

THESES OF THE DOCTORAL (PhD) DISSERTATION

GLOBAL TRADE DYNAMICS OF SOYBEANS AND SOYBEAN FLOUR: A NETWORK ANALYSIS

HENRIQUE FRIEDRICH DE OLIVEIRA

Supervisor:

Prof. Dr. József Péter Lengyel

Professor, University of Debrecen



UNIVERSITY OF DEBRECEN

Doctoral School of Management and Business

Debrecen

2026

1. INTRODUCTION OF THE TOPICS AND OBJECTIVES

The international trade of a country has been the focus of attention of many publications and can directly affect the lives of people, since a nation can benefit from selling products and those revenues may serve as a source of resources for its citizens (Bernard et al., 2007). Commerce has been present in human society since different civilisations began exchanging their surpluses with other groups. The importance of this topic is such that it has been described as one of the main pillars of the world's economic structure (Lovrić et al., 2018). Among agricultural commodities, soybeans (*Glycine max*) play a particularly prominent role in today's world economy. Soybeans were one of the agricultural products with the greatest production value in 2019 and rank as one of the most traded grains globally (FAO, 2021). In Brazil specifically, soybeans were the most exported product in 2020, with more than 28 billion USD in annual sales. China is the largest buyer of Brazilian soy, accounting for more than 70% of total Brazilian exports of this commodity (UN COMTRADE, 2021). The evolution of global soybean trade is strongly influenced by geopolitical dynamics, including the US–China trade war from 2018, the COVID-19 pandemic, and shifting regional trade agreements.

To deepen the understanding of the structure and dynamics of soybean trade, this dissertation applies Social Network Analysis (SNA) methodology, which enables the evaluation of structural features of a given network, identifying key characteristics, nodes, and their interconnections (Popp et al., 2018). The widespread availability of trade data combined with advances in computational methods has made network analysis an increasingly powerful tool for international trade research. Additionally, the Balassa Revealed Comparative Advantage (RCA) index is employed to identify countries with structural competitive advantages in soybean trade.

Research Questions

1. RQ1 – What is the global diffusion trade pattern of soybeans among countries?
2. RQ2 – How have the dynamics of trade players evolved during the last 20 years and what players have been dominating the market?
3. RQ3 – How do network analysis and the Balassa index contribute to our understanding of the international trade dynamics of soybeans?

Hypotheses

H1 – In the world international trade of soybeans, new economic countries are the exporters while developed countries are the importers of this commodity.

H2 – The exporters of soybeans have more influence within the network and are less susceptible to negative impacts caused by unforeseeable outbreaks, while importers, despite their significance, play a relatively weaker role in the network. Sub-hypotheses H2.1 through H2.6 test specific Spearman rank correlations between weighted degree measures and centrality indices (closeness, betweenness, indegree, outdegree).

H3 – The soybeans market is stable and challenging for new entrants to enter.

H4 – The international trade networks of soybeans and soy flour exhibit distinct structural patterns, reflecting different dynamics for each commodity within the global agricultural trade system. Sub-hypothesis H4.1 addresses distributional differences between layers; H4.2 addresses the correlation between country positions across the two layers.

H5 – The application of network analysis and the Balassa index in the global soybean trade network reveals certain countries with stronger network centrality and higher Balassa indices. H5.1 posits that countries with high weighted outdegree also exhibit revealed comparative advantage; H5.2 posits that distinct clusters of close trading partners can be jointly identified by both methods.

2. MATERIAL AND METHODS

The research relies on a comprehensive dataset of bilateral soybean and soybean flour/meal trade flows spanning 2000–2022. Trade data were sourced from the World Integrated Trade Solution (WITS) database maintained by the World Bank, using the Harmonized System codes HS-1201 (soya beans, whether or not broken) and HS-120810 (soya bean flour and meal). The dataset covers 231 countries and records annual bilateral trade values in USD thousands. To complement the empirical analysis, a systematic bibliographic review was conducted using the Scopus database, resulting in a corpus of 3,266 documents, subsequently refined to 382 peer-reviewed journal articles in relevant subject areas.

Network models were constructed and visualised using Gephi (version 0.10). A Python-based ETL pipeline was developed to extract, transform, and load trade data for subsequent network construction, statistical analysis, and figure generation (Matplotlib). Microsoft Office tools were used for tabular data management.

2.1. Complex Network Analysis

Complex network analysis was applied at both single-layer and multi-layer levels. The following node-level metrics were computed for each network period:

4. Node-degree (d_i): counts the number of trade partner connections of country i .
5. Node-strength (s_i): the sum of all bilateral trade values (weighted edges) for country i .
6. Degree centrality: distinguishes between outdegree (number of export destinations) and indegree (number of import sources).
7. Betweenness centrality: measures how frequently a node lies on the shortest path between other pairs of nodes – capturing intermediary or “bridge” roles.
8. Closeness centrality: the inverse of the average shortest path length from a node to all others, indicating how quickly a country can reach the entire network.
9. Eigenvector centrality: captures influence by weighting connections to highly connected neighbours more heavily than connections to peripheral nodes.
10. PageRank: analogous to Google’s algorithm; assigns scores based on incoming link volume and the importance of linking nodes.

Structural network properties computed include the clustering coefficient $C(p)$, the characteristic path length, network density, and modularity. For the multi-layer comparison of soybean and soy flour networks, the framework of Bródka et al. (2018) was applied across three approaches: (1) layer-by-layer comparison of average metric values and coefficients of variation; (2) layer distribution comparison using Jeffrey’s divergence; and (3) individual structure comparison using the Jaccard Index.

2.2. Statistical Methods

Before applying correlation analysis, the normality of all network metric distributions was tested using the Kolmogorov–Smirnov (K–S) test. Given that trade data frequently exhibits power-law distributions and significant skewness, departures from normality were anticipated. Results confirmed non-normality for all metrics ($p < 0.001$), which justified the exclusive use of non-parametric methods.

Spearman rank correlation (ρ) was applied to assess monotonic relationships between weighted degree measures and centrality indices, and between network metrics and the Balassa index. The Jaccard Index $J(W1, W2) = |W1 \cap W2| / |W1 \cup W2|$ was used to quantify structural overlap between the soybean and soy flour trade networks.

2.3. Balassa Revealed Comparative Advantage Index

The Balassa Index (RCA), originally introduced by Béla Balassa (1965), measures a country’s relative export specialisation in a specific product compared to the global average. It is calculated as: $RCA_{ij} = (X_{ij} / X_{it}) / (X_{nj} / X_{nt})$, where X_{ij} is the export value of product j by country i , X_{it} is the total exports of country i , X_{nj} is global exports of product j , and X_{nt} is total world exports. An RCA value greater than 1 indicates a revealed comparative advantage in that product.

Table 1: Summary of Analytical Methods Applied

Method	Purpose
Complex Network Analysis (Single-Layer)	Identify key nodes, trade concentration, centrality
Multi-Layer / Multiplex Analysis	Compare soybean vs. soy flour network structures
Kolmogorov–Smirnov Test	Normality diagnostics for all metric distributions

Spearman Rank Correlation (ρ)	Assess monotonic relationships between metrics
Jaccard Index	Quantify edge-overlap between commodity networks
Balassa RCA Index	Measure revealed comparative advantage in soybean export

Source: Own editing (2024)

3. MAIN FINDINGS OF THE DISSERTATION

3.1. RQ1 and H1: Global Soybean Trade Patterns

To test H1, the 231 countries participating in the soybean trade network were classified as net exporters or net importers based on the difference between their total outgoing and incoming trade flows over the 2003–2022 period. This produced 29 net exporters and 202 net importers. Countries were then classified by development status using the Human Development Index (HDI), with an HDI threshold of 0.8 applied. Countries with a mean HDI below this threshold were categorised as developing (“new economic”) countries.

Among the 29 net exporters, 67.9% are classified as developing countries, confirming that the export side of global soybean trade is predominantly composed of new economic nations. The three largest exporters by weighted outdegree are Brazil (379.62 billion USD), the United States (363.49 billion USD), and Argentina (58.10 billion USD). On the import side, China dominates overwhelmingly with a weighted indegree of 511.44 billion USD, surpassing the second-ranked country (Mexico, 35.63 billion USD) by a factor of 14. European countries (Netherlands, Spain, Germany) and Asian importers (Japan, Taiwan, Thailand) constitute the remainder of the top importers. H1 is therefore accepted.

Table 2: Top 5 Soybean Exporters and Importers by Weighted Degree (2003–2022)

Country (Export) / Weighted Outdegree (USD bn)	Country (Import) / Weighted Indegree (USD bn)
Brazil / 379.62	China / 511.44
United States / 363.49	Mexico / 35.63
Argentina / 58.10	Japan / 30.44
Paraguay / 31.20	Netherlands / 29.84
Canada / 28.69	Spain / 27.06

Source: Own editing based on UN Comtrade database (2024)

Network topology analysis reveals a total of 3,751 directed trade relationships among 231 nodes, yielding a network density of 0.071. The modularity score of 0.206 indicates the presence of four distinct trade clusters. The largest cluster (115 countries) includes the United States, China, Brazil, and Canada; the second cluster (108 countries) encompasses Germany, Italy, Thailand, and Ukraine. The average network distance of 2.28 indicates that

any two countries are on average connected through approximately two intermediary nodes, reflecting the small-world properties of global agricultural trade.

3.2. RQ2 and H2–H3: Structural Influence and Market Stability

To test H2, the eigenvector centrality distributions of exporter and importer groups were compared. Exporters exhibited a mean eigenvector centrality of 0.31, compared to 0.23 for importers. The K–S test confirmed a statistically significant difference between the two distributions ($p = 0.010$), supporting the conclusion that exporters occupy more structurally influential positions within the global soybean trade network. Furthermore, Spearman correlation analysis between the relevant degree metric and eigenvector centrality yielded $\rho = 0.98$ for importers (strong, indicating that having more import partners directly drives overall importance) versus $\rho = 0.28$ for exporters (weak, suggesting that exporters' network influence is decoupled from the mere count of their trading partners). H2 is accepted.

The sub-hypotheses were tested using Spearman correlations on the full 231-node dataset. All six sub-hypotheses were accepted based on statistically significant correlations ($p < 0.001$ throughout):

Table 3: Spearman Correlation Results for Export and Import Activity vs. Centrality Metrics

Metric Pair	Spearman ρ	Result
Weighted Outdegree – Closeness Centrality	0.909	H2.1 Accepted
Weighted Outdegree – Betweenness Centrality	0.846	H2.2 Accepted
Weighted Outdegree – Indegree	0.695	H2.3 Accepted
Weighted Indegree – Closeness Centrality	0.613	H2.4 Accepted
Weighted Indegree – Betweenness Centrality	0.673	H2.5 Accepted
Weighted Indegree – Weighted Outdegree	0.687	H2.6 Accepted

Source: Own editing (2024)

These results confirm that export activity (weighted outdegree) is more strongly associated with structural centrality than import activity (weighted indegree). Countries with high export volumes are not only better connected (higher closeness) but also serve as critical

intermediaries (higher betweenness), reinforcing their structural dominance in the global soybean supply chain.

To test H3, the trade data were divided into four consecutive five-year periods: Period 1 (2003–2007), Period 2 (2008–2012), Period 3 (2013–2017), and Period 4 (2018–2022). Spearman rank correlations of network metrics across consecutive periods were computed to assess how much the relative ranking of countries changed over time. Period 2 captures the global financial crisis; Period 4 includes the COVID-19 pandemic, US–China tariff disputes, and early impacts of the Russia–Ukraine conflict.

Table 4: Periodical Spearman Correlations for Soybean Network Metrics (HS-1201)

Metric	Periods 1–2 / 2–3 / 3–4	Periods 1–4
Degree Centrality	0.93 / 0.94 / 0.95	0.88
Betweenness Centrality	0.85 / 0.87 / 0.90	0.81
Eigenvector Centrality	0.88 / 0.90 / 0.92	0.84
Weighted Degree	0.90 / 0.88 / 0.93	0.87

Source: Own editing (2024)

All correlation coefficients for the soybean network (HS-1201) consistently exceed 0.80 across all metric comparisons and time intervals, indicating that the hierarchy of actors – in terms of connectivity, volume, and influence – remains remarkably stable throughout the 20-year period. Visual inspection of network graphs for each of the four periods confirms this: the same major nodes (United States, Brazil, China) dominate in every period, with the notable trend of Brazil increasing in dominance over the United States as the leading exporter. H3 is accepted.

3.3. RQ3 and H4–H5: Soybean and Soy Flour Networks and Comparative Advantage

The soy flour/meal network (HS-120810) involves 231 nodes and 2,483 directed trade edges, compared to 3,751 for the raw soybean network. This difference in edge count alone signals a less densely connected market. The soy flour network has a density of 0.047 (versus 0.071 for soybeans), a modularity of 0.326 (versus 0.206), and 8 clusters (versus 4), indicating a more fragmented and regionally segmented market. The leading exporter in the soy flour network is the United States (weighted outdegree: 15.25 billion USD), followed

by the Netherlands (1.61 billion USD) and Italy (0.80 billion USD) – a markedly different configuration from the raw soybean network dominated by South American producers.

Layer-by-layer comparison shows that the average degree is higher for the soybean network (32.48) compared to soy flour (21.50), and the average closeness centrality is also higher (0.31 vs. 0.26), confirming the soybean network is more tightly integrated. However, the soy flour network exhibits higher coefficients of variation for weighted outdegree (CV = 1085 vs. 875) and eigenvector centrality (CV = 97.73 vs. 87.35), indicating greater inequality in the distribution of influence across nodes in the processed product market. H4 is accepted.

Table 5: Jeffrey’s Divergence Between Soybean and Soy Flour Network Distributions

Metric	Jeffrey’s Divergence
Degree	0.862 (moderate)
Weighted Indegree	1.124 (high)
Weighted Outdegree	0.975 (moderate)
Betweenness Centrality	1.032 (high)
Closeness Centrality	0.755 (moderate)
Eigenvector Centrality	0.921 (moderate)

Source: Own editing (2024)

Jeffrey’s divergence values above 1.0 for weighted indegree and betweenness centrality indicate high dissimilarity in how nodes receive trade flows and how they function as intermediaries. Values between 0.5 and 1.0 for degree, weighted outdegree, closeness, and eigenvector centrality indicate moderate dissimilarity. These results support H4.1 (partially accepted), as differences are heterogeneous across metrics rather than uniformly significant.

The Jaccard Index between the soybean and soy flour trade networks was computed as 0.416, indicating that approximately 41.6% of unique directed trade relationships are shared between the two networks. This moderate overlap implies that while many of the same bilateral trade routes exist for both commodities, a substantial portion of trade relationships are commodity-specific. H4.2 is fully accepted, confirming that country positions in one network do not reliably predict positions in the other.

The Balassa RCA analysis covered 148 countries that exported soybeans at least once during 2003–2022. The top RCA countries are strongly concentrated in South America, reflecting the region’s natural agricultural endowments. The top 10 countries by Balassa index are presented in Table 6.

Table 6: Top 10 Countries by Balassa RCA Index for Soybeans (2003–2022)

Rank / Country	Balassa Index	Region
1. Paraguay	4,419.11	South America
2. Uruguay	2,657.09	South America
3. Brazil	1,909.58	South America
4. Argentina	710.38	South America
5. Bolivia	274.37	South America
6. Ethiopia	208.38	Africa
7. United States	203.86	North America
8. Malawi	169.33	Africa
9. Ukraine	161.71	Europe
10. Moldova	81.09	Europe

Source: Own editing based on UN Comtrade database (2024)

Spearman correlation analysis between the RCA index and network metrics yields a strong positive correlation with weighted outdegree ($\rho = 0.85$, $p < 0.001$), confirming H5.1: countries with high comparative advantage are also the most active exporters. Moderate correlations were found with closeness centrality ($\rho = 0.475$, $p < 0.001$) and degree ($\rho = 0.465$, $p < 0.001$). A weak but significant correlation was found with betweenness centrality ($\rho = 0.24$, $p = 0.003$). However, the correlations with eigenvector centrality ($\rho = 0.12$, $p = 0.146$) and PageRank ($\rho = 0.006$, $p = 0.93$) are not statistically significant. H5.2 is rejected, as the combined use of RCA and network analysis did not identify distinct, coherent trade clusters. H5 is therefore partially accepted.

Table 7: Summary of Hypothesis Outcomes

Hypothesis	Result
H1: Developing countries export; developed countries import	Accepted
H2: Exporters are more influential than importers	Accepted
H2.1: Weighted outdegree – Closeness ($\rho = 0.909$)	Accepted

H2.2: Weighted outdegree – Betweenness ($\rho = 0.846$)	Accepted
H2.3: Weighted outdegree – Indegree ($\rho = 0.695$)	Accepted
H2.4: Weighted indegree – Closeness ($\rho = 0.613$)	Accepted
H2.5: Weighted indegree – Betweenness ($\rho = 0.673$)	Accepted
H2.6: Weighted indegree – Weighted outdegree ($\rho = 0.687$)	Accepted
H3: Soybean market is stable; high barriers to entry	Accepted
H4: Soybean and soy flour networks are structurally distinct	Accepted
H4.1: Layer distributions differ significantly	Partially Accepted
H4.2: Country positions weakly correlated across networks	Accepted
H5: High centrality and high RCA jointly identifiable	Partially Accepted
H5.1: High weighted outdegree linked to RCA advantage	Accepted
H5.2: RCA + network analysis identify distinct clusters	Rejected

Source: Own editing (2024)

4. NEW AND NOVEL RESULTS OF THE DISSERTATION

Through the application of advanced network analysis techniques, this research has provided new insights into the complex and evolving dynamics of international agricultural trade. By building networks, exploring visualisation, and correlation methods, the study has illuminated previously unexplored aspects of the trade networks for soybeans trade. Additionally, by comparing soybeans and their subproducts it was possible to identify different structures and dynamics on both trades. Moreover, the present work highlights the distinctive structural features identified, the significant roles played by major exporters, and the implications of trade network stability.

1. Distinct Structural Differences Between Soybean and Soy Flour Trade Networks

This study provides the first systematic comparison of network structures between soybeans and their processed derivative (soy flour) within the same analytical framework. While both commodities originate from the same agricultural product, their trade networks exhibit fundamentally different characteristics in terms of trade concentration, hub structures, and connection patterns. This finding challenges the assumption that processed and raw agricultural commodities follow similar trade dynamics and demonstrates that value-added processing creates distinct network topologies with different vulnerabilities and opportunities for market participants.

2. Quantified Influence Patterns of Major Exporters in Soybean Networks

This research provides novel quantitative evidence of asymmetric influence patterns among major soybean exporters. Unlike previous studies that identified major players, this work demonstrates specific differences in how Brazil, the USA, and other exporters exercise network influence through different centrality measures. The findings reveal that export volume alone does not determine network influence, and that connectivity patterns create distinct strategic advantages for different exporters within the same commodity network.

3. Network Stability Despite Shifts in Major Player Roles

This research reveals an important paradox in the global soybean trade network: while the core structure of the network remains stable with consistently high barriers to entry for new participants, the relative positions and roles of established major players have shifted

significantly over the study period. Notably, Brazil emerged as the dominant exporter, surpassing the United States, while China transformed from a minor exporter to the world's largest importer, fundamentally reshaping trade flow patterns. These findings demonstrate that network stability does not imply static player positions; rather, the soybean trade exhibits structural persistence at the network level combined with dynamic repositioning among existing major actors. This suggests that while breaking into the network as a new significant player remains difficult, established participants can undergo substantial transformations in their roles, influence, and market share within the stable network framework.

4. Methodological Insights on Network Analysis and Balassa Index Integration

The combined application of network analysis and the Balassa index revealed important methodological insights about the limits of integrating these analytical approaches. While high weighted outdegree was associated with revealed comparative advantage (supporting H5.1), the expected identification of distinct trade clusters through this combined approach was not achieved (rejecting H5.2). This finding contributes to methodological knowledge by demonstrating that network centrality and comparative advantage measures do not always align, suggesting these tools capture different dimensions of trade competitiveness that require separate analytical treatment.

5. PRACTICAL APPLICABILITY OF THE RESULTS

The findings of this research have direct practical implications for policymakers and stakeholders in the global agricultural sector. The identification of key network nodes with disproportionate influence suggests that targeted policy interventions at critical trade junctures can generate significant ripple effects across the entire network. Specifically, policymakers should prioritise investment in ports, logistics, and trade infrastructure in countries that serve as network intermediaries (high betweenness centrality), as improvements here enhance efficiency across the entire trade network.

Given the high structural stability of the soybean trade network, policymakers should consider targeting support for emerging exporters and the diversification of supply chains for major importers. Countries with agricultural capacity but limited market access could benefit from trade facilitation initiatives and capacity-building programmes to enhance their competitiveness. Importing nations should strategically diversify their supplier base to reduce vulnerability and enhance resilience against geopolitical disruptions. The identification of distinct structural patterns in the soybean and soy flour networks also highlights the importance of value-added strategies, whereby countries that primarily export raw soybeans can enhance their economic resilience by developing domestic processing capacities.

6. LIST OF PUBLICATIONS RELATED TO THE DISSERTATION

Articles, studies

1. FRIEDRICH DE OLIVEIRA, H.; PÉTER, LENGYEL. Global Soybean Trade: A Complex Network Analysis of Key Exporters and Importers (2003–2023). SEA: PRACTICAL APPLICATION OF SCIENCE 12 : 36 pp. 203–216. (2024).
2. LUIZ RODRIGUES, D.; FRIEDRICH DE OLIVEIRA, H. The Use Of Social Network As A Tool To Analyse International Trade: A Systematic Literature Review. NETWORK INTELLIGENCE STUDIES 10 : 19 pp. 25–34. (2022).
3. FRIEDRICH DE OLIVEIRA, H.; LENGYEL, PÉTER. TAXATION IMPACTS ON BRAZILIAN TRADE. JOURNAL OF ECOAGRITOURISM 17 : 43 pp. 28–34. (2021).
4. FRIEDRICH DE OLIVEIRA, H.; LENGYEL, PETER. Industry sector role in the Brazilian GDP and Exportation Share. AGRÁRINFORMATIKA / JOURNAL OF AGRICULTURAL INFORMATICS 12 : 1 pp. 32–40. (2021).