making (rather than relying simply on signal volume) have been put forth by Gale and Kariv (2003), Bala and Goyal (2000), Acemoglu, Bimpikis, & Ozdaglar (2014) and prominently by DeMarzo, Vayanos, and Zweibel (2003).

In a reduced form way, my findings evidence rational information aggregation by suggesting that undergraduates account for *both* the quantity and source of signals when incorporating these signals into their decisions.

II. Data Description

This section introduces the data and institutional setting. Data for my analysis comes from de-identified administrative i) social network, ii) freshman roommate, and iii) course choice data collected from undergraduate students who graduated or are scheduled to graduate between the years of 2015 and 2020 from College of Agriculture and Life Sciences (CALS) at Cornell University. Summary statistics for these three datasets as well as for the final matched sample are described in Table 1. Cornell University is a large, highly-selective research university located in central New York. With an enrollment of roughly 3,100 undergraduate students, the College of Agriculture and Life Sciences is the second largest college at Cornell with students majoring in a variety of subjects including life sciences, pre-veterinary sciences, and economics (the largest major in the College).

Table 1: Summary Statistics Across Roommate, Social Network, and Course Choice Data Sets

	Freshman Roommate Data	Chatter Data	Course Choice Data	Matched Sample
Gender				
Female	1,794 (54.30%)	2,628 (54.31%)	2,610 (54.38%)	1,794 (54.31%)
Male	1,509 (45.67%)	2,210 (45.67%)	2,189 (45.60%)	1,508 (45.66%)
Other/Unknown	1 (0.03%)	1 (0.02%)	1 (0.02%)	1 (0.03%)
Ethnicity				
Asian	514 (13.45%)	935 (13.54%)	701 (14.13%)	449 (13.50%)
Black	340 (8.90%)	520 (7.53%)	404 (8.14%)	309 (9.29%)
Hispanic	173 (4.53%)	287 (4.16%)	222 (4.47%)	148 (4.45%)
White	1,844 (48.26%)	3,164 (45.84%)	2,278 (45.92%)	1,539 (46.29%)
Multicultural	570 (14.92%)	944 (13.68%)	712 (14.35%)	521 (15.67%)
Other	380 (9.95%)	1,053 (15.25%)	644 (12.98%)	359 (10.80%)
Citizenship				
U.S.	3.742 (96.62%)	8,522 (97.12%)	4,767 (95.02%)	3,205 (96.07%)
Other	131 (3.38%)	253 (2.88%)	250 (4.98%)	131 (3.93%)
Graduation Year				
2015	539 (16.12%)	633 (10.16%)	632 (12.60%)	539 (16.16%)
2016	617 (18.45%)	892 (14.31%)	882 (17.59%)	615 (18.44%)
2017	668 (19.98%)	1,038 (16.65%)	1,023 (20.40%)	667 (19.99%)
2018	687 (20.54%)	1,088 (17.46%)	1,073 (21.40%)	687 (20.59%)
2019	706 (21.11%)	1,157 (18.56%)	1,113 (22.20%)	701 (21.01%)
Other	127 (3.80%)	1,425 (22.86%)	291 (5.80%)	127 (3.81%)
Is First Gen. Student	238 (6.15%)	706 (8.05%)	386 (7.69%)	188 (5.64%)
Mean GPA	3.32	3.33	3.32	3.31
	(.009)	(.007)	(.008)	(.009)
N	3,873	8,775	5,017	3,336

Note: This table provides demographic summary statistics for students enrolled in the College of Agriculture and Life Sciences at Cornell. Totals may not sum due to missing/inconsistent demographic information for some students. Roommate Data exists for students who lived in campus-owned housing for at least their first year at Cornell. Standard errors of mean GPAs given below averages in parentheses.

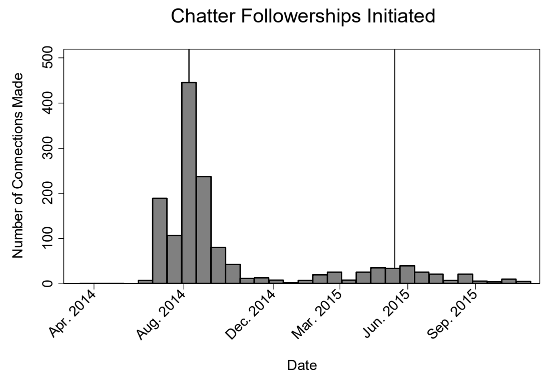
II.I *Chatter* Data

The set of social network data come from *Chatter,* a private social network used by all students and some administrators at the College. Chatter is a Salesforce-based platform used by the College to disseminate information to students concerning topics such as registration, class selection, extracurricular activities, and job search, with the self-described objective "to create a forum to assist with course enrollment, orient students to their major, decompress the Fall experience, and build a sense of community among students, faculty, and staff at the College". Students and administrators may also post advertisements, ask and answer questions of one another, and exchange public and private messages. When implemented, Chatter was explicitly designed to aggregate administrative communications which may have previously taken place via email or snail mail. Therefore, communications within this platform represent a substantial component of students' total set of information regarding administrative and academic choices at the University. There are a relatively small number of private direct messages being transmitted across the network. Instead, most users passively receive content through feeds, based on: 1) the postings of the individuals whom they choose to *follow* and 2) the *groups* to which they belong.

II.I.I Followership as a Reflection of Friendship

Electing into a *followership* relationship is a one-way relationship akin to a followership relation on the Twitter platform – it enables students to passively review the content generated by the individual they have chosen to follow. Students are much more likely to follow one another at the beginning of a school year (shown in Figures 1 & Appendix Figure A1). Summary statistics regarding students who opt-in to followership relationships are available in Table 2. The demographic composition of the subset of students for whom 'followership' relationships are observed (shown in Table 2) on the private social network is roughly comparable to the entire population of users of the Chatter platform (i.e. all students in the College).

Figure 1: Student Followership Formations Over Time



Note: This figure plots student decisions to follow another person on the Chatter platform over time. The number of connections peaks during the start of the student registration period in 2014, with over 27.3% of student-to-student ties being formed on August 8, 2014 and 17.3% of overall ties being formed on this date. The remaining hump in the distribution represents tied formed during the week of Cornell graduation 2015.

Table 2: Summary Statistics for Students Who Belong to At Least One Following Relationship

	All Chatter Users	Followership Members
Gender		
Female	2,628 (54.31%)	210 (61.58%)
Male	2,210 (45.67%)	131 (38.42%)
Other/Unknown	1 (0.02%)	0 (0.00%)
Ethnicity		
Asian	935 (13.54%)	75 (21.61%)
Black	520 (7.53%)	28 (8.07%)
Hispanic	287 (4.16%)	17 (4.90%)
White	3,164 (45.84%)	139 (40.06%)
Multicultural	944 (13.68%)	44 (12.68%)
Other	1,053 (15.25%)	44 (12.68%)
Citizenship		
U.S.	8,522 (97.12%)	326 (91.57%)
Other	253 (2.88%)	30 (8.43%)
Graduation Year		
2015	633 (10.16%)	15 (4.30%)
2016	892 (14.31%)	41 (11.75%)
2017	1,038 (16.65%)	40 (11.46%)
2018	1,088 (17.46%)	181 (51.86%)
2019	1,157 (18.56%)	65 (18.62%)
Other	1,425 (22.86%)	7 (2.01%)
Is First Gen. Student	706 (8.05%)	66 (18.54%)
Mean GPA	3.33	3.36
	(.007)	(.009)
N	8,775	356

Note: Totals may not sum due to missing/inconsistent demographic information for some students

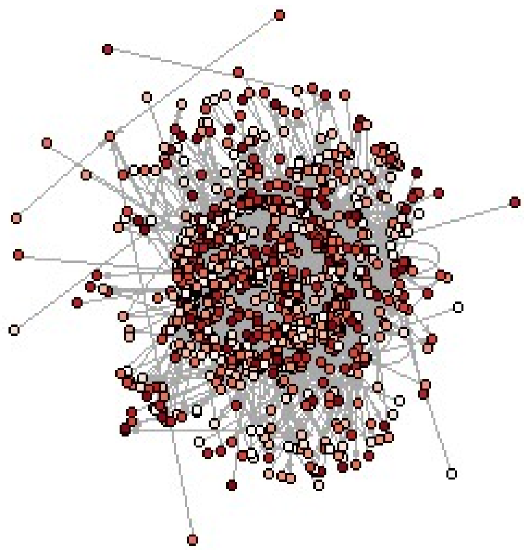
Network graphs of student followerships inclusive and exclusive of administrators are shown in Figure 2. Following administrators provides a key source of information for students. Student-to-student followership relationships are qualitatively different - unlike relationships involving administrators, users do not anticipate that these relationships will provide them access to information or resources within Chatter[4]. Instead, these followerships provide an opportunity for peer-to-peer messages and signal existing friendship relationships outside of Chatter. This symbolic meaning of followership within Chatter is consistent with the work of others studying behavior on other online social platforms. For example, in a survey of undergraduate students about their use of social networking sites, Subrahmanyam, Reich, Waechter, and Espinoza (2008) find that students use Facebook to reconnect with friends and family members and that there is large (yet imperfect) overlap between students' online and offline networks. Another example is the work of Ellison, Steinfield, & Lampe (2007) who survey first year students at Michigan State University and find that most students use Facebook to connect with people that they have already met, rather than using the platform to seek out new relationships.

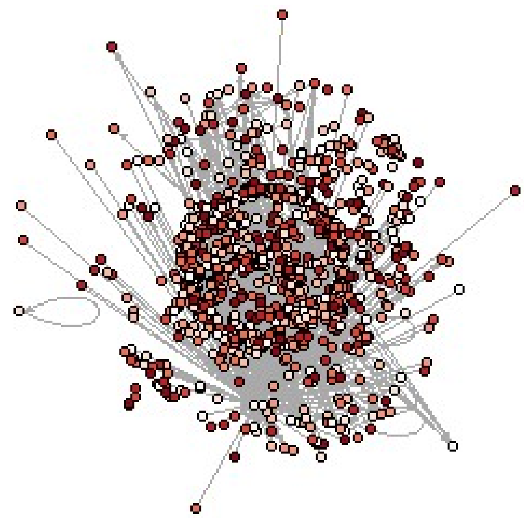
From these student-student friendships (followerships) I observe a directed graph, represented by a *356* X *356* adjacency matrix $\mathbf{G_d} = [g_{ij}]$, where $g_{ij} = 1$ if agent i is followed by agent j. For analysis, however, I follow conventions of Exponential Random Graph Modeling and collapse this graph into an undirected *356* X *356* adjacency matrix $\mathbf{G_u} = [g_{ij}]$, where $g_{ij} = 1$ if either agent i is followed by agent j or agent j is followed by agent i, 0 otherwise. Network-level summary statistics from this undirected graph are presented in Table 3. The followership network is sparse and is characterized by low levels of reciprocity and transitivity, consistent with a social network where most information transmission is passive.

Figure 2: Network Maps of Student and Administrator Connections Within the Chatter Platform

All Followerships Formed **Student-to-Student Followerships Only**

[4] This claim is based on statements made by students in focus groups I held with undergraduate students currently enrolled in the College of Agriculture and Life Sciences during the Fall of 2016.





Note: Two-dimensional network maps have been produced using the Kamada Kawai force-directed algorithm.

Table 3: Network Summary Statistics of Students Graduating 2015 - 2020

	All Students in Followership Relationships $N = 356$
Density	.002
Transitivity	.120
Reciprocity	.215

To further characterize students' peer networks, I include network-level Markov Chain Monte Carlo Maximum Likelihood parameter estimates of Exponential Random Graph Models (ERGMs) and their standard errors for the Chatter followership network for several subpopulations. Increasingly used in economics[5] to explain network structure through micro-level individual behavior, these ERGM models specify a "distribution" of networks which display specified network properties, and determine which hypothesized individual-level behaviors make the observed network structure more or less likely. A key advantage of these models is that they do not require the behavioral assumption of independence necessary for logistic regression. Using an undirected dichotomous measure of followership, this analysis explores how similarity along observable dimensions impacts the likelihood of two individuals forming a

[5] For a discussion see Chandrasekhar and Jackson 2014.

social connection, and how this likelihood differs based on students' observable characteristics. Due to the sparseness of the network (evidenced in Table 3) I include limited network-level controls.

Table 4: Exponential Random Graph Models – Salient and Non-Salient Visible Characteristics

	All 'Followerships'	Includes Profile Photo	Excludes Profile Photo
	(1)	(2)	(3)
Edges	-6.787***	-4.267***	-6.729***
	(0.314)	(0.525)	(0.368)
Gender	-0.034	-0.327	0.324*
	(0.130)	(0.170)	(0.156)
Ethnicity	-.520**	0.374	0.557***
	(0.163)	(0.142)	(0.157)
Citizenship	0.805**	-0.144	0.553*
	(0.267)	(0.472)	(0.280)
First-Gen. Status	0.776***	0.388	0.554*
	(0.183)	(0.262)	(0.250)
Matriculation Year	0.490***	0.388*	0.538***
	(0.129)	(0.262)	(0.155)
Model AIC	3,039	742	2,047
Residual deviance	3,027	730	2,035
(df)	(52,206)	(4,154)	(29,750)

Standard errors in parentheses. * Indicates coefficient significantly different from zero at 95 percent confidence; ** indicates coefficient significantly different from zero at 99 percent confidence; *** indicates coefficient significantly different from zero at 99.9 percent confidence.

Table 4 reports ERGM estimates the impact of shared gender, race, citizenship, first-generation status and cohort on the likelihood of formation of friendship ties in Chatter, estimating a uniform coefficient for each attribute's impact on network formation. Each estimated model includes controls for the density of the network but lacks a control for transitivity; this methodological choice is informed by both the low level of transitivity in all network instantiations and the failure of all models including this term to converge. Given that sorting along observables is a natural network-level tendency[6] (the sociological

[6] The statistical importance of homophily for network formation is well-documented in sociology and along observable characteristics such as race and gender, as well as along unobservable characteristics such as values and attitudes. This tendency has been observed in closely related work examining the social connectedness of college students. For example, Marmaros and Sacerdote (2006) find that race and residential proximity are important determinants of Dartmouth students' social interaction. Mayer and Puller (2008) find that two college students on Facebook are more likely to form friendships if they are of the same race, major, cohort, and or political orientation. Homophily may arise through individuals' selection of one another based on similarity ('preference'); or through

tendency of *homophily,* or "similarity breeds connection" (McPherson, Smith-Lovin, & Cook, 2001)), a robustness check on the assumption that Chatter ties represent a reflection of in-person ties follows from the role of these observables' visibility within the platform on the likelihood of tie formation.

In an online environment where individuals are less able to perceive the observable characteristics of their peers, homophilic selection mechanisms should be mitigated – individuals are less likely to select peers who are like them when comparability cannot be observed. If followership is merely a reflection of some other peer network, however, we should not see increasing sorting on observables when these observable characteristics (particularly gender and ethnicity) are saliently featured in a profile photo. In Table 4, I estimate identical ERG models, analyzing separately ties where the individual who was followed *had* uploaded a profile image of themselves (Table 4, Column 2), and ties where the individual who was followed had not uploaded a profile image of themselves (Table 4, Column 3). Column 1 of Table 4 provides ERGM parameter estimates for all peer-peer social connections in the network.

Coefficients given are log-odds, which estimate how much more or less likely a social connection is to form between two individuals who share a given characteristic, controlling for reciprocal tie formation[7]. From Column 1 we see that across the cohorts of undergraduates, consistent with homophily, having in common any of U.S. citizenship, first-generation status and cohort year make ties more likely to form between individual students. From columns two and three I observe evidence inconsistent with followerships as social connections which originated on Chatter – the social connections I observe in Chatter are more likely to include sorting on ethnicity and gender if the person being followed *did not*

---

individuals' adaptation to become more like those who are close to them ('influence'); or through the increased likelihood of similar of individuals to interact with one another and to form ties ('propinquity').

Recent work in network analysis has investigated the role of homophily in virtual environments. Huang, Shen, and Contractor (2013) find that for online gaming, offline homophily in age and in geographic space continue to have a robust effect on network formation. Tarbush and Teytelboym (2012) use data from Facebook and find that homophily is operative through the propensity of individuals to be friends with those who occupy similar social position (operationalized by similarity in number of Facebook friends).

[7] Reciprocal tie formation ("Edges" in Table 3) is a property of relational networks which has been consistently observed by sociologists. It is best practice to control for this property when estimating ERGMs.

include a profile image in their Chatter profile than if they did. While statistical power is indeed limited by the small number of individuals which have uploaded photos, we see that even point estimates contradict sorting on observables for these populations. Network analysis provides evidence supporting the assumption that Chatter ties reflect real-world relationships, or at the very least that these followership ties reflect relationships are not exclusively virtual.

II.I.II Group Membership as an Idiosyncratic Source of Information

Because of the important role of *groups* within Chatter, information transmission along followership ties alone is unlikely to be informative about the true spread of information in the network. Much like Facebook, students on Chatter may become members of *groups* organized around a particular topic. Some groups are open to all students ("public groups"), others may be joined only at the invitation of another student ("private groups"), while yet others a student may be automatically enrolled in, at the behest of platform administrators ("exogenous groups"). All users are by default members of a "Students of the College" group, several groups determined by administrators, and groups defined for their major. The number of groups on the platform is dynamic, but at the time of this analysis, there were 210 groups available on the platform. Unlike followership relationships which are opted-into by a relatively small set of students, nearly all students choose to join at least one group in Chatter. Examples of public and private groups include "Nutritional Sciences", "Career: Surprise me!", "Peace Corps", and "Colorado". Examples of exogenous groups include "Chatter Fixes", "Transfer Students", "CALS Outgoing Exchange – Spring 2015", and "HR – Term Notice System". Joining a public or a private group allows students to access a separate page within Chatter, to post and to view postings of other students and administrators who have also joined the group. Exogenous groups, however, also provide a way for platform administrators to organize particular groups of students and to send notices.

Each week, students' activity is summarized in a *Weekly Digest* which contains the contents of each message posted that week to every group to which a student belongs – *including from exogenous groups*. This *Weekly Digest* also contains messages that were sent directly to the student, and the profile image and name of each message's author. Due to an unintended feature of the platform, students receive a redundant *Weekly Digest* alert for each group to which they belong. If a student belongs to G groups, that student receives each message posted within the groups to which they belong G times. I use the idiosyncratic variation introduced by student exogenous group membership to instrument for peer message receipt. The first-stage relationship between group membership and message receipt is strongly positive (shown in Table 5).

In practice, the primary way in which students get information from Chatter is through *Weekly Digests*. This insight comes from speaking with platform administrators, and from a convenience survey I conducted during October of 2016 of 112 undergraduate students enrolled in a large undergraduate statistics course at Cornell. Students were asked directly about their platform use, followership choices, and about the informational value of these choices (see Appendix for full survey text). While students report using Chatter to get information about social activities (11.02% of respondents), registration (29.06% of respondents), and classes (33% of respondents), this information comes from administrators in the platform rather than from other students. The majority of students (n=74 or 62.71% of students who answered this question) reported that the primary way that they receive information from the platform is via their emailed posting digests, rather than through their followership relationships. An additional 10.17% (n=12) of students who answered this question reported that they do not get any information at all from the platform. This survey evidence is consistent student statements made during focus groups I conducted during the Fall of 2017.

II.II Roommate Data

I supplement this set of Chatter data with an additional measure of students' peer groups from observing the first-year on-campus housing assignments of all students attending Cornell who are graduating or are scheduled to graduate during the years 2015 through 2019. For students not requesting a specific roommate, freshman roommate assignments at Cornell are made using a third-party software which assigns students randomly to dorm rooms conditional on students' requested housing configurations (whether single, townhouse double, double, triple, quad, or quintuple), gender preferences (single gender or gender-inclusive) and their responses to a lifestyle questionnaire distributed to students with their admissions decisions at the beginning of April. The housing questionnaire asks students five-point Likert scale questions about: 1) their sleep patterns (*"I tend to go to bed at…"*, with responses ranging from *"10:00 P.M."* to *"2:00 A.M."*), 2) their musical interests (*"How often do you listen to Classical/Country/Hip-Hop/Latin/Pop/Rock?"*, with responses ranging from *"Always"* to *"Never"*), 3) their room condition (*"My room is generally…"* with responses ranging from *"Neat"* to *"Messy"*, 4) their preferred level of room sociability (*"The social condition of my room will most likely be…"*, with responses ranging from *"Lively"* to *"Quiet/Reserved"*), 5) their preferred level of sleep background noise (*"I sleep with background noise (music, TV, fan, etc.) or a light on in [the] room…"*, with responses ranging from *"Always"* to *"Never"*), 5) their smoking behaviors (*"I smoke"*, with responses of either *"Yes"* or *"No"*), 6) their preferred level of study background noise (*"I study with background noise (music, TV, fan, etc.)"* with responses ranging from *"Always"* to *"Never"*), and 7) what time they tend to wake up (*"I tend to wake up at…"* with responses ranging from *"6:00 A.M."* to *"11:00 A.M."*). Summary statistics from this survey is summarized in Appendix Table A1. In practice, staff prioritize matches between potential roommate pairs based on ambient noise during studying and smoking preferences, allowing software to make random assignments between students within these matched groups.

Since I do not have access to pre-Cornell student characteristics, I cannot empirically test the goodness of the conditional randomization of roommate assignment. The exogeneity of this peer group, however, is not key to identification.

III. Empirical Framework

Consider a simple framework where course choices are a function of the number of messages individuals receive and of unobserved individual-level characteristics and preferences. Some of the many unobserved characteristics and preferences which also influence students' course choices include preferences for a particular subject or professor, student desires to enroll in an easy or a challenging class, or the messages about a course that a student receives which I do not observe on the Chatter platform. Despite these unobserved factors, I obtain consistent estimates of the effect of *tagged* message receipt, or the receipt of a message which is endowed with the peer characteristic of the individual who delivered it, on student behavior as long as my instrument, group membership, is orthogonal to all of these factors.

I estimate the reduced form effect of *tagged* messages from student *i*'s peers about course c_k on the likelihood that student *i* chooses to enroll in course c_k. I consider separately as peers 1) roommates, 2) people individual *i* follows on Chatter and 3) people that follow individual *i* on Chatter. I consider a simplified setting in which in each semester *t,* student *i* faces a choice between one or more elective classes $c_1, c_2, c_3, \dots, c_K \in C$. For my empirical analysis, I let C be the set of the 25 most popular classes offered in the College of Agriculture and Life Sciences across all students in semester *t,* for a total set of 2,796,046 actual and potential course choices. Across all preceding time periods, student *i* may receive a message, $m_{i,j,k}$ from his peer individual *j* regarding course c_k.

- Then $m_{it,k} = \sum_{j=1}^{J} m_{i,j,k}$ defines the number of messages received by individual *i* by time *t* regarding course c_k from her peers.

- After the conversation phase, each person forms a belief about which class she prefers based on m_{it} and her own (unobserved) characteristics, $\boldsymbol{X_i}$. Each individual i then chooses one or more courses in which to enroll.

To define m_{it} in an empirically useful way, I map each raw message text from Chatter onto an indicator variable for the course to which this message pertains using textual analysis[9]. To determine whether this text was a statement (or a question), I use a popular textual analysis software, Linguistic Inquiry and Word Count (LIWC), to analyze whether or not each message includes interrogative words[10]. I define that all statements relevant to the courses in an individual's choice set (that is, the courses that an individual chose or could have chosen) belong to m_{it}. Since messages are *tagged*, this allows me to make inferences about the importance of message source in student decision-making.

Due to the presence of a discrete endogenous regressor, I use a reduced form ordinary least squares model to estimate the above:

$$y_{itc} = \alpha + \beta\widehat{m_{it,c}} + \epsilon_{itc} \qquad (1)$$

Where y_{itc} is a dummy variable which equals one if individual *i* chose course *k* in semester *t,* 0 otherwise. I use robust standard errors. In the first stage, I instrument for $m_{it,c}$, the number of unique sources of information about course *k* which individual i receives, using the individual's group membership at the time that each message was sent. Estimates of the effect of message receipt from the two-stage least squares regression are shown in Table 6, shown separately for groups of peers defined as 1) roommates, 2) people individual i follows on Chatter and 3) people that follow individual i on Chatter. Two-stage

[9] An author-written custom dictionary was used within R text mining package 'tm'. Code to replicate this text analysis is available in from the author's website.
[10] I use the variable *Interrog* to quantify a text's sentiment and whether or not it contains a question. Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

least squares estimates are not shown for roommates, given the weakness of the first stage which is driven by the fact that I observe relatively few messages transmitted on the platform (shown in Table 5).

IV. Results

Table 6 gives two-stage least squares estimates of the ordinary least squares model for the likelihood of choosing a particular class, using group membership to instrument for message receipt from each peer group. First, there is statistically significant evidence that across all messages received by students (including messages sent by administrators) the average effect of receiving an additional message about a course is roughly a 16% decrease in the likelihood of taking a particular course. This may be explained by students inferring course popularity based on message receipt and desiring to individuate, or from a disutility of receiving course related advertisements.

Estimates of the effect of peers' messages on the likelihood that a student enrolls in a given course are consistent with a model where peers' messages play a special role in student decision-making. While the effect in general of receiving a message is negative, the effect of receiving an additional message from a member of a set of peers is either a 3.38% increase in the likelihood of enrolling in a course or at 7.38% increase in likelihood on average. While I cannot jointly estimate the effect of communications from different sets of peers due to having a single instrument, coefficient comparisons suggest that the messages of peers who the individual has chosen to signal friendship with (by electing to *follow* the peer) are on average four percent more persuasive than the messages of peers who have signaled friendship with the individual (by *following* the individual). Notice that in the case of an individuals' followers, the individual has made no active choice within the platform to receive information from this peer (eliminating the possibility of motivated information seeking in this case), yet this peer's messages still influence own behavior.

Table 5: Dependent Variable: Number of Messages Received (First Stage)

	All Messages	Roommates	Follows	Followers
	OLS (1)	OLS (2)	OLS (3)	OLS (4)
Number of Exogenous Groups	0.0024*** (0.000)	0.000 (0.000)	0.222*** (0.000)	0.458*** (0.000)
N	2,834,210	2,834,210	2,834,210	2,834,210
R^2	0.077	0.000	0.228	0.228
F test: Number of Exogenous Groups Coef. = 0	16,576.63	4.000	8.90	9.92

Note: Robust standard errors are in parentheses. * Indicates coefficient significantly different from zero at 90 percent confidence; ** indicates coefficient significantly different from zero at 95 percent confidence; *** indicates coefficient significantly different from zero at 99 percent confidence.

Table 6: Effect of Message Receipt from Peers on Course Choice

	All Messages		Roommates		Follows		Followers	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)	OLS (7)	IV (8)
Number of Statements	-0.0266*** (0.000)	-16.12*** (0.125)	-0.0306*** (0.000)	--	-0.0000*** (0.000)	7.380** (2.474)	-0.0000*** (0.000)	3.380** (1.073)
Constant	0.0326*** (0.000)	1.270*** (0.009)	-0.0306*** (0.000)	--	-0.0309*** (0.000)	-52.57** (17.63)	-0.0308*** (0.000)	-49.76** (15.81)
N	2,796,046	2,796,046	2,796,046	--	2,796,046	2,796,046	2,796,046	2,796,046

Note: Robust standard errors are in parentheses. Semester-level fixed effects are included in all specifications (coefficient estimates not reported). * Indicates coefficient significantly different from zero at 90 percent confidence; ** indicates coefficient significantly different from zero at 95 percent confidence; *** indicates coefficient significantly different from zero at 99 percent confidence.

V. Discussion

Using a large novel set of deidentified undergraduate social network and course choice data from Cornell University undergraduates, I use text analysis and an instrumental variables approach to find evidence that undergraduate peers' messages are persuasive. When peers are defined by an individual's indication that she is friends with a peer (by *following* the peer), I find that peers' messages increase the individual's likelihood of taking the course by 7.4%. In the case of peers which have indicated friendship with the individual (the peer *follows* the individual), the peer's messages influence the individual's course choices by increasing the likelihood that an individual enrolls in the course by 4.4%. I am unable to identify any effect of roommates' communications on the likelihood that an individual chooses a particular course, though I am underpowered to do so.

Throughout, I have been intentionally agnostic about the specific nature of peer messages (whether statements of endorsement, advice, or information in the strict sense). An important question for planned future work remains as to whether the reduced form effect of social influence I identify is one is one of social learning. That is, do peers' signals help students to form better or more rational course choices that lead to better outcomes for the student?

References

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar. 2014. "Dynamics of Information Exchange in Endogenous Social Networks." *Theoretical Economics* 9 (1):41–97. https://doi.org/10.3982/TE1204.

Bala, Venkatesh, and Sanjeev Goyal. 2000. "A Noncooperative Model of Network Formation." *Econometrica* 68 (5):1181–1229. https://doi.org/10.1111/1468-0262.00155.

Case, Anne C., and Lawrence F. Katz. 1991. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." Working Paper 3705. National Bureau of Economic Research. https://doi.org/10.3386/w3705.

Chandrasekhar, Arun G., and Matthew O. Jackson. 2014. "Tractable and Consistent Random Graph Models." Working Paper 20276. National Bureau of Economic Research. http://www.nber.org/papers/w20276.

DeGroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69 (345):118–21. https://doi.org/10.2307/2285509.

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion Bias, Social Influence, and Unidimensional Opinions." *The Quarterly Journal of Economics* 118 (3):909–68. https://doi.org/10.1162/00335530360698469.

Ellison, Nicole B., Charles Steinfield, and Cliff Lampe. 2007. "The Benefits of Facebook 'Friends:' Social Capital and College Students' Use of Online Social Network Sites." *Journal of Computer-Mediated Communication* 12 (4):1143–68. https://doi.org/10.1111/j.1083-6101.2007.00367.x.

Gale, Douglas, and Shachar Kariv. 2003. "Bayesian Learning in Social Networks." *Games and Economic Behavior*, Special Issue in Honor of Robert W. Rosenthal, 45 (2):329–46. https://doi.org/10.1016/S0899-8256(03)00144-1.

Huang, Yun, Cuihua Shen, and Noshir S. Contractor. 2013. "Distance Matters: Exploring Proximity and Homophily in Virtual World Networks." *Decision Support Systems*, 1. Social Media Research and Applications 2. Theory and Applications of Social Networks, 55 (4):969–77. https://doi.org/10.1016/j.dss.2013.01.006.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3):531–42. https://doi.org/10.2307/2298123.

Marmaros, David, and Bruce Sacerdote. 2006. "How Do Friendships Form?" *The Quarterly Journal of Economics* 121 (1):79–119.

Mayer, Adalbert, and Steven L. Puller. 2008. "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of Public Economics* 92 (1–2):329–47. https://doi.org/10.1016/j.jpubeco.2007.09.001.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27 (1):415–44. https://doi.org/10.1146/annurev.soc.27.1.415.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M.E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Sacerdote, Bruce. 2001. “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” *The Quarterly Journal of Economics* 116 (2):681–704. https://doi.org/10.1162/00335530151144131.

———. 2011. “Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” In *Handbook of the Economics of Education*, edited by Stephen Machin and Ludger Woessmann Eric A. Hanushek, 3:249–77. Elsevier. http://www.sciencedirect.com/science/article/pii/B9780444534293000041.

Subrahmanyam, Kaveri, Stephanie M. Reich, Natalia Waechter, and Guadalupe Espinoza. 2008. “Online and Offline Social Networks: Use of Social Networking Sites by Emerging Adults.” *Journal of Applied Developmental Psychology*, Social Networking on the Internet, 29 (6):420–33. https://doi.org/10.1016/j.appdev.2008.07.003.

Tarbush, Bassel, and Alexander Teytelboym. 2015. “Social Groups and Social Network Formation.” SSRN Scholarly Paper ID 2456532. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2456532.

Zimmerman, David J. 2003. “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.” *The Review of Economics and Statistics* 85 (1):9–23. https://doi.org/10.1162/003465303762687677.

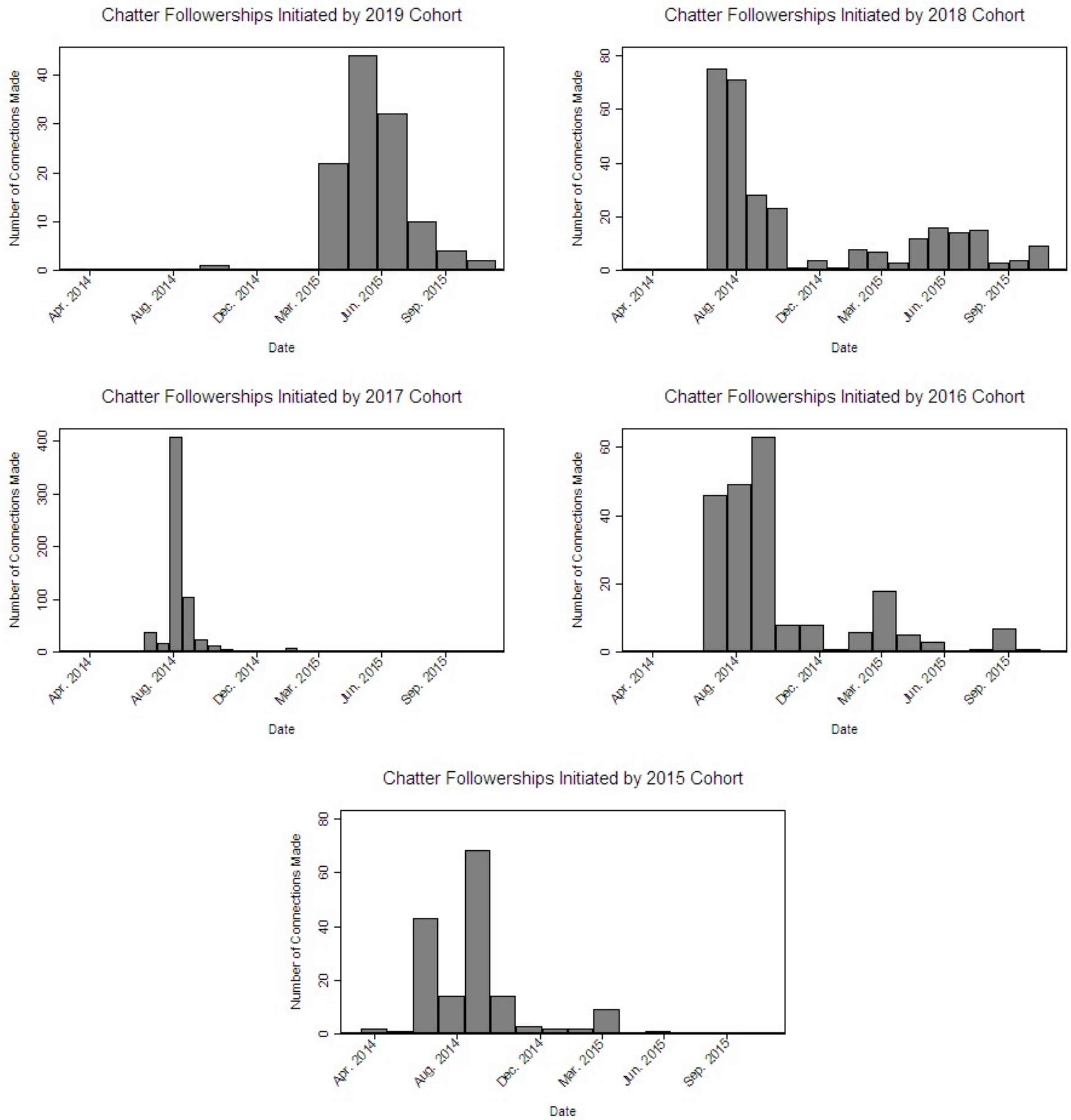
APPENDIX

Appendix Table 1: Summary Statistics of Freshman Housing Questionnaire Responses (2015-2019 Cohorts)

Cohort	2015	2016	2017	2018	2019
Going to Sleep	3.019	3.039	3.114	3.109	3.019
(1 = 10:00PM)	(.016)	(.017)	(.017)	(.017)	(.017)
Room Condition	2.372	2.368	2.437	2.435	2.465
(1 = Neat)	(.017)	(.017)	(.016)	(.017)	(.017)
Room Sociability	2.856	2.910	2.876	2.927	2.912
(1 = Lively)	(.019)	(.018)	(.018)	(.018)	(.018)
Sleep Noise	4.127	4.088	4.068	4.035	4.050
(1 = Always)	(.019)	(.019)	(.019)	(.019)	(.019)
Smoking	4.947	4.960	4.946	4.930	4.932
(1 = No)	(.008)	(.007)	(.008)	(.009)	(.009)
Study Background	3.703	3.689	3.622	3.561	3.550
(1 = Always)	(.019)	(.019)	(.019)	(.019)	(.020)
N	3,302	3,212	3,210	3,219	3,166

Note: Standard errors given in parentheses.

Figure 2. Chatter Followerships Initiated Over Time by Matriculation Cohort



Note: By cohort, this figure plots student decisions to follow another person on the Chatter platform between March 1, 2014 and November 1, 2014. Across all cohorts admitted to Cornell at the time, the number of connections peaks during the start of the student registration period in 2014, the first registration period during which the Chatter platform existed. These connections were sticky and impact student information receipt throughout their tenure as students.

Survey of Undergraduate Students Concerning Their Chatter Use (*n*=112)

Survey of Social Network Use (7 Questions)

Are you a student in the College? Yes No Don't Know

How many of your close friends do *you follow* on Chatter? _________

How many of your close friends *follow you* on Chatter? _________

If you need information **about social activities**, what social networks do you use to get answers?
[PLEASE SELECT ALL THAT APPLY]

☐ Academia.edu
☐ Chatter
☐ Facebook
☐ Google Plus
☐ LinkedIn
☐ Reddit
☐ Slack
☐ Tumblr
☐ Twitter
☐ Vine

If you need information **about registration**, what social networks do you use to get answers?
[PLEASE SELECT ALL THAT APPLY]

☐ Academia.edu
☐ Chatter
☐ Facebook
☐ Google Plus
☐ LinkedIn
☐ Reddit
☐ Slack
☐ Tumblr
☐ Twitter
☐ Vine

If you need information **about classes/coursework**, what social networks do you use to get answers?
[PLEASE SELECT ALL THAT APPLY]

☐ Academia.edu
☐ Chatter
☐ Facebook
☐ Google Plus
☐ LinkedIn
☐ Reddit
☐ Slack
☐ Tumblr
☐ Twitter
☐ Vine

In what way do you *most frequently* get information from Chatter?

- ☐ View email digests
- ☐ Seek information directly from the website
- ☐ Receive direct messages from others on the platform
- ☐ Other (Please specify): _________________