Supporting Information File (SIF)

A Survey Experiment on "Bad Bosses"

Isabella Alcañiz (**Corresponding Author**), ialcaniz@umd.edu, Associate Professor of Government and Politics, University of Maryland, 3140 Tydings Hall, College Park, MD 20742, United States

Ernesto Calvo, ecalvo@umd.edu, Professor of Government and Politics, University of Maryland, 3140 Tydings Hall, College Park, MD 20742, United States

Marcelo Escolar, marcelo.escolar@unsam.edu.ar, Professor of Politics and Government, Universidad Nacional de San Martin, Campus Miguelete, 25 de Mayo y Francia.
C.P.: 1650. San Martín, Provincia de Buenos Aires, Argentina

4/20/2020

APPENDIX A

Statistical Model Used to Measure Political Networks and Network Proximity

In this Appendix, we describe the research strategy to measure the size and structure of the personal networks of the different respondents. The model estimates three key sets of parameters measuring (i) the relative size of the voter's personal network, (ii) the relative prevalence of different group categories in the population, and (iii) the relative proximity of voters to each of these groups. We use the first set of parameters, size of the personal network, as the key independent variable of the article: "**A Survey Experiment on *Bad Bosses***". To measure the size and structure of networks, we take advantage of the survey strategy first proposed by Christopher McCarthy et. al. (McCarty et al. 2000; McCarty, Killworth, and Rennell 2007). The survey uses questions of the type "*how many X do you know*," to obtain counts of individuals belonging to different group categories. To analyze this data, we use an over-dispersed statistical model proposed by Zheng, Salganik and Gelman (2006) measuring the personal network of respondents and the prevalence of groups as a share of the respondent's personal network.[1]

a) The Questions used to measure the size of the personal network:

The strategy to measure the size of the personal network proposed by McCarty et. al. (2007) considers every respondent as an observer that provides counts of individuals they know across a variety of group categories. They include counts of names, professions, and life events, whose probability occurrences in the population are already known. In the case of Argentina, Calvo and Murillo (2013) propose a list of categories that meet the requirements suggested by McCarthy et. al., including being in the target prevalence area (0.5% of the population for name categories) to minimize deviations that result from memory over-recall (prevalence < 0.5%) or under-recall (prevalence > .5).

The question asks, "How many people do you know, who also know you and with whom you have interacted in the past year either in person, by phone, or other media [that belong to the group category]". For example, "whose name is Silvia", "that work as teachers", "that were victims of a sexual aggression", etc. The full battery of questions asked in the survey is in Figure A.1 in the next page. It is important to emphasize that "knowing someone" requires that they also know the respondent and that they had some type of interaction within the last year. The "yearly" network is defined as the personal network of the respondent and is different from the "intimate" network, which is defined as the network of individuals with whom a respondent interacts on weakly basis.

---

[1] McCarthy et.al. (2000), Zheng et al (2006).

Figure A.1: “How many X…” questions used to measure the size of the personal size

Ahora le haremos una preguntas sobre sus relaciones y contactos sociales...

52 ¿Cuántas personas conoce y que a su vez lo conocen a Ud., con quien ha interactuado en el último año ya sea en persona, por teléfono, o por algún otro medio… [LEER CATEGORÍAS]

Cantidad **[Si no conoce a nadie consignar "0" y si Ns/Nc consignar 999]**

cuyo primer nombre es Silvia?	V122
cuyo primer nombre es Patricia?	V123
cuyo primer nombre es Antonio?	V124
cuyo primer nombre es Francisco?	V125
cuyo primer nombre es Ángel?	V126
que sea maestra o maestro en una escuela primaria?	V127
que trabajen como policías?	V128
que ejerzan la profesión de abogado o abogada?	V129
ejerzan la profesión de médico o médica?	V130
que hayan tenido un hijo en el último año?	V131
que hayan fallecido en el último año?	V132
que se hayan casado en el último año?	V133
que estén discapacitadas para caminar?	V134

53 ¿Cuántas personas conoce y que a su vez lo conocen a Ud., con quien ha interactuado en el último año ya sea en persona, por teléfono, o por algún otro medio… [LEER CATEGORÍAS]

Cantidad **[Si no conoce a nadie consignar "0" y si Ns/Nc consignar 999]**

que trabajen como empleados en la municipalidad?	V135
que trabajen como empleados en alguna repartición del gobierno provincial?	V136
que realicen trabajo social en una villa de emergencia?	V137
que realicen trabajo comunitario para su iglesia o comunidad religiosa?	V138
que sean militantes de algún partido político?	V139
que hayan sido candidatos políticos para su municipio o provincia?	V140
que hayan sido candidatos a diputados nacionales o senadores?	V141
que ejerzan como jueces?	V142
que hayan sido víctimas de un crimen violento por el que fueron hospitalizados?	V143
que hayan sido víctimas de un crimen sexual?	V144
que hayan sido víctimas de robo por un moto chorro?	V145
que se encuentren actualmente en prisión?	V146

5

b) *Using Count Data to Measure the Size of the Personal Network: An intuitive description*

To measure the size and structure of the respondents’ personal networks, we use a survey design that considers every respondent in the sample as an *observer* who discloses information on the relative prevalence of different groups in the population. The survey is designed with questions of the form “how many X do you know,” asking each respondent to provide counts of groups whose frequencies in the population are known (“How many individuals do you know whose name is *Silvia*?”) and counts of groups whose frequencies in the population we seek to estimate (“How many activists from the Socialist Party do you know?”). For example, if a respondent knows two *Silvias*, given that the relative prevalence of the name *Silvia* in the population in Argentina is 0.86 per cent, a naïve estimate of the respondent’s personal network would be of approximately ≈232 individuals ($N_p = \frac{2}{.0086}$). Using a battery of questions about populations whose frequencies we know, and a slightly more sophisticated statistical model, we can estimate the size of each respondent’s personal network. In our article, we are interested in the effect of a personal network on the decision to give raises to “Bad Bosses”. This is defined as the “gregariousness” parameter, which we describe in the next section.

c) *The Statistical Strategy: An Over-Dispersed Poisson Model*

Zheng, Salganik and Gelman (2006) propose an over dispersed Poisson model that allows researchers to estimate the size of respondents’ personal network and to explore social structure in the data. The model estimates three sets of parameters that are key to understanding the network of each and all respondents: the relative size of each respondent's personal network, $\propto_i$, the relative prevalence of each group k in the population, β_k, and a parameter that explores individual-level deviations from the personal network and group prevalence. The over-dispersed Poisson model uses the count of individuals known to each respondent as the dependent variable and estimates three sets of latent parameters:

$$y_{ik} \sim Poisson(e^{\propto_i + \beta_k + \delta_{ik}}) \qquad \text{Eq. (A.1)}$$

where $\propto_i$ describes the size of the personal network of respondent i, β_k describes the expected prevalence of group k in the population, and the overdispersion parameter δ_{ik} estimates a multiplicative factor with individual and group-level deviations from the personal network $\propto_i$ and group prevalence β_k (Gelman and Hill 2007).

The vector of personal network parameters, $N \equiv \{\alpha_1^{'}, \ldots, \alpha_i^{'}\}$, provides *critical information about individual-level interaction with other individuals*. Each parameter $\alpha_{ik}^{'}$ provides information about the degree to which a respondent knows more individuals than expected from a *k-group* category, given her personal network size and group prevalence.

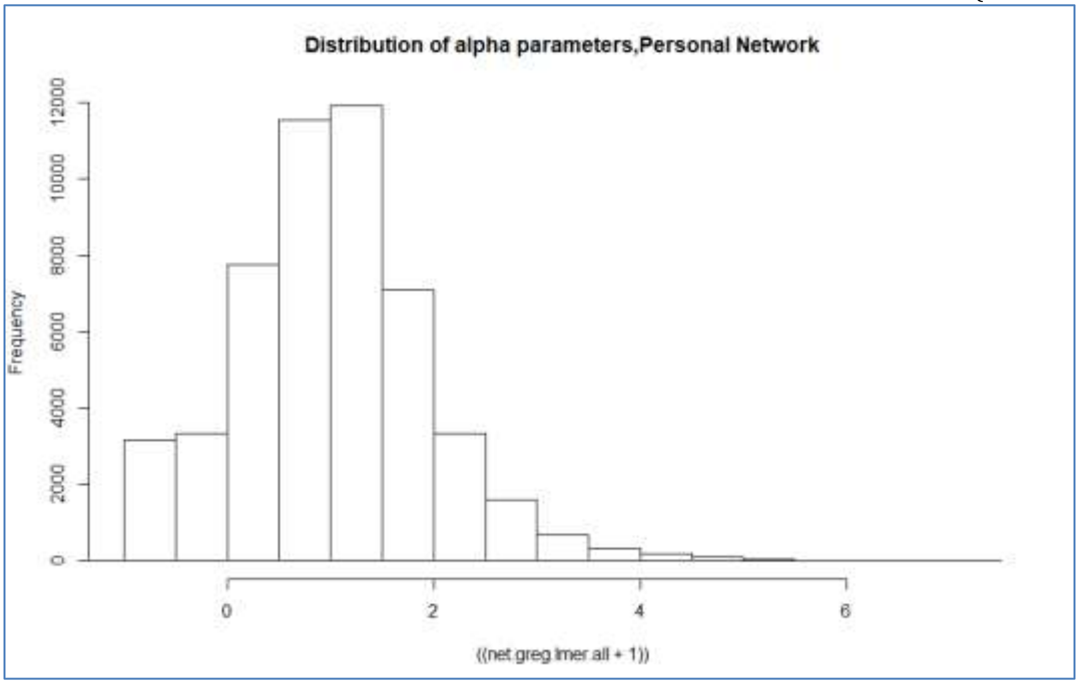
The vector of over-dispersed parameters, $H \equiv \{\delta_{11}^{'}, \ldots, \delta_{ik}^{'}\}$, provides *critical information about individual-level deviations from the overall group prevalence*. Each parameter $\delta_{ik}^{'}$ provides information about the degree to which a respondent knows more individuals than expected from a *k-group* category, given her personal network size and group prevalence. Therefore, we can study the social structure of networks—how different political categories relate to each other— by comparing the over-dispersion parameters for different groups.

Finally, to assess the social structure of networks—how different social and political categories relate to each other—we analyze the matrix of over-dispersed parameters in equation (C.3), $H \equiv \{\delta_{11}^{'}, \ldots, \delta_{ik}^{'}\}$.[2] Each parameter, $\delta_{ik}^{'}$, provides information about the degree to which a respondent knows more individuals from a particular group k than what would be expected given her personal network size and the overall group prevalence in the population.

Estimation of the model was done in R 3.5.1, using LMER with random slopes by user and group. As proposed by Gelman and Hill, the matrix of overdispersion parameters was retrieve from the residuals of the model. The vector of personal network sizes described the relative gregariousness of each respondent in the survey. The distribution of the parameters is roughly normal, modestly skewed left. The histogram plot in Figure A.2 describes the distribution of the personal network parameters.

[2] Gelman and Hill measure the absolute difference between the predicted and observed counts, because their specification does not estimate an overdispersion parameter by individual and group. Our estimation strategy provides the full matri of overdispersed parameters. As a result, we can estimate the inter-group correlation directly. Both strategies yield substantively similar inter-group correlations, clusters, and dendograms.

Figure B.2: Distribution of vector of Personal Network (Gregariousness), $N \equiv \{\alpha_1^{'}, \ldots, \alpha_i^{'}\}$



Note: Vector of personal network parameters, $N \equiv \{\alpha_1^{'}, \ldots, \alpha_i^{'}\}$, estimated using the survey questions in Table A.1 and the Eq. (A.1).

APPENDIX B:

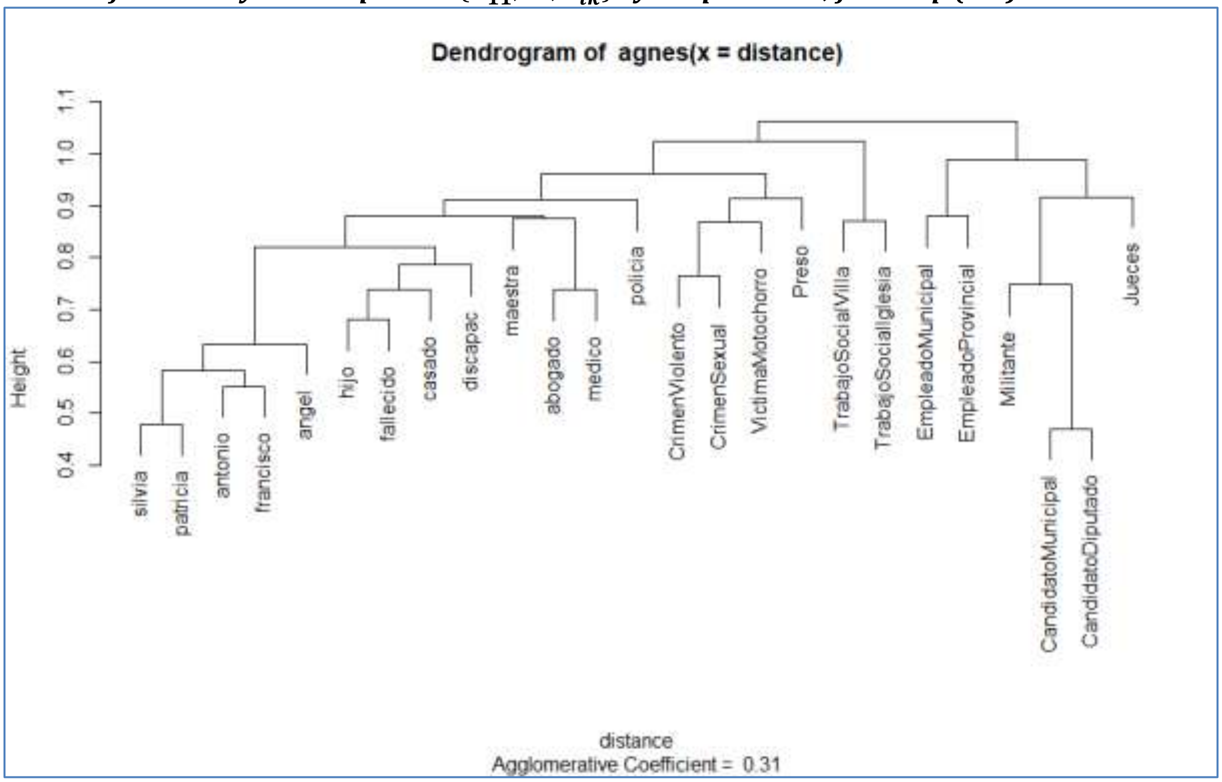
Measures of Network Structure in the Measure of Personal Network

As a check on the personal network results, we report on some of the network structure estimated from the model. As described in Appendix A, there are three key groups of parameters that are returned by the Zheng, Salganik and Gelman (2006) poisson model: the first set of parameters describes the size of each respondent's personal network, $N \equiv \{\alpha_1^{'}, \ldots, \alpha_i^{'}\}$. The second set of parameters describes the relative prevalence of different groups, $K \equiv \{\beta_1^{'}, \ldots, \beta_k^{'}\}$. Finally, the large set of parameters describes deviations from the mean network personal network size ($\hat{\alpha}$) and the mean group prevalence ($\hat{\beta}$), which is described by the set, $H \equiv \{\delta_{11}^{'}, \ldots, \delta_{ik}^{'}\}$.

The parameters $\delta_{ik}^{'}$ describe the relative proximity of each respondent *i* to the group *k*. That is, they indicate whether a respondent knows more of the group *k* than what we would expect, given the size of their personal network and the prevalence of the group in the sample. For example, we are able to compare to what extent individuals that know more victims of crime are also more likely to know people in prison. We can also assess the extent to which individuals that work in the public sector also know more people that are involved in politics.

Figure B.1 provides a dendogram that describes the level of association between the different group categories from the questions provided in Table A.1. We can see, for example, that individuals who know more public employees also know more political candidates and more judges. We can also see that individuals that know more victims of crime also know more individuals in prison. In other words, we can see how we can use the information reported by respondents to assess not only the size of their personal network but also the structure of their network.

Figure B.1: Dendrogram of association between group categories in the "Bad Boss" data, Matrix of Proximity to Groups $H \equiv \{\delta'_{11}, \ldots, \delta'_{ik}\}$ by Respondents, from Eq. (A.1)



Note: Agnes agglomerative coefficient on the matrix of respondent's proximity to each group (δ'_{ik}).