LinkedIn Economic Graph

WORKING PAPER

Measuring Gender Gaps in Economic Network Strength in the US

Matthew Baird
LinkedIn Economic Graph Research & Insights

Danielle Kavanagh-Smith
LinkedIn Economic Graph Research & Insights

Osonde Osoba
LinkedIn Responsible AI

Yuwen Wu*
Docker
*All work was done while at LinkedIn

SEPTEMBER 2023

EG WP No. #3

Measuring gender gaps in economic network strength in the US

Matthew Baird[1], Danielle Kavanagh-Smith[2], Osonde Osoba[3], and Yuwen Wu[4]

September 2023

ABSTRACT[5]

We build on prior research in both economics and network science by constructing a new framework of network strength that combines four distinct elements of network strength into a cohesive model, all in the scope of how a network could help with career advancement: (a) the value of the information your connections have for you, (b) the bandwidth of information sharing between you and your connections, (c) the redundancy of information in your network based on shared connections, and (d) the overall size of your network. We estimate the elements of this model and overall network strength using proprietary data across millions of members of LinkedIn in the United States. We use this to explore network equity between men and women. We also estimate the degree to which observable characteristics (such as education, industry, occupation, and age) can explain these differences. We then regress economic outcomes (such as employment status, recruiter outreach, and seniority) on these network strength elements to explore the extent to which observed gender gaps narrow when controlling for these network elements.

Keywords: Networks, gender, equity, labor
JEL classifications: J16, J24, I29

[1] Corresponding author. LinkedIn. mdbaird@linkedin.com
[2] LinkedIn. dkavanaghsmith@linkedin.com
[3] LinkedIn. oosoba@linkedin.com
[4] Docker. yuwen.wu@docker.com; All work was done while at LinkedIn
[5] We gratefully acknowledge the support and feedback of many individuals in this report. These include Simone Bell, Gorki De Los Santos, Meg Garlinghouse, Sakshi Jain, Karin Kimbrough, Jen Ramos, Carl Shan, and Amie Wong.

1. Introduction

Economic networks have gained increased recognition as key determinants in fostering career opportunities, knowledge exchange, and professional growth in today's knowledge-based economies. Economic networks encompass the relationships and connections individuals establish within their professional spheres, including colleagues, mentors, clients, and industry peers. These networks serve as channels for information exchange, collaboration, and opportunity sharing. Increasing digitalization, globalization, and interconnectedness has amplified the significance of networking as a strategic tool for professional growth and advancement. Economic networks provide individuals with access to a wealth of opportunities, such as job leads, referrals, skills to develop, resources for attaining goals, and social capital, mentorship, and access to influential decision-makers. The strength of an individual's economic network can significantly influence their career trajectory and economic outcomes.

Moreover, economic networks contribute to the diffusion of knowledge, innovation, and entrepreneurial activity. The exchange of information and ideas within networks can spark collaborations, promote the spread of best practices, and facilitate the transfer of industry-specific knowledge. Thus, economic networks not only benefit individuals but also have wider implications for organizational performance, industry competitiveness, and economic development at large.

Women still face more barriers in the labor market than men in the United States (US). Despite progress over the past few decades, there are persistent labor outcome disparities between men and women in the US economy for such outcomes as earnings, promotion to leadership, and general career progression (Graf et al., 2018; Lara et al., 2023; Maasoumi & Wang, 2019; Petrongolo, 2019). One potential mechanism for this persistence may be differences in the economic networks that these groups have. To examine differences in network quality between men and women, we need a useful framework for parsing or understanding the quality of member networks. In this paper, we develop a new methodology for measuring network strength which combines elements from social science literature on network strength from economics and sociology with research from network science. We apply this network strength model to the US and examine the extent of the gender gap in network strength, as well as which features of a network display the largest disparities. These gender gaps in economic network strength can stem from various underlying factors, including social norms, implicit biases, and structural barriers. Understanding and addressing these disparities are crucial steps towards achieving gender equality in the realm of professional networks.

By delving deeper into the factors that shape economic network strength and analyzing the gender gaps within each of these factors, this research aims to provide a nuanced understanding of what it means to have a strong economic network, and to examine potential challenges women face in the formation of strong networks. Moreover, by controlling for observed factors such as occupation, industry, and educational background, we can discern whether the observed gender disparities are solely a reflection of structural differences in these factors, or if additional biases and barriers persist. The findings of this study hold the potential to inform evidence-based policies and interventions that foster gender equality in economic networks, ultimately leading to more equitable economic outcomes and societal progress.

1.1. Related Literature

Our model builds on the important work of past research. Most concretely, our model builds on that of Aral (2016). Aral frames network strength in terms of information sharing, as we do here. Additionally, he describes several features of a network that we construct directly into our model, including information value (the extent to which a given connection can provide useful information), channel bandwidth (the likelihood of that information being shared), and network diversity (what we call information non-redundancy). Many papers have focused in particular on network diversity and its potential interplay with bandwidth (Bruggeman, 2016). They acknowledge the tension between the strength of a network connection—or tie—and the value of the information that tie has. The underlying concept is that stronger ties—people to whom we are closest—have high information bandwidth and are thus more likely to share information. But they have less valuable information, because it is information, people are likely to already possess or receive through other channels. Thus, weak ties—people with whom we share few mutual connections—are potentially the most valuable even though people may have lower average bandwidth communication with them. Recent work examining LinkedIn networks provided strong, empirical evidence that supports the idea that weak ties connections are the most beneficial to career progression (Rajkumar et al., 2022).

There is also a strong history of evaluating network disparities by gender and other dimensions. McDonald et al. (2009) find that men receive more job leads than women from routine conversations (as well as a racial divide for the same outcome). Further, networks with higher proportions of White men have more total job leads shared than women and minority dominated networks. McGuire (2002) examined financial services employees and found that, within their firms, women received less help from their network than men do. These differences persisted even when controlling for worker characteristics as well as other network characteristics including measures of network strength, closeness, and bandwidth. This supports a larger conceptualization of what it means to have a strong network.

An additional body of work has examined the extent to which gaps in network features between groups are associated with gaps in economic conditions, as well as when even having similar features of networks can result in different outcomes. For example, Pedulla & Pager (2019) find that Black and White workers leverage their networks at similar rates and even receive job leads at similar rates, but that the networks have differing levels of effectiveness. When each group leverages their network to seek jobs, having a higher number of White connections is associated with a higher likelihood of receiving a job offer than having more Black connections. Barbulescu (2015) finds that having diverse networks benefits workers at different stages of the job search process, although having more focused groups of connections by occupation benefits workers in moving forward in the interview process. Childers & Kaplan (2023) investigate the role of networking among women in leadership in business and find that women report networking being critical to their success—and that they indeed network often, many of them daily. However, only 32% of the respondents reported having a mentor. The economic connectedness of social networks has also been linked to economic outcomes such as income mobility (Chetty et al., 2022a, 2022b).

Thus, the literature has established that networks are important, and that there are critical divides by gender. However, we build on the literature by creating a cohesive model of economic network strength and estimating it using proprietary LinkedIn network data.

2. Model

Our model of network strength is built around the concept of the quality and amount of information an individual is likely to receive from their network that could help them in their career. We model it borrowing from the literature where possible, and base it on four components which we bring together into one cohesive metric. The four elements are information value, information bandwidth, information redundancy, and network size.

2.1. Information value

We start with information value—how useful the information that a potential connection node would be to a person's career advancement. A node could possess information about job opportunities, skills in demand, connections to make, and generally provide mentoring and advice to help careers. Let V_{ij} be the potential information value that node j (the alter, or potential connection) has to offer node i (the ego, or person whose network strength we seek to estimate). Intuitively, these are observable employment measures that would make any given connection attractive to helping a member in their career.

Proxying for the above types of information value, we use the following 5 measures, $V_{ij}^1, \ldots, V_{ij}^5$, which we explain in more detail below

1. Alter j is in a senior position
2. Alter j does not have the "Open to Work" status on in their profile
3. The number of skill endorsements alter j has received
4. The occupation similarity of alter j's work to the ego i
5. The industry similarity of alter j's work to the ego i

Note that measures 1-3 are node-level measures (the information value alter j possesses would be the same for any ego i), whereas measures 4-5 are edge-level measures (the information value depends on both who the alter and who the ego is, and a given alter would have different value for different egos). We explain the inclusion of each measure in turn. First, we measure whether an alter is working in a senior position. Senior positions are standardized within LinkedIn's employment data based on titles and keywords used by the member in their profile. We consider as senior any of the following positions: senior, manager, director, VP, CXO, partner, and owner.[6] We consider an alter who is in a more senior position to offer higher information value than one in a junior position because they are likely to have more insights into hiring and skills in demand, and to be able to offer more useful recommendations (informally and formally) for the ego.

With regards to information value from not being open to work, members on LinkedIn are able to create a flag for them being open to work, which they can allow to be visible to all other members or invisible but pass the filter that recruiters would put on only looking for workers open to work. We use this as an indicator of job stability and

[6] Note that owners may include self-employed workers with no employees that may not have as much valuable information for a member's career advancement. We chose to include owners in the list of senior positions however, given it will include owners of businesses with employees who may be of more help.

position. We consider an alter who is not open to work as adding more information value on average because they would have more connections within their work and across other employers and be in a better position to assist the ego.

For number of skill endorsements, LinkedIn platform allows members to endorse others for skills. A member can have a higher total number of endorsements either by having more skills and by, for each skill, receiving more endorsements. We consider the total number of endorsements an alter has as a measure of both people and social capital, and thus consider an alter who has more endorsed skills to be in a better position to help the ego.

Occupation similarity captures the extent to which an alter's occupation is similar to that of the ego.We calculate occupation similarity based on the frequency of observed job transitions in the US using LinkedIn profile histories. $\Pr(occ_{mt+1} = 1|occ_{nt} = 1)$ estimates the probability of working in occupation m given you worked in occupation n in the prior position. This is $\frac{\Pr(occ_{mt+1}=1 \cap occ_{nt}=1)}{\Pr(occ_{nt}=1)}$. We calculate this using sample analogues from the US data of all job transitions. Two things are of note—first, m can equal n: we also estimate the probability of transitioning from a job in a given industry to another job in the same industry. Thus, a given occupation can have a higher self-similarity than another occupation if people are likely to transition within the occupation at a higher rate (as is typically the case). Second, the similarity measure is not symmetric in our model. Indeed, we hypothesize that $\Pr(occ_{mt+1} = 1|occ_{nt} = 1) \neq \Pr(occ_{nt+1} = 1|occ_{mt} = 1)$. In other words, some occupations may serve as launching pads or gateway occupations into other occupations. For example, one common transition we observe is from pharmacy technician to pharmacist. In our methodology, a pharmacist would have a high information value to a pharmacy technician in terms of occupation similarity (given we observe transitions from pharmacy technician to pharmacist with a relatively high conditional probability), but a pharmacy technician may not have as high information value for a pharmacist (given those transitions are much more rare). Thus, an ego who is a pharmacy technician would likely have greater value from being connected to a pharmacist than a pharmacist would have for being connected to a pharmacy technician.

We calculate industry similarity in the same manner as occupational similarity. Note that in the example above, this allows for the law clerk to have valuable information potentially for the lawyer (being in the same industry), but not as much as a lawyer would have for a law clerk (being in the same industry and in an occupation, they are likely to transition into).

With these five measures of information value, we normalize for comparison and aggregation. To do so, we need to get everything on a common scale. We create z-scores for each one (i.e., subtracting the mean and dividing by the standard deviation), yielding $\widetilde{V_{ij}^k} = \left(V_{ij}^k - mean\left(V_{ij}^k\right)\right)/sd\left(V_{ij}^k\right)$. We then create an aggregate measure of information value $\widetilde{V_{ij}} = \sum_k \widetilde{V_{ij}^k}$. Note that we are taking an unweighted sum, and assuming that each of the five measures are equally important. We do this for simplicity, in absence of guidance on which measures would be more important. In follow-up research, this could be revisited as we pursue analysis examining which elements of a strong network impact employment outcomes the most.

However, we do not want the information value in the end to be on a z-scale. It is both harder to interpret and the negative values (when later combined in the full framework) would imply that an ego would prefer not having a connection with a member (value of zero) with below-average information value (negative value), an undesirable feature with respect to our model which assumes all connections provide non-negative value. We take the average of the standardized values, $\widetilde{V_{ij}}$, and estimate the percentile of it across the entire sample using $\widehat{V_{ij}} = 1/(1 + \exp(\widetilde{V_{ij}}))$. This is a logistic distribution transformation, which puts the information value on a positive, bounded scale. Under the assumption that $\widetilde{V_i}$ is distributed logistically, this also means that $\widehat{V_{ij}}$ is the cumulative distribution function value, or percentile of $\widetilde{V_{ij}}$. Specifically, under the distributional assumption, a value of for example 0.8 for an edge between an ego and alter would imply that its total information value is in the 80th percentile across all edges. We use this simplifying parametric assumption for tractability of the estimates, given we are measuring this across billions of edges.

2.2. Information bandwidth

Information is only helpful if shared. We consider a network stronger if there is higher communication along edges the ego has. Following the naming from Aral (2016), we call this network feature information bandwidth, B_{ij}. High information bandwidth implies high probability of sharing the information, and makes an edge higher value. A network with many high-bandwidth edges is all-else-equal a stronger economic network.

We model information bandwidth using observed messaging behavior between nodes. Specifically, an edge has higher bandwidth (probability of sharing information) if the ego *received* more messages in the past year from the alter. Given some connections may be less than a year old, we in practice use the average number of messages received from the alter within an ego's network each month over the past year (or across the time frame in which they have been connected). Call this average number of messages MR_{ij}. Note that we do not use the total number of messages sent and received along an edge, only messages received. This is aimed at controlling for non-symmetric relationships, such as if an ego is connected to an alter who has many (e.g., 30,000) connections and does not know the ego well. If there are no observed messages received from that alter, we predict a low likelihood of message transfer. Also note that this implies that bandwidth is not symmetric along an edge.

We need for the measure using received messages to reflect the probability of an alter who has useful information about a job opportunity to share it. As it is now, the estimate of MR_{ij} will range from zero to some positive value larger than one. Also note that there is a large mass point at zero. Indeed, over 90% of all edges in our data had a value of zero (did not receive a message in the past year). We use a simple transformation of the percentile of MR_{ij} to put it on a scale from 0.01 to 0.5 that is, $B_{ij} = perc\left(MR_{ij}\right) *$

.49 + 0.01 (where $perc(MR_{ij})$ is the percentile and $perc(0) = 0$). This assumes that alters who send an ego the most messages (across the entire sample, not just for that given ego) have a 50% probability of sharing valuable information, while alters who send no messages have a 1% probability of sharing valuable information.

2.3. Information non-redundancy

Next, we care about how insulated or redundant a network is. Intuitively, a new connection that is highly redundant, also referred to as a strong tie (connected to other alters with whom the ego is already connected to) is less valuable all-else-equal than a low-redundancy/weak tie. The information that the alter has for the ego has already-existing pathways for sharing. We choose a single measure of information redundancy given by the local closure coefficient, ρ_i (Yin et al., 2019). The local closure coefficient measures the fraction of an ego's "closed wedges where the ego is the head of the wedge", or in other words the fraction of second-degree connections that are already first-degree connections for an ego, or the fraction of my friends' friends that I am also friends with. We calculate this directly using the network data. Note that for tractability reasons, for now instead of calculating this as an edge-level measure (for each alter, the fraction of their connections that the ego is connected with), we calculate it as a node-level measure (the average across all edges for a given ego).

2.4. Network size

Our fourth input to network strength is network size. Intuitively, a larger network offers more opportunities for information to be shared with the ego regarding career advancement. We calculate network size directly from the data and is given by the number of alters connected to the ego, $\mathcal{N}_i$. It is reflected in the total, or aggregate, network strength measure by allowing for more edges in the final summation.

2.5. Network strength

We bring together the four elements of the model to create an aggregated statistic of network strength S_i. The information value that alter j for ego i has ($\widehat{V_{ij}}$) will be shared from alter to ego with probability B_{ij}. Thus, the expected information value from that edge is given by $E\left[\widehat{V_{ij}}.\,|B_{ij}\right] = \widehat{V_{ij}}B_{ij}$. Each alter provides an expected information value to the ego, and so we can sum across all edges connected to the ego to get a total measure of network strength. However, this is only the first-degree network strength.[7] We label this as S_i^1, and it is given by $S_i^1 = \sum_{j \in \mathcal{N}_i} B_{ij}\widehat{V_{ij}}$

Thus, first-degree network strength is increasing in network size $\mathcal{N}_i$, information bandwidth B_{ij}, and information value $\widehat{V_{ij}}$. However, we have not yet incorporated information redundancy. This concept is implicitly related to second-degree connections. We want to model the extent to which an ego's connections' connections can pass on information to the ego. The local closure coefficient ρ_i offers a model-consistent way to incorporate information redundancy. Recall that ρ_i is the fraction of second-degree connections that are already first-degree connections. Thus, it offers a natural scaling of the connections' network values, say $\widehat{V_{jk}}$ for alter j's connection to node k. We scale down the value of each of the second-degree connections by the probability that the ego already has that information value using the local closure coefficient. Then we can calculate the second-degree network value from alter j to ego i as follows.

$$S_{ij}^2 = B_{ij}(1 - \rho_i) \sum_{k \in \mathcal{N}_j} B_{jk}\widehat{V_{jk}}$$

And thus

$$S_i^2 = \sum_{j \in \mathcal{N}_i} S_{ij}^2 = \sum_{j \in \mathcal{N}_i} B_{ij}(1 - \rho_i) \sum_{k \in \mathcal{N}_j} B_{jk}\widehat{V_{jk}}$$

[7] First degree connections are alters to whom the ego is connected. Second degree connections are nodes that the alter is connected to (the alters' alters), and thus to whom the ego is connected through a given alter.

$\sum_{k\in\mathcal{N}_j} B_{jk}\widehat{V_{jk}}$ is familiar, and is the first-degree network strength that node *k* has for alter *j*. The value of each of these to ego i (that is, the value the second degree connection has to the ego) is then scaled by B_{ij} (the probability that alter j shares the information they received from their connection k back to ego i) as well as by $1-\rho_i$, the fraction of these second degree connections that are already connected to ego i and thus already included (with likely higher sharing probabilities) in the first-degree connections network strength S_i^1.

With the first-degree network strength and second-degree network strength, we can calculate the total network strength. This is given by

$$S_i = S_i^1 + S_i^2 = \sum_{j\in\mathcal{N}_i} B_{ij}\widehat{V_{ij}} + \sum_{j\in\mathcal{N}_i} B_{ij}(1-\rho_i)\sum_{k\in\mathcal{N}_j} B_{jk}\widehat{V_{jk}}$$
$$= \sum_{j\in\mathcal{N}_i} B_{ij}\widehat{V_{ij}} + \sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j}(1-\rho_i)B_{ij}B_{jk}\widehat{V_{jk}}$$

Operationally, we estimate the network strength for each person in this order

1. Calculate $\widehat{V_{ij}}$ and B_{ij} for all first- and second-degree edges
2. Calculate S_i^1 based on $\widehat{V_{ij}}$ and B_{ij}
3. Calculate S_i^2 based on the prior-estimated $\widehat{V_{ij}}$ with the associated bandwidths and redundancy measure

From this, it is straightforward then to estimate gaps in network strength (or the components of network strength) by group, such as gender. For example, the gender gap would be given by the difference in average network strength between groups, i.e.

$$Gender\ Gap = \frac{\sum_{i=male} S_i}{\sum_{i=male} 1} - \frac{\sum_{i=female} S_i}{\sum_{i=female} 1}$$

Note too that individual elements and measures of information value, bandwidth, redundancy, and network size each have intuitive interpretations and help us understand the sources of network strength disparities between groups. We can first contrast S_i^1 and S_i^1 within and between genders to determine the extent to which disparities are driven by first-degree connections versus second-degree connections. Additionally, focusing now on the first-degree connections, we can calculate each of the factors. For example, $S_i^{1V} = \frac{1}{|\mathcal{N}_i|} * \sum_{j \in \mathcal{N}_i} \widehat{V_{ij}}$ is a measure of the average network information value, and shows what the network strength would be if individuals had the same network size (dividing by the network size) and had 100 percent bandwidth, or all information was shared. $S_i^{1B} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} B_{ij}$ calculates the first-degree network strength that would result if the information was shared at the 99th percentile (top score) and had the same network size, given actual bandwidth. $|\mathcal{N}_i|$ is not only the network size but is the total network strength under perfect information sharing where information value was in the 100th percentile.

3. Data and Context

Our data is drawn from a random sample of 1 million LinkedIn members residing in the US in 2023, as well as all first-degree and second-degree connections of each of these members. From that 1 million drawn, we end up with a sample size of just over 818,000 members (with attrition from the sample being due to not having a reliable gender measure, having an account with no connections, and other related missing information reasons). This results in billions of edges that we examine. LinkedIn is a platform for professional networking, and thus offers an ideal scenario for investigating economic network strength at a large scale.

Network size and redundancy measures are derived directly from the network structure. We observe each active connection as of the date of investigation (June 2023). Information bandwidth is drawn from the observed messaging behavior between each member connection. Given we investigate bandwidth at the edge-level, this excludes such messaging as from recruiters or others outside of a member's network. Finally, information value is derived from several inputs. Occupational and industry similarity are derived from the primary, current job position held by each member, as self-entered by the member as part of their LinkedIn profile. Seniority level is standardized given the job title each member lists for their primary position. We consider a position to be senior if it is in a management position. Open to work status is a flag that each member can activate (publicly or privately).

Additionally, we recognize that gender identity is not binary. Some LinkedIn members identify beyond the traditional gender constructs of "man" and "woman." However, for this analysis we evaluate members based on this binary construct. Some members have opted to self-identify their gender. For those who have not, LinkedIn infers the gender of members included in this analysis either by the pronouns used on their LinkedIn profiles or inferred on the basis of first name. Members whose gender could not be inferred as either man or woman were excluded from this analysis.

4. Results

4.1. Summary statistics on measures

We first examine the different inputs into the model. Figure 1 presents the information value measures between men and women using boxplots. We transform the z-values to percentiles to more easily view differences. The dark line in the center represent the median values. The boxes present the 25th to 75th percentile range, while the line presents the minimum to maximum range (excluding outliers).

We find that men have on average higher information value for all five measures. The advantage shown at the median for men is largest for being more likely to be connected to other members who have more endorsed skills (gap of 7.5 percentile points), who are in senior positions (6.8 percentile points), and who are not open to work (5.1 percentile points). The gaps in information value derived from industry and occupation similarity are substantially smaller (2.6 and 1.1, respectively). Additionally, note that the variation in network information value is much larger within gender than between gender, as shown by the large spread in the box charts for each group. Appendix Table A.1 presents the averages and medians for these measures.

Figure 1: Network Information Value Measures

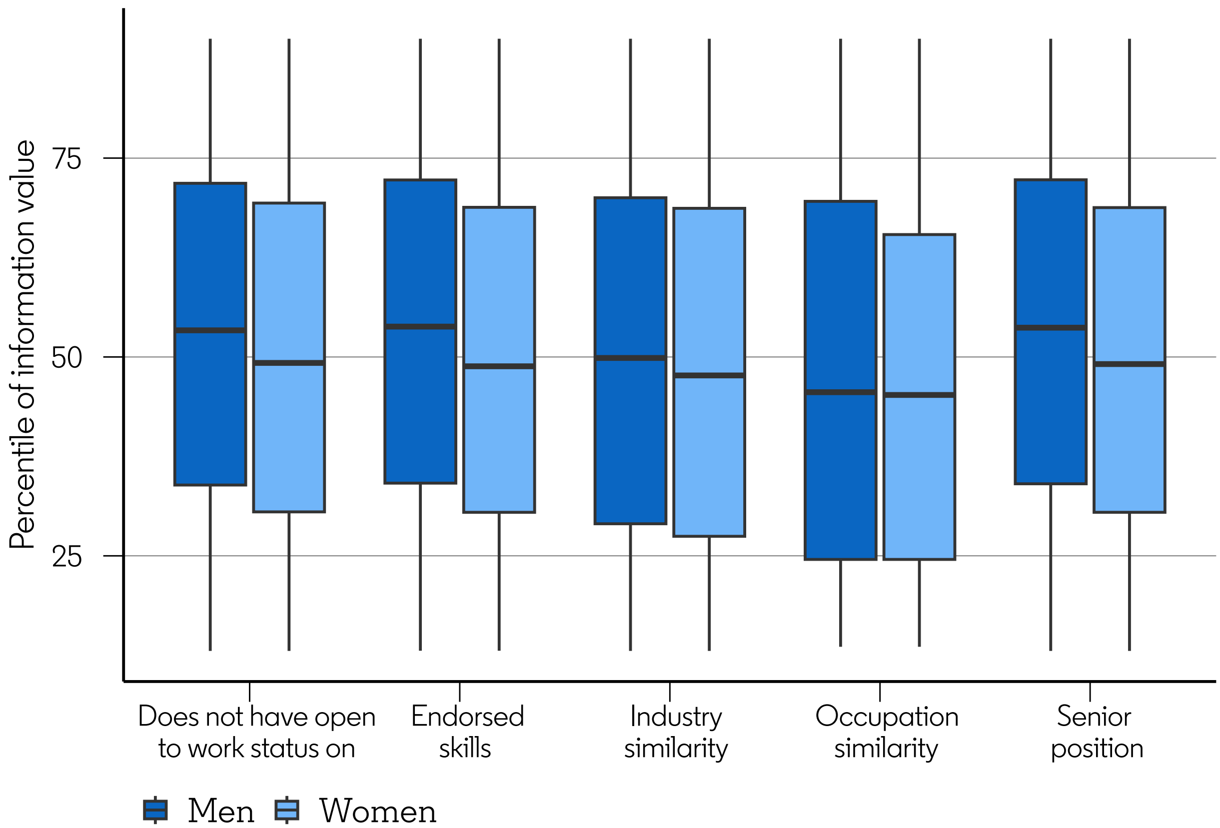


Figure 2 presents the elements of total network strength. Men have higher average values for three of the four inputs: network size (8.2 percentile points—men have larger networks on average), information value (7.6 percentile points—the summary of figure 1 across all information value measures, where men are more likely to be connected to individuals who may be able to help their career progress), and information bandwidth (4.8 percentile point gap—men communicate more on platform with their connections). Women hold an advantage in information non-redundancy of 6.1 percentile points—

woman are more likely than men to have a network with more weak ties, or in other words, connections who are in turn connected to other people that the members are not themselves connected to.

Figure 2: Network Strength Inputs

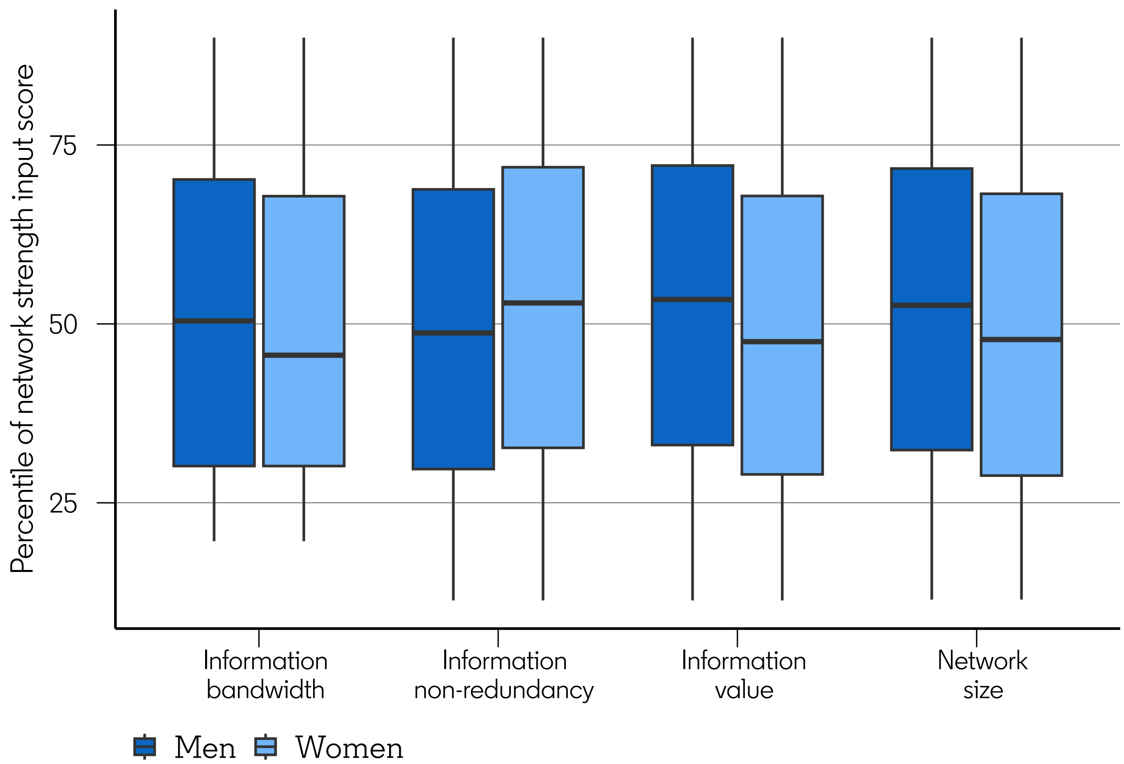
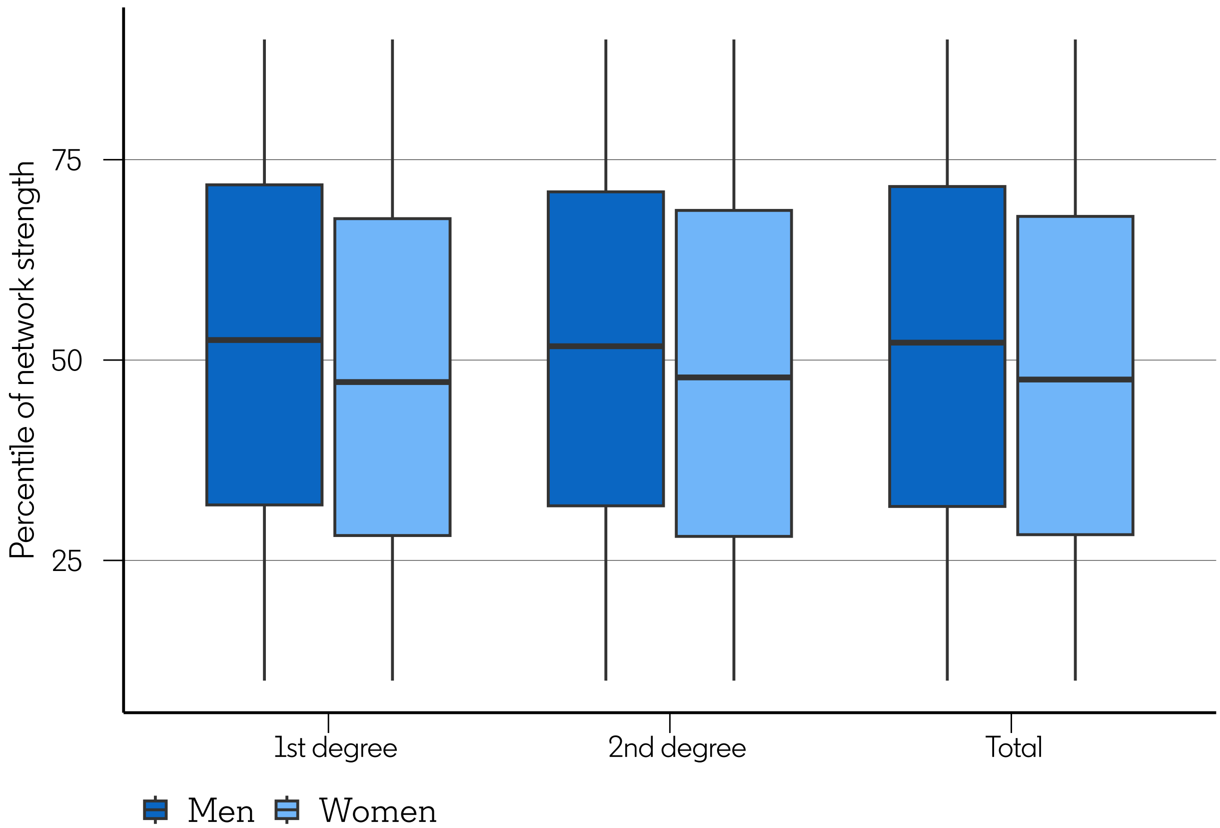


Figure 3 presents the total network strength. The 1st degree network strength gap is largest. The second-degree network strength gap is smaller, likely driven by the advantage women have in information non-redundancy. For overall network strength, the average percentile for men is in the 53.9th percentile, whereas for women it is 45.6th percentile, for a gap of 8.3 percentile points. This is a significant divergence between the two groups.

Figure 3: Overall network strength



4.2. Regression analysis

We next test the extent to which observed gaps can be explained by observable characteristics. For each outcome examined (translated to percentiles to increase interpretability and comparability), we regress the outcome on gender alone, with each of a set of control variables, and with all control variables. The results are reported in Table 1. All coefficients reported are statistically significant at the $p<0.01$ level with the exception of -0.11 for industry similarity, all covariates. This coefficient is significant at $p<0.05$. We find that covariates in all cases either narrow or flip the gender gap. For example, for total network strength, the average percentile gap without controlling for any covariates is 5.976 percentile points higher for men than women; controlling for all covariates decreases the gap to 3.174. Of the tested covariates, controlling for the occupation of the member has the largest reduction in the gap. Examining the other factors, information value and network size in particular are reduced significantly by controlling for the member characteristics. For information value, the gaps decreases from 4.469 to a gap of 1.331. Occupation again decreases this gap more than the other covariates. Geography and age—factors least likely to be correlated with gender—tend to have the smallest impact on narrowing the gaps.

Table 1: Regression results

	Covariates included						
	None	Industry	Occupation	Geography	Age	Grad. Year	All
Inputs to information value							
Ind. similarity	-1.493	-0.625	-0.639	-1.347	-1.198	-1.053	-0.11
Occ. similarity	-0.75	-0.404	0.466	-0.627	-0.548	-0.43	0.528
OTW	-2.756	-2.677	-1.833	-2.822	-1.186	-1.231	-0.749
Seniority	-4.196	-3.691	-2.321	-4.192	-3.056	-2.781	-1.995
Skills	-4.753	-3.963	-2.719	-4.633	-2.922	-2.964	-1.753
Inputs to total network strength							
Inf. value	-4.469	-3.515	-2.216	-4.378	-3.192	-3.012	-1.331
Inf. bandwidth	-2.536	-2.048	-1.804	-2.428	-2.796	-2.859	-1.832
Inf. redundancy	3.946	3.867	2.899	3.846	3.305	3.438	2.487
Network size	-5.529	-5.05	-3.202	-5.382	-4.574	-4.602	-2.856
Total network strength							
Total	-5.976	-5.232	-3.464	-5.811	-5.275	-5.254	-3.174

5. Discussion

Economic networks can help people advance in their career, through identification of job opportunities and information regarding in-demand skills and people to connect with. Unfortunately, inequality in the labor market between groups may be reinforced by underlying disparities in economic networks. In this paper we develop a new model of network strength and estimate it using LinkedIn data. We evaluate differences in network strength between men and women. We find that there are notable divergences, with men having larger networks filled with people in better positions to help their careers than women—in particular, people in more senior positions, with more endorsed skills, who are more likely to not be open to work. This all translates into a total network strength gap of 8.3 percentile points. Put another way, if you randomly selected 100 people and ordered them from weakest to strongest economic network, the average women would be 8 people lower in the line than the typical man.

However, when we control for observable characteristics of the worker, we find that each of these gaps narrow, or even reverse. Occupation is the most important in explaining gender gaps in network strength. For example, for overall network strength, the mean gap decreases from 0.8 percentile points advantage for men to 0.5 percentile point advantage if we only control for occupation. If we control for other factors, it remains at around 0.5 percentage points. The overall network strength narrows from 6 percentile points at the average to 3.2 percentile points—roughly falling in half. Most of this reduction appears to arise from controlling for the occupation that the worker is in.

Future research is needed to determine which of these elements of network strength impact economic outcomes the most. Doing so will help us understand if these gaps are meaningful, how to benchmark the 6 percentile point gap overall, and which features of a network are most important to improve economic outcomes. This will result in policy implications for both social network platforms seeking to improve equity, but also for public policy by thinking about which dimensions of network disparities can be addressed with the most efficacy. In the meantime, these results suggest particular attention may be paid to policies that encourage women to grow their networks, especially if they can reach out to more senior and more skills workers. Mentoring activities and outreach from more senior workers can help address these gaps.

REFERENCES

Aral, S. (2016). The future of weak ties. *American Journal of Sociology*, *121*(6), 1931–1939.

Barbulescu, R. (2015). The strength of many kinds of ties: Unpacking the role of social contacts across stages of the job search process. *Organization Science*, *26*(4), 1040–1058.

Bruggeman, J. (2016). The strength of varying tie strength: Comment on Aral and Van Alstyne. *American Journal of Sociology*, *121*(6), 1919–1930.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., & Jacob, M. (2022a). Social capital I: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., & Jacob, M. (2022b). Social capital II: Determinants of economic connectedness. *Nature*, *608*(7921), 122–134.

Childers, C., & Kaplan, L. (2023). *The Network Effect: How Women Leaders Leverage Connections to Achieve Power and Drive Business Success* (Chief). https://join.chief.com/rs/702-CII-507/images/Chief%20Networking%20Study.pdf

Graf, N., Brown, A., & Patten, E. (2018). The narrowing, but persistent, gender gap in pay. *Pew Research Center*, *9*.

Lara, S., Baird, M., & Hood, R. (2023). Progress and barriers in global gender leadership. *LinkedIn Economic Graph White Paper*. https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/global-gender-representation.pdf

Maasoumi, E., & Wang, L. (2019). The gender gap between earnings distributions. *Journal of Political Economy*, *127*(5), 2438–2504.

McDonald, S., Lin, N., & Ao, D. (2009). Networks of opportunity: Gender, race, and job leads. *Social Problems*, *56*(3), 385–402.

McGuire, G. M. (2002). Gender, race, and the shadow structure: A study of informal networks and inequality in a work organization. *Gender & Society*, *16*(3), 303–322.

Pedulla, D. S., & Pager, D. (2019). Race and Networks in the Job Search Process. *American Sociological Review*, *84*(6), 983–1012. https://doi.org/10.1177/0003122419883255

Petrongolo, B. (2019). The gender gap in employment and wages. *Nature Human Behaviour*, *3*(4), 316–318.

Rajkumar, K., Saint-Jacques, G., Bojinov, I., Brynjolfsson, E., & Aral, S. (2022). A causal test of the strength of weak ties. *Science*, *377*(6612), 1304–1310.

Yin, H., Benson, A. R., & Leskovec, J. (2019). The local closure coefficient: A new perspective on network clustering. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 303–311.

APPENDIX

Table A.1: Percentile of information value inputs

	Mean		*Median*	
	Men	Women	Men	Women
Industry	50.90	49.41	51.20	48.58
Occupation	51.60	50.85	50.56	49.50
Not open to work	51.50	48.75	52.26	47.18
Seniority	52.09	47.90	53.16	46.38
Endorsed skills	52.22	47.47	53.48	45.96

Table A.2: Percentile of network strength inputs

	Mean		*Median*	
	Men	Women	Men	Women
Bandwidth	52.43	49.90	52.11	47.30
Redundancy	52.79	47.26	53.91	45.76
Network size	48.46	52.40	47.16	53.28
Information value	52.08	47.61	53.54	45.91

Table A.3: Percentile of total network strength

	Mean		*Median*	
	Men	Women	Men	Women
1st degree	52.82	46.75	54.18	45.40
2nd degree	51.85	47.86	52.63	46.68
Total network strength	52.78	46.80	53.93	45.60