Brazilian Timber's Transportation Analysis

Felipe A. Moreno and Tiago da Silva
Network Science

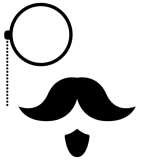
Table of contents

Introduction

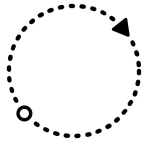
The objectives

What are the agents that take the most part on deforestation?

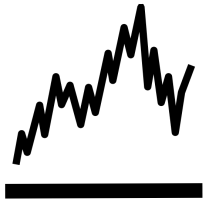
Can we identify an unusual asymmetry within the customers' market?



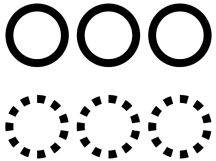
Are there oligopolies (or even monopolies) in this market? What are the main *players* therein?



How frequent, if at all, are internal and cyclic transactions in an enterprise? Are there isolated and perpetual markets?



Which of these aspects were modified through time? Is the market *stable*?



The same enterprise enjoys different labels; is this rectifiable? The correspondent nodes should be similar.

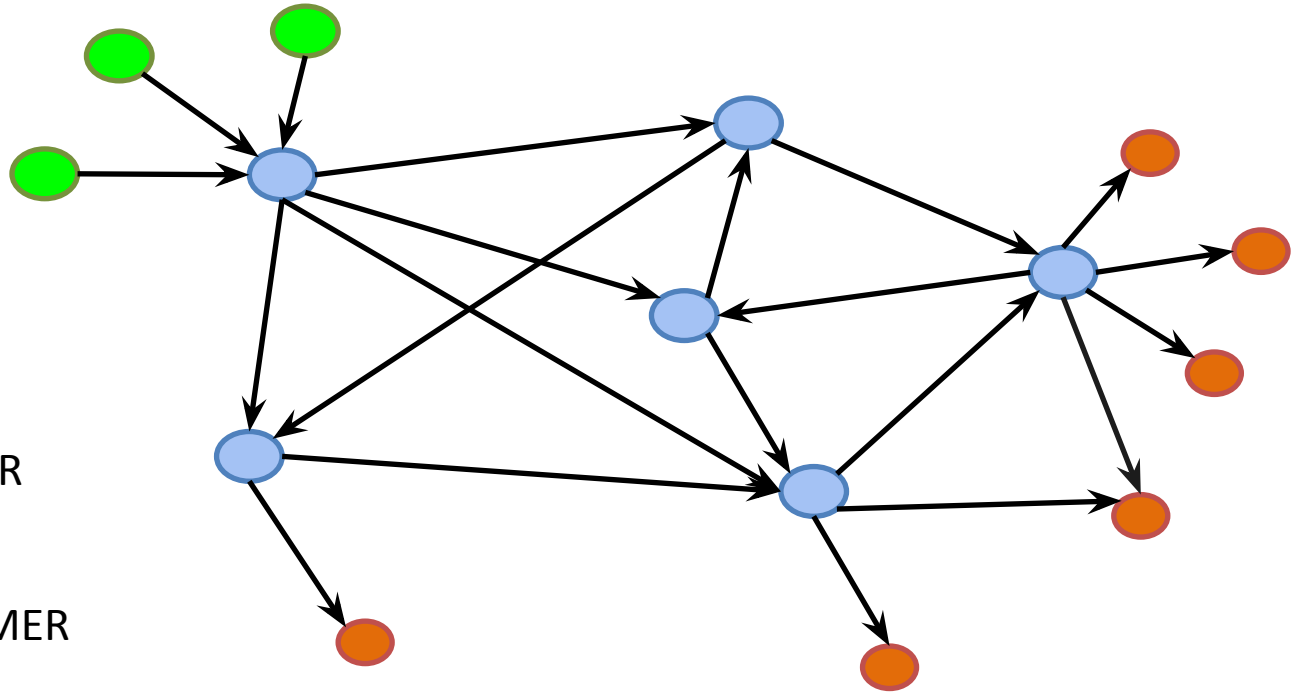
TIMBER MARKET: AN OVERVIEW

MANEJO (SUPPLIER)

PÁTIO (TRADER)

FINAL (CONSUMER)

Directed Graph Diagram



SUPPLIER

TRADER

CONSUMER

Data set

Structural description

- A relational database of transactions between (legal) persons

	OrigemID	Volume	DestinoID
0	PTO-04006133000126	0.2300	FINAL-03056752906
1	PTO-02746443000151	0.2433	FINAL-04171467950
2	PTO-02746443000151	0.2260	FINAL-19171535000153
3	PTO-22018942000130	0.5090	PTO-03776594000115
4	PTO-31446851000196	3.3413	PTO-18401124000144

- Contemplating geographical and temporal attributes

	UFOrigem	UFDestino	Ano
0	PR	PR	2019
1	PR	PR	2019
2	PR	PR	2019
3	PR	PR	2019
4	PR	PR	2019

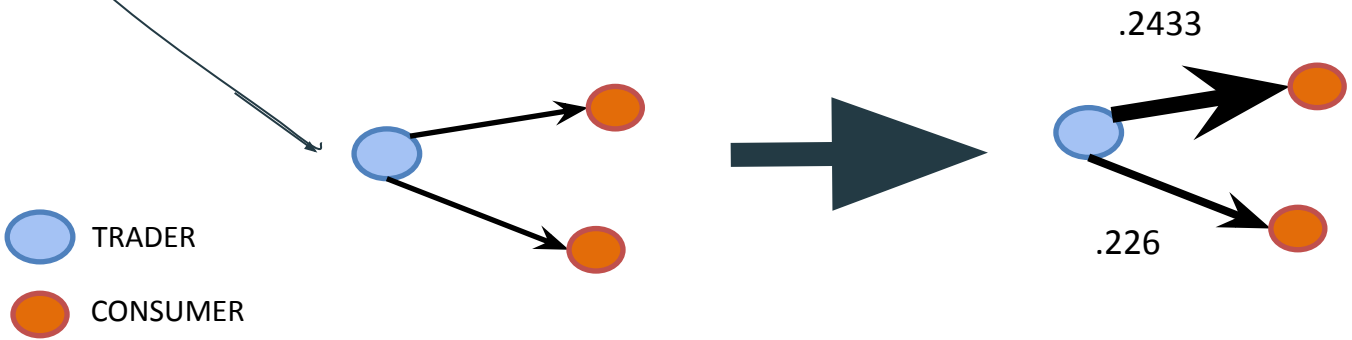
2017
2018
2019
2020

Network instantiation

- Each edge corresponds to an aggregated transaction

	OrigemID	Volume	DestinoID
0	PTO-04006133000126	0.2300	FINAL-03056752906
1	PTO-02746443000151	0.2433	FINAL-04171467950
2	PTO-02746443000151	0.2260	FINAL-19171535000153
3	PTO-22018942000130	0.5090	PTO-03776594000115
4	PTO-31446851000196	3.3413	PTO-18401124000144

- The edges' weights equal the (sum of the) traded volumes



Global (intertemporal) statistics

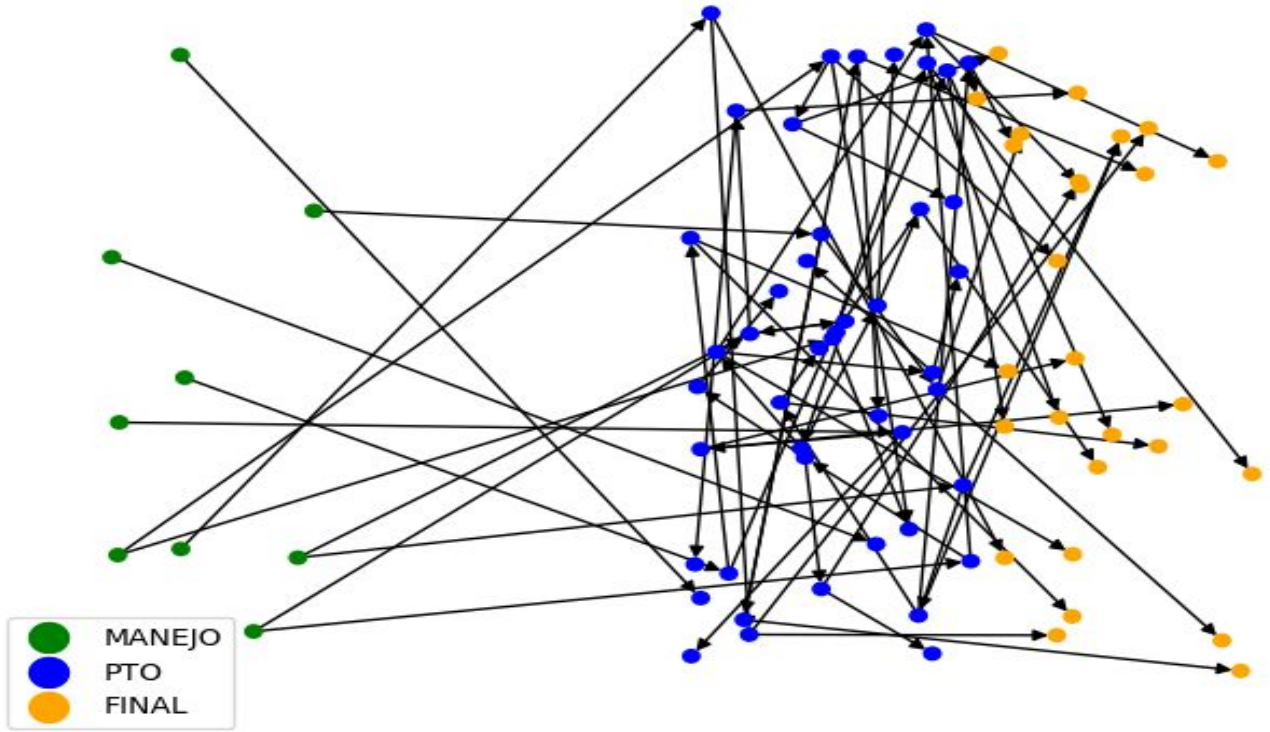
Year	$\bar{d}$	$\bar{d}_{\text{in}}$	Number of nodes	Number of links
2017	2.21	1.11	286,857	317,627
2018	2.21	1.11	286,655	317,596
2019	2.21	1.10	306,974	339,191
2020	2.17	1.09	317,791	345,616

Global (intertemporal) statistics

Year	$\bar{d}$	$\bar{d}_{\text{in}}$	Number of nodes	Number of links
2017	2.21	1.11	286,857	317,627
2018	2.21	1.11	286,655	317,596
2019	2.21	1.10	306,974	339,191
2020	2.17	1.09	317,791	345,616

VERY LARGE!

A glimpse on the network's topology: Few samples



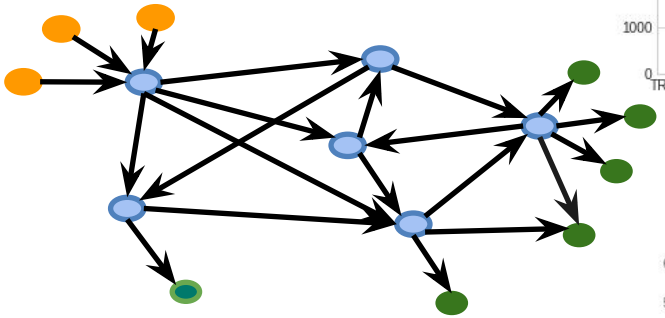
Choosing a subgraph

- We choose a subgraph in which each node executed more than 12 transactions.

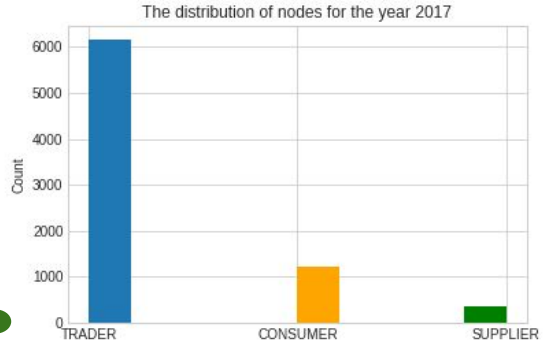
Year	Average degree	Number of nodes	Number of links	Giant component
2017	6.49	7736	25118	7686
2018	6.33	8026	25434	7864
2019	6.30	8421	26553	8149
2020	5.66	7757	21969	7528

Thus, the subsequent assertions concern around 3% of the network.

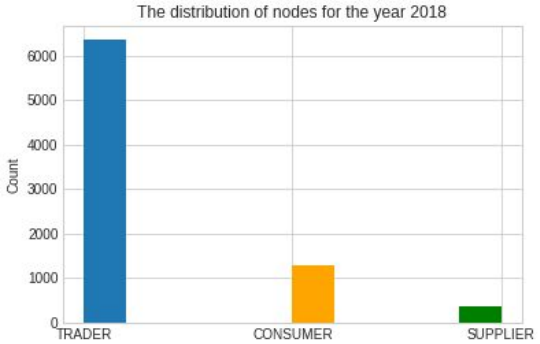
The variability within each subgraph



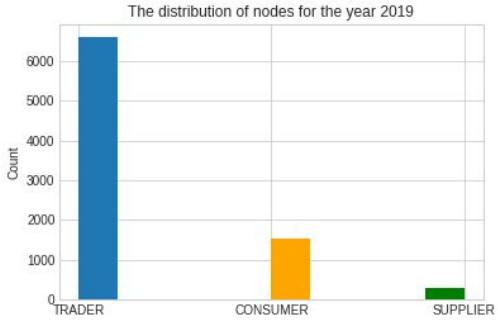
The complexity of the network is effectively contained in the subgraph induced by the traders!



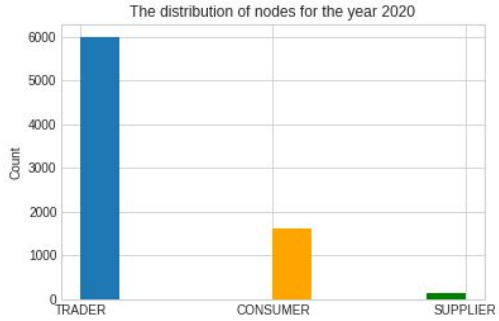
2017



2018

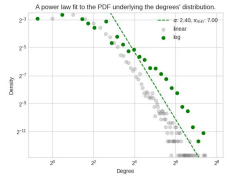


2019



2020

The objectives

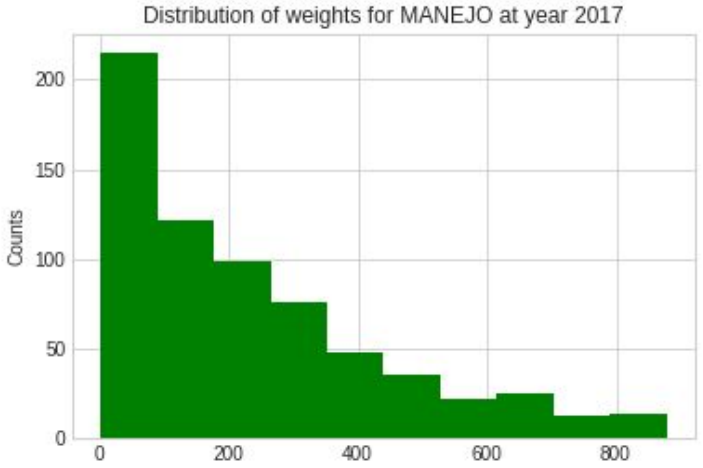
Structural objectives	Epistemic objectives
• What is the underlying probabilistic behavior of the network?  • And how is it different from a randomly sampled topology?	• These objectives concern the analysis of the network as a central planner. • Suppose we wish to introduce policies to modify the market's equlibrium. *The most important players.* *The unexpected irregularities.*

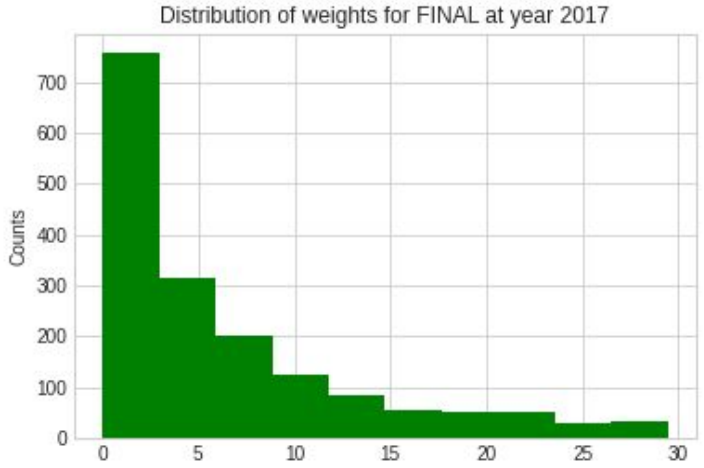
Epistemic analysis

A question: how are the weights – the traded volumes – distributed within the consumers and the suppliers?

Is the market homogeneous?

(different scales!)





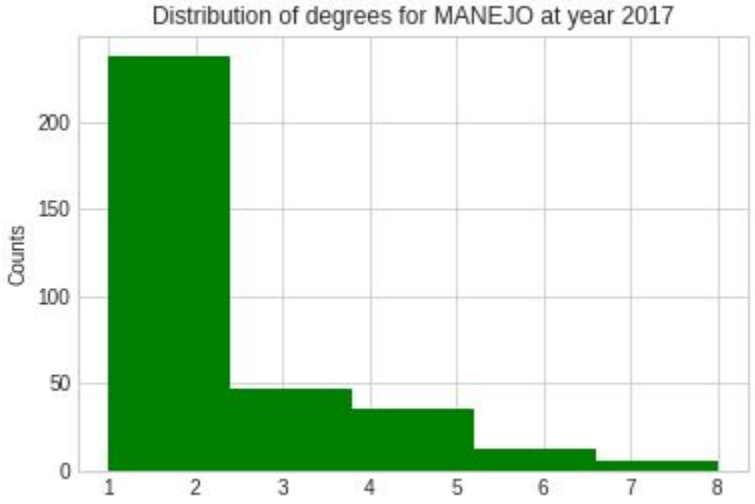
The market is beyond any homogeneity; and this is preserved through time.

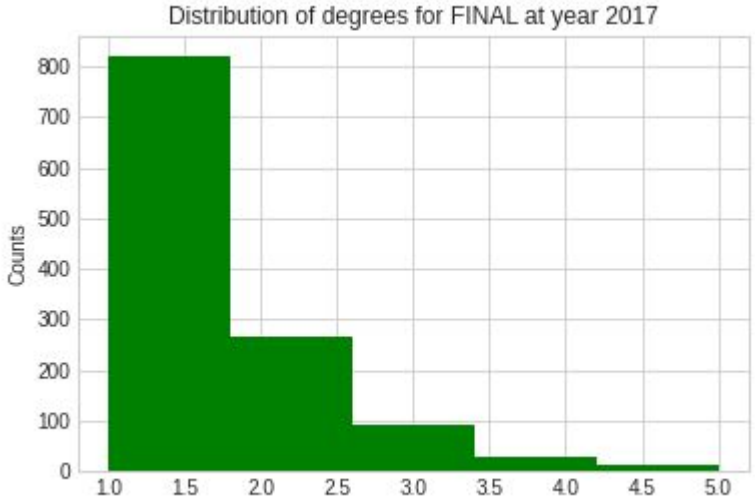
Epistemic analysis

A question: how are the weights – the traded volumes – distributed within the consumers and the suppliers?

Is the market homogeneous?

Same phenomenon for the degrees!





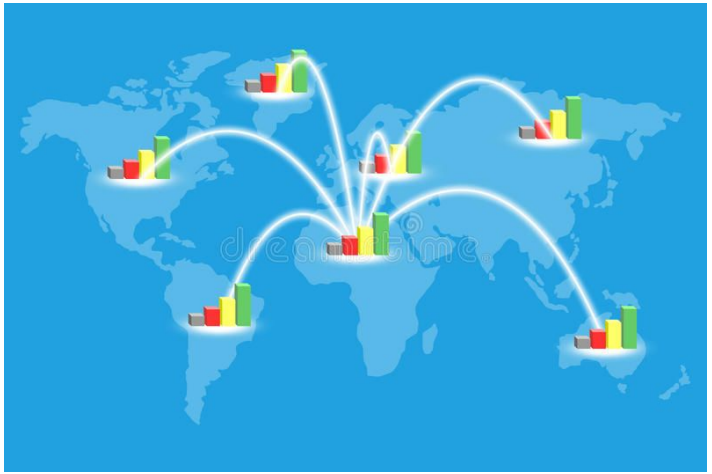
Epistemic analysis

A question: how are the weights – the traded volumes – distributed within the consumers and the suppliers?

Is the market homogeneous?

It is not!

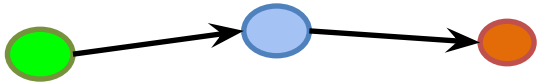
This is consistent with the expansion of markets: the most important agents attract the newcomers, leading to the formation of highly concentrated hubs.



Epistemic analysis

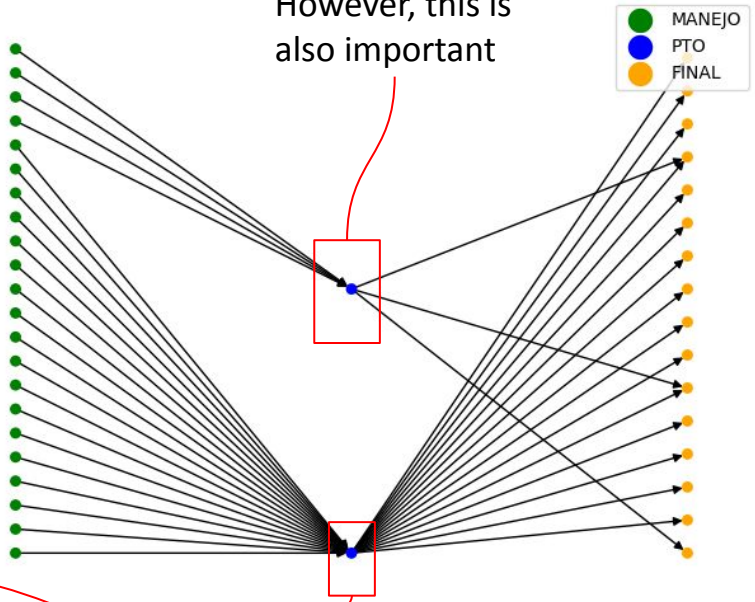
A question: The market is highly asymmetric; however, nodes globally introvert may be locally crucial.

Are there oligopolies?



Monopoly!
It controls
the market.

However, this is
also important



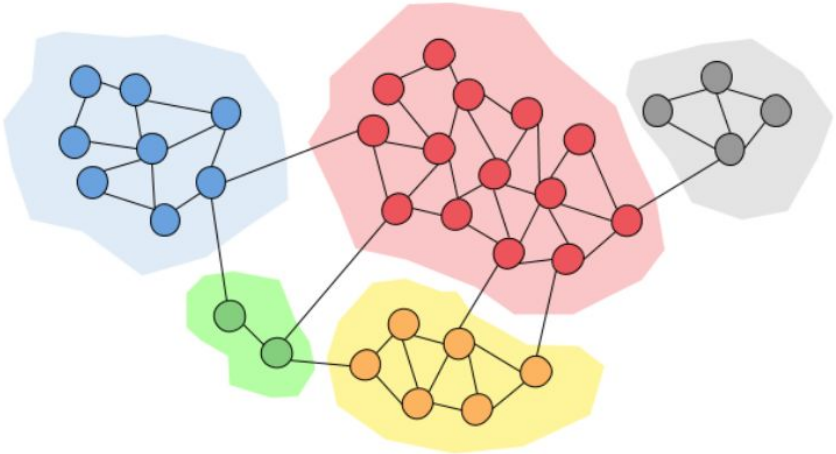
This node is
important globally

Epistemic analysis

A question: The market is highly asymmetric; however, nodes globally introvert may be locally crucial.

Are there oligopolies?

- We should identify locally dense groups – communities



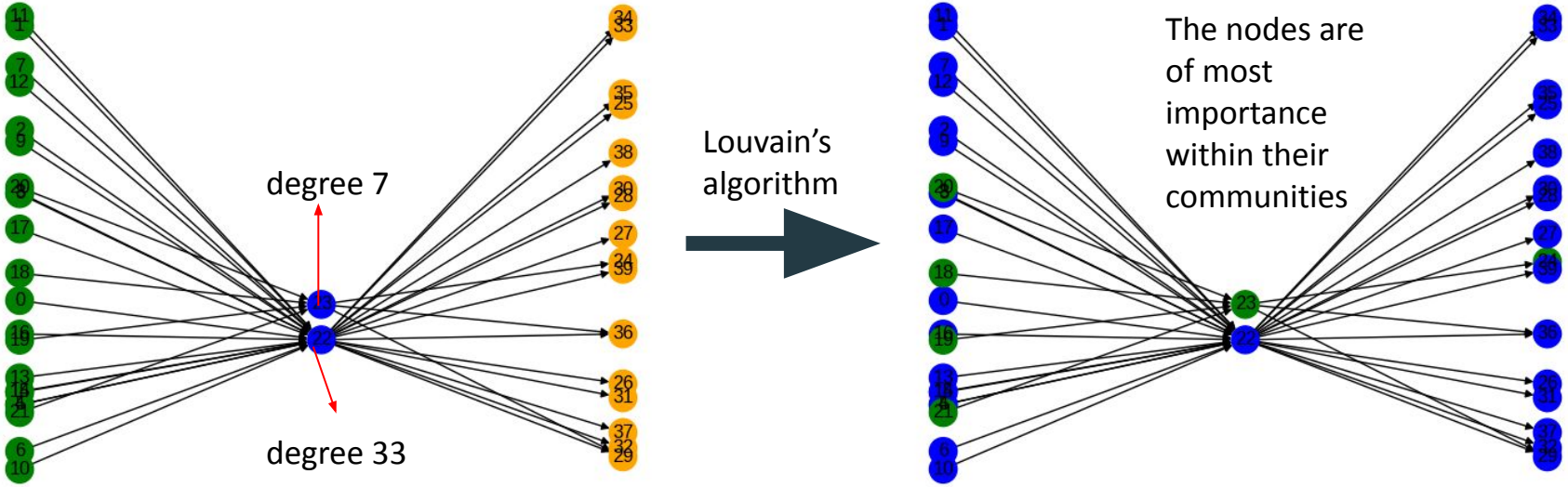
- Louvain's algorithm
  - Agglomerative (approximate) modularity maximization clustering

Scalable: it tackles large networks!

Epistemic analysis

A question: The market is highly asymmetric; however, nodes globally introvert may be locally crucial.

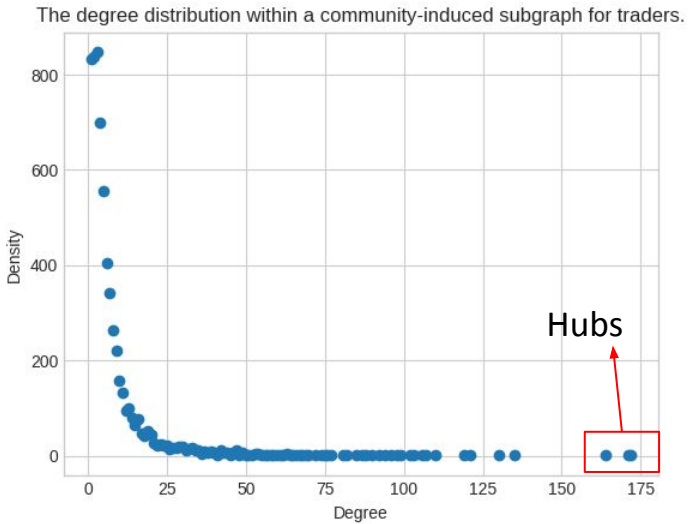
Are there oligopolies?

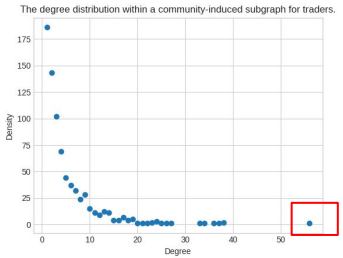


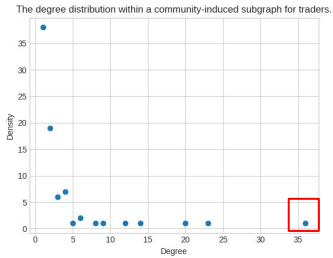
Epistemic analysis

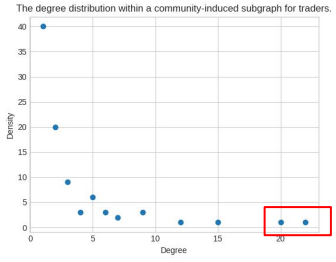
A question: The market is highly asymmetric; however, nodes globally introvert may be locally crucial.

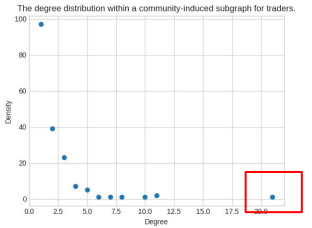
Are there oligopolies?









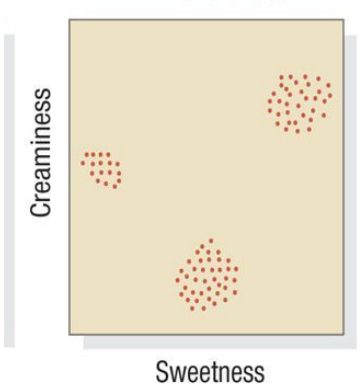


Epistemic analysis

A question: The market is highly asymmetric; however, nodes globally introvert may be locally crucial.

Are there oligopolies?

- Louvain's algorithm characterized a highly granular market:
  - for each year, around 800 communities were identified



A market with strongly clustered preferences

This is consistent with the verification that players generally stick with their suppliers and consumers

Which is emphatic in a large scale market; structural modifications are costly

Relevant agents in deforestation

Steps:

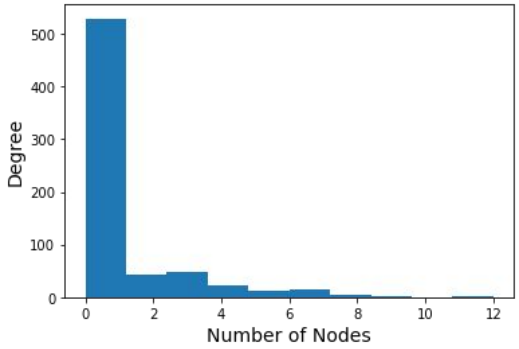
- First, calculate degree
- Second, saw quantity of volume were transported through time

Year	More active	Degree	Volume	Presence
2017	MANEJO-09446400953	12	2443.976	2017
2018	MANEJO-19989128847	15	20903.66	2018, 2019
2019	MANEJO-84639632000	13	3193.94	2019, 2020
2020	MANEJO-10823775000	7	3393.45	2020

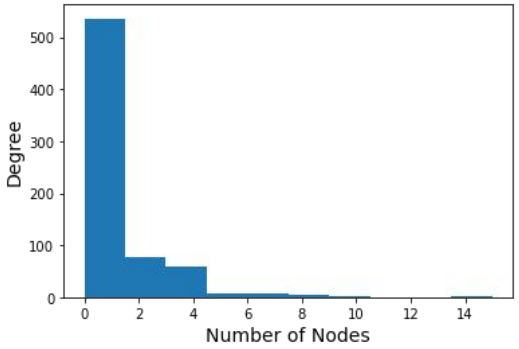
Year	More active	Degree	Volume	Presence
2017	FINAL-79080602003503	12	2443.976	2017
2018	FINAL-79080602003503	16	20903.66	2018, 2019
2019	FINAL-79080602003503	18	3193.94	2019, 2020
2020	FINAL-79080602003503	16	3393.45	2020

Nodes out-Degree

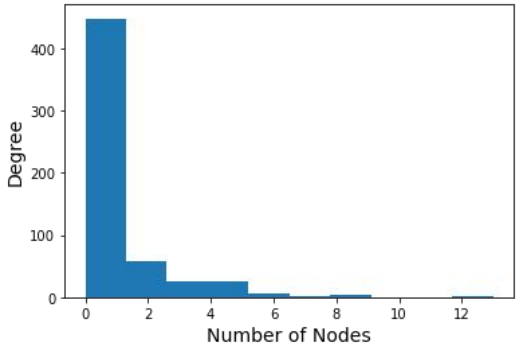
2017



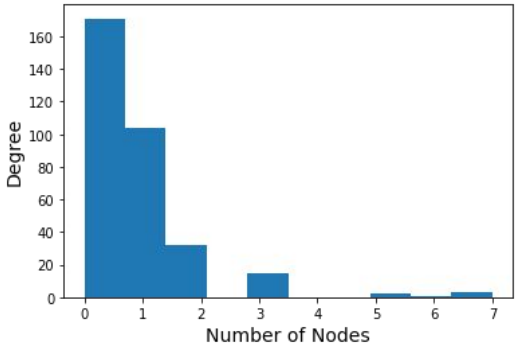
2018



2019

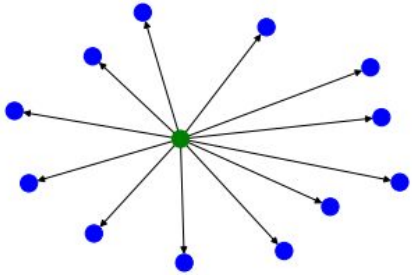


2020

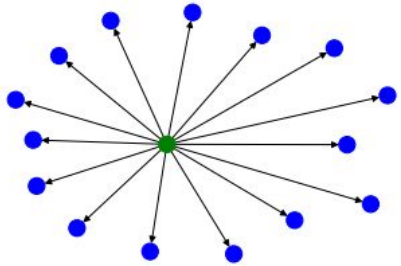


Behavior through time

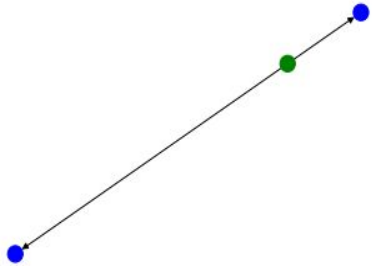
MANEJO-09446400953(-8.877,-62.702)-_2017



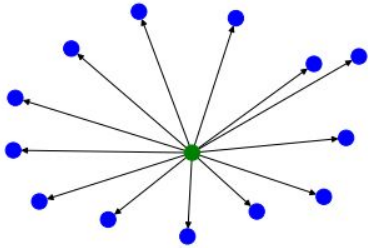
MANEJO-19989128847(-8.906,-66.04)-_2018



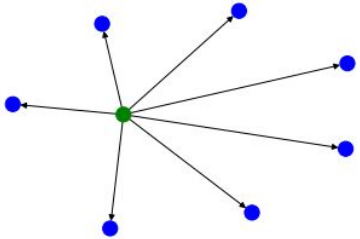
MANEJO-19989128847(-8.906,-66.04)-_2019



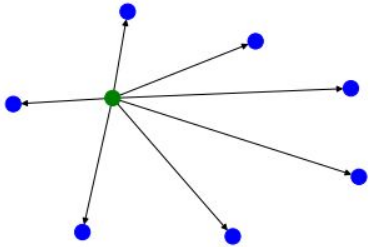
MANEJO-84639632000270(-8.876,-62.704)-_2019



MANEJO-84639632000270(-8.876,-62.704)-_2020

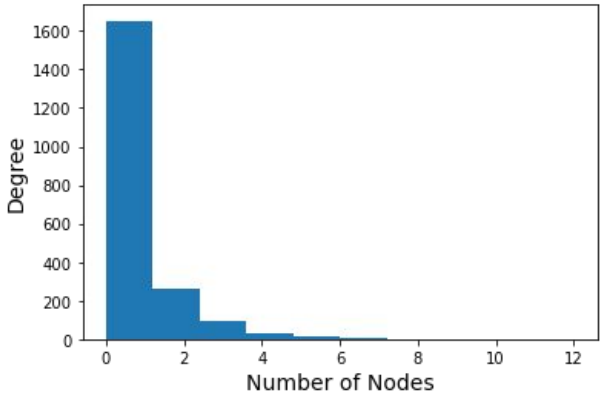


MANEJO-10823775000110(-9.397,-68.195)-_2020

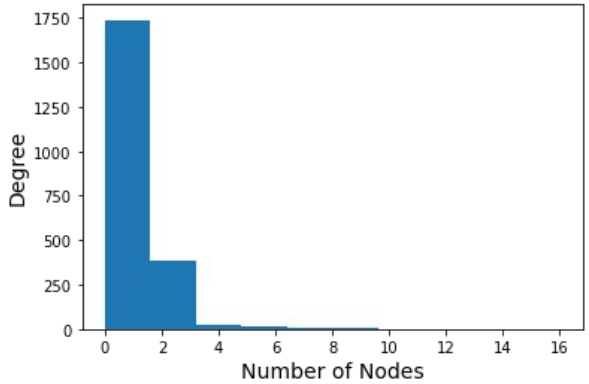


Nodes in-Degree

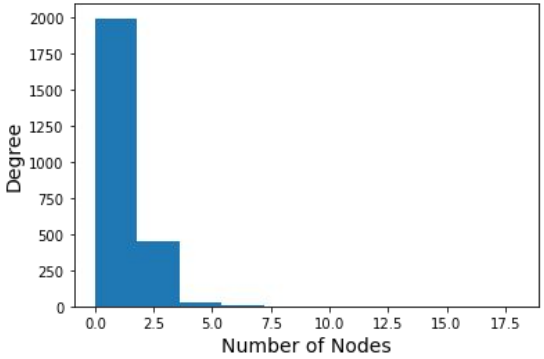
2017



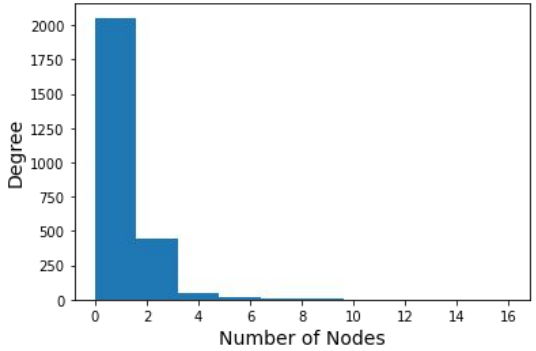
2018



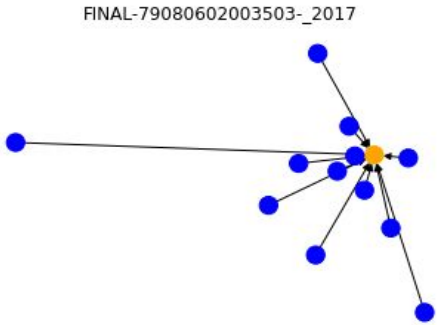
2019

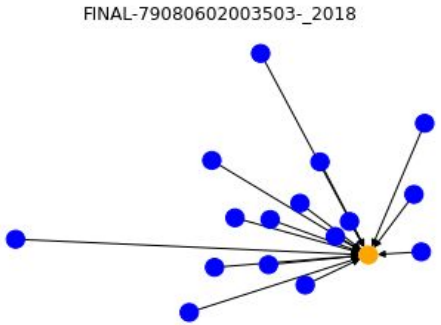


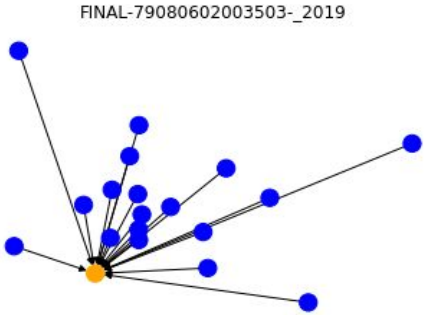
2020

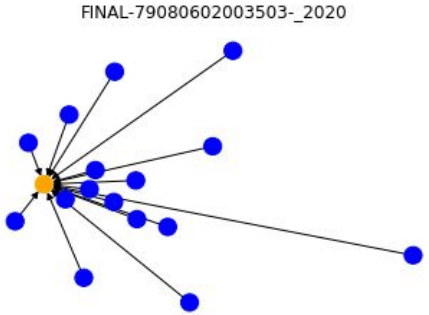


Behavior through time



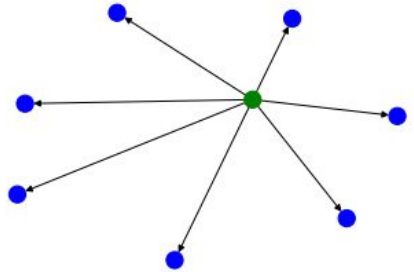




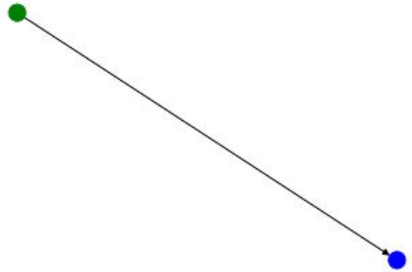


Special cases

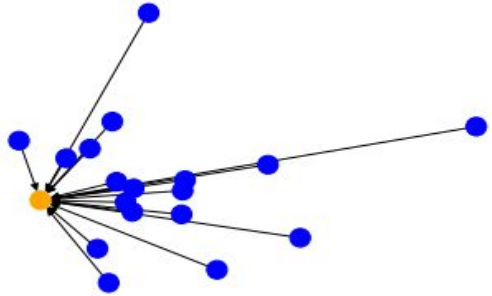
_2020 - MANEJO-10823775000110(-9.397,-68.195) - 7 - 3393.4497999999994



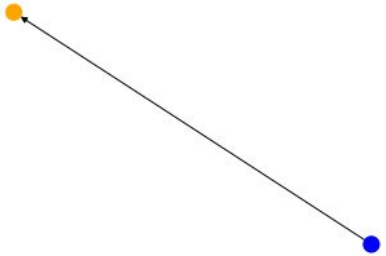
_2020 - MANEJO-26579253000155(-3.062,-55.528) - 1 - 5827.9346



_2019 - FINAL-79080602003503 - 18 - 21.012999999999998



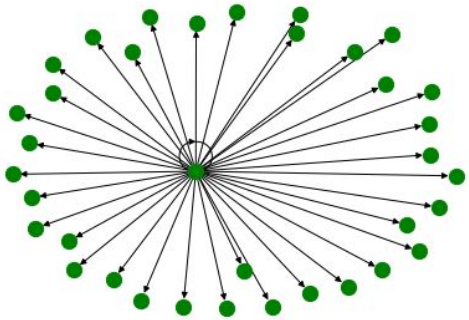
_2019 - FINAL-01902323408 - 1 - 745.7871



Cycle analysis

Steps:

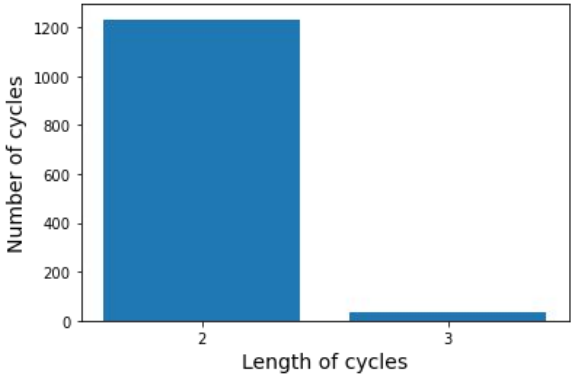
- First, identify self-cycles !!!
- Second, recursive DFS along all nodes.



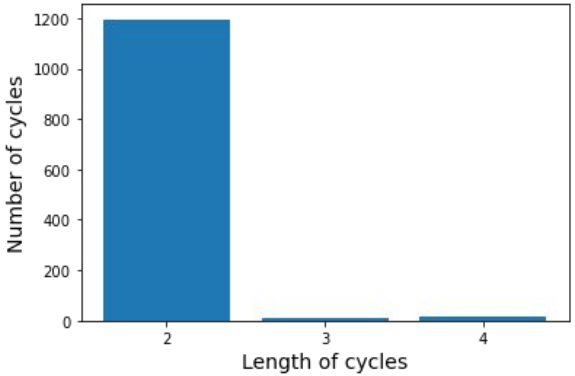
Year	Self-cycles	Cycles	Max length	How many?
2017	990	1269	3	36
2018	11,712	1224	4	17
2019	24,687	1248	4	5
2020	10,694	967	5	41

Cycle Frequency

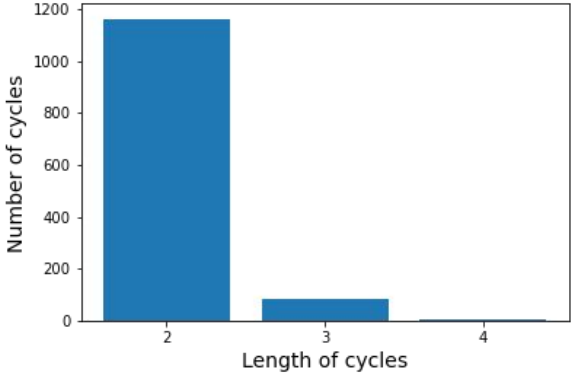
2017



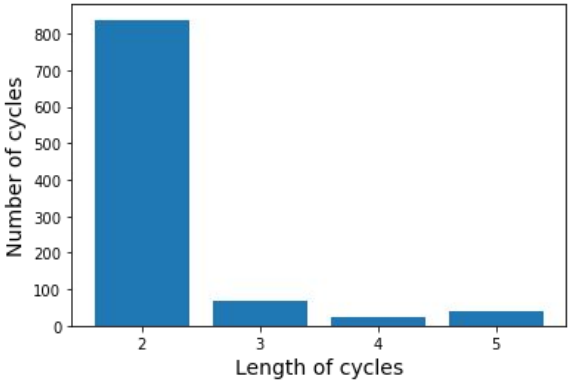
2018



2019



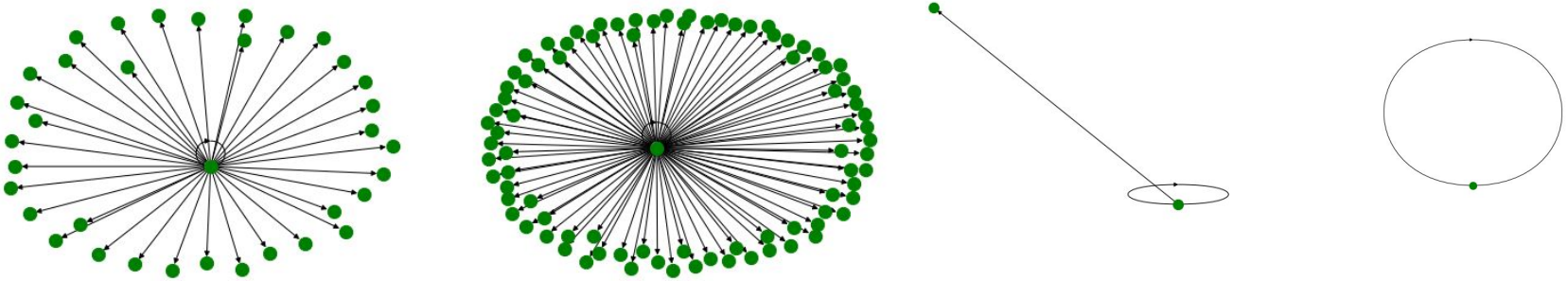
2020



Node sample

Node: "PTO-08242838000101" behavior analysis.

2017 2018 2019 2020

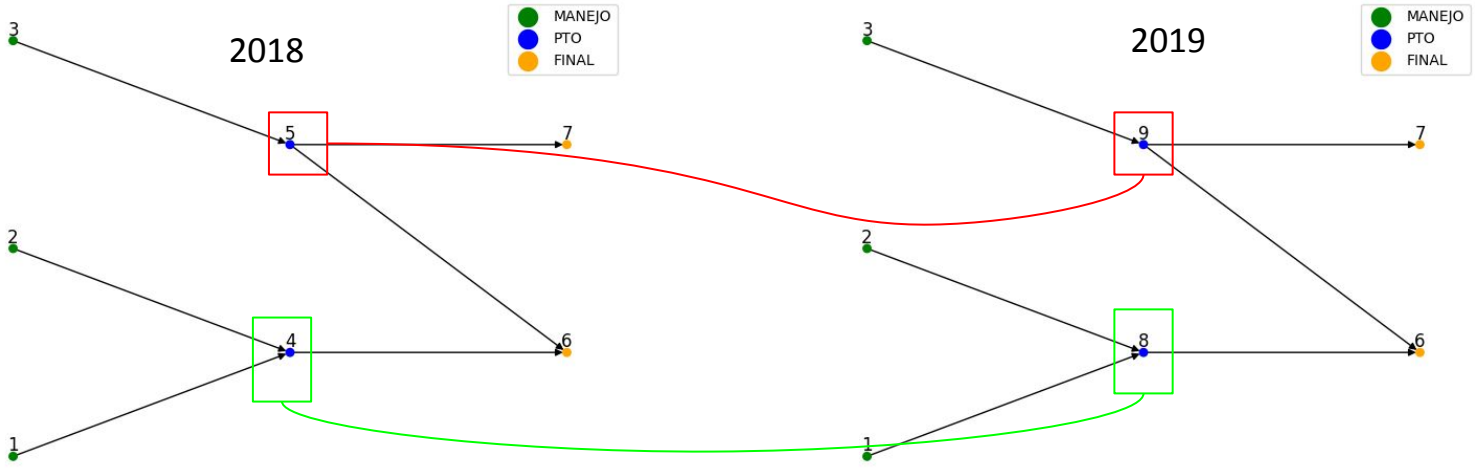


- Covid?
- Trocou os dados?
- Não foram cadastrados os transportes?

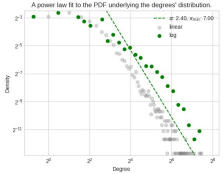
Epistemic analysis

A question: The behaviors we identified are consistently verified through time.
Can we measure the node-similarity between different temporal instantiations?

- Enterprises' names are volatile; moreover, the registration of data is subjected to human flaws



The objective was to incorporate an equivalent of regular similarity between graphs; however, it was an unsuccessful endeavor.

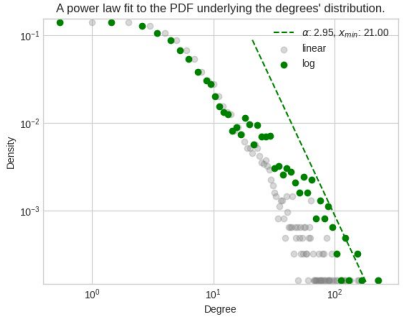


Structural analysis

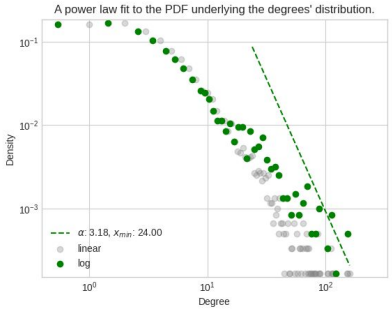
- This section is devoted to the verification of properties that, in spite of its (a priori) epistemological unsuitability, are appropriate for network analysis.

- **Scale-free property** of the networks induced by the traders

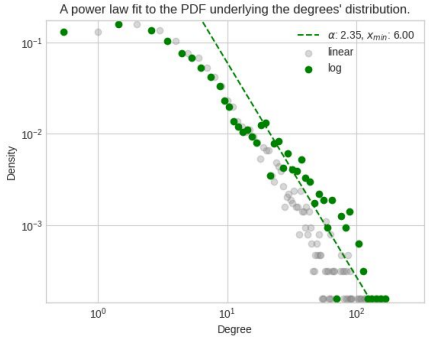
2017



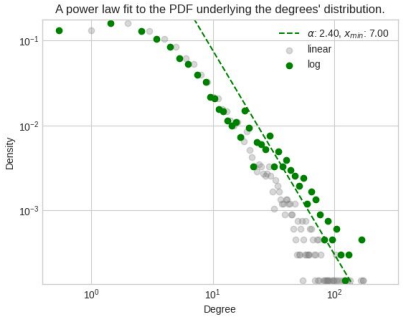
2020



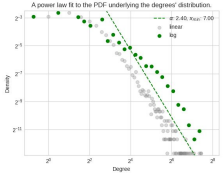
2018



2019



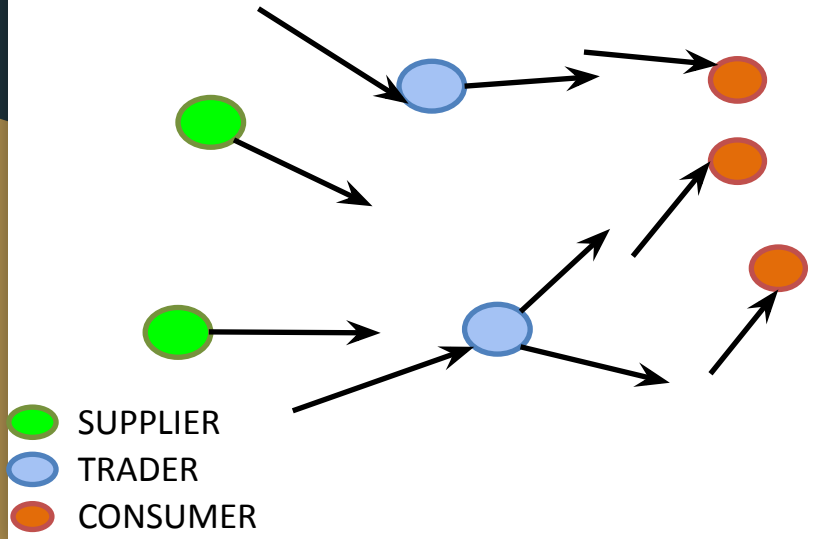
The network is generally appropriately captured by a power-law distribution; it is endowed with heavy-tailed properties.



Structural analysis

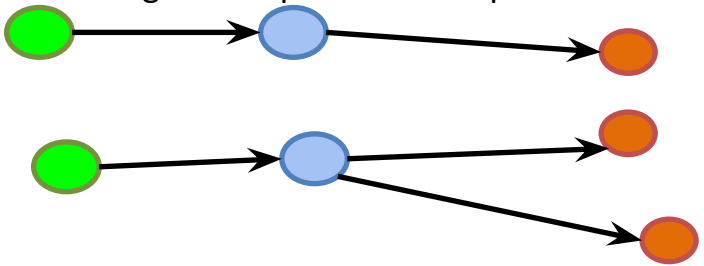
If we fix the degree distribution, what properties would be verified by chance, instead of by the particular network's topology?

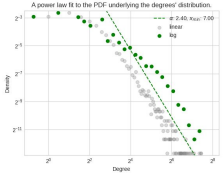
- A null model against which we should evaluate our expectations



A directed configuration model

- Stubs (in and out) are assigned uniformly at random
- As the network is large, the probability of arising self-loops is not disruptive



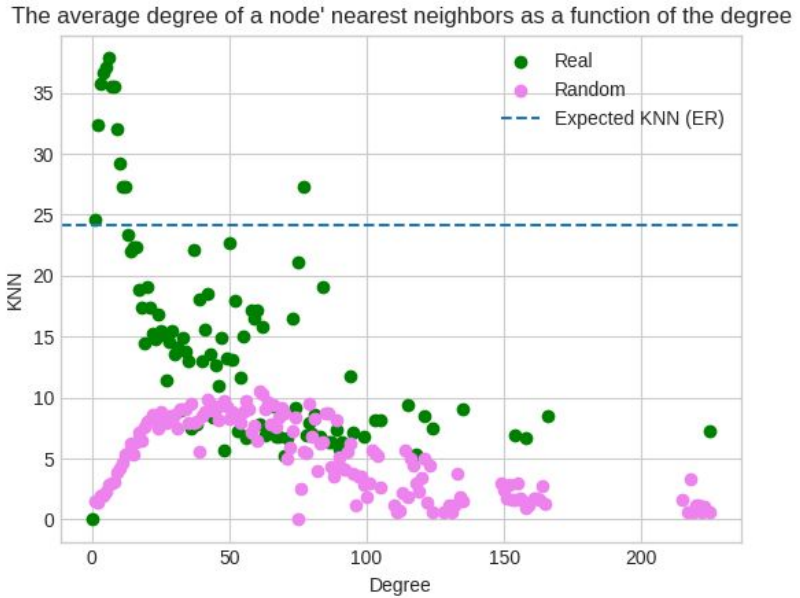


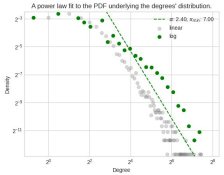
Structural analysis

- This section is devoted to the verification of properties that, in spite of its (a priori) epistemological unsuitability, are appropriate for network analysis.

Disassortative mixing!

- **Assortative mixing**: is there a preferential attachment between locally important nodes?

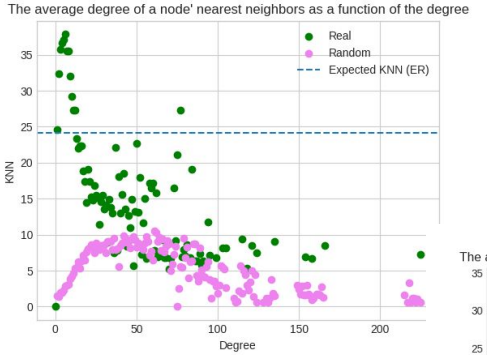




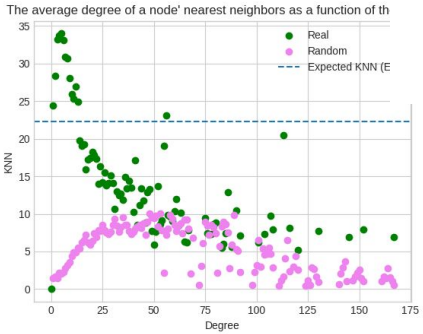
Structural analysis

This is reasonable as an attribute of growing markets!

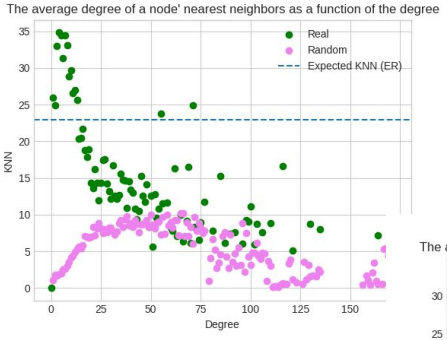
- This section is devoted to the verification of properties that, in spite of its (a priori) epistemological unsuitability, are appropriate for network analysis.



2017

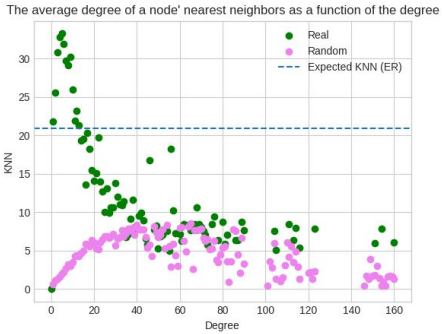


2018



2019

2020



The random model does exhibit a **structurally disassortative** behavior, though

Conclusions

- The timber market conforms a trade model with intermediaries
- It is endowed with systematic **heterogeneous** and **disassortative** properties
- Its topology is seemingly characterized by the **existence of submarkets** with often, voluminous trades

Limitations

- Nonetheless, we did inspect around 3% of each network.
- An extensive investigation would concomitantly require more computational power and time.
- Moreover, the data contain disrupted registers.
- They were not identifiable through the network's topology

	OrigemID	CPF_CNPJ_Rem	NomeOrigem
2614	PTO-396********	396********	JGOMC
3531	PTO-396********	396********	JGOMC
4270	PTO-396********	396********	JGOMC
70397	PTO-012********	012********	AOG

Questions?

Thanks!