Interactions and Networks

Paolo Zacchia

Microeconometrics

Lecture 16

Measuring externalities

- Economic theory posits that *externalities* occur in a variety of different settings. Some types of externalities are framed as *spillover* effects (for example, knowledge/skill diffusion).

- Sociology is also traditionally interested in the consequences of social interactions, like *peer effects* at school.

- Who "spills over" on whom, to what extent, and why them? All these questions call for a suitable econometric framework.

- Empirical observation and everyday experience suggest that social interactions have a nuanced *networked* structure.

- The original interest for *spillover* and *peer effects* prompted the development of econometric models for interactions that occur in networks. In parallel (and loosely relatedly), models for the study of *network formation* have also emerged.

A set of interrelated frameworks

This Lecture overviews econometric models for studying **social effects** and **interactions**, and it proceeds as follows.

1. The traditional **“linear-in-means”** model, its related issues and some early empirical literature are reviewed first.

2. A self-contained, brief introduction to **networks** follows.

3. The standard framework for the study of **spillover** and **peer effects** using networks is presented, alongside its connection with *spatial* econometrics and related empirical studies.

4. A brief introduction to linear *penalized estimators* allows to review models that *infer* networks from **panel data**.

5. Lastly, models of **network formation** are overviewed along a discussion of related theoretical and econometric issues.

The linear-in-means model (1/5)

- In a seminal article, inspired by some literature in sociology, Manski (1993) discussed a simple model for the analysis of "social effects" (a broad, encompassing term).

- Let observations be partitioned across **groups**: $\mathcal{C}(i)$ denotes the set of all group mates of observation i (excluding i), for $i = 1, \ldots, N$. Let $|\mathcal{C}(i)|$ be the number of i's group mates.

- This group structure is **transitive**: if $j \in \mathcal{C}(i)$ and $k \in \mathcal{C}(j)$, it holds that $k \in \mathcal{C}(i)$ too, $i \in \mathcal{C}(k)$, *et cetera*.

- Let there be a **dependent variable** Y_i which can possibly depend on the effect of social interactions (e.g. consumption choices, grades at school, technology adoption, *et cetera*).

- Let there also be some **independent variable** X_i capturing some predetermined individual characteristic.

The linear-in-means model (2/5)

- The typical **linear-in-means** model is as follows.

$$Y_i = \alpha + \beta \frac{1}{|\mathcal{C}(i)|} \sum_{j \in \mathcal{C}(i)} Y_j + \gamma X_i + \delta \frac{1}{|\mathcal{C}(i)|} \sum_{j \in \mathcal{C}(i)} X_j + \varepsilon_i$$

- The **group means** of both Y_i and X_i, for every observation, enter linearly into the model (on the right-hand side).

- Here, parameter β measures the **endogenous effect** on Y_i that is due to "social," "peer" or other "spillover" effects. It captures co-movement of Y_i in the group.

- Instead, parameter δ measures the **exogenous effect**, also called **contextual effect**, that captures the impact upon Y_i of a group's average characteristic X_i.

- The model easily extends to multiple independent variables X_i (with exogenous effects that are also multidimensional).

The linear-in-means model (3/5)

- What is the theoretical rationale for the endogenous effect? It is typically postulated that β captures social mechanisms that lead to correlated behavior.

- An example is a model where individuals *choose* Y_i while it causes externalities, and individual **utility** is **quadratic**.

$$U_i\left(Y_1, \dots, Y_N; X_1, \dots, X_N; \varepsilon_i\right) = \beta \frac{1}{\left|\mathcal{C}\left(i\right)\right|} \sum_{j \in \mathcal{C}(i)} Y_j Y_i + \\ + \left(\alpha + \gamma X_i + \delta \frac{1}{\left|\mathcal{C}\left(i\right)\right|} \sum_{j \in \mathcal{C}(i)} X_j + \varepsilon_i\right) Y_i - \frac{1}{2} Y_i^2$$

  The Nash equilibrium delivers the linear-in-means model.

- Different specifications of utility, e.g. one where individuals have a *preference for conformity* in the choice of Y_i (Blume, Brock, Durlauf and Jayaraman; 2015) yield a similar model.

The linear-in-means model (4/5)

- A key issue of this model, as observed by Manski, is that it is **not identified** in many relevant settings.

- In particular, while both parameters β and δ are interesting in their own right, often they cannot be disentangled.

- Specifically, expectations of Y_i *that condition on each group's* X_j, for $j \in \mathcal{C}(i)$, cannot be exploited for identification as it is typical in linear models. With a liberal use of notation:

$$\mathbb{E}[Y_i | \mathcal{C}(i)] = \left(\frac{\alpha}{1-\beta}\right) + \left(\frac{\gamma+\delta}{1-\beta}\right) \mathbb{E}[X_j | \mathcal{C}(i)]$$

  thus, the parameters $(\alpha, \beta, \gamma, \delta)$ are not separately identified.

- This was called **reflection problem** by Manski. Intuitively, the exogenous effect of X_j onto Y_i is "reflected" back on Y_j via the endogenous effect: the two are indistinguishable.

The linear-in-means model (5/5)

- Some *non-linear* specifications of social effects do not suffer from the reflection problem (Brock and Durlauf, 2001, 2007), though others do (Manski, 2003). Non-linearities might pose additional problems of specification choice.

- In his paper, Manski also analyzed **correlated effects**: that is, correlation structures ε_i of the error term *within groups*.

- Group-level **"common shocks"** are a special case of those (they can be thought as group-level "fixed effects" $\alpha_{\mathcal{C}(i)}$).

- If correlated effects do exist *and* ε_i also correlates with X_j for $j \in \mathcal{C}(i)$, they lead to a particular omitted variable bias.

- Intuitively, both social effect types cannot be distinguished from unobserved characteristics (background, environment) that are common to or shared by all observation in a group.

Early empirical results on social effects (1/2)

- These identification challenges related to the linear-in-means model did not prevent the gradual emergence of an empirical literature about "spillover" and "peer" effects.

- An early notable example is the paper by Sacerdote (2001) who exploits a natural **experiment**: the random assignment of roommates at Dartmouth University.

- Sacerdote finds that a roommate's "ability" improves one's own grade point average (GPA), arguably due to peer effects.

- In this setting, groups are always bidimensional *pairs* of two roommates; Y_i is the GPA, and X_i is an ability index.

- Sacerdote's model is essentially a linear-in-means one where the endogenous effect is restricted ($\beta = 0$) and the parameter of interest is that for the contextual effect: δ.

Early empirical results on social effects (2/2)

- Other early empirical studies adopted approaches similar to Sacerdote's; most focused on peer effects in the classroom.

- Hoxby (2000) estimates a model without endogenous effects where the key X_i variables are gender and race, arguing that between-cohort demographic composition is exogenous.

- Hanushek, Kain, Markman and Rivkin (2003) exploit *lagged values* of Y_i to attempt capturing the endogenous effect.

- Carrel, Fullerton and West (2009) utilize random assignment to "squadrons" in the U.S. Air Force academy to estimate a model similar to Sacerdote's, but with larger peer groups.

- Even Ammermueller and Pischke (2009) estimated a model without endogenous effect; they used within-school variation in primary school class composition.

Towards the identification of social effects

- Eventually, econometricians and empiricists found avenues for the identification of both social effect types.
- Following different approaches, Moffitt (2001), Lee (2007) as well as Graham (2008) noted that **higher-order moments** enable identification of key social effects.
- More famously, Bramoullé, Djebbari and Fortin (2009; BDF) proved that identification holds if social interactions occur in a **non-transitive** structure, e.g. in social **networks**.
- The identification result by BDF, which is reviewed later, is actually more general: it nests the heterogeneous-group-size argument by Lee, and it also allows for common shocks.
- The paper by Blume, Brock, Durlauf and Jayaraman (2015) concludes that reflection is more an exception than a rule.

Higher-order identifying moments (1/5)

- The identification of the endogenous effect via higher-order moments can be illustrated as follows.

- Write the model in compact matrix notation:

$$\mathbf{y} = \alpha\boldsymbol{\iota} + \beta\mathbf{C}\mathbf{y} + \gamma\mathbf{x} + \delta\mathbf{C}\mathbf{x} + \boldsymbol{\varepsilon}$$

  where $\mathbf{C}$ is an $N \times N$ matrix with $c_{ii} = 0$ for $i = 1, \dots, N$ and $c_{ij} = \left(\left|\mathcal{C}\left(i\right)\right|\right)^{-1} \cdot \mathbb{1}\left[j \in \mathcal{C}\left(i\right)\right]$ for all (i, j) pairs.

- Assume *homoscedasticity*: $\mathbb{V}\text{ar}\left[\boldsymbol{\varepsilon}|\,\boldsymbol{x}\right] = \sigma^2\mathbf{I}$. This implies the following **covariance restriction**:

$$\mathbb{V}\text{ar}\left[\boldsymbol{y}|\,\boldsymbol{x}\right] = \sigma^2\left(\mathbf{I} - \beta\mathbf{C}\right)^{-2}$$

  which is enough to identify β if σ^2 is known/estimated.

- Note that $|\beta| < 1$ for the conditional variance to be bounded.

Higher-order identifying moments (2/5)

- Intuition can be gained through the typical development of the Leontiev inverse expansion:

$$(\mathbf{I} - \beta\mathbf{C})^{-1} = \sum_{r=0}^{\infty} \beta^r \mathbf{C}^r$$

  which shows that the covariance between two outcomes Y_i and Y_j of two observations i and j from the same group is *enhanced* (as one would expect) by the endogenous effect.

- This result obtains under strong restrictions upon $\mathbb{V}\text{ar}\left[\boldsymbol{\varepsilon}|\,\boldsymbol{x}\right]$, yet Lee (2007) developed a GMM estimation procedure that applies under more general conditions.

- Crucially, Lee's approach works if *there is variation in group size*. This is best reviewed later as a particular case of BDF.

- This result was later generalized further by Rose (2017).

Higher-order identifying moments (3/5)

In a celebrated paper, Graham (2008) adopts a related approach. He considers a model without exogenous variables X_i:

$$Y_{ci} = \alpha_c + (\gamma_0 - 1)\,\bar{\varepsilon}_c + \varepsilon_{ci}$$

where $c = 1, \ldots, C$ indexes *classes* (e.g. in a school), and:

- α_c is a class-specific random effect (due e.g. to teachers);
- ε_{ci} is a generally unobserved random variable subsuming all individual heterogeneity (ability, family background, etc.);
- $\bar{\varepsilon}_c \equiv M_c^{-1}\left(\sum_{j\in\mathcal{C}(i)} \varepsilon_{cj} + \varepsilon_{ci}\right)$ is the average value of ε_{ci} in class c, while $M_c = |\mathcal{C}(i)| + 1$ is the *size* of class c.

The parameter of interest is the so-called **social multiplier** γ_0: with peer effects $\gamma_0 > 1$. It relates to Manski's endogenous effect.

Higher-order identifying moments (4/5)

- To identify γ_0, Graham exploits variation in class **type** W_c. His result, while more general, is best illustrated by a binary distinction: $W_c = 1$ if c is *large* and $W_c = 0$ if c is *small*.

- The probability distributions of both α_c and ε_{ci} depend on W_c. Hence, $\mathbb{V}\text{ar}\left(\alpha_c, \varepsilon_{c1}, \dots, \varepsilon_{cM_c} \middle| M_c, W_c\right)$ and consequently $\mathbb{V}\text{ar}\left(Y_{c1}, \dots, Y_{cM_c} \middle| M_c, W_c\right)$ conditionally depend on W_c. Let these be the "primitive" moments of this model.

- By developing these moments, Graham notes that it is more convenient to work with the *within* and *between* components of the conditional variance of Y_{ci}, which he defines as follows.

$$G_c^w = \frac{1}{M_c} \frac{1}{M_c - 1} \sum_{i=1}^{M_c} \left(Y_{ci} - \bar{Y}_c\right)^2$$

$$G_c^b = \left(\bar{Y}_c - \mathbb{E}\left[Y_{ci} \middle| W_c\right]\right)^2$$

Higher-order identifying moments (5/5)

- By studying the expectations of G_c^w and G_c^b conditional on W_c, Graham develops a **Wald estimator** based on:

$$\frac{\mathbb{E}\left[G_c^b \middle| W_c = 1\right] - \mathbb{E}\left[G_c^b \middle| W_c = 0\right]}{\mathbb{E}\left[G_c^w \middle| W_c = 1\right] - \mathbb{E}\left[G_c^w \middle| W_c = 0\right]} =$$

$$= \gamma_0^2 + \frac{\tau_0^2(1) - \tau_0^2(0)}{\mathbb{E}\left[G_c^w \middle| W_c = 1\right] - \mathbb{E}\left[G_c^w \middle| W_c = 0\right]}$$

where $\tau_0(W_c)$ is a function of the primitive moments.

- This Wald estimator identifies γ_0 if $\tau_0^2(1) = \tau_0^2(0)$. Graham develops assumptions to sustain this result: the central one is that *both* students and teachers are **randomly assigned** to classes (as in the dataset he uses in his application).

- Intuitively, the **between** variance of Y_{ci} is amplified in large classes through peer effects, and *possibly* also by unobserved heterogeneity across classes – as summarized by $\tau_0(W_c)$.

Brief introduction to networks (1/5)

- To best illustrate the network approach to the identification of social effects, it is useful to first introduce networks briefly.

- A **network** is defined by a pair $(\mathcal{I}, \mathcal{G})$.

- Here, $\mathcal{I}$ is a set of **nodes**, which has dimension $|\mathcal{I}| = N$.

- Instead, $\mathcal{G} = \{g_{ij} : (i, j) \in \mathcal{I}^2\}$ is a set of **edges**, representing pairwise connections between nodes indexed by i and j.

- Unless *self-links* are allowed, $g_{ii} = 0$ for $i = 1, \dots, N$.

- Edges can be arrayed in an $N \times N$ **adjacency matrix G**.

$$\mathbf{G} = \begin{bmatrix} 0 & g_{12} & \dots & g_{1N} \\ g_{21} & 0 & \dots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N1} & g_{N2} & \dots & 0 \end{bmatrix}$$

Brief introduction to networks (2/5)

- Edges are typically **normalized**: $g_{ij} \in [0,1]$ for all $g_{ij} \in \mathcal{G}$.
- A network is **undirected** if $g_{ij} = g_{ji}$ for all $(i,j) \in \mathcal{I}^2$; it is otherwise **directed**.
- A network is **unweighted** if $g_{ij} \in \{0,1\}$ for all $g_{ij} \in \mathcal{G}$; it is otherwise **weighted**.
- A network is **bipartite** if $\mathcal{I}$ can be partitioned in two parts: $\mathcal{I}_1$ and $\mathcal{I}_2$, such that $\mathcal{I}_1 \cup \mathcal{I}_2 = \mathcal{I}$ and $g_{ij} = 0$ if $(i,j) \in \mathcal{I}_1$ or if $(i,j) \in \mathcal{I}_2$ (edges only occur across the two "parts").
- A **path** is a sequence of *nonzero* edges $(g_{ij}, g_{jk}, \ldots, g_{\ell m}, g_{mn})$ that connects two nodes $(i,n) \in \mathcal{I}^2$.
- The dimension of a path is its **length**. The **shortest path length** between two nodes $(i,n) \in \mathcal{I}^2$ is self-explanatory.

Brief introduction to networks (3/5)

- In directed networks, the **in-degree** $d_i^i = \sum_{h=1}^{N} \mathbb{1}\left[g_{hi} > 0\right]$ is the total number of edges directed towards node $i \in \mathcal{I}$, while the **out-degree** $d_i^o = \sum_{j=1}^{N} \mathbb{1}\left[g_{ij} > 0\right]$ is the total number of edges departing from it.

- In undirected networks, a node i's in-degree and out-degree coincide, and they are more simply called **degree** d_i.

- In weighted networks, the two quantities $s_i^i = \sum_{h=1}^{N} g_{hi}$ and $s_i^o = \sum_{j=1}^{N} g_{ij}$ are called **in-strength** and **out-strength** (or simply **strength** s_i if the network is undirected).

- These quantities match sums over columns or rows of $\mathbf{G}$.

- In an unweighted network, the in-strength, out-strength and strength of a node equal the in-degree, out-degree and degree (respectively).

Brief introduction to networks (4/5)

- The adjacency matrix $\mathbf{G}$ is **row-normalized** if $\mathbf{G}\boldsymbol{\iota} = \boldsymbol{\iota}$ (that is, all its rows sum up to one; $s_i^o = 1$ for all $i \in \mathcal{I}$).

- In unweighted networks, the entries of the r-th **power** $\mathbf{G}^r$ of the adjacency matrix – for $r \in \mathbb{N}$ – count the number of paths of length r that connect each pair of nodes.

- In an undirected network, a **triad** $(i, j, k) \in \mathcal{I}^3$ is **closed** if $g_{ij} g_{jk} g_{ki} > 0$ (all nodes are connected), **open** otherwise.

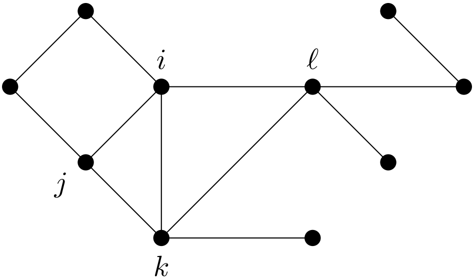
- Multiple measures of **centrality** – more "recursive" versions of degree and strength – exist. For example, the sequence of *Katz-Bonacich* centralities $\mathbf{k} = (k_1, \dots, k_N)$ is as follows:

$$\mathbf{k} = (\mathbf{I} - a\mathbf{G})^{-1} \boldsymbol{\iota} = \sum_{r=0}^{\infty} a^r \mathbf{G}^r \boldsymbol{\iota}$$

  given some "attenuation factor" $a \in [0, 1]$.

Brief introduction to networks (5/5)

Here is a graphical illustration of a simple, stylized network. In it, nodes are represented by circles and edges by lines.



- This network is undirected and weights are not reported.
- Yet the network displays substantial degree heterogeneity.
- Nodes i, j, k and ℓ appear to be especially "central."
- Moreover, two closed triads are found among them.

Social effects in networks (1/8)

- The central result by BDF was to show that if spillover, peer and more generally social effects occur in a non-trivial social **network**, all of Manski's effects are identified.

- Their starting point is the model

$$Y_i = \alpha + \beta \sum_{j=1}^{N} g_{ij} Y_j + \gamma X_i + \delta \sum_{j=1}^{N} g_{ij} X_j + \varepsilon_i$$

  where g_{ij} is a **network edge** between observations i and j. It is assumed that $\mathbb{E}\left[\varepsilon_i \middle| X_1, \dots, X_N; \mathcal{G}\right] = 0$.

- In compact matrix notation, the model writes as follows.

$$\mathbf{y} = \alpha \iota + \beta \mathbf{G} \mathbf{y} + \gamma \mathbf{x} + \delta \mathbf{G} \mathbf{x} + \boldsymbol{\varepsilon}$$

- BDF assume $\mathbf{G}$ to be row-normalized, but their results hold more generally. This model nests the linear-in-means one.

Social effects in networks (2/8)

- Their main result (Proposition 1) has shown that $(\alpha, \beta, \gamma, \delta)$ are globally point-identified if matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$ are **linearly independent** *and* $\gamma\beta + \delta \neq 0$.

- This requires an **intransitive network**: at least some triads must be open (that is, non-closed)! This rules out the group structure of the linear-in-means model with equal group size.

- To gather some intuition, let matrix $(\mathbf{I} - \beta\mathbf{G})$ be invertible (which requires $|\beta| < 1$). Thus, the *reduced form* of $\mathbf{y}$ is:

$$\begin{aligned}\mathbf{y} &= (\mathbf{I} - \beta\mathbf{G})^{-1} (\alpha\boldsymbol{\iota} + \gamma\mathbf{x} + \delta\mathbf{G}\mathbf{x} + \boldsymbol{\varepsilon}) \\ &= \frac{\alpha}{1-\beta}\boldsymbol{\iota} + \gamma\mathbf{x} + (\gamma\beta + \delta)\sum_{r=0}^{\infty} \beta^r \mathbf{G}^{r+1}\mathbf{x} + \sum_{r=0}^{\infty} \beta^r \mathbf{G}^r \boldsymbol{\varepsilon}\end{aligned}$$

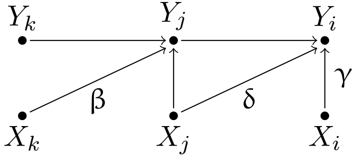
which implies that all **instruments** of the form $\mathbf{G}^{r+1}\mathbf{x}$ for $r \in \mathbb{N}_0$ can be leveraged for identification.

Social effects in networks (3/8)

- To substantiate, consider any three nodes (i, j, k) in an open triad of an undirected, unweighted network.

- Nodes i and j are linked ("friends") to one another, and so are j and k ($g_{ij} = g_{jk} = 1$), but i and k are not ($g_{ik} = 0$).

- Social effects can be expressed through the following graph.



- Parameter γ is identified off variation in one's *own* X_i; the exogenous effect δ is identified off variation in *friends'* X_j; the endogenous effect is identified off variation in the X_k of *friends of friends* (*if* $\gamma\beta + \delta \neq 0$: the effects do not offset one another). Triad openness yields **exclusion restrictions**.

Social effects in networks (4/8)

- The second result by BDF (Proposition 2) shows that even if $\mathbf{G}$ has a *group structure* – like $\mathbf{C}$ previously – social effects are identified if groups have **different sizes** *and* $\gamma\beta + \delta \neq 0$.

- This results nests Lee's (2007) and is a special case of BDF's Proposition 1. With a group structure, the reduced form is:

$$Y_i = \frac{\alpha}{1-\beta} + \left[\gamma + \frac{\beta\left(\gamma\beta+\delta\right)}{\left(1-\beta\right)\left(M_i - 1 + \beta\right)}\right] X_i + \\ + \frac{\gamma\beta+\delta}{\left(1-\beta\right)\left(1+\dfrac{\beta}{M_i - 1}\right)} \bar{X}_i + \nu_i$$

  where M_i is the size of i's group, $\bar{X}_i$ is the average of X_i in i's group, and ν_i is a composite error term.

- Variation in M_i leads, in turn, to variation in reduced form coefficients across groups, as explained elaborately by BDF.

Social effects in networks (5/8)

- The third result by BDF (Proposition 3) is about **directed** networks: for identification it suffices that $\mathbf{G}^2 \neq \mathbf{0}$ *or* $\alpha \neq 0$.

- Their last results (Propositions 4-5) cover **common shocks**. Suppose that observations are split across C separate groups or networks, and the model reads:

$$Y_{ci} = \alpha_c + \beta \sum_{j=1}^{N} g_{ij} Y_{cj} + \gamma X_{ci} + \delta \sum_{j=1}^{N} g_{ij} X_{cj} + \varepsilon_{ci}$$

  where $c = 1, \ldots, C$, $c(i)$ identifies the group or network of observation i, $g_{ij} = 0$ if $c(i) \neq c(j)$, while α_c is a random shock shared in group/network c.

- Identification requires that the four matrices $\mathbf{I}$, $\mathbf{G}$, $\mathbf{G}^2$, $\mathbf{G}^3$ are all **linearly independent**. This is best understood via a representation of the model in "local differences."

Social effects in networks (6/8)

- Taking "local differences" here amounts to pre-multiplying the model (in compact matrix notation) by $\mathbf{I} - \mathbf{G}$:

$$(\mathbf{I} - \mathbf{G})\,\mathbf{y} = \beta\,(\mathbf{I} - \mathbf{G})\,\mathbf{G}\mathbf{y} + (\gamma\mathbf{I} + \delta\mathbf{G})\,(\mathbf{I} - \mathbf{G})\,\mathbf{x} + (\mathbf{I} - \mathbf{G})\,\boldsymbol{\varepsilon}$$

  and since $\mathbf{G}$ is row-normalized, common shocks vanish.

- An examination of the reduced form clarifies why in this case linear independence from $\mathbf{G}^3$ is required (Proposition 4).

- BDF develop an analogous result also for "global" differences (pre-multiplication by $\mathbf{I} - \iota\iota^{\mathrm{T}}/N$; Proposition 5).

- It may appear at a first sight that these conditions are quite demanding, but they apply to a vast set of networks.

- Yet this analysis does not address correlated effects that are not "common" across well-identified subgroups.

Social effects in networks (7/8)

- The analysis of identification easily extends to a model with multiple independent variables:

$$\mathbf{y} = \boldsymbol{\alpha} + \beta \mathbf{G} \mathbf{y} + \mathbf{X} \boldsymbol{\gamma} + \mathbf{G} \mathbf{X} \boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

  where $\mathbf{X}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are multidimensional and $\boldsymbol{\alpha}$ is a vector of common shocks. Consider a locally differenced version of it.

- Estimation is based on Kelejian and Prucha (2008) and Lee (2003). Let here $\boldsymbol{\theta} = (\beta, \boldsymbol{\gamma}, \boldsymbol{\delta})$.

- First, a 2SLS estimator $\widehat{\boldsymbol{\theta}}_{2SLS}$ is calculated, with $(\mathbf{I} - \mathbf{G})\,\mathbf{X}$, $(\mathbf{I} - \mathbf{G})\,\mathbf{G}\mathbf{X}$ and $(\mathbf{I} - \mathbf{G})\,\mathbf{G}^2\mathbf{X}$ as instruments.

- Then, a *more efficient* 2SLS estimator $\widehat{\boldsymbol{\theta}}_{LEE}$ obtains using

$$\mathbf{G}\left(\mathbf{I} - \widehat{\beta}_{2SLS}\mathbf{G}\right)^{-1}\left[(\mathbf{I} - \mathbf{G})\left(\mathbf{X}\widehat{\boldsymbol{\gamma}}_{2SLS} + \mathbf{G}\mathbf{X}\widehat{\boldsymbol{\delta}}_{2SLS}\right)\right]$$

  as well as $(\mathbf{I} - \mathbf{G})\,\mathbf{X}$ and $(\mathbf{I} - \mathbf{G})\,\mathbf{G}\mathbf{X}$ as instruments.

Social effects in networks (8/8)

- BDF showcase their approach in Monte Carlo experiments as well as in an empirical application based on the *Add Health* data (a sample of U.S. schools collected between 1994 and 1995, with information about students' friendships). In their analysis, Y_i is the consumption of "recreational activities."

- The paper by BDF has been influential beyond expectations, and motivated novel empirical applications. While it allowed to overcome Manski's reflection problem, the BDF approach is however problematic in three main respects.

- First, it requires accurate data on the network(s) $\mathbf{G}$, which are typically very costly to acquire. Second, it assumes that the network (specifically, its edge set $\mathcal{G}$) is *exogenous*. Third, it restricts correlated effects to common shocks only.

- Research is currently active along all these three dimensions.

A connection with spatial econometrics

The extended framework developed by BDF is tightly connected to the "General Nesting Model" in **spatial econometrics**:

$$\mathbf{y} = \alpha\boldsymbol{\iota} + \beta\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{W}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} = \rho\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\upsilon}$$

where $\mathbf{W}$ is a zero-diagonal matrix of "spatial weights" (often inverse functions of geographical distance, e.g. "distance decay") and the error terms $\boldsymbol{\varepsilon}$ follow a spatially autoregressive structure, with $\boldsymbol{\upsilon}$ typically being a vector of i.i.d. shocks.

- Restrictions on this model take various names. For example, $\boldsymbol{\delta} = \mathbf{0}$ and $\rho = 0$ yield a "Spatially Autoregressive" model.
- The spatial econometrics literature has generally focused on MLE and GMM approaches to estimate versions of the GNS model, typically assuming that $\mathbf{W}$ and $\mathbf{X}$ are exogenous.

Later empirical results on social effects (1/2)

- It is worth to briefly review some empirical analyses about social effects in networks that followed BDF.

- The first actual application of a "friends of friends" approach was developed by De Giorgi, Pellizzari and Redaelli (2010; DGPR) independently of BDF.

- They studied peer effects on the choice of major (economics versus management: a *binary* Y_i) at Bocconi University.

- At Bocconi, future students of both majors used to take some "common courses" in their first year and a half of education. Students are allocated **randomly** into common courses.

- DGPR consider two students i and j as "friends" ($g_{ij} > 0$) if they attended at least four common courses together. This delivers an exogenous *and intransitive* network.

Later empirical results on social effects (2/2)

- Patnam (2013) studies the effect of firm networks shaped by board interlocks on firm policy decisions in India.

- More recently, De Giorgi, Frederiksen and Pistaferri (2020) studied peer effects in *household consumption* using Danish administrative data. They allow for different exogenous and endogenous effects that depend on either the husband's or the wife's *coworker network*; this helps break transitivity.

- Arduini, Patacchini and Rainone (2020) study a model *à la* BDF where $\boldsymbol{x}_i$ features an external *treatment*, and where the exogenous and endogenous effects are *heterogeneous*. Using this model they evaluate a Mexican program against poverty, finding that 50% of its effect is due to externalities.

- Others adopt different approaches; Bramoullé, Djebbari and Fortin (2020) provide a recent extended survey.

Towards more general models of network effects

The research aimed at solving the three main issues of the BDF framework can be summarized as follows.

1. While network data are costly to collect, $\mathbf{G}$ may be *inferred* from panel data. The intuition is that network effects reveal themselves as cross-correlation in $\boldsymbol{y}$ over time. Extant efforts (including Manresa, 2016; de Paula, Rasul and Souza, 2020; Rose, 2020) are based on *linear penalized estimators*.

2. Attempts to address endogeneity of the edge set $\mathcal{G}$ include control function methods (Arduini, Patacchini and Rainone, 2015; Johnsson and Moon, 2019) as well as GMM for panel data (Kuersteiner and Prucha, 2020).

3. Methods suited to more general correlated effects structures (in different settings) have been proposed by Zacchia (2020) and Pereda-Fernández and Zacchia (2021).

Linear penalized estimators: a summary (1/3)

A brief summary of linear penalized estimators (developed next) later helps overview methods for inferring unknown networks $\mathbf{G}$.

Consider a linear model $y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i$. An **elastic net** estimator of $\boldsymbol{\beta}$ is defined as follows:

$$\widehat{\boldsymbol{\beta}}_{EN} = \underset{\boldsymbol{\beta} \in \mathbb{R}^K}{\arg\min} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \left(\kappa \|\boldsymbol{\beta}\|_1 + (1-\kappa) \|\boldsymbol{\beta}\|_2 \right)$$

where the OLS objective function $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2$ is "penalized" by a term increasing in $\boldsymbol{\beta}$ governed by the **penalization parameter** $\lambda \geq 0$. Here, $\kappa \in [0, 1]$ controls the "type" of elastic net estimator.

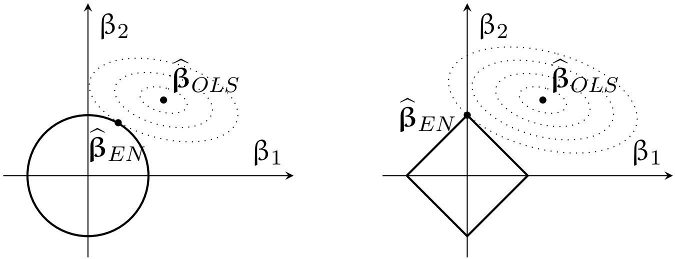
At the extremes:

- if $\kappa = 0$, $\widehat{\boldsymbol{\beta}}_{EN}$ is a **ridge regression** estimator;
- if $\kappa = 1$, $\widehat{\boldsymbol{\beta}}_{EN}$ is instead a **Least Absolute Shrinkage and Selection Operator** (LASSO) estimator.

Linear penalized estimators: a summary (2/3)

- Both the values of λ and κ are chosen by the researcher.

- The penalty is a constraint to the least squares minimization problem. This is visualized in the figure below, for a stylized case where $|\boldsymbol{\beta}| = 2$. The continuous lines are the constraints for ridge regression (left panel) and LASSO (right panel).



- By construction, these estimators "shrink" OLS estimates of $\boldsymbol{\beta}$ *towards* zero, and increasingly so the higher is λ.

Linear penalized estimators: a summary (3/3)

- In ridge regression, the penalty is based on the L^2 norm for $\boldsymbol{\beta}$: this shrinks all coefficients uniformly.

- With the LASSO, the penalty is based on the L^1 norm for $\boldsymbol{\beta}$: most coefficients are shrunk *at* zero; only few are "selected."

- Internals values of $\kappa \in (0, 1)$ lead to intermediate behavior.

- By construction, all elastic net estimators are **inconsistent**. However, the coefficients selected by the LASSO can be used in "post-LASSO" consistent OLS estimation.

- These estimators are especially popular in machine learning settings where $K >> N$ and good prediction matters. The LASSO is especially suited to a so-called **sparsity** condition.

$$\sum_{k=1}^{K} \mathbb{1}\left[\beta_k \neq 0\right] << N$$

Identification of spillover networks (1/8)

- Manresa (2016) applied a penalized estimation approach to the following model about Research and Development (X_{jt}) "spillovers" on firm productivity (Y_{it}).

$$Y_{it} = \alpha_i + \gamma_{ii} X_{it} + \sum_{j \neq i} \gamma_{ji} X_{jt} + \varepsilon_{it}$$

$$\Rightarrow \mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\Gamma} \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

- She assumed a **sparsity condition** $\sum_{j \neq i} \mathbb{1}\left\{\gamma_{ji} \neq 0\right\} << T$ for all firms i in the panel.

- She estimated the model via the LASSO (and post-LASSO subsequent OLS) using data on U.S. public companies.

$$\left(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\Gamma}}\right) = \underset{(\boldsymbol{\alpha}, \boldsymbol{\Gamma})}{\arg\min} \sum_{t=1}^{T} \|\mathbf{y}_t - \boldsymbol{\alpha} - \boldsymbol{\Gamma} \mathbf{x}_t\|_2^2 - \lambda \|\mathrm{vec}\left(\boldsymbol{\Gamma}\right)\|_1$$

- She extended the model to more covariates, time effects and examined the **asymptotics** of the ultimate OLS estimates.

Identification of spillover networks (2/8)

- Rose (2020) extends the framework to spatial lags of $\mathbf{y}_t$:

$$\mathbf{\Phi y}_t = \alpha \iota + \mathbf{\Gamma x}_t + \varepsilon_t$$

  and treats both $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ as **unknown estimands**.

- He suggests – and this is indeed helpful – to think of a linear simultaneous equations (SEM) representation of this model.

$$\begin{aligned} \phi_{11} y_{1t} + \ldots + \phi_{1N} y_{Nt} &= \alpha + \gamma_{11} x_{1t} + \ldots + \gamma_{1N} x_{Nt} + \varepsilon_{1t} \\ \phi_{21} y_{1t} + \ldots + \phi_{2N} y_{Nt} &= \alpha + \gamma_{21} x_{1t} + \ldots + \gamma_{2N} x_{Nt} + \varepsilon_{2t} \\ \ldots &= \ldots \\ \phi_{N1} y_{1t} + \ldots + \phi_{NN} y_{Nt} &= \alpha + \gamma_{N1} x_{1t} + \ldots + \gamma_{NN} x_{Nt} + \varepsilon_{Nt} \end{aligned}$$

- Rose postulates **sparsity** conditions on both $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ akin to the classical SEM **rank/order conditions** (Lecture 9).

- For estimation's sake, he penalizes both $\text{vec}(\mathbf{\Phi})$ and $\text{vec}(\mathbf{\Gamma})$.

Identification of spillover networks (3/8)

- De Paula, Rasul and Souza (2020; DPRS) instead apply the basic idea to the full-fledged BDF framework.

$$\mathbf{y}_t = \alpha\boldsymbol{\iota} + \beta\mathbf{G}\mathbf{y}_t + \gamma\mathbf{x}_t + \delta\mathbf{G}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

- The objective is to estimate the parameters $\boldsymbol{\theta} \equiv (\mathbf{G}, \beta, \gamma, \delta)$. Observe that here only one matrix is unknown: $\mathbf{G}$.

- As in BDF, it is useful to exploit the model's *reduced form*:

$$\begin{aligned}\mathbf{y}_t &= (\mathbf{I} - \beta\mathbf{G})^{-1}(\alpha\boldsymbol{\iota} + \gamma\mathbf{x}_t + \delta\mathbf{G}\mathbf{x}_t + \boldsymbol{\varepsilon}_t) \\ &= \mu\boldsymbol{\iota} + \mathbf{\Pi}\mathbf{x}_t + \boldsymbol{\nu}_t\end{aligned}$$

  where μ is a composite parameter, $\mathbf{\Pi}$ is a matrix of reduced form parameters, and $\boldsymbol{\nu}_t$ is a vector of composite error terms.

- The key question that DPRS address is whether there is a **unique** mapping $\mathbf{\Pi} \mapsto (\mathbf{G}, \beta, \gamma, \delta)$.

Identification of spillover networks (4/8)

- The key contribution by DPRS is to show that their model is **identified** under loose conditions. Indeed, *sparsity is not necessary*, but their model is more restricted than Rose's.

- The *inverse* mapping between structural and reduced form parameters is more easily derived in a *demeaned* model, and it is expressed as follows.

$$\mathbf{\Pi}(\boldsymbol{\theta}) = \gamma \mathbf{I} + (\beta\gamma + \delta) \sum_{k=0}^{\infty} \beta^{k-1} \mathbf{G}^{k}$$

- Their Theorem 1 proves **local** point identification of $\boldsymbol{\theta}$ under the assumptions detailed next. The proof is an application of the implicit function theorem, as in Rothenberg (1971).

- Their Theorem 2 proves **global** point identification of $\boldsymbol{\theta}$ if the sign of $\beta\gamma + \delta$ is known *a priori*.

Identification of spillover networks (5/8)

The assumptions that support identification in Theorems 1 and 2 by DPRS are as follows. They refer to the *true* parameter values.

1. $g_{ii} = 0$ for $i = 1, \ldots, N$: this is a standard normalization.

2. $\sum_{j=1}^{N} |g_{ij}| \leq 1$ for all $i \in \mathcal{I}$ while $|\beta| \leq 1$: social interactions are "stationary" in the sense that the variance-covariance of $\boldsymbol{y}_t$ is bounded.

3. $\beta\gamma + \delta \neq 0$: the effects do not offset one another, as in BDF.

4. $\sum_{j=1}^{N} |g_{ij}| = 1$ for at least an observation i: this is another normalization.

5. $\operatorname{diag}\left(\mathbf{G}^2\right)$ is not a multiple of ι: **their crucial assumption** (its interpretation is that nodes have different "popularity").

Identification of spillover networks (6/8)

The model by DPRS is easily extended to:

- individual fixed effects $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^{\mathrm{T}}$,

$$\mathbf{y}_t = \boldsymbol{\alpha} + \beta \mathbf{G} \mathbf{y}_t + \gamma \mathbf{x}_t + \delta \mathbf{G} \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

- time-varying shocks α_t for $t = 1, \ldots, T$,

$$\mathbf{y}_t = \alpha_t \boldsymbol{\iota} + \beta \mathbf{G} \mathbf{y}_t + \gamma \mathbf{x}_t + \delta \mathbf{G} \mathbf{x}_t + \boldsymbol{\varepsilon}_t$$

- higher-dimensional exogenous characteristics $\mathbf{X}_t$,

$$\mathbf{y}_t = \alpha \boldsymbol{\iota} + \beta \mathbf{G} \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\gamma} + \mathbf{G} \mathbf{X}_t \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t$$

- or any combination of the above.

As usual, fixed effects are purged by appropriate transformations ("within" or "between") of the data.

Identification of spillover networks (7/8)

While unnecessary for identification, **sparsity** comes back from the backdoor for the sake of **estimation**.

- Given any element g_{ij} of $\mathbf{G}$, DPRS assume the following.

$$\sum_{j \neq i} \mathbb{1}\left\{g_{ij} \neq 0\right\} << T$$

- Their estimation approach adapts the **Elastic Net GMM** by Caner and Zhang (2014). Denoting by $\mathbf{y}$ and $\mathbf{x}$ the vectors that stack *all* the observations of $\mathbf{y}_t$ and $\mathbf{x}_t$:

$$\widehat{\boldsymbol{\theta}}_{ENGMM} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \overline{\mathbf{g}}_{NT}^{\mathrm{T}}\left(\boldsymbol{\theta}; \mathbf{y}, \mathbf{x}\right) \mathbf{A}_{NT} \overline{\mathbf{g}}_{NT}\left(\boldsymbol{\theta}; \mathbf{y}, \mathbf{x}\right) - \\ - \lambda\left[\kappa \|\boldsymbol{\theta}\|_1 + (1-\kappa)\|\boldsymbol{\theta}\|_2\right]$$

where, as usual with GMM, $\overline{\mathbf{g}}_{NT}\left(\boldsymbol{\theta}; \mathbf{y}, \mathbf{x}\right)$ is a set of sample analogues of the population moment conditions, while $\mathbf{A}_{NT}$ is a weighting matrix.

Identification of spillover networks (8/8)

- The population moment conditions set by DPRS are:

$$\boldsymbol{g}_{NT}\left(\theta;\boldsymbol{y}_1,\dots,\boldsymbol{y}_T,\boldsymbol{x}_1,\dots,\boldsymbol{x}_T\right)=\mathbb{E}\begin{bmatrix}\widetilde{\boldsymbol{x}}_1\left(\widetilde{\boldsymbol{y}}_1-\boldsymbol{\Pi}\left(\theta\right)\widetilde{\boldsymbol{x}}_1\right)\\ \vdots\\ \widetilde{\boldsymbol{x}}_T\left(\widetilde{\boldsymbol{y}}_T-\boldsymbol{\Pi}\left(\theta\right)\widetilde{\boldsymbol{x}}_T\right)\end{bmatrix}$$

  where $\widetilde{\boldsymbol{x}}_t$, $\widetilde{\boldsymbol{y}}_t$ denotes appropriate transformation of the data aimed at purging constants or fixed effects (for $t=1,\dots,T$) and $\boldsymbol{\Pi}\left(\theta\right)$ is the structural-to-reduced parameters mapping.

- This can be a computationally demanding problem: to find a global minimum, DPRS adapt the so-called *particles swarm* algorithm (Kennedy and Eberhart, 1995).

- Good asymptotics here depend on N and T to "grow about at the same rate." This is quite a strong data requirement! The empirical application illustrated by DPRS – about tax competition between U.S. states – is informed accordingly.

Introduction to network formation

- In parallel with the use of networks to estimate social effects, interest for models of **network formation** also grew.

- Specifically, these are models aimed at explaining the entries of an adjacency matrix $\mathbf{G}$. Their key features are a **dyadic** level of variation – the unit of observation is a pair $(i, j) \in \mathcal{I}^2$ – and **discrete outcomes**.

- Ideally, these econometric models are predicated on explicit theoretical (behavioral) models of network formation. This, however, leads to econometric complications: most notably, unobserved heterogeneity and multiple equilibria.

- There is not a tight connection between these models and the study of social effects. However, control function approaches to correct for endogeneity of $\mathbf{G}$ in the BDF model demand a suitable model of network formation.

Dyadic regression in brief

- A **dyadic regression** is a statistical or econometric model for the expectation of a dependent variable Y_{ij} conditional on some independent variables $\boldsymbol{x}_{ij}$; these are all observed at the level of a *pair* (i, j), with $i, j = 1, \dots, N$ and $i \neq j$.

$$\mathbb{E}\left[Y_{ij} \middle| \boldsymbol{x}_{ij}\right] = h\left(\boldsymbol{x}_{ij}; \theta\right)$$

- Such a model has a complex structure of **cross-observation dependence**. Let the error term be defined as follows:

$$\varepsilon_{ij} = Y_{ij} - h\left(\boldsymbol{x}_{ij}; \theta\right)$$

  one can reasonably expect that $\mathbb{C}\text{ov}\left(\varepsilon_{ij}, \varepsilon_{k\ell} \middle| \boldsymbol{x}_{ij}, \boldsymbol{x}_{k\ell}\right) \neq 0$ if the two pairs (i, j) and (k, ℓ) have any elements in common.

- Inference on estimators of θ requires suitable estimators of their variance-covariance. Cameron and Miller (2014) as well as Tabord-Meehan (2021) develop them for linear $h\left(\boldsymbol{x}_{ij}; \theta\right)$ functions; Graham (2020) studies the general case.

Theories of network formation: a summary (1/2)

- The baseline network formation model in graph theory and statistical physics is the **"random graph"** model by Erdős and Rényi (1959) for undirected, unweighted networks. This model treats g_{ij} as the realization of a random variable G_{ij} that follows a simple Bernoulli distribution.

$$\mathbb{P}\left(G_{ij} = g_{ij}; p\right) = p^{g_{ij}} \left(1 - p\right)^{1 - g_{ij}} \cdot \mathbb{1}\left[g_{ij} \in \{0, 1\}\right]$$

- While simple, the random graph model can reproduce many topological properties of real-world networks, though not all.

- To an economist's eye, this model is devoid of any behavioral component: this led to the development of **strategic** models of network formation that are modeled as a *game*.

- Strategic games of network formation are not simple: edges result from the decisions taken by two players (e.g. i and j), whose preferences may depend on the whole topology $\mathcal{G}$.

Theories of network formation: a summary (2/2)

- One famous solution concept is **pairwise stability** (Jackson and Wolinsky, 1996) which applies to undirected, unweighted networks: edges occur if *both* i and j "agree" *in equilibrium*.

- Let $\mathcal{G}$ be the equilibrium edge sequence, $\mathcal{G}_{-ij}$ be the one that obtains by deleting the link between i and j (if $g_{ij} = 1$) and $\mathcal{G}_{+ij}$ the one resulting by adding that link (if $g_{ij} = 0$). Thus:

$$g_{ij} = 1 \iff U_i(\mathcal{G}) \geq U_i(\mathcal{G}_{-ij}) \textbf{ and } U_j(\mathcal{G}) \geq U_j(\mathcal{G}_{-ij})$$
$$g_{ij} = 0 \iff U_i(\mathcal{G}) > U_i(\mathcal{G}_{+ij}) \textbf{ or } U_j(\mathcal{G}) > U_j(\mathcal{G}_{+ij})$$

  where $U_i(\cdot)$, $U_j(\cdot)$ denote individual utilities that depend on the *entire* edge sequence $\mathcal{G}$ (also about other nodes).

- However, this solution concept rules out **transfers** between nodes. An alternative model provides the following.

$$g_{ij} = 1 \iff U_i(\mathcal{G}) + U_j(\mathcal{G}) \geq U_i(\mathcal{G}_{-ij}) + U_j(\mathcal{G}_{-ij})$$

Conditional network formation (1/7)

- Graham (2017) developed the baseline econometric model to study the formation of undirected, unweighted networks.

- Let G_{ij} be a **dyadic** random variable with support on $\{0, 1\}$ and whose realization is g_{ij}. Graham's model reads:

$$G_{ij} = \mathbb{1}\left[\alpha_i + \alpha_j + \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{ij} \geq 0\right]$$

  where α_i and α_j are two individual "attributes" that encode **individual heterogeneity**, $\boldsymbol{x}_{ij}$ are **dyadic** characteristics (arrayed as $\boldsymbol{X}$), and ε_{ij} is a **dyadic** idiosyncratic **shock**.

- The model is **symmetric** across dyads: with N nodes it has $N(N-1)$ effective cross-sectional observations.

- This network formation rule implicitly allows for "transfers" of utility, but there are **no strategic** considerations at play (the latent variable does not depend on other edges).

Conditional network formation (2/7)

- The original motivation for this model is a classical problem in the statistical analysis of networks: the need to distinguish "homophily" from "unobserved heterogeneity."

- Why are central nodes typically connected to one another? One explanation is **homophily**: node similarity (potentially encoded in $\boldsymbol{x}_{ij}$) is conducive to higher edge probabilities.

- An alternative explanation is that nodes are **heterogeneous** in their tendency to generate edges (and therefore acquire a higher *degree*), as encoded in α_i and α_j.

- This is a standard problem for an econometrician: Graham's contribution is a solution for this particular dyadic setting.

- For simplicity's sake, Graham removes complications related to cross-dependence: ε_{ij} is assumed i.i.d. across dyads.

Conditional network formation (3/7)

- More specifically, Graham assumes $\varepsilon_{ij} \sim \text{Logistic}(0,1)$. This leads to the following *edge formation* conditional probability.

$$\mathbb{P}\left(G_{ij}=1\middle|\,\boldsymbol{x}_{ij}\right)=\frac{\exp\left(\alpha_i+\alpha_j+\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right)}{1+\exp\left(\alpha_i+\alpha_j+\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right)}$$

- Conceptually, MLE estimation of this augmented logit model is straightforward. Shall one here worry about the *incidental parameter problem* due to the “fixed effects” α_i and α_j?

- Graham argues that one should not: **consistent** estimation of $\boldsymbol{\beta}$ *and* $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_N)$ is possible. Intuitively, there are $N-1$ observations that identify each fixed effect α_i.

- Graham does propose a “Joint Maximum Likelihood” (JML) estimator for this model, but this is not his key contribution. This JML estimator may be computationally demanding.

Conditional network formation (4/7)

- Graham's key contribution is the so-called **"Tetrad Logit"** (TL) conditional MLE estimator for this model.

- To build motivation, write the **observed degree sequence** as $\mathbf{d} = (d_1, \ldots, d_N)$, the corresponding random vector as $\boldsymbol{d}$, and the set of adjacency matrices with degree sequence $\mathbf{d}$ as:

$$\mathbb{G}_{\mathbf{d}} = \{\mathbf{H} : \mathbf{H} \in \mathbb{G}, \mathbf{H}\boldsymbol{\iota} = \mathbf{d}\}$$

  where $\mathbb{G}$ is defined as the support of $\boldsymbol{G}$: the random matrix that generates $\mathbf{G}$. Matrices $\mathbf{H} \in \mathbb{G}_{\mathbf{d}}$ have entries h_{ij}.

- The degree sequence $\mathbf{d}$ is a *sufficient statistic* for $\boldsymbol{\alpha}$. Indeed, the probability to observe the network or adjacency matrix $\mathbf{G}$ *conditional* on $\mathbf{d}$ does not depend on $\boldsymbol{\alpha}$.

$$\mathbb{P}\left(\boldsymbol{G} = \mathbf{G} \middle| \boldsymbol{X}, \boldsymbol{d} = \mathbf{d}\right) = \frac{\exp\left(\sum_{i=1}^{N}\sum_{j<i} g_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right)}{\sum_{\mathbf{H}\in\mathbb{G}_{\mathbf{d}}}\exp\left(\sum_{i=1}^{N}\sum_{j<i} h_{ij}\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right)}$$

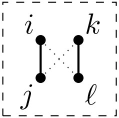
Conditional network formation (5/7)

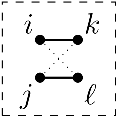
- This approach is reminiscent of "conditional" estimation of logit models with fixed effects in panel data. However, it is complicated to apply in practice: $\mathbb{G}_{\mathbf{d}}$ is difficult to construct.

- Thus, Graham proposes a variation of it which exploits the identifying power of **tetrads**: combinations of four distinct nodes (i, j, k, ℓ). He defines the following random variable.

$$S_{ij,k\ell} = G_{ij}G_{k\ell}\left(1 - G_{ik}\right)\left(1 - G_{j\ell}\right) - \\ - \left(1 - G_{ij}\right)\left(1 - G_{k\ell}\right)G_{ik}G_{j\ell}$$

- It is $S_{ij,k\ell} = 1$ if $G_{ij} = G_{k\ell} = 1$, $G_{ik} = G_{j\ell} = 0$; $S_{ij,k\ell} = -1$ if the opposite occurs; $S_{ij,k\ell} = 0$ otherwise. The two edges $G_{i\ell}$ and G_{jk} ("dotted" in the figure below) do not affect $S_{ij,k\ell}$.

$S_{ij,k\ell} = 1 \Longleftrightarrow$ 

$S_{ij,k\ell} = -1 \Longleftrightarrow$ 

Conditional network formation (6/7)

- By defining $\widetilde{\boldsymbol{x}}_{ij,k\ell} \equiv \boldsymbol{x}_{ij} + \boldsymbol{x}_{kl} - \boldsymbol{x}_{ik} - \boldsymbol{x}_{j\ell}$, it is:

$$\mathbb{P}\left(S_{ij,k\ell} = 1 \middle| \widetilde{\boldsymbol{x}}_{ij,k\ell}, S_{ij,k\ell} \in \{-1, 1\} \right) = \frac{\exp\left(\widetilde{\boldsymbol{x}}_{ij,k\ell}^{\mathrm{T}} \boldsymbol{\beta}\right)}{1 + \exp\left(\widetilde{\boldsymbol{x}}_{ij,k\ell}^{\mathrm{T}} \boldsymbol{\beta}\right)}$$

  which identifies $\boldsymbol{\beta}$: neither this expression depends on $\boldsymbol{\alpha}$.

- Graham thus defines the following log-likelihood index.

$$\begin{aligned} & l_{ij,k\ell}\left(\boldsymbol{\beta} \middle| s_{ij,k\ell}, \widetilde{\mathbf{x}}_{ij,k\ell} \right) = \\ & \quad = \left| s_{ij,k\ell} \right| \left\{ s_{ij,k\ell} \widetilde{\mathbf{x}}_{ij,k\ell}^{\mathrm{T}} \boldsymbol{\beta} - \log\left[1 + \exp\left(s_{ij,k\ell} \widetilde{\mathbf{x}}_{ij,k\ell}^{\mathrm{T}} \boldsymbol{\beta} \right) \right] \right\} \end{aligned}$$

- However, this index is **not** permutation-invariant. Graham defines a **symmetric** one that averages over all the $4! = 24$ permutations of the tetrad, which are collected in the set $\mathbb{I}_4$.

$$\gamma_{ij,k\ell}\left(\boldsymbol{\beta}\right) = \frac{1}{24} \sum_{(i',j',k',\ell') \in \mathbb{I}_4} l_{i'j',k'\ell'}\left(\boldsymbol{\beta} \middle| s_{ij,k\ell}, \widetilde{\mathbf{x}}_{ij,k\ell} \right)$$

Conditional network formation (7/7)

- The TL estimator is thus defined as follows.

$$\widehat{\boldsymbol{\beta}}_{TL} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^K} \binom{N}{4}^{-1} \sum_{i=1}^{N} \sum_{j<i} \sum_{k<j} \sum_{\ell<k} \gamma_{ij,k\ell}(\boldsymbol{\beta})$$

- The intuition that supports this estimator is that the tetrad index $l_{ij,k\ell}(\boldsymbol{\beta})$ conditions on the *local* degree sequence in the tetrad, not unlike an unfeasible estimator that conditions on the entire degree sequence $\mathbf{d}$ would do.

- However, identification of $\boldsymbol{\beta}$ requires that conditional on the local degree sequence, a tetrad (i, j, k, ℓ) can assume multiple configurations. Hence, many tetrads in the data would **not** contribute to the MLE problem. Graham discusses a useful "taxonomy of tetrads" that helps appreciate this feature.

- Graham shows that the Tetrad Logit estimator is consistent and asymptotically normal: certainly not an easy task.

Strategic network formation: an open issue

- In Graham's model, the formation of an edge is conditionally independent of the rest of the edge sequence $\mathcal{G}$. Thus, it can be seen as a restriction of a more general model, such as the following (for $\delta = 0$) where *"friends in common"* matter.

$$G_{ij} = \mathbb{1}\left[\alpha_i + \alpha_j + \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \delta \sum_{k=1}^{N} G_{ik}G_{kj} + \varepsilon_{ij} \geq 0\right]$$

- As a game, this model features **multiple equilibria**. While strategic considerations in network formation are important, their econometric evaluation is still a conceptual challenge.

- The literature on the econometrics of games treats multiple equilibria under the framework of *set identification*. This is hard to apply here: the dimension of $\mathbb{G}$ is in the order of 2^N.

- See de Paula, Richards-Shubik and Tamer (2018) and Sheng (2020) for recent advances that rely on specific restrictions.