Propensity Score Matching with Network Data

Evaluating the effect of GATT on bilateral trade

Bruno Arpino[1] Luca De Benedictis[2] Alessandra Mattei[3]

[1]Universitat Pompeu Fabra (Spain), bruno.arpino@upf.edu
[2]University of Macerata (Italy), luca.debenedictis@unimc.it
[3]University of Florence (Italy), mattei@disia.unifi.it

ECONOMICS OF GLOBAL INTERACTIONS 6.0
Bari, September 7-8, 2015

Outline of the talk

- Causal inference with confounders
  - a brief review of Rubin's causal model
- Strong ignorability and propensity score matching
- The Network as a confounder
  - Network data and network's statistics
  - Degree centrality: local
  - Eigenvector centrality: global
- Application: GATT membership and the causal effect on bilateral trade
  - A brief history of GATT
  - Data: treatment, pre-treatment covariates, and network structure.
  - Propensity Score Matching with Network Data
  - The effect of GATT
- Summing-up

Part I

Causal inference with confounders

The Rubin's causal model: A primer

- Given a random sample of n units (country pairs): $ij = 1, \ldots, n$
- and a Treatment T_{ij} = Participation in the General Agreement on Tariffs and Trade (GATT)

$$T_{ij} = \begin{cases} 0 & \text{If dyad } ij \text{ is a } \textit{mixed dyad}\text{: Either only one country or} \\ & \text{ no country in dyad } ij \text{ is a GATT participant} \\ 1 & \text{If both countries in dyad } ij \text{ are GATT participants} \end{cases}$$

- Outcome variable: Bilateral trade = Average imports and exports in a given year for a dyad
- Under SUTVA, each dyad ij has two potential outcomes for bilateral trade:

$$\begin{aligned} Y_{ij}(0) &= \text{Potential bilateral trade if dyad } ij \text{ is a mixed dyad} \\ Y_{ij}(1) &= \text{Potential bilateral trade if both countries in dyad } ij \\ & \quad \text{are GATT participants} \end{aligned}$$

- Causal estimand: Average Treatment Effect

$$ATE = \mathbb{E}\left[Y_{ij}(1) - Y_{ij}(0)\right]$$

The Rubin's causal model: Observed data

For each dyad *ij* we observe

- A vector of pre-treatment variables, which includes country-specific characteristics, dyad-specific characteristics and network characteristics:

$$\mathbf{X} = \mathbf{X_i} \cup \mathbf{X_j} \cup \mathbf{X_{ij}} \cup \mathbf{N_i} \cup \mathbf{N_j}$$

- The treatment actually received: T_{ij}
- Only one of the potential outcomes for each unit, either $Y_{ij}(0)$ or $Y_{ij}(1)$, depending on the treatment actually received:

$$Y_{ij} = Y_{ij}(T_{ij}) = Y_{ij}(0)(1 - T_{ij}) + Y_{ij}(1)T_{ij}$$

This is the fundamental problem of causal inference

- Statistical inference for causal effects requires the specification of an assignment mechanism (who gets treated?)

The Rubin's causal model: The assignment mechanism

- In observational studies the assignment mechanism is unknown
- Strong Ignorability (Rosenbaum and Rubin, 1983)

  Unconfoundedness: $\big(Y_{ij}(0), Y_{ij}(1)\big) \perp T_{ij} | \mathbf{X}$

  Positivity: $0 < P(T_{ij} = 1 | \mathbf{X}) < 1$

- Unconfoundedness (selection on observable or conditional independence assumption) ensures that potential outcomes and treatment are conditionally independent given pre-treatment variables
  - ✓ Unconfoundedness amounts to assuming that, within cells defined by the values of pre-treatment variables **X**, the treatment is randomly assigned
- The positivity assumption implies that in large samples we can find units exposed to different treatment levels for all values of the pre-treatment variables

The Rubin's causal model: The assignment mechanism

- Under unconfoundedness,

$$\mathbb{E}[Y_{ij}(t)|\mathbf{X}=\mathbf{x}] = \mathbb{E}[Y_{ij}(t)|T_{ij}=t, \mathbf{X}=\mathbf{x}] = \mathbb{E}[Y_{ij}|T_{ij}=t, \mathbf{X}=\mathbf{x}]$$

- Under unconfoundedness, we can estimate ATT by first estimating the average treatment effect on the treated for a subpopulation of treated dyads with covariates $\mathbf{X}=\mathbf{x}$:

$$\begin{aligned}&\mathbb{E}[Y_{ij}(1)-Y_{ij}(0)|T_{ij}=1, \mathbf{X}=\mathbf{x}] = \\ &\qquad \mathbb{E}[Y_{ij}|T_{ij}=1, \mathbf{X}=\mathbf{x}] - \mathbb{E}[Y_{ij}|T_{ij}=0, \mathbf{X}=\mathbf{x}]\end{aligned}$$

  and

$$\begin{aligned}&\mathbb{E}[Y_{ij}(1)-Y_{ij}(0)|T_{ij}=1] = \\ &\qquad \mathbb{E}\left[\mathbb{E}[Y_{ij}|T_{ij}=1, \mathbf{X}=\mathbf{x}] - \mathbb{E}[Y_{ij}|T_{ij}=0, \mathbf{X}=\mathbf{x}]|T_{ij}=1\right]\end{aligned}$$

  where the outer expetation is over the distribution of $\mathbf{X}|T_{ij}=1$

- We need to estimate $\mathbb{E}[Y_{ij}|T_{ij}=t, \mathbf{X}=\mathbf{x}]$ for all values of t and $\mathbf{x}$ in the support of $\mathbf{X}$
- If the positivity assumption is violated at $\mathbf{X}=\mathbf{x}$, it would be infeasible to estimate $\mathbb{E}[Y_{ij}|T_{ij}=t, \mathbf{X}=\mathbf{x}]$ for each $t \in \{0,1\}$

Part II

Strong ignorability and Propensity Score Matching

The role of the propensity score (Rosenbaum and Rubin, 1983)

- Removing all biases, by adjusting for differences in observed covariates, may be difficult to implement with a large number of covariates
- The conditional probability of receiving the treatment given the pre-treatment variables, $e_{ij}(\mathbf{X}) \equiv P(T_{ij} = 1|\mathbf{X})$, is the propensity score
- The propensity score is a balancing score: $\mathbf{X} \perp T_{ij}|e_{ij}(\mathbf{X})$
- Strong ignorability given the propensity score: If treatment assignment is strongly ignorable given pre-treatment variables, then it is strongly ignorable given the propensity score, that is

$$\textit{If} \quad \big(Y_{ij}(0), Y_{ij}(1)\big) \perp T_{ij}|\mathbf{X} \quad \textit{and} \quad 0 < e_{ij}(\mathbf{X}) < 1$$

then

$$\big(Y_{ij}(0), Y_{ij}(1)\big) \perp T_{ij}|e_{ij}(\mathbf{X}) \quad \textit{and} \quad 0 < P(T_{ij} = 1|e_{ij}(\mathbf{X})) < 1$$

- All biases due to observable covariates can be removed by conditioning solely on the propensity score

Propensity score methods

- The true propensity score is generally unknown: The propensity score needs to be estimated
- The goal is to obtain estimates of the propensity score that statistically balance the covariates between treated and control groups
- We estimate the propensity score using a logit regression model

  $$e_{ij}(\mathbf{X}) = \frac{\exp\{g(\alpha, \mathbf{X})\}}{1 + \exp\{g(\alpha, \mathbf{X})\}}$$

  where g is a function of covariates with linear and higher order terms
- The choice of which higher order terms to include is determined solely by the need to obtain an estimate of the propensity score that satisfies the balancing property
- We use matching on the estimated propensity score to create treated and control groups with adequate balance in the covariate distributions
- After matching the average treatment effect is estimated comparing matched treated and matched control dyads

Part III

The Network as a confounder

The Rubin's causal model: The choice of the pre-treatment variables

- In observational studies it is crucial to think very carefully about why some units (e.g., dyads) receive the active treatment condition (e.g., both countries in the dyad are GATT participants) versus the control treatment condition (e.g., a dyad is a mixed dyad)
- The choice of the pre-treatment variables conditional on which the strong ignorability assumption holds is a critical issue
- In our study, we argue that it is crucial to have information on
  - $\mathbf{X}_i$ and $\mathbf{X}_j$: Country-specific characteristics;
  - $\mathbf{X}_{ij}$: Dyad-specific characteristics;
  - $\mathbf{N}_i$ and $\mathbf{N}_j$: Network structure
- Controlling for the pre-treatment Network structure of trade flows allows us to account for trade interdependence

The Rubin's causal model: The choice of the pre-treatment variables

- The central assumption of our analysis is that with network data, treatment assignment is strongly ignorable given country-specific characteristics, dyad-specific characteristics and network characteristics:

  but strongly ignorability holds given country-specific characteristics, dyad-specific characteristics and network structure:

  $$\big(Y_{ij}(0), Y_{ij}(1)\big) \perp T_{ij} | \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{ij}, \mathbf{N}_i, \mathbf{N}_j$$

  and

  $$0 < P(T_{ij} = 1 | \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_{ij}, \mathbf{N}_i, \mathbf{N}_j) < 1$$

  it would not be so if network characteristics where not included among the potential confounders.

- Network structure is reasonably correlated both with the potential outcomes for bilateral trade as well as with GATT participation: Ignoring network structure may induce bias.
- We investigate the importance to account for network structure by estimating *ATE* including and not including N in the set of matching variables

Technical issues: interdependence

- *Network centrality measures*:

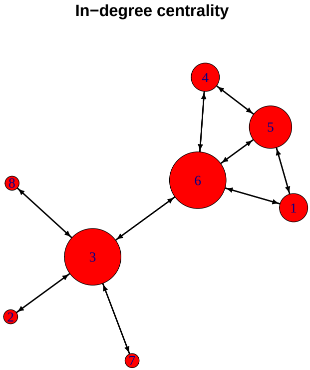
✓ *Degree centrality*: it measures how a node is connected to others. Since the PTA network is unweighted, the degree centrality measures the centrality of a node by the number of connections the node has. The degree centrality is essentially a *local centrality measure*. It takes into consideration only the *direct links* of a node, its *nearest neighborhood*, respectless to the position of the node in the network's structure.

✓ *Eigenvector centrality*: it measures how important, central, influential or tightly clustered a node's neighbors are. It measures a country's centrality looking at the importance of its neighbors. The Eigenvector centrality is essentially a *global centrality measure*.

Unweighted Network: Degree Centrality - Example

The node **Degree centrality** is

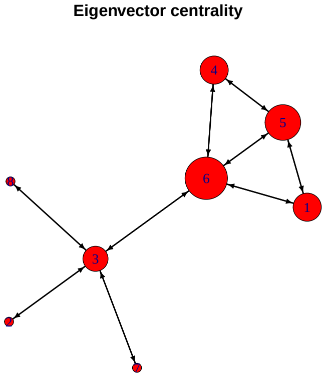
- **Node 1**: 2 $\rightsquigarrow \frac{2}{N-1} = \frac{2}{7} = 0.28$
  $\rightsquigarrow \frac{2}{M} = \frac{2}{18} = 0.11$
- **Node 2**: 1
- **Node 3**: 4
- **Node 4**: 2
- **Node 5**: 3
- **Node 6**: 4
- **Node 7**: 1
- **Node 8**: 1



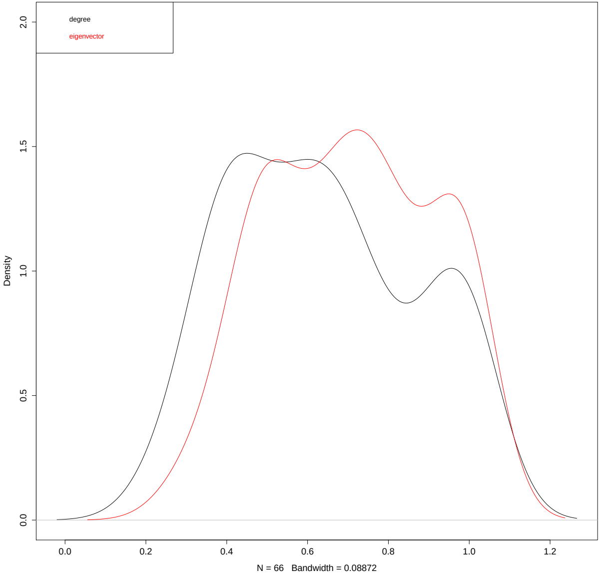
Unweighted Network: Eigenvector Centrality - Example

The node **Eigenvector centrality** is

- **Node 1**: 0.378
- **Node 2**: 0.122
- **Node 3**: 0.338
- **Node 4**: 0.378
- **Node 5**: 0.479
- **Node 6**: 0.569
- **Node 7**: 0.122
- **Node 8**: 0.122



Centralities: degree and eigenvector



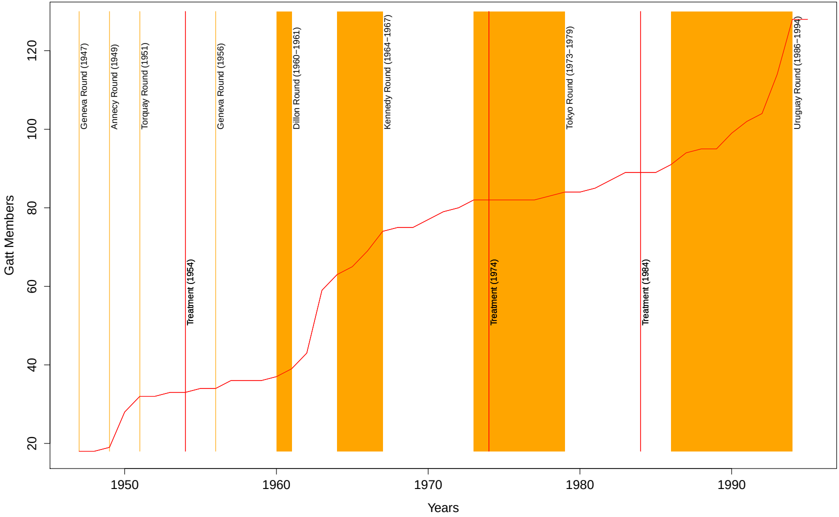
Part IV

Application: GATT membership and the causal effect on bilateral trade

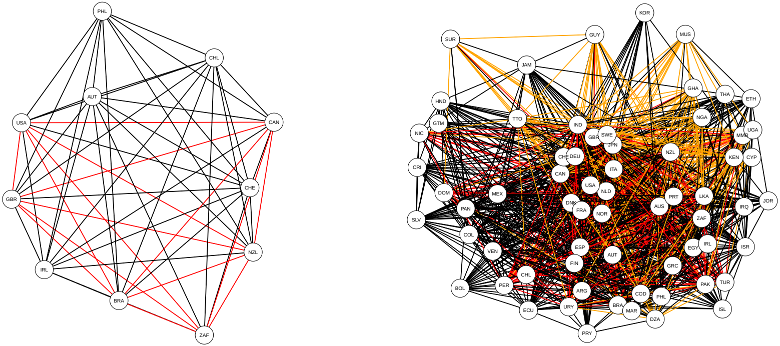
A brief history of the GATT

- The *General Agreement on Tariffs and Trade* (GATT) was established in 1947 (and ended in 1994, superseded by the WTO) having the goal "to remove or diminish barriers which impede the flow of international trade and to encourage by all available means the expansion of commerce" (Irvin, 1995)
  - defined rules to govern trade policy;
  - pursued a binding non-discriminatory tariff-reduction multilateral strategy: (best) concessions between any two participants automatically get passed to others according to the most-favored-nation (MFN) principle;
  - decisions where taken in eight rounds of trade negotiations.
- GATT's "members":
  - Formal members, as classified in Rose (2004)
  - GATT's rules applied also to Nonmember Participants, as classified in Tomz et al. (2007): Colonies and overseas territories (Art. XXVI of GATT), Newly independent states, and Provisional members.
- Treatment
  - We examine the causal effect of GATT membership when both countries in the dyad are: Formal members or Nonmember Participants in 1954.

The GATT: members and negotiation rounds



GATT formal members & informal participants: 1948 - 1954



Source: Authors' elaboration on Rose (2004) and Tomz et al. (2007) data. Nodes are countries (Iso3 codes) and edges are trade links ($y_{ij} > 0$). Red edges indicate that both countries are formal members of GATT; orange edges indicate that both countries are formal members or nonmember participants.

Outcome, treatment and covariates

- Outcome. Is the logarithm of average imports and exports in a given year (1955) for each dyad: y_{ij}.
- Treatment. Both countries in the dyad are GATT's formal members or nonmember participants (1955): T_{ij}.
- Covariates are selected according to the gravity model of international trade (Baier and Bergstand, 2008).
  - X_{ij}: log product real GDP; log product real GDP per capita; log product land area, log distance, land border, common language; GSP, regional FTA, currency union;
  - X_i and X_j: landlocked, island, currently colonized, common colonizer, past colonial relation;
  - N_i and N_j: degree and eigenvector centrality measures (1954).

Common support and balance

- In order to satisfy the positivity assumption we drop treated (control) dyads whose propensity score is higher than the maximum or less than the minimum of propensity score of the control (treated) dyads

	Without N Common Support		With N Common Support		
Treatment	Off	On	Off	On	*Total*
Control dyads	0	766	28	738	766
Treated dyads	5	548	15	538	553
Total	5	1 314	43	1 276	1 319

- Summary of the distribution of the abs bias before and after matching

	Without N		With N	
Covariates	Mean bias	Median bias	Mean bias	Median bias
Standard covariates (without N)				
Before matching	18.8	17.2	18.8	17.2
After matching	1.2	0.6	4.7	2.9
All covariates (with N)				
Before matching	23.2	21.7	23.2	21.7
After matching	6.2	1.2	5.2	4.1

Balance before and after matching: Percent Bias

Variable	Before matching	After matching Without N	After matching With N
Currency Union	15.5	-0.1	-4.4
Log distance	18.0	0.4	-1.1
Log product real GDP	39.1	5.2	8
Log product real GDP pc	16.4	2.1	15.6
Common language	-12.7	1.4	-1.9
Land border	-8.0	-0.2	0.1
Landlocked	-28.8	0.1	-2
Island	24.9	0.7	-5.1
Log product land areas	21.6	0.9	3.8
Common colonizer	21.8	-0.3	-10.9
Currently colonizer	15.3	-2.7	-1.8
Past colonial relationship	3.1	0.1	1.1
Degree_i	40.3	26.8	9.4
Degree_j	34.0	17.6	4.8
Eigenvector_i	39.4	26.1	9.7
Eigenvector_j	32.5	15	3.8

Average treatment effect

	ATE	Bootstrap Std. Err.	Normal-based 95% Conf. Interval	
Without N	0.415	0.107	0.205	0.626
With N	0.304	0.117	0.075	0.534

Part V

Summing-up and conclusion

Summing-up

- GATT had a positive influence on bilateral trade (in 1955);
- The estimated causal effect of GATT is a bilateral trade flow 35% higher between GATT's members wrt the control group of country-pairs;
- Ignoring network characteristics would lead to an upward bias (48% more than the "true" effect);
- Ignoring network characteristics would lead to an under-estimation of the positive effect of GATT on non-member countries through interdependence.

More to do

- Are centrality measures the “right statistics” to include in the Propensity Score?
- Are there feasible alternatives to capture network effects?
- Should we take into consideration monadic or dyadic network statistics (centrality vs dissimilarity)?
- How to deal with the violation of the SUTVA (Stable Unit Treatment Value Assumption) due to Network interference in a Rubin’s causal model framework ?
- More?

Propensity Score Matching with Network Data

Evaluating the effect of GATT on bilateral trade

Bruno Arpino[1] Luca De Benedictis[2] Alessandra Mattei[3]

[1]Universitat Pompeu Fabra (Spain), bruno.arpino@upf.edu
[2]University of Macerata (Italy), luca.debenedictis@unimc.it
[3]University of Florence (Italy), mattei@disia.unifi.it

ECONOMICS OF GLOBAL INTERACTIONS 6.0
Bari, September 7-8, 2015