The Strength of Weak Ties: Causal Evidence using People-You-May-Know Randomizations

Guillaume Saint-Jacques, MIT and LinkedIn
Sinan Aral, MIT
Erik Brynjolfsson, MIT
Ya Xu, LinkedIn
Edoardo M. Airoldi, Harvard

Abstract

The causal relationship between tie strength and labor market outcomes is of interest to a large variety of actors, from individual workers seeking to optimally allocate their resources as they develop their own social network to firms seeking to leverage candidates' networks in their recruitment process. It is also of interest to a social planner or a professional social network platform interested in increasing efficiency in labor market matching processes or increasing equality of opportunity. Using a number of "People You May Know" experiments (testing recommendation algorithms) conducted at LinkedIn between 2014 and 2016, we seek to identify the sign of the causal relationship between tie strength and labor market mobility in two different ways. First, by conducting an edge-level regression of job transmission on tie strength using a PYMK randomization as an instrument. Then, with an individual-level regression of number of jobs reported on individual network clustering coefficient, using over 700 past treatments as instruments with regularization. Both sets of results point to decreasing returns in the relationship between structural tie strength and mobility. These results indicate that a strong tie is not always individually more useful than a weak one, and that the most useful ties are likely not the weakest or the strongest, but the ones that strike a good compromise between strength and diversity.

Introduction

Tie strength may impact worker's labor market outcomes in a variety of ways. A commonly studied phenomenon is job transmission over social ties. If an individual works at a certain firm, she may inform her social ties of job openings, or leverage her personal knowledge of her friends and their abilities to help her firm quickly (and cheaply) identify promising candidates - something many companies encourage through a referral bonus. This results in an individual's connections having a higher likelihood of ending up working at the same firm as her. In a survey, (Granovetter 1973) finds that more people report obtaining their job through a weak social tie than through a strong one. This observation, however, leaves open the question of whether weak ties are cited more often by respondents simply because they are more numerous, or because they are also individually more valuable (for the purposes of finding a job) than strong ones. Two main challenges stand in the way of answering this question. First, until recently, the data required to observe individual's networks and job transmission was not available. Estimating the strength of a tie requires rich data, measuring the intensity of interactions between any pair of individuals, and complete social network data, allowing to compute structural measures of tie strength (like number of friends in common).

Furthermore, as job transmission is a rare event, large-scale data collection is necessary.

The second, more fundamental challenge is one of simultaneity and endogeneity. Individual's labor market outcomes both determine and are determined by their social networks, and the evolution of both is very likely to be correlated with a number of unobservable factors confounding correlational analyses. An individual's network will likely influence the individual's job market options, but an individual's endeavor toward switching jobs may also lead the individual to grow his/her network in a certain way. In this paper, we address this second challenge by using experimental data, rather than observational data.

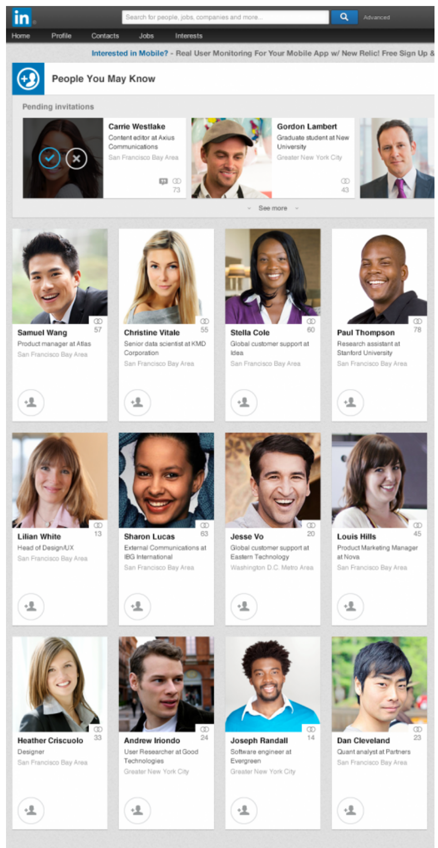

Figure 1 Screen Capture of People-You-May-Know page on LinkedIn

Strategy no. 1: Edge-level regression with one past randomization as an instrument

An in-depth observational investigation of the relationship between tie strength and job transmission can be found in (Gee, Jones, and Burke 2016; Gee et al. 2017) construct a proxy measure of job transmission based on three criteria:

- User A reports working at company *c* at date D_1.
- User B report working at that same company *c* at a later date D_2, with D_2 and D_1 being at least one year apart.
- User A and user B were friends on the social network at least one full year before D_2.

When these three criteria are met, a tie is then tagged as a “sequential job” tie. The paper then shows a series of edge-level regressions (where the unit of analysis is the tie itself, as opposed to the individual), using the above measure as the dependent variable and various measures of tie strength as regressors. The authors find a positive correlation between tie strength (as measured by interactions on Facebook and number of friends in common) and job transmission. We are interested in a somewhat different question: causally, at the margin, when adding a new tie, does tie strength have positive or negative impact on the probability of job transmission? Where is one's energy better spent: developing strong ties or weak ties? To answer this question, we rely on a past randomized experiment as a source of exogenous network variation.

Data

To carry out the edge-level analysis, we choose to remain close to Gee et al's (2017) approach of using a regression, in which the unit of analysis is the social tie, the dependent variable is a binary indication of a *sequential job tie*, as defined above, and various quantifications of tie strength are included among the regressors. In particular, we contrast two quantifications of tie strength: interaction intensity and number of connections in common. Departing from a purely observational approach, in order to obtain causal identification of the role of strong ties on job transmission, we rely on a two-week experiment conducted by LinkedIn's People You May Know (PYMK) service[6], which recommends possible new ties to users when they log in to the site, conducted in early 2015. The experiment we exploit was carried out to test several different (randomly allocated) tie recommendation algorithms. We construct a sample of edges two years after the experiment, and compute a *sequential job* binary indicator using the same definition as in Gee et al. . However, we also include in the sample all individuals with ties that were created as a result of the experiment, and not only the ones that ended up resulting in a job transmission[1].

As a first measure of tie strength, we compute the intensity of interactions among users along various dimensions, such as interactions through the feed, messaging, recommendations, and others. These scores along these dimensions are then averaged and scaled to produce a variable labeled *interaction intensity*. As a second measure of

[1] Because of this, the frequency of sequential jobs observed in the data set we construct is about ten times smaller than in Gee et al (2017-1). If we were to restrict the sample to only individuals with at least one recorded sequential tie, the estimated frequency of sequential job ties would be about 3%, which is close to the one reported by the authors.

tie strength, we count the number of friends any two connected individuals had in common when the tie was created. Individual's degrees are also entered as controls in our regression.

In the randomized experiment that provides exogenous variation in our analysis, several tie recommendation variants were allocated using Bernoulli randomization across users of the platform. Relatively to the control variants of the recommendation algorithm, treatment variants recommend more triad-closing ties. Millions of invitations were sent which ended up being accepted. We restrict our analysis to these accepted edges, and use, as an instrument for our two measures of tie strength, the several treatment variant assigned to the invitation sender.

Results

Table 1 shows an excerpt of the regression results. The first column shows the results of a simple OLS regression of the indicator of whether each tie resulted in a sequential job on our measures of tie strength and a number of control variables. Controls, such as whether both individuals went to the same school, or whether they are located in the same region or city, are not shown; neither is their age difference. The second column shows the results of the same model specification, but employing a Probit model instead. Finally, the third column shows the results of a two-stage least square (2SLS) estimation. The results of the first two columns are broadly consistent with the observational results of (Gee et al. 2017): more interaction and more structural closeness (as defined by the percentage of friends in common) is associated with a higher likelihood of job transmission. Also similar with their observational results are the coefficients on the controls: similarity between members increases the likelihood of job transmission, and dissimilarity decreases that likelihood. Having gone to the same school, or living in the same region or the same city is associated with more job transmissions, whereas greater age differences are associated with fewer.

We conduct a number of tests in order to check whether our instrument is weak or not. Our 2SLS estimates are very similar to other available instrumental variable estimators, such as Limited Information Maximum Likelihood or Fuller. Our first-stage F-statistic, which is well over 10, and an Anderson-Rubin test for weak instruments rejects the null that the instrument is weak.

Table 1 Dependent variable: Sequential Job indicator variable

Dependent variable: Sequential Job dummy			
	OLS	Probit	2SLS
Interactions (messages)	**.0002*****	**.018*****	**-.005**
% friends in common	**.081*****	**8.779*****	**1.882*****
…High % friends in common (dummy)	**.002*****	**.178*****	**-.061*****
… Interaction term	**-.079*****	**-8.740*****	**-.884****
Same school	.0003***	.019***	-.002
Age difference	-.00001***	-.001***	-.00001
Same region	.001***	.130***	.001
Same city	.001***	.088***	.001
Ego connection count control	Included	Included	included
Alter connection count control	Included	Included	included
Constant	included	included	included
Same-gender controls	included	included	included

Observations	19,763,317	19,763,317	19,763,317

Nonlinear relationships.

Figure 2 and Figure 3 (axes scales hidden for confidentiality) illustrate nonlinear relationships between the two measures of tie strength we consider and the probability of job transmission. Figure 1 shows the relationship between interaction intensity and the probability of observing a sequential job. The black dots show the unconditional relationship as present in the raw data, and the blue dots show the interaction intensity as predicted from the first stage equation (all edges in our dataset are binned into ten groups which are then shown on this scatterplot).

Figure 2 shows the relationship between interaction intensity and job transmission: the more people interact on the platform, the more likely it is that one will end up working for the same company as the other one.

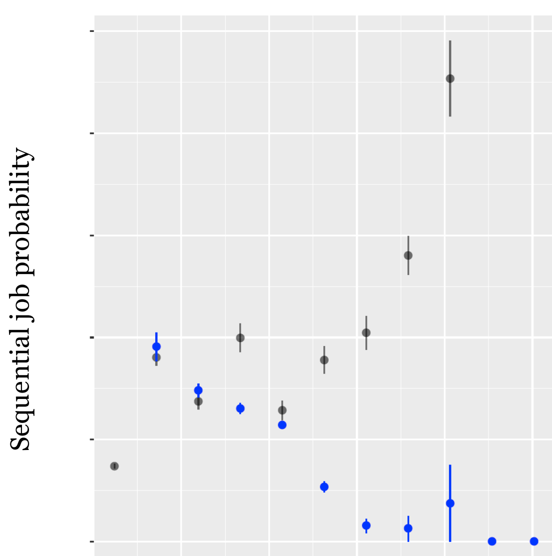

Figure 3 Interaction Intensity

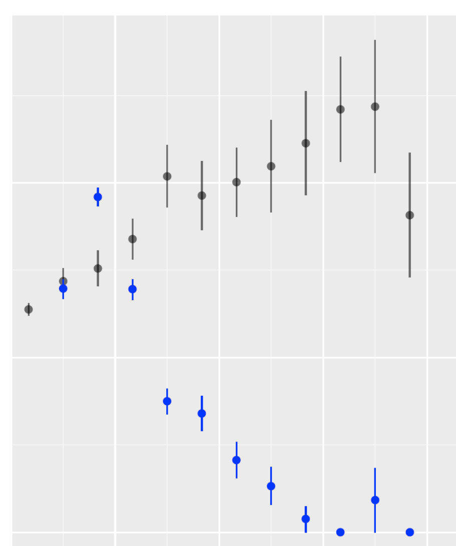

Figure 3 Number of connections in common

Figure 3 shows the relationship between probability of sequential jobs and number of connections that the ego and the alter had in common before the tie was created; i.e. the number of open triangles the tie closed when it was created. In both figures, the difference between the blue dots and the black ones is striking. When looking at a pure correlation (black dots), we see a positive relationship; descriptively, when people have more friends in common, a sequential job is more likely to be observed. However, leveraging the experimental data and using fitted values from the first stage (blue dots) reveals a more complex relationship. When disconnected individuals already have many friends in common, the tie that results from nudging them to connect has a lower probability of leading to a sequential job than if they had few friends in common. One can hypothesize two mechanisms that might lead to this result. First, it is possible that such a tie is redundant, i.e., the individual most likely already has access to most of the information about job openings and recommendations from the preexisting friends in common. Second, it is possible that triangles that are still open after the network has been evolving for a long time may be ones that both individuals are reluctant to create, for example, if they know but dislike each other. In both cases, one would expect that nudging the relevant individuals to connect on may be relatively unproductive. Similarly, artificially creating ties, or suggesting ties, between individuals with no friends in common also seems to have low value. This may be because high network distance is the sign of large differences between individuals, so that they have little to gain from becoming connected.

Strategy no. 2: Individual-level regression with many past randomizations as instruments.

In this approach, we shift approaches from edge-level regression to individual-level regression. The LinkedIn teams developing new tie recommendation algorithms carry out randomized experiments to evaluate the performance of the new algorithms routinely.

Here, we use many of these experiments as instruments. For the purpose of this analysis, we restrict our attention to a subset of the LinkedIn graph, namely all users reporting a location within the San Francisco Bay area. At the end of each PYMK experiment, we construct a graph consisting of the edges that existed at that date, and compute the clustering coefficient for each member. The clustering coefficient, computed at the individual level, is the ratio of the number of ties existing in an individual's 1.5-out ego network to the number of ties that would exist if the 1.5-out ego network were fully connected (i.e. if all of the focal individual's peers were connected to each other). In other words, this gives us the proportion of closed triads around the focal individual. We use this as our variable of interest. Each experiment has many variants, leaving us with over 700 experiment-variant combinations, which are all potential instruments.

Table 2 Individual-level regression coefficient, with and without instrument selection

	All Instruments	**With Instrument Selection**
Clustering coefficient	**-10.25**	**-15.83**
s.d.	5.07	8.66

P-value	0.0433	0.0682

As a dependent variable, for each member, we count the number of different positions that are listed on her profile. We collect this value at the end of each experiment. Using different outcome variables, such as progress in seniority levels, or number of firms worked for, does not significantly change the outcome. We follow the approach of instrumental variable cross validation proposed by (Peysakhovich and Eckles 2017) in order to select the strongest instruments. The procedure selects 493 instruments, and reveals a negative effect of increasing clustering coefficient (closing triangles around individuals) on number of jobs reported. In this specification, increasing the clustering coefficient by 10% would decrease the number of positions reported by 1.5. It is likely that closing too many triangles around an individual fills her ego-network with relatively less useful ties, and reduce her exposure to novel information and to different firms, therefore reducing her labor market mobility options, resulting in a lower number of reported jobs and positions over time.

References

Gee, Laura K., Jason J. Jones, Christopher J. Fariss, Moira Burke, and James H. Fowler. 2017. “The Paradox of Weak Ties in 55 Countries.” *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2016.12.004.

Gee, Laura K, Jason Jones, and Moira Burke. 2016. “Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook’s Social Network.” *Journal of Labor Economics* 35 (2). The University of Chicago Press:485–518. https://doi.org/10.1086/686225.

Granovetter, Mark S. 1973. “Granovetter - 1973 - The Strength of Weak Ties.” *American Journal of Sociology*. https://doi.org/10.1037/a0018761.

Peysakhovich, Alexander, and Dean Eckles. 2017. “Learning Causal Effects from Many Randomized Experiments Using Regularized Instrumental Variables,” January.

http://arxiv.org/abs/1701.01140.