

# Social capital estimation with exponential random graphs in the labor market

## Empirical Industrial Organization

---

Carlos Lezama

Instituto Tecnológico Autónomo de México

### Introduction

It is well known that inequalities in the labor market lead up to income disproportions given an initial endowment of opportunities. Said this, networks of agents that start with a worse wage status will have higher drop-out rates and there will be persistent differences in wages between groups according to the starting states of their networks. In addition, if we observe that these inequalities do not appear by magic, we may consider a previous and important stage for the establishment of the idealization of the Mexican middle class: a college degree.

Hereby, it is mandatory to question the public and private higher education and their relationship with firms within the labor market. By such, we purport that it is true that there exists a higher probability to be hired by a top-level firm if you graduated from a private university. Thus, we model a relationship between schools, individual and firms using random graphs, and aim to build a brand new inequality index based on education to be potentially replicated in developing economies.

All of this is based on the importance of the networks of relationships between the people who live, study and work in a particular society and allow the society to function effectively, also referred as **social capital**.

### Literature Review

A wide range of empirical studies of labor markets have shown that a significant fraction of all jobs are found through social networks. The role of informal social networks in labor markets has been emphasized first by Granovetter (1995). He found that over 50% of jobs were found through personal contacts in the US at the time. In a recent paper, Jackson and Calvó-Armengol (2005) introduce a network model of job information transmission. The model reproduces the empirically stylized fact that the employment situation of individuals that are connected, either directly or indirectly, is correlated. Further, they show that the topology of the network influences the length and correlation of unemployment among individuals.

More recently, in the field of statistical models of random graphs, Jackson and Chandrasekhar (2012) address the problem that standard techniques of estimating Statistical Exponential Random Graph Models (SERGMs) have been shown to have exponentially slow mixing times for many specifications, they show that by reformulating network formation as a distribution over the space of sufficient statistics instead of the space of networks, the size of the space of estimation can be greatly reduced, making estimation practical and easy, and also show how choice-based (strategic) network formation models can be written as such.

### Model

We will consider networks that connect (or not) universities, individuals and firms — which may form links or triangles (given this setup there cannot appear isolated nodes). This way, the probability of the network is assumed to be proportional to

$$\exp(\beta \cdot S(g)) \tag{1}$$

where  $\beta$  is a vector of model parameters. Turning the above expression into a probability of observing network  $g$  requires normalizing this expression by summing across all possible networks, and so the probability of observing  $g$  is

$$P_{\beta}(g) = \frac{\exp(\beta \cdot S(g))}{\sum_{g'} \exp(\beta \cdot S(g'))}. \tag{2}$$

For simplicity, we build  $S(g) = [S_{\circ}(g), S_{-}(g), S_{\triangle}(g)]^T$  which stand for the network's number of isolated nodes, links and triangles, respectively — which describe how linked individuals are. Future statistical and algorithm details will be adapted from Jackson and Chandrasekhar (2012). Finally, we may potentially consider wage as an independent variable.

## Data

The data may not be publicly available, so we need to conduct a survey across multiple universities to obtain it. The survey structure has to be simple to avoid noisy answers. First, we need to create a list of  $n$  top-tier firms — potentially, with  $n = 50$  to improve survey search and readability. Second, we may require a sufficiently large sample from  $k$  faculties (namely — and potentially because of their well-known popularity — business- and STEM-related careers) that strongly represent their own population. Finally, they will be asked four (4) questions:

1. Where do/did you study? — selection of a curated list that contains ITAM and/or UNAM, to mention a few.
2. To which of the following firms have you applied for a job? ( $A$ ) — selection of a curated list that contains Amazon and/or Google, to mention a few.
3. From which of the above have you received a job offer? ( $B$ )
4. In which of the above firms have you worked? ( $C$ )

Note that  $A \supseteq B \supseteq C$ .

## References

- Granovetter, M.S. (1995) Getting a Job: A Study of Contacts and Careers. Harvard University Press, Cambridge, MA. <https://doi.org/10.7208/chicago/9780226518404.001.0001>
- Calvó-Armengol, Antoni & Jackson, Matthew. (2005). Networks in Labor Markets: Wage and Employment Dynamics and Inequality. Journal of Economic Theory. 132. 27-46. <https://doi.org/10.1016/j.jet.2005.07.007>
- G a, Arun & Jackson, Matthew. (2012). Tractable and Consistent Random Graph Models. SSRN 2150428. <https://doi.org/10.2139/ssrn.2150428>