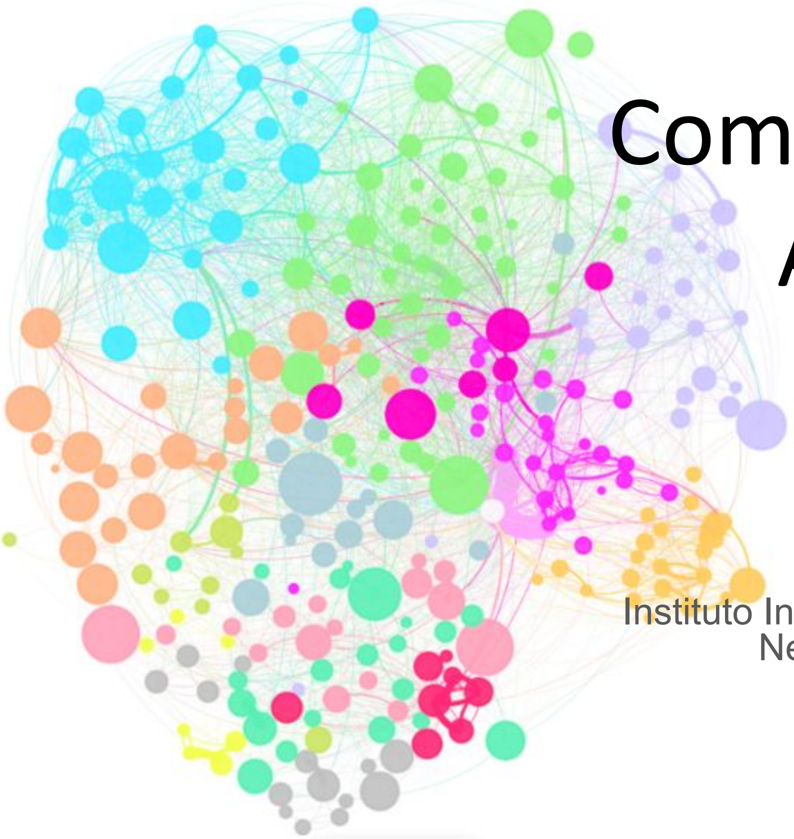
Instituto Interdisciplinario de Economía Política de Buenos Aires (IIEP-BAIRES)

NetLab Laboratorio de Redes y Sistemas Complejos

CONICET UBA IIEP

ICTP SAIFR



Complex Economic Networks: Analysis, Applications and Data

Viktoriya Semeshenko

Instituto Interdisciplinario de Economía Política (IIEP-UBA-CONICET)
Networks and Complex Systems Laboratory (**Netlab**)

vika.semeshenko@gmail.com
@vsemesh @Netlab_IIEP

NetSciX23 @NetSciX23

Key dates

Abstract Submission: October 28th, 2022 (23:59:59 AoE)

Author Notification: November 25th, 2022 (23:59:59 AoE)

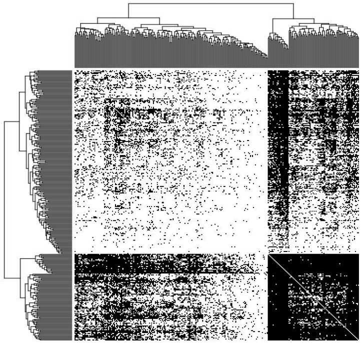
Early Registration: December 12th, 2022 (23:59:59 AoE)

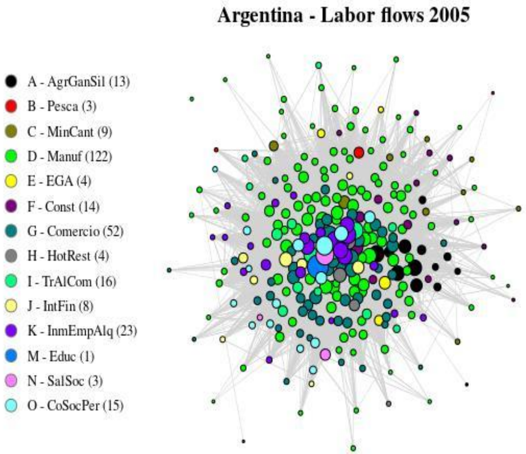
Clase 3: Network Reduction Techniques

Sources used in the course:

- Barabási: Albert-László Barabási. [Network Science]
- Menczer et al: [Filippo Menczer, Santo Fortunato, Clayton A. Davis, A First Course in Network Science. Cambridge University Press 2020]

Flow Network: Predominant year-to-year structure





Source: Elaboración propia SIPA, año 2005

Metrics	Average
Nodes (Order)	**287**
Edges (size)	26.961
Weights	278.117
CComp	1,10
Densidad	0,33
⟨k⟩	188
⟨kin⟩	94
⟨kout⟩	94
⟨knn⟩	**285**
Assortativity	-0,26
Reciprocidad	0,69
CCoef	0,65
LCP	1,68
Diámetro	3,40

What does it mean in terms of labor flows?

Sectors

Links &/industries

… of labor

… of links

Average ties by sec

… de lazos

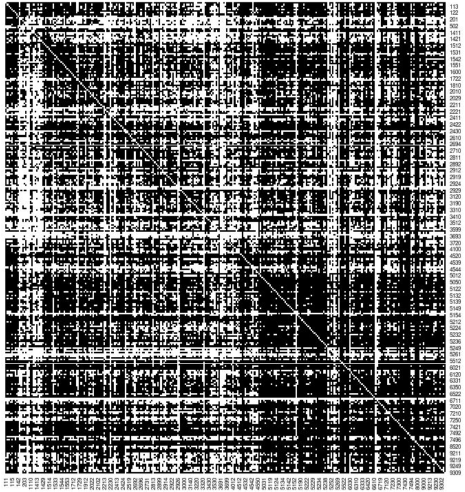
Triángles

Caminos e/sec

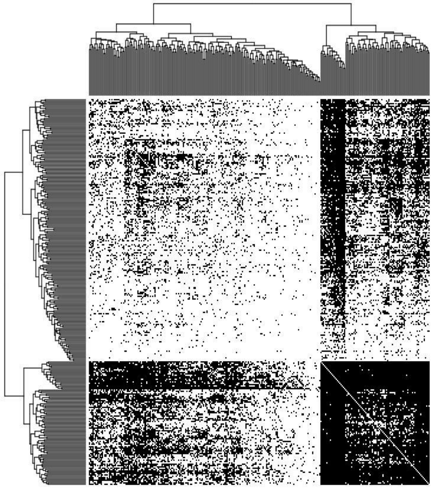
- Single large connected component
- Relevant link density (non-sparse matrices)
- High Degree (*k*)
- High reciprocity of ties
- High clustering coefficient
- Short paths (reduced diameter: 3 steps)

→high reciprocity (0.69) in the connections and a high transitivity or global clustering (0.65), meaning that 2 out of 3 of the possible connections between three nodes (triples) constituted closed triangles.

C. Heatmaps



C1. Lexicographic,
in alphabetical order



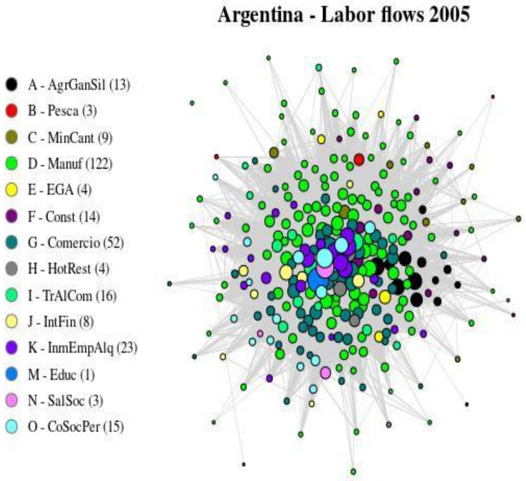
C2. Ordered,
using hierarchical clustering

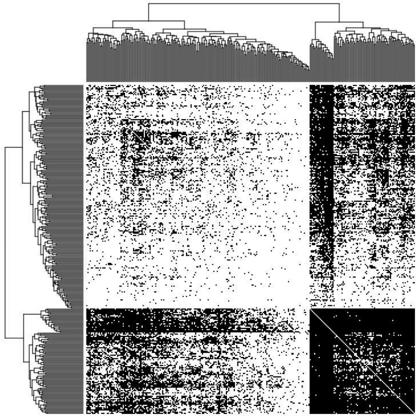
reorganize the transition matrices into groups of nodes that interact with each other with greater intensity (blocks), groups of nodes that interact with each other and with the entire network (center, core), or groups that only interact with subsets of nodes with greater connectivity (periphery)

Flow Network (cont.)

- **Properties**: Small world
  - High clustering coefficient (transitivity)
  - Short average paths

- Core-periphery structure
  - Core: subset of high degree nodes, highly interconnected
  - Periphery: subset of low degree nodes, connected (almost only) with core nodes





Centrality Measures

- Node centrality : **degree**
- **Closeness**: What if it's not so important to have many direct friends? Or be "between" others. But one still wants to be in the "middle" of things, not too far from the center.
  - Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph
- **Betweenness**: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?

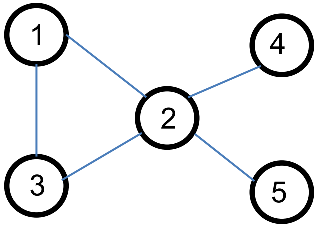
Degree Centrality

- We saw that the **degree** is a very natural measure of *centrality* in social networks
- The average degree of a network indicates how connected the nodes are on average
  - it may not be representative of the actual distribution of degree values, like is the case when the nodes have heterogeneous degrees, as in many real-world networks

Degree Centrality

- Degree Centrality: of node *i* is defined using its adjacency matrix, *A*:

$$k_i = \sum_{j=1}^{n} a_{ij} = (e^T A)_i = (Ae)_i$$



Since the adjacency matrix of the network is

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

the node degree vector is

$$\mathbf{k} = A\mathbf{e} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 2 \\ 1 \\ 1 \end{bmatrix}.$$

That is, the degrees of the nodes are: $k(1) = k(3) = 2$, $k(2) = 4$, $k(4) = k(5) = 1$, indicating that the most central node is 2.

Closeness

- Another way to measure the centrality of a node is by determining how “close” it is to the other nodes
- Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

Closeness

- This can be done by summing the distances from the node to all others
- If the distances are short on average, their sum is a small number and we say that the node has high centrality
- This leads to the definition of **closeness centrality**, which is simply the inverse of the sum of distances of a node from all others
- The closeness centrality of node i: $g_i = \frac{1}{\sum_{j \neq i} l_{ij}}$

  where is the distance from *i* to *j* and the sum runs over all the nodes of the network, except *i* itself.

Closeness Centrality

- The closeness of the node *i* in an undirected network *G* is defined as:

$$CC(i) = \frac{n-1}{S(i)}$$

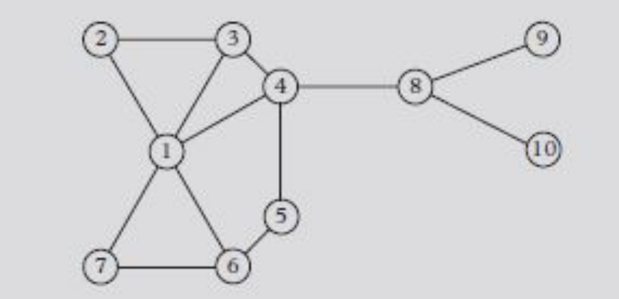
where $S(i)$ is the distance sum, calculated from the shortest path distances:

$$S(i) = \sum_{j \in V(G)} d(i,j)$$

- In a directed network a node has in- and out-closeness centrality.
- In-closeness corresponds to how close this node is to nodes it is receiving information from.
- Out-closeness centrality indicates how close the node is from those it is sending information to.

Closeness Centrality

- Ejemplo:



the distance matrix of this network, which is given by

$$D = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 & 1 & 1 & 2 & 3 & 3 \\ 1 & 0 & 1 & 2 & 3 & 2 & 2 & 3 & 4 & 4 \\ 1 & 1 & 0 & 1 & 2 & 2 & 2 & 2 & 3 & 3 \\ 1 & 2 & 1 & 0 & 1 & 2 & 2 & 1 & 2 & 2 \\ 2 & 3 & 2 & 1 & 0 & 1 & 2 & 2 & 3 & 3 \\ 1 & 2 & 2 & 2 & 1 & 0 & 1 & 3 & 4 & 4 \\ 1 & 2 & 2 & 2 & 2 & 1 & 0 & 3 & 4 & 4 \\ 2 & 3 & 2 & 1 & 2 & 3 & 3 & 0 & 1 & 1 \\ 3 & 4 & 3 & 2 & 3 & 4 & 4 & 1 & 0 & 2 \\ 3 & 4 & 3 & 2 & 3 & 4 & 4 & 1 & 2 & 0 \end{bmatrix}$$

- The vector of distance-sum of each node is then

$$s = De = (e^T D)^T = [15\ 22\ 17\ 14\ 19\ 20\ 21\ 18\ 26\ 26]^T$$

- For node 1 : $CC(1) = \frac{10-1}{S(1)} = \frac{9}{15}$ =0,6
- The full vector of closeness centralities is:

CC = [0.600 0.409 0.529 **0.643** 0.474 0.450 0.428 0.500 0.346 0.346]*T*

Betweenness

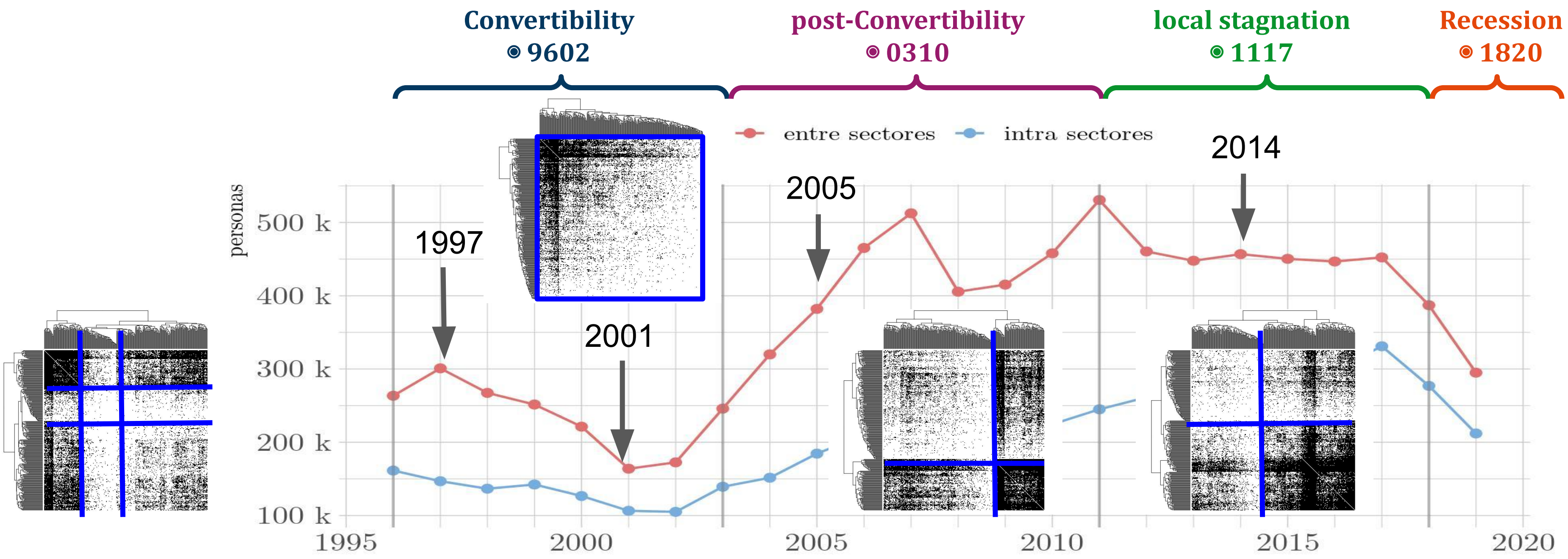
- Many phenomena taking place in networks are based on diffusion processes
- Betweenness shows that a node is the more central, the more often it is involved in these processes

Betweenness Centrality

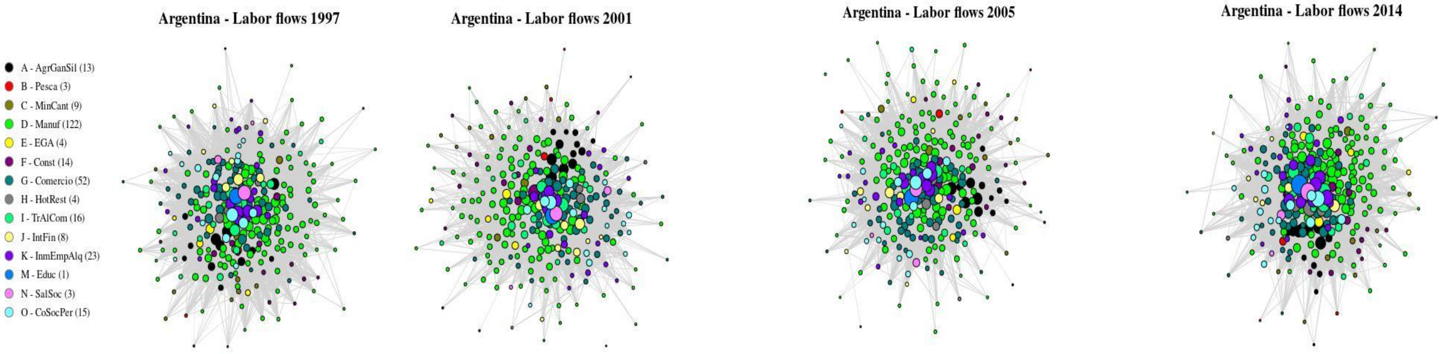
- The betweenness of the node *i* in an undirected network *G* is defined as:

$$BC(i) = \sum_i \sum_k \frac{\rho(j,i,k)}{\rho(j,k)}, \; i \neq j \neq k$$

- where $\rho(j,k)$ is the number of shortest paths connecting the node *j* to the node *k*
- $\rho(j,i,k)$ is the number of these shortest paths that pass through node *i* in the network.



Convertibility
◉ 9602
post-Convertibility
◉ 0310
local stagnation
◉ 1117
Recession
◉ 1820
entre sectores
intra sectores
personas
500 k
400 k
300 k
200 k
100 k
1995
2000
2005
2010
2015
2020
1997
2001
2005
2014

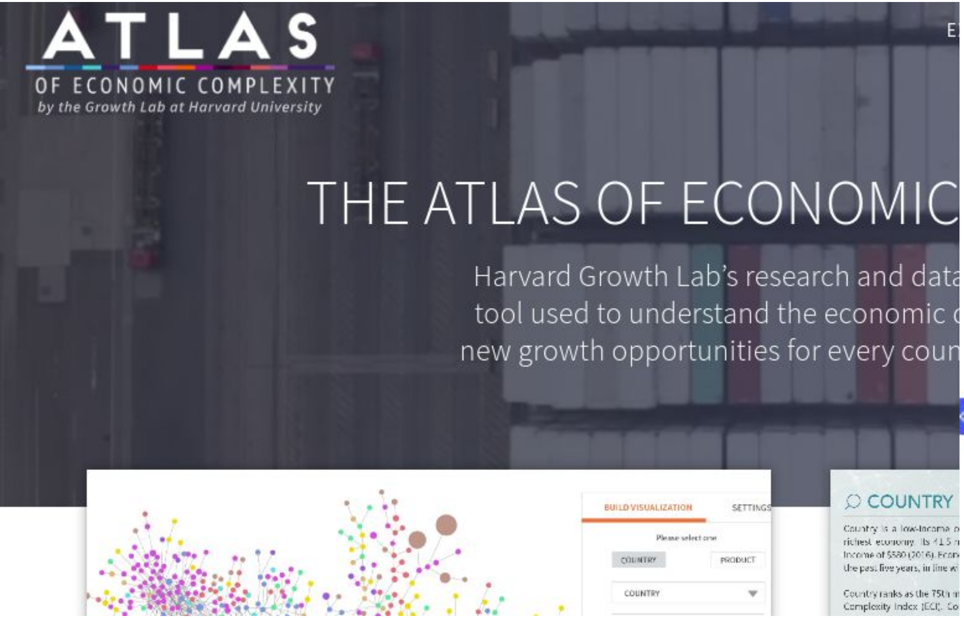


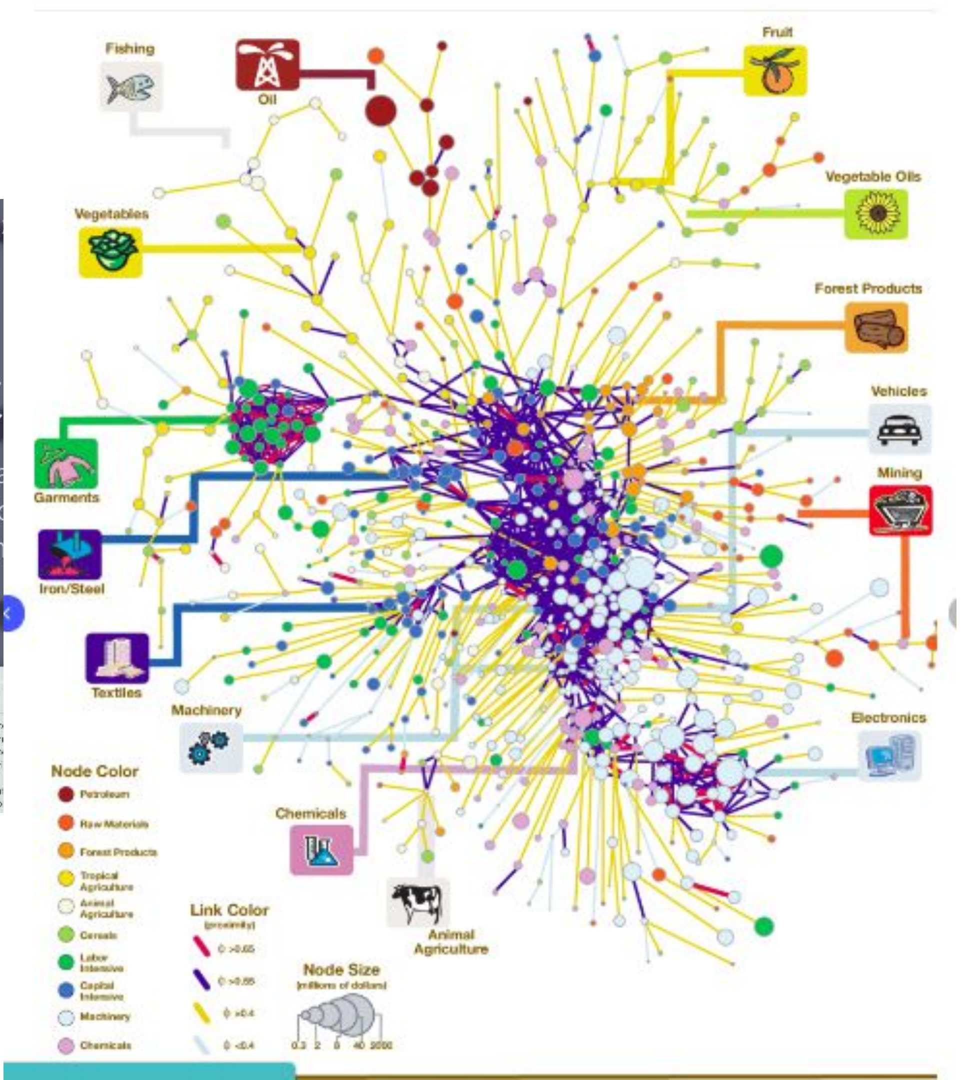
Argentina - Labor flows 1997
Argentina - Labor flows 2001
Argentina - Labor flows 2005
Argentina - Labor flows 2014
A - AgrGanSil (13)
B - Pesca (3)
C - MinCant (9)
D - Manuf (122)
E - EGA (4)
F - Const (14)
G - Comercio (52)
H - HotRest (4)
I - TrAlCom (16)
J - IntFin (8)
K - InmEmpAlq (23)
M - Educ (1)
N - SalSoc (3)
O - CoSocPer (15)

What other network can we extract?

- Individuals change jobs
  - capabilities and knowledge exchange between industries
  - skill-relatedness: describe how strongly the industries are connected through labor flows (reflect similarities in the industries'; human capital requirements)

Industry Space vs Product Space





Escuela Giambiagi XXIV, 12 al 14 de Octubre. Buenos Aires, Argentina.

Industry Space analogy to Product Space

MAPPING STRATIFICATION: THE INDUSTRY-OCCUPATION SPACE REVEALS THE NETWORK STRUCTURE OF INEQUALITY

Dominik Hartmann
University of Hohenheim/Fraunhofer IMW Leipzig/University of São Paulo

Cristian Jara-Figueroa
MIT Media Lab Cambridge

Mary Kaltenberg
Maastricht University/Brandeis University

Paulo Gala
Fundação Getulio Vargas

Research Area INEPA

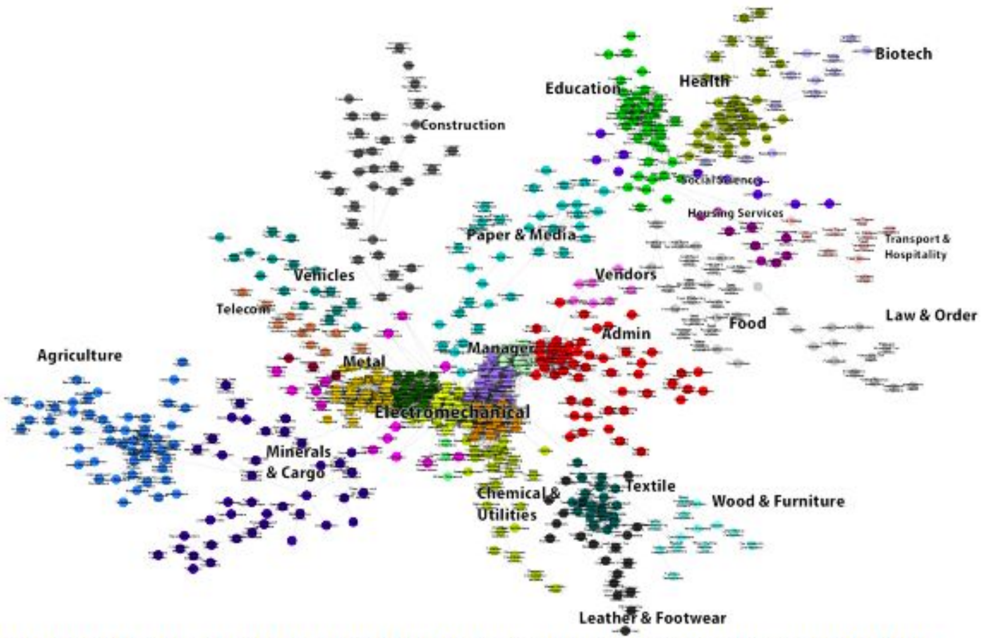
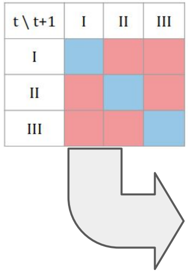


Figure 1. The Brazilian Industry-Occupation Space. Each node presents one of 600 CBO occupations and is colored according to 28 network communities identified by a combination of Peixoto (2014) and Louvain community detection algorithms. Links between the nodes depict to which extent 600 occupations co-appear in 558 different types of industries.

1st reduction option: Industry Space

Skill-Relatedness indicator (SR)

- ➢ There are mobility costs associated with specificities (firm, industry, task) in the accumulation of worker human capital
- ➢ Labor flows reveal degree to which human capital moves between industries (Neffke *et al.* (2017))
- ➢ **SR** Indicator shows divergence between observed and expected flows (similar to an χ^2 independence test in contingency tables)



Observed Flows

$$\begin{array}{cccc|c} f_{11} & f_{12} & \dots & f_{1n} & F_{1.} \\ f_{21} & f_{22} & \dots & f_{2n} & F_{2.} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{n1} & f_{n2} & \dots & f_{nn} & F_{n.} \\ \hline F_{.1} & F_{.2} & \dots & F_{.n} & F_{..} \end{array}$$

Expected values (in cell)

$$f^e_{ij} = \frac{F_{i.} \, . \, F_{.j}}{F_{..}}$$

SR + normalization

$$R_{ij} = \frac{\text{observed}}{\text{expected}} = \frac{f_{ij}}{f^e_{ij}}$$

$$\hat{R}_{ij} = \frac{R_{ij} - 1}{R_{ij} + 1}$$

$\hat{R}$ matrix

$$\begin{bmatrix} \hat{R}_{11} & \hat{R}_{12} & \dots & \hat{R}_{1n} \\ \hat{R}_{21} & \hat{R}_{22} & \dots & \hat{R}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{n1} & \hat{R}_{2n} & \dots & \hat{R}_{nn} \end{bmatrix}$$

Skill-relatedness

4.3. The skill-relatedness structure of labor-flow matrices

So far, we have documented patterns in raw labor flows. However, the size of labor flows will depend on the sizes and flow rates (i.e., the fraction of employees switching jobs) of the industries involved. To isolate the structure underlying inter-industry labor flows, we calculate the ratio between the observed volume of labor flows, and the one that would be expected from industries flow rates. If workers switched industries with probabilities proportional to the total outflow of the industry of origin, $F^s_{i.}$, and the total inflow into the destination industry, $F^s_{.j}$, the expected labor flow between i and j is given by $\widehat{F}^s_{ij} = \frac{F^s_{i.}F^s_{.j}}{F^s_{..}}$ and the ratio of observed to expected flows by[21]:

$$R^s_{ij} = \frac{F^s_{ij}F^s_{..}}{F^s_{i.}F^s_{.j}} \qquad (1)$$

Industry Space (cont.)

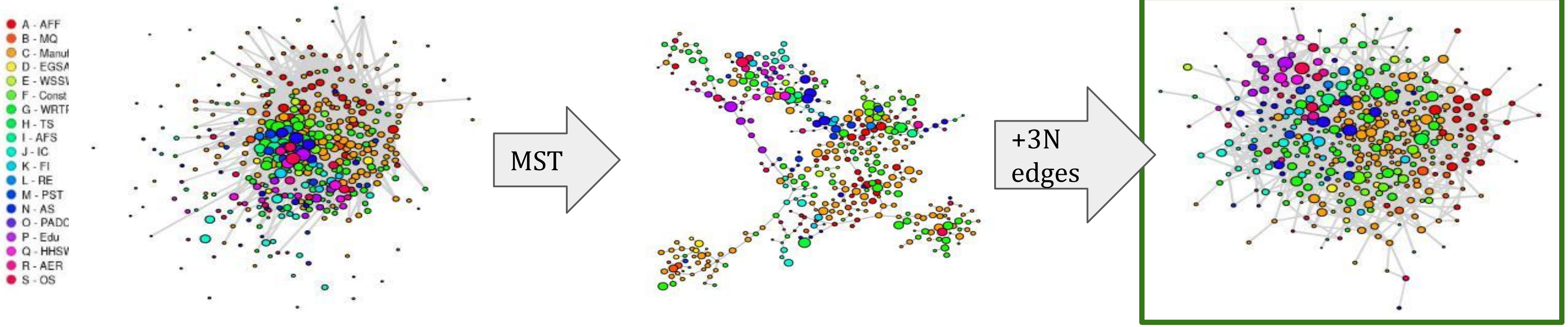
Labor flows

t \ t+1	I	II	III
I			
II			
III			

Indicator SR

$$\begin{matrix} \hat{R}_{11} & \hat{R}_{12} & \dots & \hat{R}_{1n} \\ \hat{R}_{21} & \hat{R}_{22} & \dots & \hat{R}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{R}_{n1} & \hat{R}_{2n} & \dots & \hat{R}_{nn} \end{matrix}$$

- A network reduction technique to obtain industry space:
  - SR indicator matrix symmetrization (undirected graph)
  - Aplicación del árbol de expansión máxima (MST)
  - Link addition: MST + 3N strongest SR links



De Raco, S. & Semeshenko, 2019. Labor Mobility in Argentina. JDGames, doi: 10.3934/jdg.2019008

Core-Periphery decomposition

We can decompose the network to reveal its core–periphery structure. This is accomplished by iteratively filtering out shells of low-degree nodes and focusing on the remaining, denser and denser cores.

The degree of each node can be used to separate a network into distinct portions, called shells, based on their position in the core–periphery structure of the network.

Low-degree outer shells correspond to the periphery.

As they are removed, or peeled away, what remains is a denser and denser inner subnetwork, the core.

Core-Periphery decomposition

Formally, the k-core decomposition algorithm starts by setting $k=0$. Then it proceeds iteratively. Each iteration corresponds to a value of k and consists of a few simple steps:

1. Recursively remove all nodes of degree k, until there are no more left.

2. The removed nodes make up the k-shell and the remaining nodes make up the ($k+1$)-core, because they all have degree $k+1$ or larger.

3. If there are no nodes left in the core, terminate; else, increment k for the next iteration.

Industry Space [cont.]

Argentina

k-core	Employment		Sectors		Degree			
	#	%	#	%	Min	Max	Avg	SDev
1	28,620	0.5	28	6.8	1	2	1	0.31
2	76,385	1.4	31	7.6	2	4	2	0.62
3	164,689	2.9	36	8.8	3	5	3	0.55
4	470,714	8.4	62	15.1	4	9	5	1.30
5	776,840	13.9	62	15.1	5	12	7	1.48
6	4,079,098	72.9	191	46.6	6	31	12	4.29
Total	**5,596,346**	**100.0**	**410**	**100.0**				

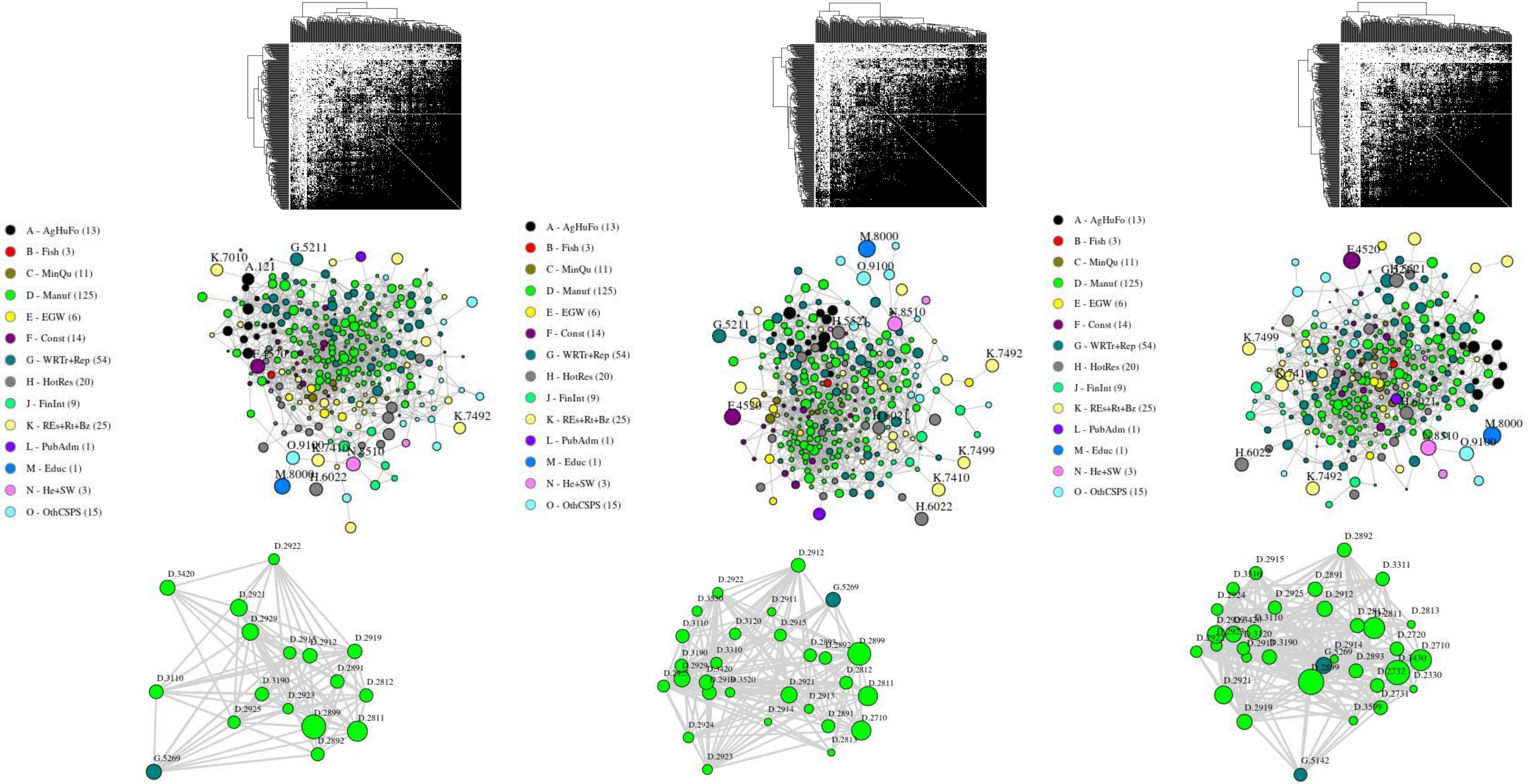
- ➢ Adds a higher level of significance to the observed structure
- ➢ Utility: diversification of productive activities

- k-core decomposition for Arg spanned in 6 nested subgraphs
- The 191 economic activities in the k6-core were connected with at least 6 and most 31 activities
- The size of this maximum core (47% of the nodes of the network) reflect a quite heterogeneous subset of activities, spanned over 17 grand divisions of classifications, and captured almost 73% of total private employment
- 59 sectors contained in the k1-core and k2-cores, which can be considered as periphery, presented a maximum of 2 and 4 degrees respectively, and a total of almost 2% of total employment

1996-2003 **2003-2011** **2011-2018**

Espacio industrial

Core de máxima conectividad



- Centro: Sectores de mayor conectividad son de industria manufacturera
- Periferia: Sectores débilmente conectados

2nd reduction option: Global Thresholds

- look for labor flow corridors
- dashboard visualization allows to explore the reduction of these labor networks through global thresholds on the minimum number of exchanges between sectors

HANDS-ON

1. Implement the proposed Standing Ovation (model 2) assuming spatial structure. You can try to implement some synthetic networks (random, small-worlds), o directly try some real world data networks from the repository sources offered in the course.
2. Implement PD game on the spatial structure. Change the rule of updating strategy: instead of imitation mechanism suggest another and see what happens with the global level of cooperation.
3. References on financial networks:
    a. Network models and financial stability
    b. Systemic risk in banking ecosystems