econet: An R Package for Parameter-Dependent Network Centrality Measures

Marco Battaglini
Cornell University, EIEF, NBER

Valerio Leone Sciabolazza
Sapienza University of Rome, CEIS

Eleonora Patacchini
Cornell University, EIEF

Sida Peng
Microsoft Research

Abstract

The R package **econet** provides methods for estimating parameter-dependent network centrality measures with linear-in-means models. Both nonlinear least squares and maximum likelihood estimators are implemented. The methods allow for both link and node heterogeneity in network effects, endogenous network formation and the presence of unconnected nodes. The routines also compare the explanatory power of parameter-dependent network centrality measures with those of standard measures of network centrality. Benefits and features of the **econet** package are illustrated using data from Battaglini and Patacchini (2018) and Battaglini, Leone Sciabolazza, and Patacchini (2020).

Keywords: network econometrics, heterogeneous peer effects, endogenous network formation, least-square estimators, maximum likelihood estimators, R.

1. Introduction

Since its inception, network analysis has mostly focused on the discovery of topological properties of network structures. This has changed dramatically over the past ten years. An emerging literature in economics has shown that network centrality measures, which were traditionally viewed as descriptive, have an interpretation within equilibrium models of behavior. The pioneer paper is Ballester, Calvó-Armengol, and Zenou (2006). This paper considers a model in which an agent's effort is triggered by the effort of his/her socially connected peers. It shows that the equilibrium levels of effort are linear functions of the agent's position in the network as measured by an indicator within the family of Katz-Bonacich centralities. Katz-Bonacich centralities (Katz 1953; Bonacich 1972, 1987) are network centrality measures that count all nodes that can be reached through a direct or indirect path, penalizing in different ways the contributions of distant nodes in determining a given node's centrality. The discount factor is captured by a parameter, thus making the measures of centrality parameter-dependent. The sociological literature has been treating this parameter as a nuisance parameter, and arbitrarily setting it to any value smaller than one. However, the contribution of Ballester *et al.* (2006) is to show that this parameter captures the strength of peer effects or social interactions that stem from the aggregation of dyadic peer influences. More specifically, the empirical counterpart of the Ballester *et al.* (2006) equilibrium condition is a linear model of social interactions, where the individual levels of effort are linear functions

of the levels of effort of the connected agents. The parameter capturing this influence is then used to measure the individual's importance in the network. Since then, a burgeoning empirical literature has used the linear-in-means model to prove that peer effects and the individual position in the social network play an important role in explaining many social and economic outcomes, including consumer behavior, voting patterns, job search, information diffusion, innovation adoption, international trade and risk sharing (see e.g., An 2011, 2015a; Jackson, Rogers, and Zenou 2017; Hsieh, Lin, and Patacchini 2020; Zenou 2016, for recent reviews).

Very recently Battaglini and Patacchini (2018) present a new theory of competitive vote-buying to study how interest groups allocate campaign contributions when legislators care about the behavior of other legislators to whom they are socially connected. The model provides an alternative microfoundation for network measures within the family of Katz-Bonacich centralities. This theory predicts that, in equilibrium, campaign contributions are proportional to a parameter-dependent measure of network centrality, similar to the one proposed by Ballester *et al.* (2006). While Ballester *et al.* (2006) is a purely theoretical paper, Battaglini and Patacchini (2018) test the theory with data from five recent United States (US) Congresses. In doing so, it confronts a variety of empirical challenges. For example, the theories described above show the importance of combining the information on network centrality with additional information on characteristics of the agents, since these characteristics can magnify or reduce the role played by a central agent. Moreover, the theories provide a framework to study the role of network endogeneity. In fact, when agents are strategic in choosing their peers, and omitted variables (such as social skills) drive both agent's behavior and social connectedness, the estimation of the peer effects parameter might be flawed (Manski 1993).[1] Battaglini *et al.* (2020) derive a model to control for network endogeneity allowing for a two-stage correction à la Heckman (Heckman 1979) and demonstrate the relevance of alumni connections in shaping politicians' legislative effectiveness. The common trait between all these theoretical models is that they establish a link between observable outcomes associated to a network node (for example, educational attainments as proxies of effort levels in Ballester *et al.* (2006); the money received by politicians in Battaglini and Patacchini (2018); the levels of legislative effectiveness in Battaglini *et al.* (2020)) and the respective centrality of the node. This theoretical link can be used to estimate the parameter in the centrality measure using the observable outcomes. This is useful because, for example, it allows to test for network effects or to acquire a deeper knowledge of the topological features of the respective networks.

The routines contained in the package **econet** (Battaglini, Leone Sciabolazza, Patacchini, and Peng 2021) for R (R Core Team 2021) allow for implementation of a number of variations of the linear-in-means model to obtain alternative centrality measures within the family of Katz-Bonacich centrality. Both nonlinear least squares (NLLS) and maximum likelihood (ML) estimators are provided. Several methods for dealing with the identification of network effects are implemented. Moreover, the **econet** package allows for comparison of the explanatory power of parameter-dependent network centrality measures with those of standard measures of network centrality (Wasserman and Faust 1994). As a result, **econet** expands the large set of tools available to R users interested in network analysis. Specifically, it has at least four merits. First, it complements the R packages implementing traditional individual-level centrality measures for binary networks, **igraph** (Csárdi and Nepusz 2006) and **sna** (Butts 2008), and

[1] See also An (2015a) and VanderWeele and An (2013) for a discussion on the difficulties in studying peer effects.

weighted networks, **tnet** (Opsahl 2009), and group-level centrality measures for both binary and weighted networks, **keyplayer** (An and Liu 2016), by introducing new eigensolutions-based techniques to rank individual agents' centrality. Second, whereas previous packages, such as **btergm** (Leifeld, Cranmer, and Desmarais 2018b), **hergm** (Schweinberger and Luna 2018), the **statnet** suite (Goodreau, Handcock, Hunter, Butts, and Morris 2008), and **xergm** (Leifeld, Cranmer, and Desmarais 2018a), created environments for modeling the statistical processes underlying network formation, **econet** provides the first framework to investigate the socio-economic processes operating on networks (i.e., peer effects). Third, it completes the collection of functions for modeling spatial dependence in cross-sectional data provided by **spdep** (Bivand 2022) and **splm** (Millo and Piras 2012), by allowing the users to: i) consider the presence of unconnected nodes, and ii) address network endogeneity. Finally, it equips the R archive with routines still unavailable in other commonly used software for the investigation of relational data, such as MATLAB (The MathWorks, Inc. 2011), Pajek (Batagelj and Mrvar 2003), Python (Van Rossum *et al.* 2011) and Stata (StataCorp 2017). The example we use to showcase the functionality of our R package is taken from Battaglini and Patacchini (2018) and Battaglini *et al.* (2020). The R package **econet** (Battaglini *et al.* 2021) is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=econet`

The rest of the paper is organized as follows. Section 2 briefly reviews the theoretical background of the different approaches used to model the socio-economic processes operating on the network. Section 3 discusses the key elements for the estimation of parameter-dependent centralities, and presents a general taxonomy of the models implemented by the R package **econet**. Section 4 lays out various models and methods to deal with network endogeneity. Section 5 demonstrates the use of the main functions of the package **econet** to determine agent's centrality with examples. Section 6 concludes.

2. Microeconomic foundation

This section provides a theoretical background for the network models of peer effects implemented by the R package **econet**. Many network centrality measures have been introduced in the literature, each capturing different aspects of network topology. Which is the correct way to measure how central is an agent in a network? In this section we describe three economic models that derive conditions under which the Katz-Bonacich centralities that can be estimated using econet are the correct measures of an agent centralities. The aim is to acquaint the researcher interested in working with **econet** of the different theoretical premises of these models, so that he/she can choose the model and the relative **econet** functions most appropriate for conducting his/her investigation.

Model A describes the competition between two or more lobbyists who distribute monetary contributions among n legislators to influence their votes. Each lobbyist aims at maximizing the number of legislators that vote for his/her own preferred policy option. Two legislators are socially connected if they derive utility from voting in the same way. The model provides conditions under which, in the unique Nash equilibrium of the game, the money promised to legislator i is proportional to the Katz-Bonacich centrality of i. The model therefore provides a clear economic interpretation to the Katz-Bonacich centrality and illustrates it relevance in this context. This model was first presented in Battaglini and Patacchini (2018).

Model B studies the extent to which social connections influence the legislative effectiveness of members of the US Congress. By legislative effectiveness we mean the ability of a legislator to pass legislation. In this model, the effectiveness in passing legislation of Congress member i is described by a "production function" in which the inputs are the Congress member i's effort and the effectiveness of all other socially connected Congress members, each weighted by the strength of their social link with i. To determine the optimal level of effort, a legislator needs to predict the equilibrium effectiveness of all the socially linked Congress members, who in turn need to do the same in a Nash equilibrium. Here too, the model provides conditions under which the effectiveness of a Congress member in equilibrium is proportional to a weighted version of the Katz-Bonacich centrality of the legislator, in which the weights are a specific function of the legislator's characteristics. This model was first presented in Battaglini *et al.* (2020).

Models A and B are not alternative ways to represent the same economic problem. Models A and B describe completely different social interactions: the first, a competitive game between two lobbyists; the second, a cooperative game between n legislators.[2] Models A and B are therefore relevant to our discussion because they show how the similar centrality measure can emerge as relevant in a completely different context. Since the tools provided in **econet** can be used to estimate both generalities, these examples illustrate how **econet** can be useful in studying completely different social problems.

Section 2.3 also shortly discusses the popular network model of peer effects by Ballester *et al.* (2006). Interestingly, the prediction of this model is the same of that derived from Model B. This implies that the functions contained in **econet** to estimate Model B can also be used to test the predictions of the model by Ballester *et al.* (2006). The difference between model B and Ballester *et al.* (2006) is in the way the strategic environment is modeled. In model B, the productivity of i is affected by the productivity of the other socially connected players; in Ballester *et al.* (2006), it is assumed that the cost of effort of i depends on the effort level of the other players. In Ballester *et al.* (2006), the actions (i.e., the levels of effort) are predicted to be equal to the Katz-Bonacich centralities; in model B, it is the outcomes that are predicted equal to the centralities. Model B is better suited for empirical analysis as the effectiveness can often be observed and measured, but not effort. Models that have attempted to test the predictions of Ballester *et al.* (2006) have approximated effort with output, but this approximation is not possible if other unobserved factors affect output. The model in Ballester *et al.* (2006), however is an important reference point because it is one of the first models to study these issues. The tools provided by **econet** are also useful in estimating the parameters of the model in Ballester *et al.* (2006).

2.1. Setup of model A

Battaglini and Patacchini (2018), BP henceforth, consider a model in which a legislature with n members chooses between one of two alternatives: a new policy, denoted by A, and a status quo policy, denoted by B.[3] Legislator i's utility of voting for policy p, denoted by $U^i(\mathbf{x}(p))$,

[2] The difference is not just in the interpretation. The models are formally different games in a game theoretic sense: the set of players are different, the strategy space is different, the payoffs are different.

[3] We present here a simplified version of the model in BP for brevity. We refer to the original paper for the more general version.

is:

$$U^i(\mathbf{x}(p)) = \omega\left(s^i(p)\right) + \phi \sum_j g_{i,j} x_j(p) + \varepsilon_p^i \tag{1}$$

The first term in Equation 1 is the utility of the interest groups' contributions: $s^i(p)$ is the contribution pledged in exchange for a vote for p, and $\omega(s)$ is the utility that legislator i's receives from a contribution s. The function $\omega(\cdot)$ is increasing, concave and differentiable with $\lim_{s\to 0}\omega'(s) = \infty$, $\lim_{s\to\infty}\omega'(s) = 0$. The second term in Equation 1 describes the social interaction effects. The social network is described by a $n \times n$ matrix G with generic element $g_{i,j} > 0$, $x_j(p)$ is an indicator function equal to one if legislator j chooses p and zero otherwise, and $g_{i,j}$ measures the strength of the social influence of legislator j on legislator i. The final term in Equation 1 represents other exogenous factors that may affect i's preference for or aversion to voting for p. The terms are normalized so that $\varepsilon_A^i = \varepsilon^i$, where ε^i can be positive or negative, and ε_B^i is set at zero.[4]

Two interest groups, also denoted A and B, attempt to influence the policy outcome. Interest group A is interested in persuading as many legislators as possible to chose policy A; interest group B, instead, is interested in persuading the legislators to choose policy B. Each interest group is endowed with a budget W and promises a contingent payment to each legislator that follows its recommendation. Specifically, interest group A promises a vector of payments $\mathbf{s}_A = (s_A^1, \ldots, s_A^n)$ to the legislators, where s_A^i is the payment received by legislator i if he chooses A; similarly, interest group B promises a vector of payments $\mathbf{s}_B = (s_B^1, \ldots, s_B^n)$ to the legislators, where s_B^i is the payment received by legislator i if he votes for B.

Legislator i is willing to vote for A if and only if $E\left[U_B^i(x) - U_A^i(x)\right] \leq 0$. It is assumed that the interest groups do not know with certainty the legislators' preferences, and so are unable to perfectly forecast how payments affect their voting behavior: so ε^i is assumed to be an independent, uniformly distributed variable with mean zero and density $\Psi > 0$, whose realization is observed only by i. Observing that the probability that legislator i votes for A is $\varphi_i = \mathsf{E}(x_i(A))$, we therefore have that i votes for A only if:

$$\varepsilon^i \leq \omega(s_A^i) - \omega(s_B^i) + \phi \sum\nolimits_j g_{i,j}\left(2\varphi_j - 1\right), \tag{2}$$

From Equation 2, we have that in an interior solution in which all probabilities are in $(0,1)$, the legislators' probabilities of choosing A, $\varphi = (\varphi_1, \ldots, \varphi_n)$, are characterized by the non linear system:

$$\begin{pmatrix} \varphi_1 \\ \cdots \\ \varphi_n \end{pmatrix} = \begin{pmatrix} 1/2 + \Psi\left(\omega(s_A^1) - \omega(s_B^1) + \phi \sum_j g_{1,j}\left(2\varphi_j - 1\right)\right) \\ \cdots \\ 1/2 + \Psi\left(\omega(s_A^n) - \omega(s_B^n) + \phi \sum_j g_{n,j}\left(2\varphi_j - 1\right)\right) \end{pmatrix} \tag{3}$$

that gives a unique vector of equilibrium probabilities $\varphi(s) = \{\varphi_1(s), \ldots, \varphi_n(s)\}$.

The game proceeds as follows. In stage 1, the lobbyists simultaneously commit to a vector of payments $\mathbf{s}_A$ and $\mathbf{s}_B$, without observing ε_p^i. A strategy for interest group l is a probability

[4]Obviously, it is natural to assume that the legislators care about the outcome of the vote. This effect of their vote is proportional to the probability of being pivotal: that is, the case in which A and B votes are tied, or one of them is one vote below the other. The exact pivotal probabilities are computed and incorporated in the legislators' expected utilities in the analysis in BP. Here we omit these terms for simplicity, since in any case they are very small in an election with hundreds of voters.

distribution over the set of feasible transfers S, that is:

$$S = \{s : \sum_i s^i \leq W,\ s^i \geq 0 \text{ for } i = 1, \ldots, n\}.$$

In stage 2, the Congress members see the vector of payments and the shocks ε_p^is and optimally decide how to vote. The lobbyists therefore expect the Congress members to vote with probabilities $\varphi(s) = \{\varphi_1(s), \ldots, \varphi_n(s)\}$ given by Equation 3.

A pair of strategies constitute a Nash equilibrium if they are mutually optimal: the strategy of interest group A maximizes the expected number of legislators who adopt A given φ and interest group B's strategy; and the strategy of interest group B minimizes the expected number of legislators who adopt A given φ and interest group A's strategy.

Interest group A solves:

$$\max_{\mathbf{s}_A \in S} \left\{ \sum_i \left[\varphi_i(\mathbf{s}_A, \mathbf{s}_B) \right] \right\} \tag{4}$$

taking $\mathbf{s}_B$ as given. Interest group B's problem is the mirror image of A's problem, as it attempts to minimize the objective function of Equation 4 taking $\mathbf{s}_A$ as given. The equilibrium solution must satisfy the first order condition:

$$\sum_j \partial \varphi_j(\mathbf{s}_A, \mathbf{s}_B)/\partial s_l^i = \lambda_l \text{ and } \sum_{j=1}^n s_l^j = W \text{ for } i = 1, \ldots, n,\ l = A, B \tag{5}$$

where λ_l is the Lagrangian multiplier associated with the budget constraints $\sum_i s_l^i \leq W$ in interest group l's problem. BP show that the problem of Equation 4 is well behaved and fully characterized by Equation 5; in equilibrium, moreover, A and B have the same Lagrangian multipliers $\lambda_A = \lambda_B = \lambda_*$, so the first order condition is:

$$D\varphi^\top \cdot 1 = \lambda_*$$

where $D\varphi^\top = (\partial \varphi_1^*/\partial s_A^i, \ldots, \partial \varphi_n^*/\partial s_A^i)$.

To understand the relationship between lobbying and centralities we need to "unpack" the voting probabilities. Differentiating Equation 3 and rearranging, we obtain:

$$D\varphi = \Psi \left[I - 2\Psi\phi \cdot \boldsymbol{G} \right]^{-1} D\omega \tag{6}$$

where $D\varphi$ and $D\omega$ are the Jacobians of, respectively, φ and ω. Using Equation 6, we can rewrite the first order condition for the optimality of the lobbyists as:

$$\begin{aligned} D\varphi^\top \cdot 1 &= \Psi \cdot D\omega^\top \cdot \left(I - \phi^* \cdot \boldsymbol{G}^\top \right)^{-1} \cdot \mathbf{1} = \lambda_* \\ &\Rightarrow \quad D\omega^\top \cdot \mathbf{b}\left(\phi^*, \boldsymbol{G}^\top \right) = \lambda_*/\Psi \end{aligned} \tag{7}$$

where $\phi^* = 2\Psi\phi$ and $\mathbf{b}\left(\phi^*, \mathbf{G}^\top\right)$ is the vector of Bonacich centralities of the matrix $\boldsymbol{G}^\top$ with parameter ϕ^*. Note that $D\omega$ is a vector of zeros except for its i-th element that is equal to $\omega'(s_*^i)$. We can therefore write our necessary and sufficient condition of Equation 7 as:

$$b_i\left(\phi^*, \boldsymbol{G}^\top \right) \cdot \omega'(s_*^i) = \lambda_* \text{ for } i = 1, \ldots, n \tag{8}$$

where, without loss in generality, we have incorporated the constant Ψ in the Lagrangian multiplier λ_*. The necessary and sufficient condition of Equation 8 shows the determinants

of the interest group's monetary allocation. The interest group chooses s_*^i equalizing the marginal cost of resources to its marginal benefit. The marginal cost is measured by the Lagrangian multiplier λ_* of Equation 4. The marginal benefit is measured by the increase in expected votes for A. Equation 8 makes clear that, because of network effects, the direct benefit of making a transfer to i is magnified by a factor that is exactly equal to $b_i\left(\phi^*, \boldsymbol{G}^\top\right)$, the Bonacich centrality of i in $\boldsymbol{G}^\top$ with a constant ϕ^*.

2.2. Setup of model B

Battaglini, Leone Sciabolazza, and Patacchini (2020), BLP henceforth, consider a congress comprised of n legislators, where $\mathcal{N} = \{1, \ldots, n\}$ is the set of legislators. Each legislator has a pet legislative project that she/he wants to implement. The goal of each legislator is to maximize her/his legislative effectiveness, measured by the probability of implementing the project. They assume that legislator i's legislative effectiveness at the r-th congress $\mathbf{y}_{\mathbf{r},i}$ is a function of i's characteristics, her/his effort and the legislative effectiveness of all the legislators that i has befriended. Specifically, the technology is assumed to be:

$$\mathbf{y}_{\mathbf{r},i} = A_{r,i} + \varphi \sqrt{\sum_j g_{i,j}\mathbf{y}_{\mathbf{r},j}} \cdot l_i \tag{9}$$

Equation 9 represents the "production function" for legislative effectiveness. The first term, $A_{r,i}$, is a fixed effect idiosyncratic to i.[5] The second term, which is new in our model, captures the importance of social connections. The social network is described by a $n \times n$ matrix $\boldsymbol{G}$ with the generic element $g_{i,j}$ that measures the strength of the social influence of legislator j on legislator i. The adjacency matrix can be as simple as tracking the connections among legislator i and j, for example, $g_{ij} = 1$ if i is connected to j ($j \neq i$) and $g_{ij} = 0$ otherwise. We set $g_{ii} = 0$. The level of effort is l_i; the cost of exerting a level of effort l_i is $(l_i)^2/2$.

A strategy for a legislator is described by a function $l_i : \mathcal{T} \to [0, 1]$, mapping i's type A_i to an effort level. It is assumed that when the floor opens for business, each legislator i chooses her/his own level of effort l_i simultaneously, taking as given the social network and her/his own expectations of the other legislators' effectiveness. Given the optimal reaction functions, the levels of effectiveness are endogenously determined by Equation 9.[6]

The optimal level of effort l^i by a type $i = 1, \ldots, n$ solves the problem:

$$\max_{l_i} \left\{ A_i + \varphi \left(\sqrt{\sum_{j=1}^{n} g_{i,j}\mathbf{y}_{\mathbf{r},j}} \right) \cdot l_i - (l_i)^2/2 \right\},$$

taking $\mathbf{y}_{\mathbf{r}} = (\mathbf{y}_{\mathbf{r},1}, \ldots, \mathbf{y}_{\mathbf{r},n})$ as given. Substituting the solution to this maximization problem in condition of Equation 9, we obtain that the equilibrium levels of legislative efficiency for

[5] This term may include a variety of characteristics that have been highlighted in the existing literature as important for effectiveness: the legislator's seniority, gender and race (potentially in the presence of discrimination) and the legislator's position in the committee system and party hierarchy.

[6] This approach is similar to that in general equilibrium theory in economics where consumers choose their optimal consumption taking prices as given: here legislators choose their levels of effort taking the other legislators' effectiveness as given. As in general equilibrium theory (where prices are endogenous since they need to clear markets), here the levels of effectiveness are endogenous since they must satisfy the externality Equation 9 given the optimal effort levels.

a type $i = 1, \ldots, n$ are given by: $\mathbf{y}_{\mathbf{r},i} = A_{r,i} + \frac{\varphi^2}{2} \sum_{j=1}^{n} g_{i,j} \mathbf{y}_{\mathbf{r},j}$. These equations can be expressed in matrix form as:

$$\left[I - \left(\varphi^2/2\right) \cdot \boldsymbol{G}\right] \cdot \mathbf{y}_{\mathbf{r}} = \mathbf{A}_{\mathbf{r}} \tag{10}$$

where $\mathbf{y}_{\mathbf{r}} = (\mathbf{y}_{\mathbf{r},1}(\boldsymbol{G}, \mathbf{A}_{\mathbf{r}}), \ldots, \mathbf{y}_{\mathbf{r},n}(\boldsymbol{G}, \mathbf{A}_{\mathbf{r}}))'$ is the vector of legislative effectiveness $\mathbf{y}_{\mathbf{r},i}(\boldsymbol{G}, \mathbf{A}_{\mathbf{r}})$ solving Equation 10, and $\mathbf{A}_{\mathbf{r}} = (A_{r,1}, \ldots, A_{r,n})'$ is the vector of types' characteristics. The equilibrium levels of effectiveness are therefore uniquely defined as:

$$(\mathbf{y}_{\mathbf{r},1}(\boldsymbol{G}, \mathbf{A}_{\mathbf{r}}), \ldots, \mathbf{y}_{\mathbf{r},n}(\boldsymbol{G}, \mathbf{A}_{\mathbf{r}}))' = \left[I - \left(\varphi^2/2\right) \cdot \boldsymbol{G}\right]^{-1} \mathbf{A}_{\mathbf{r}}.$$

that is the weighted Katz-Bonacich centrality of legislator i in network $\boldsymbol{G}$ with discount factor $\varphi^2/2$ and weights $\mathbf{A}_{\mathbf{r}} = (A_{r,1}, \ldots, A_{r,n})'$. In the presence of social spillovers among connected legislators (i.e., $\varphi > 0$), however, the effectiveness of any legislator depends on the characteristics of all other legislators, with each legislator weighted using their distance in the network (the weights given by the rows of $\left[I - \left(\varphi^2/2\right) \cdot \boldsymbol{G}\right]^{-1}$). The standard model is nested as a special case of the more general model (with $\varphi = 0$), and so we are able to test if social connections improve the fit of our estimates of $\mathbf{E}$.

2.3. Alternative setup

Alternative microfoundations are games with linear-quadratic utilities that capture linear externalities in agents' actions. A popular setup is a social network model of peer effects with conformity preferences. Let $\mathbf{y}_{\mathbf{r},i}$ denote the legislator i's legislative effectiveness at the r-th congress. Denote by $\overline{\mathbf{y}}_{\mathbf{r},i}$ the average effort of individual i's peers, given by:

$$\overline{\mathbf{y}}_{\mathbf{r},i} = \frac{1}{\bar{g}_i} \sum_{j=1}^{n} g_{i,j} \mathbf{y}_{\mathbf{r},j,}$$

Each legislator i at the congress r selects an effort $\mathbf{y}_{\mathbf{r},i}$, and obtains a payoff $u_{r,i}(\mathbf{y}_{\mathbf{r}})$ that depends on the effort profile $\mathbf{y}_{\mathbf{r}}$ in the following way:

$$u_{r,i}(\mathbf{y}_{\mathbf{r}}) = \left(a_{r,i} + \eta_r + \varepsilon_{r,i}\right) \mathbf{y}_{\mathbf{r},i,} - \frac{1}{2} \mathbf{y}_{\mathbf{r},i,}^2 - \frac{d}{2} \left(\mathbf{y}_{\mathbf{r},i,} - \overline{\mathbf{y}}_{\mathbf{r},i}\right)^2 \tag{11}$$

where $d > 0$. The benefit part of this utility function is given by $\left(a_{r,i} + \eta_r + \varepsilon_{r,i}\right) \mathbf{y}_{\mathbf{r},i,}$ while the cost is $\frac{1}{2} \mathbf{y}_{\mathbf{r},i,}^2$; both are increasing in own effort $\mathbf{y}_{\mathbf{r},i,}$. In this part, $a_{r,i}$ denotes the agent's ex-ante *idiosyncratic heterogeneity*, which is assumed to be deterministic, perfectly *observable* by all individuals in the network and corresponds to the observable characteristics of individual i and to the observable average characteristics of individual i's peers. To be more precise, $a_{r,i}$ can be written as:

$$a_{r,i} = \sum_{m=1}^{M} \beta_m x_{r,i}^m + \frac{1}{\bar{g}_i} \sum_{m=1}^{M} \sum_{j=1}^{n} \theta_m g_{ij}\, x_{r,i}^m$$

where $x_{r,i}^m$ is a set of M variables accounting for observable differences in individual characteristics of individual i, and β_m, θ_m are parameters. In the utility function of Equation 11) η_r denotes the unobservable network characteristics and $\varepsilon_{r,i}$ is an error term, meaning that there is some uncertainty in the benefit part of the utility function. Both η_r and $\varepsilon_{r,i}$ are observed by the individuals but not by the researcher. The second part of the utility function

$\frac{d}{2}\left(\mathbf{y}_{\mathbf{r},i,} - \overline{\mathbf{y}}_{\mathbf{r},i}\right)^2$ reflects the influence of peers' behavior on own behavior. It is such that each individual wants to minimize the *social distance* between herself and her reference group, where d is the parameter describing the *taste for conformity*. Here, the individual loses utility $\frac{d}{2}\left(\mathbf{y}_{\mathbf{r},i,} - \overline{\mathbf{y}}_{\mathbf{r},i}\right)^2$ from failing to conform to others. This is the standard way economists have been modeling conformity (see, among others, Akerlof 1980, Bernheim 1994, Kandel and Lazear 1992, Akerlof 1997, Fershtman and Weiss 1998, Patacchini and Zenou 2012, Patacchini, Rainone, and Zenou 2012).[7] The social norm $\overline{\mathbf{y}}_{\mathbf{r},i}$ can be interpreted as peers' social status. Observe that the social norm here captures the differences between individuals due to network effects. It means that individuals have different types of friends and thus different reference groups $\overline{\mathbf{y}}_{\mathbf{r},i}$. As a result, the social norm each individual i faces is endogenous and depends on her location in the network as well as the structure of the network.

In this game where agents choose their effort level $y_{i,k} \geq 0$ simultaneously, there exists an unique Nash equilibrium (see, e.g., Patacchini and Zenou (2012) given by:

$$\mathbf{y}^*_{\mathbf{r},i,} = \phi \frac{1}{\bar{g}_i} \sum_{j=1}^{n_k} g_{ij} \mathbf{y}^*_{\mathbf{r},j,} + (1-\phi)\left(a_{r,i} + \eta_r + \varepsilon_{r,i}\right)$$

where $\phi = d/(1+d)$. The optimal effort level depends on the individual ex ante heterogeneity $(a_{r,i})$, on the unobserved network characteristics (η_r) and it is increasing with the average effort of the reference group.

3. Network models of peer effects

Following the notation defined in the previous section, for each network r with adjacency matrix $\boldsymbol{G} = [g_{ij}]$, the k-th power of $\boldsymbol{G}$ given by $\boldsymbol{G}^k = \boldsymbol{G}^{(k\text{times})}_{\cdots}\boldsymbol{G}$ keeps track of direct and indirect connections in r. More precisely, the (i,j)-th cell of $\boldsymbol{G}^k$ gives the number of paths of length k in r between i and j. In particular, $\boldsymbol{G}^0 = \boldsymbol{I}$.

Definition 1 (Katz 1953; Bonacich 1987)**.** *Given a vector $\boldsymbol{u} \in \mathbb{R}^n_+$, and $\phi \geq 0$ a small enough scalar, the vector of Katz-Bonacich centralities of parameter ϕ in network g is defined as:*

$$\boldsymbol{b}(g,\phi) = (\boldsymbol{I} - \phi \boldsymbol{G})^{-1}\boldsymbol{u} = \sum_{p=0}^{\infty} \phi^p \boldsymbol{G}^p \boldsymbol{u}. \tag{12}$$

The reduced form of the first order necessary and sufficient condition for optimality in the behavioral model A developed by BP (see Equation 8), can be written as:

$$\mathbf{y}_{\mathbf{r}} = \alpha \cdot (\boldsymbol{I} - \phi \boldsymbol{G})^{-1} + X_r^{\top}\beta + \epsilon_r, \tag{13}$$

where $\mathbf{y}_{\mathbf{r}}$ is the vector of outcomes for the n agents in network r,[8] X_r is a matrix collecting the characteristics of the agents and ϵ_r is a random error term. The coefficients α, ϕ and β

[7] Ballester *et al.* (2006) and Calvó-Armengol, Patacchini, and Zenou (2009) present similar microfoundations for peer effects where agents' behavior depends on the aggregate (rather than average) behavior of peers.

[8] In the context investigated by BLP, $\mathbf{y}$ is the amount of money received by legislators from interest groups in support for their electoral campaign. Observe that model A can be applied to the study of peer effects in other contexts where the behavior of the agents under study is consistent with the theory underlying this model. The same reasoning applies for model B. The interested readers is referred to the recent reviews by An (2011, 2015a); Jackson *et al.* (2017); Hsieh *et al.* (2020); Zenou (2016) for a comprehensive review of the many empirical applications of network models of peer effects.

are the parameters to estimate. Model 13 can be written as

$$\mathbf{y}_r = \alpha \cdot \boldsymbol{b}_1(g,\phi) + X_r^\top \beta + \epsilon_r,$$

where $\boldsymbol{b}_1(g,\phi) = \boldsymbol{b}(g,\phi)$ from Equation 12 when $\boldsymbol{u} = 1$.
For a sample with $\bar{r}$ networks, one can stack up the data by defining $y = (\mathbf{y}_1^\top, \cdots, \mathbf{y}_{\bar{r}}^\top)^\top$, $\epsilon = (\boldsymbol{\epsilon}_1^\top, \cdots, \boldsymbol{\epsilon}_{\bar{r}}^\top)^\top$, $\mathbf{b}(\phi) = \left(\boldsymbol{b}(g,\phi)^\top, \ldots, \boldsymbol{b}(g,\phi)^\top\right)^\top$, $X = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_{\bar{r}}^\top)$ and $G = \text{diag}\{\boldsymbol{G}_r\}_{r=1}^{\bar{r}}$. Observe that the generic matrix $\boldsymbol{G}_r$ has dimension $n_r \times n_r$, and G has dimension $n \times n$, with $n = \sum_{r=1}^{\bar{r}} n_r$. For the entire sample, the model is:

$$y = \alpha \cdot \mathbf{b}(\phi) + X\beta + \epsilon \tag{14}$$

We extend Model 14 by accounting for heterogeneity in network spillovers. Specifically, Model 14 becomes:

$$y = \alpha[I - G(\phi I + \gamma\Lambda)]^{-1}1 + X\beta + \epsilon \tag{15}$$

where $\Lambda = I \otimes z$ is a matrix with the values in the vector z on the diagonal, and all other values are 0. Matrix I has dimension $n \times n$ and the vector z has dimension has dimension $1 \times n$. The vector z represents a given characteristic of the agents. Consistently, the term γ allows for the possibility that agents with different characteristics may be more or less susceptible to social spillovers.

The reduced form of the first order necessary and sufficient condition for optimality in the behavioral model developed by BLP (see Equation 10), is:

$$\mathbf{y}_r = (I - \phi G)^{-1}(\alpha + X_r^\top \beta) + \epsilon_r$$

which can be rewritten as:

$$\mathbf{y}_r = \boldsymbol{b}_2(g,\phi) + \epsilon_r \tag{16}$$

where $\boldsymbol{b}_2(g,\phi) = \boldsymbol{b}(g,\phi)$ from Equation 12 when $\boldsymbol{u} = (\alpha + \boldsymbol{X}_r^\top \beta)$. Equation 16 shows that, in this case as well, the optimal behavior is proportional to a centrality measure within the Family 12. When we consider the case where the parameter ϕ associated with network externalities is not constant across agents, Model 16 in matrix formulation becomes:

$$y = (I - \theta\Lambda G)^{-1}(\alpha + X\beta) + \epsilon \tag{17}$$

$$y = (I - \eta G\Lambda)^{-1}(\alpha + X\beta) + \epsilon \tag{18}$$

In Equations 17 and 18, Λ is an identity matrix with dimension $n \times n$. In Equation 17, $\theta = \theta_0 + \theta_1 z$, where θ_0 is a rescaling factor, and θ_1 quantifies the interaction between the adjacency matrix G and the vector z. Consequently, θ_1 measures the extent to which the peers of an agent with a given characteristic are susceptible to her/his influence. Similarly, in Equation 18, $\eta = \eta_0 + \eta_1 z$, where η_0 is a rescaling factor, and η_1 measures the extent to which an agent with a given characteristic z is more susceptible to the influence of her/his peers.[9]

In addition, BLP also consider the possibility of heterogeneous links (rather than nodes). We consider the case in which agents belong to two different groups and interactions are different between and within groups. To allow for group effects, one can reorder the matrix G so that

[9]For additional details on these models see the online appendix of the paper by Battaglini *et al.* (2020).

the first n_1 columns refer to agents in the first group, and the other $n_2 = n - n_1$ columns refer to agents in the second group. The matrix G can now be divided into four submatrices. The submatrices in the main diagonal (of dimensions $n_1 \cdot n_1$ and $n_2 \cdot n_2$) collect the interactions within groups, whereas the remaining two submatrices (of dimensions $n_1 \cdot n_2$ and $n_2 \cdot n_1$) collect interactions between groups. G can thus be decomposed in two $n \cdot n$ matrices G_{wit} and G_{btw}, with $G = G_{wit} + G_{btw}$. G_{wit} is a matrix that has the same top left and bottom right components as G and it is zero otherwise, and G_{btw} is a matrix that has the same bottom left and top right components as G and it is zero otherwise. As a result, Model 16 becomes

$$y = (I - \phi_1 G_{wit} - \phi_2 G_{btw})^{-1}(\alpha + X\beta) + \epsilon \tag{19}$$

where ϕ_1 captures within-group spillovers, and ϕ_2 between-group spillovers. The taxonomy of models outlined above shows that centrality measures should be calculated according to the most appropriate behavioral model describing the agents' behaviors and the requested level of heterogeneity of agents and network links. In BP and BLP we bring these theories to the data and find support for their predictions in different contexts.

3.1. Estimation

Models 14 and 16 cannot be estimated by a simple OLS regression in which $\mathbf{y}$ represents the dependent variables and $\mathbf{b}(\phi)$ and X are the independent variables because $\mathbf{b}(\phi)$ is a nonlinear function of a parameter to be estimated, ϕ. We can, however, obtain estimates for α, ϕ and β using NLLS or ML.

The NLLS requests solving the nonlinear least-squares problem for Equation 14 and 16. This task is performed by **econet** using the Levenberg-Marquardt algorithm implemented by Box, Davies, and Swann (1969) in the R package **minpack.lm** developed by Elzhov, Mullen, Spiess, and Bolker (2022). The details of how the NLLS works in practice can be found in More (1978).

The ML estimation requests ML functions that can be derived as follows, by assuming $\varepsilon \sim N(0, \sigma^2 I)$. For Equation 14, assume $\varepsilon \sim N(0, \sigma^2 I)$, the log likelihood function is

$$\ln(L) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{1}{2}\left[y - (I - \phi G)^{-1}1 - X\beta\right]'\left[y - (I - \phi G)^{-1}1 - X\beta\right]/\sigma^2$$

where n is the total sample size. Equation 15 will have the same form as Equation 14, except substituting ϕG with $G(\phi I + \gamma\Lambda)$.

In a similar fashion, we consider the following maximum likelihood functions for Equation 16, 17, 18 and 19. For Equation 16, the log likelihood function corresponding to Equation 20 is:

$$y = \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\left[y - (I - \phi G)^{-1}X\beta\right]'\Omega^{-1}\left[y - (I - \phi G)^{-1}X\beta\right] \tag{20}$$

where $\Omega = \sigma^2(I - \phi G)^{-1}(I - \phi G')^{-1}$. Equation 17, 18 and 19 have the same likelihood function as above except substituting ϕG with $\theta\Lambda G$ or $\eta G\Lambda$ or $\phi_1 G_{wit} - \phi_2 G_{btw}$. When the network is sparse, isolates (i.e., individuals with no neighbors) will exist.In this case, we can modify the likelihood function to expedite the process. Rewrite

$$G = \begin{pmatrix} G^c_{n_1 \times n_2} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & \mathbf{0}_{n_2 \times n_1} \end{pmatrix}$$

where n_1 is the size of all connected individuals and n_2 is the sizes of isolates.

$$(I - \phi G)^{-1} = \begin{pmatrix} (I - \phi G^c_{n_1 \times n_2})^{-1} & \mathbf{0}_{n_1 \times n_2} \\ \mathbf{0}_{n_2 \times n_1} & I_{n_2 \times n_1} \end{pmatrix}$$

Define $y = \left(y^{c'}, y^{u'}\right)'$ and $X = \left(X^{c'}, X^{u'}\right)'$. The likelihood function (Equation 20) can be transformed into:

$$\begin{aligned} y = & \frac{n_1}{2} \ln(2\pi) - \frac{1}{2} \ln|\Omega_1| - \frac{1}{2} \left[y^c - (I - \phi G^c)^{-1} X^c \beta\right]' {\Omega_1}^{-1} \left[y^c - (I - \phi G^c)^{-1} X^c \beta\right] + \\ & \frac{n_2}{2} \ln(2\pi) - \frac{n_2}{2} \sigma^2 - \frac{1}{2} \left[y^u - X^u \beta\right]' \left[y^u - X^u \beta\right] \end{aligned} \tag{21}$$

Where $\Omega_1 = \sigma^2 (I - \phi G^c)^{-1} (I - \phi G^c)^{-1}$.

Our package allows the adjacency matrix to be used as a direct input. This setting simplifies the data processing procedure compared with other R packages like **spdep** when dealing with social network data. Packages like **spdep** are designed for spatial data. In this environment, the network data is required to be imported as neighbor pairs. However, social network data differs from spatial data since isolates may exists (nodes that have no connection with all other nodes). Networks containing isolates are not compatible with the data structure for packages like **spdep**.

Our package not only provides a way to get around this problem but also proposes an efficient algorithm when including those isolates. Instead of inverting the entire adjacency matrix, we show in the algebra above that one only needs to invert the adjacency matrix for connected nodes. The likelihood function can be written as a sum of the spatial auto-regressive (SAR) likelihood function for connected nodes and a standard linear likelihood function of isolates (see Equation 21).

4. Addressing network endogeneity

In many real-world contexts, the network topology is the result of the choices of the agents as much as their behavior over the observed topologies. As a result, the data structure can be endogenous, and inference neglecting this issue would be invalid. The simplest way to tackle the problem is to model network formation using a homophily model (see e.g., Fafchamps and Gubert 2007; Mayer and Puller 2008; Lai and Reiter 2017; Apicella, Marlowe, Fowler, and Christakis 2012; Attanasio, Barr, Cardenas, Genicot, and Meghir 2012) where the existence of a link between i and j, $g_{i,j}$, is explained by the distance between i and j in terms of characteristics, according to the model

$$g_{i,j} = \delta_o + \sum_l \delta_{l+1} |x_i^l - x_j^l| + u_{i,j} \tag{22}$$

where x_i^l for $l = 1, \ldots, L$ are i's characteristics. As standard in the literature on dyadic link formation, the main assumption underlying Model 22 is dyadic independence, i.e., the assumption that each agent's choices are not influenced by others' decisions, and therefore each link in the network occurs with the same probability.[10]

[10]This approach can be extended to dyadic dependence using latent space models or exponential random graph models (see An 2011, for a discussion).

Fafchamps, Leij, and Goyal (2010) and Graham (2015) suggest a variation of this model where this assumption can be tested. They suggest to include in the model the length of the shortest distance between i and j (Fafchamps *et al.* 2010), or the number of shared friends between i and j (Graham 2015).

Let us denote this additional variable, $\kappa_{i,j}$:

$$g_{i,j} = \delta_o + \delta_2 \kappa_{i,j} + \sum_l \delta_{l+w} |x_i^l - x_j^l| + u_{i,j} \tag{23}$$

A statistically significant estimate of the parameter δ_2 would suggest that the presence of a link depends on the presence of links at path lengths higher than 2 (or on the number of shared friends), thus indicating a violation of the hypothesis of dyadic independence.

A different variation of Model 22 is proposed by Graham (2016). The model in Graham (2016) accounts for agents' unobserved heterogeneity by adding fixed effects for agents i and j:

$$g_{i,j} = \delta_o + \delta_1 \omega_{i,j} + \sum_l \delta_{l+1} |x_i^l - x_j^l| + \iota_i + \iota_j + u_{i,j}, \tag{24}$$

These models of network formation can help mitigate concerns about network endogeneity in linear models of peer effects, such as Models 14 and 16, in one of two ways: i) they can be used to predict network connections on the basis of exogenous agents' characteristics and then use the predicted network topology as an instrument for the actual network structure, or ii) they can be used as a first step selection equation to derive a correction for network endogeneity à la Heckman.[11]

BLP follow approach ii. Under the assumption that $\varepsilon_r = (\varepsilon_{r,1}, \ldots, \varepsilon_{r,n})'$ and $\{(u_{i,j,r})\}_{i,j}$ are jointly normal with $\mathsf{E}(\epsilon_{i,r}^2) = \sigma_\epsilon^2$, $\mathsf{E}(\epsilon_{i,r} u_{i,j,r}) = \sigma_{\epsilon u}$ for all $i \neq j$, $\mathsf{E}(u_{i,j,r} u_{ik,r}) = \sigma_u^2 \ \forall j = k$, and $\mathsf{E}(u_{i,j,r} u_{i,k,r}) = 0 \ \forall j \neq k$, the expected value of the error term conditional on the link formation is $\mathsf{E}(\epsilon_{i,r} \mid \{u_{i,j,r}\}_{j \neq i}) = \psi \cdot \sum_{j \neq i} u_{i,j,r}$, where $\psi = \sigma_{\epsilon u} / \sigma_u^2$. It follows that Model 16 can be written as:

$$\mathbf{y}_r = [I - \phi \boldsymbol{G}]^{-1} \cdot [\alpha \cdot \mathbf{1} + \boldsymbol{X}_r \beta + \psi \xi_r + \varepsilon_r] \tag{25}$$

where $\xi_{i,r} = \sum_{j \neq i} u_{i,j,r}$ with $\xi_r = (\xi_{i,r}, \ldots, \xi_{n,r})'$. The term $\psi \xi_r$ now captures the selection bias. The model can then be estimated with NLLS (see Section 3). Under these assumptions, approach i) and approach ii) should produce similar results since both deliver consistent estimators.

Under approach i), however, inference is complicated because the selectivity term ξ is a generated regressor from a previous estimation and no closed form solution is available for the NLLS standard errors estimates in a network context. For this reason, BLP use bootstrapped standard errors. Because of the inherent structural dependency of network data, the design of the resampling scheme for this bootstrap procedure needs special consideration. The residual vector in Equation 25 does not contain i.i.d. elements, and one cannot sample with replacement from this vector. Then, BLP use the residual bootstrap procedure, which is common in spatial econometrics (see Anselin 1990), where resampling is performed on the structural errors, under the assumption that they are i.i.d. In practice, the vector of structural errors are derived from $\varepsilon = [I - \widehat{\phi} G] \boldsymbol{u}$, where $\boldsymbol{u}$ is the residual vector from Equation 24.

[11] See Heckman (1979), who first proposed this technique. For different applications of the Heckman approach in spatial statistics and for network models see Johnnson and Moon (2021); Qu and Lee (2015).

An important challenge in using the two step procedures for both approach i) and ii) is finding exclusion restrictions, that is factors that affect network formation only.[12] This is notoriously a difficult task. BLP use an original instrument: connections between agents which are made during adolescence. Those connections are powerful predictors of social contacts later on in life, but clearly predetermined to decisions taken in the adulthood.

5. Implementing econet

We now turn the discussion to the implementation of the functions contained in **econet**. **econet** implements the set of linear models of social interactions introduced in Section 2, where an agent's outcome is a function of the outcomes of the connected agents in the network. The routines provide both NLLS and ML estimators. The possible sources of endogeneity that could hinder the identification of a causal effect in the model can be addressed by implementing the two-step correction procedures described in Section 3. The estimated parameter capturing the impact of the social network on an agent's performance is then used to measure the individual importance in the network, obtaining a weighted version of Katz-Bonacich centrality. Finally, the explanatory power of the parameter-dependent centrality can be compared with those of standard measures of network centrality. It is worth emphasizing that **econet** allows the inclusion of unconnected agents, for whom the Katz-Bonacich centrality is constant.

Specifically, **econet** provides four functions. The first one is `net_dep`, which allows one to estimate a model of social interactions and compute the relative weighted Katz-Bonacich centralities of the agents. Different behavioral models can be chosen (i.e., those provided by BP and BLP). Moreover, the hypothesis of homogeneous or heterogeneous spillovers can be tested. The second function is `boot`, which is built to obtain valid inference when the NLLS estimator with Heckman correction is used. The third function is `horse_race`, which allows one to compare the explanatory power of parameter-dependent centralities relative to other centrality measures. The fourth function is `quantify`, and it is used to assess the effect of control variables in the framework designed by BLP.

5.1. Detailing the functions

The modeling choices presented so far are implemented by the function `net_dep`. The first three arguments of this function are: i) `formula`, an object of class '`formula`' which specifies the independent variable and the controls; ii) `data`, an object of class '`data.frame`' containing the values of the variables included in *formula*, iii) `G`, an object of class '`Matrix`' where the generic element g_{ij} is used to track the connection between i and j in the social network. `G` can be unweighted (i.e., $g_{ij} = 1$ if i and j are connected, and 0 otherwise), or weighted (e.g., g_{ij} collects the intensity of the relations between i and j) and column-normalized. The matrix must be arranged in the same order of the data, and row and column names must indicate agents' ids.

[12] Technically, the two step model is identified even using exactly the same set of regressors in both stages since the dyad-specific repressors used in the first stage (the network formation stage) are expressed in absolute values of differences. These differences do not appear in the outcome equation. Identification is thus achieved by exploiting non-linearities specific to the network structure of our model. While this strategy has been used in the applied network literature (see e.g., Goldsmith-Pinkham and Imbens 2013; Hsieh and Lee 2016), it may be a tenous source of identification in some cases.

Model	Hypothesis	Equation	Centrality measure $\boldsymbol{b}(g,\phi)$	
A	`lim`	14	ϕ: homogeneous	$\boldsymbol{b}(g,\phi) = (I - \phi G)^{-1}1$
	`het`	15	ϕ: heterogeneous by node type	$\boldsymbol{b}(g,\phi) = [I - G(\phi I + \gamma\Lambda)]^{-1}1$
B	`lim`	16	ϕ: homogeneous	$\boldsymbol{b}(g,\phi) = (I - \phi G)^{-1}1$
	`het_l`	17	ϕ: heterogeneous outgoing influence by node type	$\boldsymbol{b}(g,\phi) = (I - \theta\Lambda G)^{-1}1$
	`het_r`	18	ϕ: heterogeneous ingoing influence by node type	$\boldsymbol{b}(g,\phi) = (I - \eta G\Lambda)^{-1}1$
	`par`	19	ϕ: heterogeneous by link type	$\boldsymbol{b}(g,\phi) = (I - \phi_1 G_{wit}\phi_2 G_{btw})^{-1}1$

Table 1: Field specification in `net_dep`.

The next two arguments in `net_dep`, `model` and `hypothesis`, are used to specify the model to be estimated through an object of class '`character`', as documented in Table 1. Specifically, the argument `model` indicates the framework to be applied: i.e., it is set to `Model_A` or `Model_B` to implement respectively the framework by BP and BLP. The argument `hypothesis` is necessary to indicate whether peer effects are assumed to be homogeneous `model = "lim"` or heterogeneous (`model = c("het", "het_l", "het_r")`).

The argument `z` is used to specify the source of heterogeneity for the peer effects parameter ϕ (i.e., the variable z in Equations 15, 17, 18), or the groups in which the network should be partitioned when `model = "par"`. Specifically, `z` is a numeric vector where the generic element i refers to agent's i characteristic (e.g., it takes 1 if i is a female, and 0 otherwise).

In order to correct for the potential bias arising from network endogeneity, we include in the function four arguments: i) `endogeneity`, a logical object equal to `TRUE` if `net_dep` should implement the two-step correction procedure as e.g., in Model 25, and `FALSE` otherwise; ii) `correction`, an object of class '`character`' that is set to indicate whether `net_dep` should implement a Heckman correction (`correction = "heckman"`), or an instrumental variable approach (`correction = "iv"`); iii) `exclusion_restriction`, a object of class '`Matrix`' used to specify the matrix to instrument the endogenous network; iv) `first_step`, an object of class '`character`' which specifies the network formation model to be used in the first step of the procedure for both `correction = "heckman"` and `correction = "iv"`. This argument can be equal to `standard` (Equation 22), `shortest` (Equation 23, as in Fafchamps *et al.* 2010), `coauthor` (Equation 23, as in Graham 2015), `fe` (Equation 24), and `degree`, where the difference in degree centrality between agents is an additional regressor of Equation 22.

Finally, the argument `estimation` allows the user to choose the estimation technique. This is an object of class '`character`' that can be one of two options: `NLLS` (nonlinear least squares), which implements the Levenberg-Marquardt optimization algorithm for solving the nonlinear least-squares problem using the function `nlsLM` contained in the R package **minpack.lm** (Elzhov *et al.* 2022) or `MLE` (maximum likelihood), which uses the function `mle2` of the R package **bbmle** to implement the quasi-Newton method by Bolker and R Core Team (2021) for maximum-likelihood bound constrained optimization.

The complete list of all the inputs of the function `net_dep` is available in the help page of the package **econet**, and it is accessible from R by running the code `?net_dep`.

The output of `net_dep` consists of a list of three objects: i) the point estimates and relative standard errors of the model's parameters; ii) the vector of agents' network centrality; iii) the point estimates and relative standard errors of the parameters of the first stage model, if `endogeneity = TRUE`.

We provide below a list of examples to illustrate the functionality of `net_dep`. Since in all examples we reject the hypothesis of normality of the errors, the NLLS estimation method is used.[13]

Exercise 1: Katz-Bonacich centrality with parameter constant across agents

In the first example, we estimate the association between a Congress member's network centrality and the amount of dollars he received from interest groups to finance his electoral campaign for the 111th Congress using Model 14. The variables used to control for the effect of legislators' characteristics are: party affiliation (`party`); gender (`gender`); chairmanship (`nchair`); whether or not the Congress member has at least one connection in the network (`isolate`).

The network used for this exercise represents the connections between agents which are made during adolescence. The network is constructed using information on the educational institutions attended by the Congress members. Specifically, we assume that a tie exists between two Congress members if they graduated from the same institution within eight years of each other. We set a link between two Congress members, g_{ij}, to be equal to the number of schools they both attended within eight years of each other; then we row normalize the social weights so that $\sum_i g_{ij} = 1$ for any i. This analysis is a simplified version (in terms of both data and controls included in the model specification) of the analysis in BP.

```
R> library("econet")
R> set.seed(2)
R> data("a_db_alumni", package = "econet")
R> data("G_alumni_111", package = "econet")
R> db_model_A <- a_db_alumni
R> G_model_A <- a_G_alumni_111
R> are_factors <- c("party", "gender", "nchair", "isolate")
R> db_model_A[are_factors] <- lapply(db_model_A[are_factors], factor)
R> db_model_A$PAC <- db_model_A$PAC/1e+06
R> f_model_A <- formula("PAC ~ gender + party + nchair + isolate")
R> starting <- c(alpha = 0.47325, beta_gender1 = -0.26991,
+    beta_party1 = 0.55883, beta_nchair1 = -0.17409,
+    beta_isolate1 = 0.18813, phi = 0.21440)
R> lim_model_A <- net_dep(formula = f_model_A, data = db_model_A,
+    G = G_model_A, model = "model_A", estimation = "NLLS",
+    hypothesis = "lim", start.val = starting)
R> summary(lim_model_A)


Call:
Main Equation: PAC ~ alpha * solve_block(I - phi * G) %*% Ones +
beta_gender1 * gender1 + beta_party1 * party1 +
beta_nchair1 * nchair1 + beta_isolate1 * isolate1
```

[13]Observe that in the examples presented in the paper, a set of starting values is used to estimate NLLS with the function `net_dep`. When these are not provided by the user, `net_dep` uses some random values, and NLLS takes significantly more time to converge in this case. The reader interested in learning more about how starting values should be used when running NLLS estimations can refer to Box *et al.* (1969). Additional details on how to specify starting values with `net_dep` can be found by running the code `?net_dep` from R.

```
              Estimate Std. Error t value Pr(>|t|)
alpha          0.47325    0.17969   2.634  0.00876 **
beta_gender1  -0.26991    0.10504  -2.570  0.01052 *
beta_party1    0.55883    0.08363   6.682  7.5e-11 ***
beta_nchair1  -0.17409    0.19004  -0.916  0.36016
beta_isolate1  0.18813    0.18196   1.034  0.30178
phi            0.21440    0.27005   0.794  0.42768
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1042.55  loglik: -514.28
```

In this model, the estimate of the spillover effect (ϕ), assumed to be the same for all Congress members, is positive (even though is not statistically significant in this example): The estimated value of ϕ is used to calculate the weighted Katz-Bonacich centrality of the agents (as mentioned, this value can be extracted from the second object stored in the output of `net_dep`: e.g., `lim_model_A$centrality`). The estimate of the intercept α directly measures the impact of this network centrality measure. Its interpretation is akin to the one of an estimated coefficient in a linear regression model. We find a positive effect of a legislator's centrality on campaign contributions, showing that more connected Congress members are likely to receive more attention from interest groups. Specifically, when the Katz-Bonacich centrality of agent i increases by one unit, the amount of dollars received by i from interest groups increases by $0.47 \times 1,000,000 = 470,000$\$. The same logic can be applied to interpret the other estimated coefficients.

Let us now repeat the same exercise for Model 16. We use a simplified version (in terms of both data and controls included in the model specification) of the analysis presented in BLP. Our goal here is to investigate the association of legislative networks with Congress members' legislative effectiveness score (LES). Network ties are defined here as the number of bills that j cosponsored with i. Also in this case, we impose that $\sum_i g_{ij} = 1$. The underlying idea is that legislators' productivity is affected by the productivity of the other legislators with whom they interact. However, while alumni connections are formed during adolescence and can thus be reasonably assumed to be exogenous to a Congress member's political activity, cosponsorships are instead endogenous, since legislators are clearly strategic in choosing with whom to cosponsor a bill. The function `net_dep` allows the users to control for network endogeneity using the Heckman correction procedures described in Section 3.1. The legislators' characteristics used in this context are the same used in the previous example except the variable `isolate`, since all Congress members have at least one cosponsorship link.

```
R> data("db_cosponsor", package = "econet")
R> data("G_alumni_111", package = "econet")
R> db_model_B <- db_cosponsor
R> G_model_B <- G_cosponsor_111
R> G_exclusion_restriction <- G_alumni_111
R> are_factors <- c("gender", "party", "nchair")
R> db_model_B[are_factors] <- lapply(db_model_B[are_factors] , factor)
R> f_model_B <- formula("les ~gender + party + nchair")
R> starting <- c(alpha = 0.23952, beta_gender1 = -0.22024,
```

```
+    beta_party1 = 0.42947, beta_nchair1 = 3.09615,
+    phi = 0.40038, unobservables = 0.07714)
R> lim_model_B <- net_dep(formula = f_model_B, data = db_model_B,
+    G = G_model_B, model = "model_B", estimation = "NLLS",
+    hypothesis = "lim", endogeneity = TRUE,
+    correction = "heckman", first_step = "standard",
+    exclusion_restriction = G_exclusion_restriction,
+    start.val = starting)
R> summary(lim_model_B)

Call:
Main Equation: les ~ solve_block(I - phi * G) %*% (alpha * Ones +
beta_gender1 * gender1 + beta_party1 * party1 +
beta_nchair1 * nchair1 + beta_unobservables * unobservables)
                   Estimate Std. Error t value Pr(>|t|)
alpha               0.23952    0.07130   3.359 0.000851 ***
beta_gender1       -0.22024    0.14052  -1.567 0.117787
beta_party1         0.42947    0.10111   4.247 2.65e-05 ***
beta_nchair1        3.09615    0.25951  11.931  < 2e-16 ***
phi                 0.40038    0.06136   6.525 1.90e-10 ***
beta_unobservables  0.07714    0.06138   1.257 0.209519
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1342.51  loglik: -664.25

R> summary(lim_model_B, print = "first.step")

First step:  y ~ exclusion_restriction + gender1 + party1 + nchair1
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.481e-03  3.809e-05  38.887  < 2e-16 ***
exclusion_restriction 4.657e-03  4.129e-04  11.279  < 2e-16 ***
gender1              -6.638e-05  2.135e-05  -3.109  0.00188 **
party1                2.366e-03  1.940e-05 121.922  < 2e-16 ***
nchair1              -4.180e-04  3.436e-05 -12.163  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R2: 0.07
```

The output of function `summary(lim_model_B, print = "first.step")` reports the estimates of the first step Model 22. The interpretation of the results from the first step model follows the standard interpretation of the results of a linear probability model. Being connected in the alumni network (`G_exclusion_restriction`) has a positive and significant impact on the probability that two legislators will cosponsor a bill, hence the alumni network is an important predictor of the cosponsorship network (`G_Model_B`).

The output of the function `summary(lim_model_B)` presents the estimation results of Model 25. Because of the two-step procedure, standard errors of Model 16 are bootstrapped using the function `boot`. This function takes the following arguments: i) `fit`, the first object of `net_dep`'s output; ii) `group`, a numeric vector used to specify if the resampling should be performed within specific groups; iii) `niter`, an object of class '`numeric`' which indicates the iterations of the bootstrap; iv) `weights`, a logical object equal to `TRUE` if the object `fit` is estimated with the argument `to_weight` different from `NULL`, and `FALSE` otherwise (the second is the default option); `parallel`, a logical object equal to `TRUE` if the user wants to use parallel computation, and `FALSE` otherwise (the second is the default option); `ncores`, an object of class '`numeric`' which indicates the number of cores to be used for running parallel computation.[14]

```
R>  boot_lim_estimate <- boot(object = lim_model_B, hypothesis = "lim",
+    group = NULL, niter = 2, weights = FALSE, parallel = FALSE,
+    ncores = NULL)
R>  boot_lim_estimate
```

```
                   coefficient boot.Std.Error boot.t.value boot.p.value
alpha                0.23951648     0.08747214     2.738203 6.432331e-03
beta_gender1        -0.22023753     0.14368757    -1.532753 1.260671e-01
beta_party1          0.42946569     0.11875662     3.616352 3.339708e-04
beta_nchair1         3.09614813     0.24416950    12.680323 1.487650e-31
phi                  0.40038484     0.06001437     6.671483 7.756648e-11
beta_unobservables   0.07714043     0.05965438     1.293123 1.966581e-01
```

The results of this second exercise show a positive and significant network effects (ϕ) on the effectiveness of agents, meaning that Congress members benefit from their interactions with the colleagues conscripted to their own causes. It is worth noting that network endogeneity does not seem to be a major concern in this simple example, since the correlation between the unobservables of link formation and outcome equation (ϕ in Model 16) is not statistically significant.

Observe that while in Model 14 the estimated effect of network centrality is captured by one parameter α, in Model 16 it requires further elaboration since it varies with individual characteristics. For the k-th covariate in Model 16, if $\phi > 0$, the centrality's marginal effect is $(I - \varphi G)^{-1}(I\beta_k)$, which is a $n \cdot n$ matrix with its (i, j)-th element representing the effect of a change in the characteristic k for agent j on the outcome of agent i. The diagonal elements capture the direct effect of a marginal change in the characteristic k for agent i. The elements outside the diagonal instead capture the indirect effects, that is the effects on the outcome of i triggered by variation of the characteristic k in other agents. The direct effects are comparable to the OLS estimated effects without considering the network effects. The important difference in comparing the estimates of the covariates in the models with and without network effects is precisely that when $\phi > 0$, the marginal effect of the k-th covariate in Model 16 is not just β_k but it also depends on the individual's position in the network (i.e., on the individual's network centrality).

[14]Please note that this function can take considerable time to run. In addition, the argument `parallel` has been fully tested only within the Windows operating system. We plan in the future to make this option available across other major operating systems.

The function `quantify` allows the user to run this task. Specifically, it provides the estimated impacts of the agents' characteristics with network effects (which is equivalent to the estimated impacts of agents' network centrality by characteristic) and compares them with the OLS estimates. Because, as we said before, the marginal effects of characteristics are different for different agents, the function `quantify` reports the mean, standard deviation, maximum and minimum for both direct and indirect effects. It requires the argument `fit`, which is the output of either the function `net_dep` or `horse_race` (discussed below).

```
R> quantify(fit = lim_estimate_model_B)
```

```
                     beta Direct_mean Direct_std Direct_max Direct_min
beta_gender1      -0.2202     -0.2205     0.0002    -0.2202    -0.2217
beta_party1        0.4295      0.4299     0.0003     0.4323     0.4295
beta_nchair1       3.0961      3.0992     0.0025     3.1162     3.0961
beta_unobservables 0.0771      0.0772     0.0001     0.0776     0.0771
                   Indirect_mean Indirect_std Indirect_max Indirect_min
beta_gender1             -0.0003       0.0005       0.0000      -0.0149
beta_party1               0.0007       0.0009       0.0290       0.0000
beta_nchair1              0.0047       0.0068       0.2091       0.0000
beta_unobservables        0.0001       0.0002       0.0052       0.0000
```

The estimation results show the mean, standard deviation, maximum and minimum for both direct and indirect effects. Perhaps unsurprisingly, it appears that the indirect effects are smaller than the direct effects.[15]

As already shown, agents' centrality measures can be accessed using the operator `$` with `net_dep`'s output, e.g., `lim_model_B$centrality`, and it can be used for different applications. For example, we can use it to rank agents' positions in the Congress social space, or to study the centrality distribution in the Republican and the Democratic party, as we do in Figure 1.

Exercise 2: Katz-Bonacich centrality with heterogeneous by node parameter

The code below estimates Model 15 using gender as the relevant dimension of heterogeneity. Gender is a dummy variable which takes 1 if the legislator is a female, and 0 otherwise.

```
R> z <- as.numeric(as.character(db_model_A[, "gender"]))
R> f_het_model_A <- formula("PAC ~ party + nchair + isolate")
R> starting <- c(alpha = 0.44835, beta_party1 = 0.56004,
+    beta_nchair1 = -0.16349, beta_isolate1 = 0.21011,
+    beta_z = -0.26015, phi = 0.34212, gamma = -0.49960)
R> het_model_A <- net_dep(formula = f_het_model_A, data = db_model_A,
+    G = G_model_A, model = "model_A", estimation = "NLLS",
+    hypothesis = "het", z = z, start.val = starting)
R> summary(het_model_A)
```

[15] For additional details on the interpretation of the estimated parameters of network models with peer effects see LeSage and Pace (2009), Chapter 2.7.

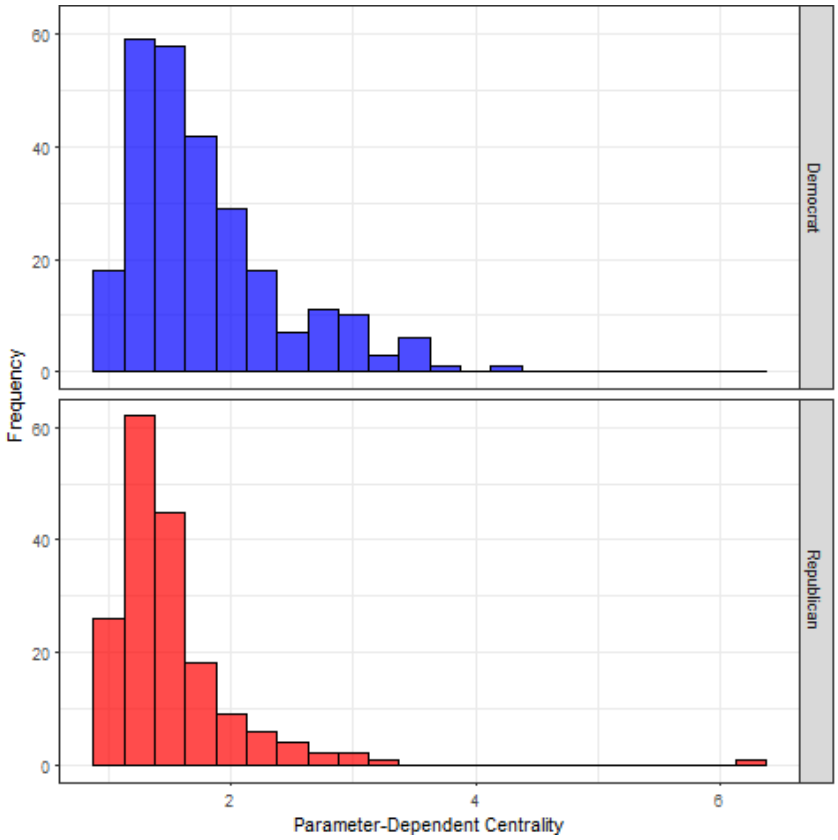

Figure 1: BLP model: Distribution of Parameter-Dependent Network Centrality.

```
Call:
Main Equation:  PAC ~ alpha * solve_block(I - G %*% (phi * I +
gamma * G_heterogeneity)) %*% Ones + beta_party1 * party1 +
beta_nchair1 * nchair1 + beta_isolate1 * isolate1 + beta_z * z
              Estimate Std. Error t value Pr(>|t|)
alpha          0.44835    0.17942   2.499   0.0128 *
beta_party1    0.56004    0.08342   6.713 6.21e-11 ***
beta_nchair1  -0.16349    0.18984  -0.861   0.3896
beta_isolate1  0.21011    0.17927   1.172   0.2418
beta_z        -0.26014    0.10655  -2.441   0.0150 *
phi            0.34212    0.25377   1.348   0.1783
gamma         -0.49960    0.34662  -1.441   0.1502
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1042.34  loglik: -513.17
```

The results show that in this simple example γ is not significant, suggesting that female Congress members do not have a different level of influence than their male peers. Note that the weighted Katz-Bonacich centrality derived from the spillover effect are stored in the object `het_model_A$centrality`.

If we use Equation 17 and 18, we distinguish between outgoing and incoming influence, that is if females are more (or less) able to influence and to be influenced by their peers respectively. The code below implements this analysis. For ease of exposition, we do not consider the possible endogeneity of the social network.

```
R> z <- as.numeric(as.character(db_model_B[, "gender"]))
R> f_het_model_B <- formula("les ~ party + nchair")
R> starting <- c(alpha = 0.23952, beta_party1 = 0.42947,
+    beta_nchair1 = 3.09615, beta_z = -0.12749,
+    theta_0 = 0.42588, theta_1 = 0.08007)
R> het_model_B_l <- net_dep(formula = f_het_model_B, data = db_model_B,
+    G = G_model_B, model = "model_B", estimation = "NLLS",
+    hypothesis = "het_l", z = z, start.val = starting)
R> starting <- c(alpha = 0.04717, beta_party1 = 0.51713,
+    beta_nchair1 = 3.12683, beta_z = 0.01975,
+    eta_0 = 1.02789, eta_1 = 2.71825)
R> het_model_B_r <- net_dep(formula = f_het_model_B, data = db_model_B,
+    G = G_model_B, model = "model_B", estimation = "NLLS",
+    hypothesis = "het_r", z = z, start.val = starting)
R> summary(het_model_B_l)

Call:
Main Equation:  les ~ solve_block(I - (theta_0 * I -
theta_1 * G_heterogeneity) %*% G) %*% (alpha * Ones +
beta_party1 * party1 + beta_nchair1 * nchair1 + beta_z * z)
             Estimate Std. Error t value Pr(>|t|)
alpha         0.22740    0.07584   2.998  0.00287 **
beta_party1   0.41382    0.10198   4.058 5.88e-05 ***
beta_nchair1  3.07797    0.26148  11.771  < 2e-16 ***
beta_z       -0.12749    0.21199  -0.601  0.54789
theta_0       0.42587    0.07418   5.741 1.77e-08 ***
theta_1       0.08007    0.12320   0.650  0.51611
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1343.68  loglik: -664.84

R> summary(het_model_B_r)

Call:
Main Equation: les ~ solve_block(I - G %*% (eta_0 * I -
eta_1 * G_heterogeneity)) %*% (alpha * Ones +
beta_party1 * party1 + beta_nchair1 * nchair1 + beta_z * z)
             Estimate Std. Error t value Pr(>|t|)
alpha         0.04717    0.06867   0.687  0.49251
beta_party1   0.51713    0.09259   5.585 4.13e-08 ***
beta_nchair1  3.12683    0.25680  12.176  < 2e-16 ***
```

```
beta_z        0.01976    0.15098   0.131  0.89595
eta_0         1.02790    0.22400   4.589 5.85e-06 ***
eta_1         2.71825    0.91547   2.969  0.00315 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1336.63  loglik: -661.31
```

The results shows that females do not seem to be able to influence their socially connected peers (θ is not significant), but they are helpful to their colleagues ($\eta_1 > 0$). In terms of network analysis, this implies that all else being equal, Congress members located close to female colleagues benefit from their position since they can leverage females to be more effective in their legislative activity. In this case, the weighted Katz-Bonacich centralities are stored in `het_model_B_l$centrality` and `het_model_B_r$centrality`.

Exercise 3: Katz-Bonacich centrality with heterogeneous by link parameter

In this last exercise, we show how to explore the hypothesis that relations within and between parties might have a different impact on the Congress members' LES. The code below estimates Model 19 where the within and between effects are shaped by party membership.

```
R> z <- as.numeric(as.character(db_model_B[, "party"]))
R> starting <- c(alpha = 0.242486, beta_gender1 = -0.229895,
+    beta_party1 = 0.42848, beta_nchair1 = 3.0959,
+    phi_within = 0.396371, phi_between = 0.414135)
R> party_model_B <- net_dep(formula = f_model_B, data = db_model_B,
+    G = G_model_B, model = "model_B", estimation = "NLLS",
+    hypothesis = "par", z = z, start.val = starting)
R> summary(party_model_B)

Call:
Main Equation:  les ~ solve_block(I - phi_within * G_within -
phi_between * G_between) %*% (alpha * Ones + beta_gender1 * gender1 +
beta_party1 * party1 + beta_nchair1 * nchair1)
             Estimate Std. Error t value Pr(>|t|)
alpha         0.24249    0.08988   2.698 0.007251 **
beta_gender1 -0.22990    0.14120  -1.628 0.104221
beta_party1   0.42848    0.12623   3.394 0.000751 ***
beta_nchair1  3.09590    0.26008  11.904  < 2e-16 ***
phi_within    0.39637    0.07306   5.425 9.66e-08 ***
phi_between   0.41414    0.24175   1.713 0.087420 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1344.1  loglik: -665.05
```

The estimation results show that connections within one's own party and between parties are both significant for advancing a piece of legislation. The weighted Katz-Bonacich centrality can be found in the object `party_model_B$centrality`.

5.2. Centrality measure comparison

Network centrality measures adopt different criteria for ranking the importance of an agent in a network. As a result, it might be the case that a network centrality measure robustly predicts how the individual's importance in the network determine her/his outcome, while other measures of centrality do not do so as well. Therefore, one can expect that the outcome of an agent may be significantly predicted by different centrality measures, but only the measures which better explain the agents' outcome will remain significant when these are included together in the same regression model, while the others will not be distinguished from zero.

The R package **econet** allows to evaluate the explanatory power of parameter-dependent centralities relative to other centrality measures, which depend on network topology only. More specifically, **econet** considers the following measures, which are computed using the R package **igraph** (Csárdi and Nepusz 2006): indegree centrality, outdegree centrality, degree centrality, betweenness centrality, incloseness centrality, outcloseness centrality, and closeness centrality. It also reports eigenvector centrality, which is calculated using the R package **sna** (Butts 2008).[16] We then implement an augmented version of Models 14 and 16 where we add one (or more) centrality measures in the matrix of individual characteristics $X_r^\top$. By doing so we can run an horse race across different centrality measures.[17]

The arguments of the function `horse_race` are similar to the `net_dep` ones. The different ones are: i) `centralities`, an object or a vector of class '`character`' specifying the names of the centrality measure(s) to be used; ii) `directed`, a logical object which is set to `TRUE` if the network is directed, and `FALSE` otherwise; iii) `weighted`, a logical object equal to `TRUE` if links between agents have weights, and `FALSE` otherwise; and iv) `normalization`, an object of class '`character`' which can be used to normalize centrality measures before the estimations.[18]

The output of this function is also similar to `net_dep`, i.e., a list containing the results of the main estimates in the first object, a `data.frame` listing the centrality measures in the second object, and the results of the first step estimation in the third object (if `endogeneity = TRUE`). An example of how this model is implemented and results are stored is provided in the following example, where we use a linear regression model where betweenness centrality is a regressor.

```
R> starting <- c(alpha = 0.214094, beta_gender1 = -0.212706,
```

[16] Indegree centrality is the number of incoming links of one node; outdegree centrality is the number of outgoing links of one node; degree centrality is the sum of in and out degree; betweenness centrality is the number of times a node falls on the shortest path between two other nodes; incloseness centrality is the inverse of the average distance of one node from all the other nodes passing through incoming links; outcloseness centrality is the inverse of the average distance of one node from all the other nodes passing through outcoming links; closeness centrality is the sum of in and out closeness, and eigenvector centrality is proportional to the sum of the centrality of agent's neighbors (see Jackson 2010, for further details).

[17] It is worth noting that centrality measures are computed assuming that the underlying network is fixed. However, as pointed out by An (2015b), social networks are easily malleable. How robust the centrality measure is to the potential changes in the network is worth further studying.

[18] The options available are: `NULL`, no normalization; `bygraph`, divide by the number of nodes in the network minus 1 (for degree and closeness) or the number of possible links in the network (betweenness), `bycomponent`, divide by the number of nodes in agent's component minus 1 (for degree and closeness) or the number of possible links in agent's component (betweenness); `bymaxgraph`, divide by the maximum centrality value in the network; `bymaxcomponent`, divide by the maximum centrality value in agent's component.

```
+    beta_party1 = 0.478518, beta_nchair1 = 3.09234,
+    beta_betweenness = 7.06287e-05, phi = 0.344787)
R> horse_model_B <- horse_race(formula = f_model_B,
+    centralities = "betweenness", directed = TRUE, weighted = TRUE,
+    normalization = NULL, data = db_model_B, G = G_model_B,
+    model = "model_B", estimation = "NLLS", start.val = starting)
R> summary(horse_model_B, centrality = "betweenness")

Call:
Main Equation:  les ~ gender + party + nchair + betweenness
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.300e-01  9.042e-02   3.650 0.000295 ***
gender1     -8.904e-02  1.445e-01  -0.616 0.538011
party1       7.054e-01  1.139e-01   6.194 1.36e-09 ***
nchair1      3.202e+00  2.644e-01  12.112  < 2e-16 ***
betweenness  1.851e-04  3.985e-05   4.645 4.51e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1362.19  loglik: -675.09

> summary(horse_model_B)

Call:
Main Equation: les ~ solve_block(I - phi * G) %*% (alpha * Ones +
beta_gender1 * gender1 + beta_party1 * party1 +
beta_nchair1 * nchair1 + beta_betweenness * betweenness)
                   Estimate Std. Error t value Pr(>|t|)
alpha              2.141e-01  7.676e-02   2.789  0.00552 **
beta_gender1      -2.127e-01  1.408e-01  -1.511  0.13162
beta_party1        4.785e-01  1.105e-01   4.331 1.85e-05 ***
beta_nchair1       3.092e+00  2.593e-01  11.927  < 2e-16 ***
beta_betweenness   7.063e-05  4.561e-05   1.549  0.12219
phi                3.448e-01  7.101e-02   4.856 1.68e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

AIC: 1341.61  loglik: -663.8
```

The results of this estimation show evidence that being able to broker connections in the network, as measured by betweenness centrality in the object `summary(horse_model_B)` is associated with higher legislative effectiveness. The effect disappears when we include network effects, that is when we add betweenness centrality as an additional regressor in the linear model of social interactions (Equation 16). This suggests that Katz-Bonacich centrality is a more robust predictor of effectiveness in this context. Estimates are stored in the object `horse_estimate_model_B`, whereas centrality measures in `horse_model_B$centrality`.

6. Conclusions

We have described the key elements for the estimation of parameter-dependent centralities derived from equilibrium models of behavior, and discussed the use of the package **econet** for the implementation of such metrics. The methods described in the paper are derived from several modifications to the linear-in-means model – for which both nonlinear least squares and maximum likelihood estimators are provided – and they allow one to model both link and node heterogeneity in network effects, endogenous network formation and the presence of unconnected nodes. Furthermore, they provide the means to compare the explanatory power of parameter-dependent network centrality measures with those of standard measures of network centrality. A number of examples are used to walk the reader through the discussion and orientate the application of these methods to new potential directions of research.

References

Akerlof GA (1980). "A Theory of Social Custom of Which Unemployment May Be One Consequence." *Quarterly Journal of Economics*, **94**(4), 749–775. doi:10.2307/1885667.

Akerlof GA (1997). "Social Distance and Social Decisions." *Econometrica*, **65**(5), 1005–1027. doi:10.2307/2171877.

An W (2011). "Models and Methods to Identify Peer Effects." In J Scott, PJ Carrington (eds.), *The Sage Handbook of Social Network Analysis*, chapter 34, pp. 514–532. The Sage Publications, London.

An W (2015a). "Instrumental Variables Estimates of Peer Effects in Social Networks." *Social Science Research*, **50**, 382–394. doi:10.1016/j.ssresearch.2014.08.011.

An W (2015b). "Multilevel Meta Network Analysis with Application to Studying Network Dynamics of Network Interventions." *Social Networks*, **43**, 48–56. doi:10.1016/j.socnet.2015.03.006.

An W, Liu YH (2016). "**keyplayer**: An R package for Locating Key Players in Social Networks." *The R Journal*, **8**(1), 257–268. doi:10.32614/rj-2016-018.

Anselin L (1990). "Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis." *Journal of Regional Science*, **30**(2), 185–207. doi:10.1111/j.1467-9787.1990.tb00092.x.

Apicella CL, Marlowe FW, Fowler JH, Christakis NA (2012). "Social Networks and Cooperation in Hunter-Gatherers." *Nature*, **481**(7382), 497. doi:10.1038/nature10736.

Attanasio O, Barr A, Cardenas JC, Genicot G, Meghir C (2012). "Risk Pooling, Risk Preferences, and Social Networks." *American Economic Journal: Applied Economics*, **4**(2), 134–167. doi:10.1257/app.4.2.134.

Ballester C, Calvó-Armengol A, Zenou Y (2006). "Who's Who in Networks. Wanted: The Key Player." *Econometrica*, **74**(5), 1403–1417. doi:10.1111/j.1468-0262.2006.00709.x.

Batagelj V, Mrvar A (2003). "Pajek – Analysis and Visualization of Large Networks." In M Jönger, P Mutzel (eds.), *Graph Drawing Software*, chapter 3, pp. 77–103. Springer-Verlag, Berlin.

Battaglini M, Leone Sciabolazza V, Patacchini E (2020). "Effectiveness of Connected Legislators." *American Journal of Political Science*, **64**(4), 739–756. doi:10.1111/ajps.12518.

Battaglini M, Leone Sciabolazza V, Patacchini E, Peng S (2021). ***econet**: Estimation of Parameter-Dependent Network Centrality Measures*. R package version 0.1.94, URL https://CRAN.R-project.org/package=econet.

Battaglini M, Patacchini E (2018). "Influencing Connected Legislators." *Journal of Political Economy*, **126**(6), 2277–2322. doi:10.1086/700074.

Bernheim BD (1994). "A Theory of Conformity." *Journal of Political Economy*, **102**, 841–877.

Bivand R (2022). ***spdep**: Spatial Dependence: Weighting Schemes, Statistics*. R package version 1.2-4, URL https://CRAN.R-project.org/package=spdep.

Bolker B, R Core Team (2021). ***bbmle**: Tools for General Maximum Likelihood Estimation*. R package version 1.0-24, URL https://CRAN.R-project.org/package=bbmle.

Bonacich P (1972). "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology*, **2**(1), 113–120. doi:10.1080/0022250x.1972.9989806.

Bonacich P (1987). "Power and Centrality: A Family of Measures." *American Journal of Sociology*, **92**(5), 1170–1182. doi:10.1086/228631.

Box MJ, Davies D, Swann WH (1969). *Non-Linear Optimisation Techniques*. Oliver & Boyd.

Butts CT (2008). "Social Network Analysis with **sna**." *Journal of Statistical Software*, **24**(6), 1–51. doi:10.18637/jss.v024.i06.

Calvó-Armengol A, Patacchini E, Zenou Y (2009). "Peer Effects and Social Networks in Education." *Review of Economic Studies*, **76**(4), 1239–1267. doi:10.1111/j.1467-937x.2009.00550.x.

Csárdi G, Nepusz T (2006). "The **igraph** Software Package for Complex Network Research." *InterJournal*, **Complex Systems**(1695), 1–9.

Elzhov TV, Mullen KM, Spiess AN, Bolker B (2022). ***minpack.lm**: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in **MINPACK**, Plus Support for Bounds*. R package version 1.2-2, URL https://CRAN.R-project.org/package=minpack.lm.

Fafchamps M, Gubert F (2007). "The Formation of Risk Sharing Networks." *Journal of Development Economics*, **82**(7), 326–350. doi:10.1016/j.jdeveco.2006.05.005.

Fafchamps M, Leij MJ, Goyal S (2010). "Matching and Network Effects." *Journal of the European Economic Association*, **8**(1), 203–231. doi:10.1111/j.1542-4774.2010.tb00500.x.

Fershtman C, Weiss Y (1998). "Social Rewards, Externalities and Stable Preferences." *Journal of Public Economics*, **70**(1), 53–73. doi:10.1016/s0047-2727(98)00060-7.

Goldsmith-Pinkham P, Imbens GW (2013). "Social Networks and the Identification of Peer Effects." *Journal of Business & Economic Statistics*, **31**(3), 253–264. doi:10.1080/07350015.2013.801251.

Goodreau SM, Handcock MS, Hunter DR, Butts CT, Morris M (2008). "A **statnet** Tutorial." *Journal of Statistical Software*, **24**(9), 1–26. doi:10.18637/jss.v024.i09.

Graham BS (2015). "Methods of Identification in Social Networks." *Annual Review of Economics*, **7**(1), 465–485. doi:10.1146/annurev-economics-080614-115611.

Graham BS (2016). "An Econometric Model of Network Formation with Degree Heterogeneity." *Econometrica*, **85**(4), 1033–1063. doi:10.3982/ecta12679.

Heckman J (1979). "Sample Selection Bias as a Specification Error." *Econometrica*, **47**(1), 153–161. doi:10.2307/1912352.

Hsieh C, Lin X, Patacchini E (2020). "Social Interaction Methods." In K Zimmerman (ed.), *The Handbook of Labor, Human Resources and Population*, pp. 1–30. Springer-Verlag.

Hsieh CS, Lee L (2016). "A Social Interactions Model with Endogenous Friendship Formation and Selectivity." *Journal of Applied Econometrics*, **31**(2), 301–319. doi:10.1002/jae.2426.

Jackson MO (2010). *Social and Economic Networks*. Princeton University Press.

Jackson MO, Rogers BW, Zenou Y (2017). "The Economic Consequences of Social-Network Structure." *Journal of Economic Literature*, **55**(1), 49–95. doi:10.1257/jel.20150694.

Johnnson I, Moon HR (2021). "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach." *The Review of Economics and Statistics*, **103**(2). doi:10.1162/rest_a_00870.

Kandel E, Lazear EP (1992). "Peer Pressure and Partnerships." *Journal of Political Economy*, **100**, 801–817.

Katz L (1953). "A New Status Index Derived from Sociometric Analysis." *Psychometrica*, **18**(1), 39–43. doi:10.1007/bf02289026.

Lai B, Reiter D (2017). "Democracy, Political Similarity, and International Alliances, 1816–1992." *Journal of Conflict Resolution*, **44**(2), 203–227. doi:10.1177/0022002700044002003.

Leifeld P, Cranmer SJ, Desmarais BA (2018a). ***xergm**: Extensions of Exponential Random Graph Models*. R package version 1.8.3, URL https://CRAN.R-project.org/src/contrib/Archive/xergm/.

Leifeld P, Cranmer SJ, Desmarais BA (2018b). "Temporal Exponential Random Graph Models with **btergm**: Estimation and Bootstrap Confidence Intervals." *Journal of Statistical Software*, **83**(6), 1–36. doi:10.18637/jss.v083.i06.

LeSage J, Pace RK (eds.) (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

Manski CF (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies*, **60**(3), 531–542. doi:10.2307/2298123.

Mayer A, Puller SL (2008). "The Old Boy (and Girl) Network: Social Network Formation on University Campuses." *Journal of Public Economics*, **92**(1), 329–347. doi:10.1016/j.jpubeco.2007.09.001.

Millo G, Piras G (2012). "**splm**: Spatial Panel Data Models in R." *Journal of Statistical Software*, **47**(1), 1–38. doi:10.18637/jss.v047.i01.

More JJ (1978). "The Levenberg-Marquardt Algorithm: Implementation and Theory." In GA Watson (ed.), *Lecture Notes in Mathematics 630: Numerical Analysis*, pp. 105–116. The Sage Publications.

Opsahl T (2009). *Structure and Evolution of Weighted Networks*. Ph.D. thesis, University of London (Queen Mary College), London. URL http://toreopsahl.com/publications/thesis/.

Patacchini E, Rainone E, Zenou Y (2012). "Strong versus Weak Ties in Education." *Discussion paper 8223*, CEPR.

Patacchini E, Zenou Y (2012). "Juvenile Delinquency and Conformism." *Journal of Law, Economic, and Organization*, **28**, 1–31. doi:10.1093/jleo/ewp038.

Qu X, Lee L (2015). "Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix." *Journal of Econometrics*, **184**(2), 209–232. doi:10.1016/j.jeconom.2014.08.008.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL https://www.R-project.org/.

Schweinberger M, Luna P (2018). "**hergm**: Hierarchical Exponential-Family Random Graph Models." *Journal of Statistical Software*, **85**(1), 1–39. doi:10.18637/jss.v085.i01.

StataCorp (2017). *Stata Statistical Software: Release 15*. StataCorp LLC, College Station. URL http://www.stata.com/.

The MathWorks, Inc (2011). *MATLAB – The Language of Technical Computing, Version 7.5*. The MathWorks, Inc., Natick. URL http://www.mathworks.com/products/matlab/.

Van Rossum G, *et al.* (2011). *Python Programming Language*. URL https://www.python.org/.

VanderWeele TJ, An W (2013). "Social Networks and Causal Inference." In S Morgan (ed.), *Handbook of Causal Analysis for Social Research*, chapter 17, pp. 353–374. Springer-Verlag, New York.

Wasserman S, Faust K (1994). *Social Network Analysis: Methods and Applications*. 1st edition. Cambridge University Press, Cambridge.

Zenou Y (2016). "Key Players." In Y Bramoullé, A Galeotti, B Rogers (eds.), *Oxford Handbook on the Economics of Networks*, chapter 10, pp. 244–274. Oxford University Press, Oxford.

Affiliation:

Marco Battaglini
Department of Economics
Cornell University
Ithaca, NY 14850, United States of America
and EIEF *and* NBER
E-mail: battaglini@cornell.edu

Valerio Leone Sciabolazza
Department of Economics and Law
Sapienza University of Rome
Rome, 00161, Italy
and CEIS
E-mail: valerio.leonesciabolazza@uniroma1.it

Eleonora Patacchini
Department of Economics
Cornell University
Ithaca, NY 14850, United States of America
and EIEF
E-mail: ep454@cornell.edu

Sida Peng
Office of Chief Economist
Microsoft Research
Redmond, WA 14865, United States of America
E-mail: sidpeng@microsoft.com