

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет экономических наук

Исследовательская работа

АНАЛИЗ ЭКСПОРТА И ИМПОРТА ГАЗА НА СТРАНОВОМ УРОВНЕ В УСЛОВИЯХ ГЛУБОКОЙ НЕОПРЕДЕЛЕННОСТИ

по направлению подготовки Экономика

образовательная программа «Экономика и статистика»

Выполнил:

Студент группы БСТ213

Еськов Михаил Михайлович

Руководитель:

Профессор, Алескеров Фуад Тагиевич

Москва 2023

Оглавление

| | |
|---|-----------|
| Введение | 3 |
| Подготовка данных | 4 |
| Выгрузка и предобработка | 4 |
| Визуализация | 6 |
| Круговые диаграммы | 6 |
| Кривые экспорта/импорта | 10 |
| Проблема данных | 11 |
| Обзор литературы | 12 |
| Подсчёт индексов | 15 |
| In-degree centrality | 15 |
| Out-degree centrality | 16 |
| Degree centrality | 17 |
| Betweenness centrality | 18 |
| Closeness centrality | 19 |
| PageRank | 20 |
| Short-range interaction centrality | 23 |
| Long-range interaction centrality | 24 |
| Заключение | 25 |
| Список литературы | 26 |

Введение

Газ является одним из ключевых энергетических ресурсов в современной мировой экономике, и его экспорт и импорт играют важную роль в глобальных энергетических рынках. Сетевой анализ является мощным инструментом для изучения взаимосвязей и взаимодействия между странами в контексте газовых поставок.

Целью данной работы является проведение сетевого анализа экспорта и импорта газа с использованием индексов центральности. Мы сосредоточимся на исследовании структуры и динамики газовых транспортных потоков, а также выявлении ключевых участников на рынке.

Сетевой анализ позволит нам исследовать географическое распределение газовых поставок, анализировать связи между странами и определить наиболее значимые узлы в газовой сети. Для этого мы будем использовать индексы центральности, такие как центральность посредничества, центральность по степени, центральность близости, PageRank, SRIC и LRIC.

Полученные результаты помогут нам понять, какие страны являются ключевыми поставщиками и потребителями газа, и какие маршруты транспортировки газа наиболее важны с точки зрения стабильности и эффективности. Это позволит принять более обоснованные решения в области энергетики и торговли газом.

Однако следует отметить, что в данной работе я не буду углубляться в детали инфраструктуры поставок газа, так как мой основной фокус будет на анализе сетевых связей между странами на основе данных об объемах экспорта и импорта газа.

В дальнейшем я буду применять методы сетевого анализа, основанные на индексах центральности, для исследования и визуализации глобальных газовых сетей. Результаты этого исследования могут быть полезными для принятия решений в энергетическом секторе, разработки стратегий торговли газом и оптимизации транспортных маршрутов.

Подготовка данных

Выгрузка и предобработка

Данные были взяты с сайта [UN Comtrade](https://comtrade.un.org/) по коду 2711 - Petroleum gas and other gaseous hydrocarbons. В него входят природный газ, пропан, бутан и остальные углеводородные газы, в том числе и в сжиженном виде. Данные были выгружены с 2001 года по 2020 год. Они были объединены в 4 блока по пять лет: 2001-2005гг, 2006-2010гг, 2011-2015гг и 2016-2020гг. И агрегированы по парам вида страна1-страна2. Например: Россия поставила в Китай в 2001г X\$ газа, в 2002г Y\$ газа и так далее, тогда в таблице в строке Россия – Китай будет значение $X + Y + \dots$ \$. Данные были взяты конкретно по импорту, поскольку значение импорта отражено более точно, а расхождение с экспортом может достигать до 10-20%.

В данных (отправитель и получатель) присутствуют зоны nes (not elsewhere specified): Areas, nes; Other Asia, nes и т. д. Также есть Special Categories, которые не раскрываются по просьбе партнёра страны. Все эти наблюдения были удалены, поскольку нас интересует сетевой анализ на страновом уровне.

Помимо этого, в данных присутствуют пропуски в столбце объёма (Qty в кг). Чтобы исказить выборку как можно меньше, каждый пропуск был заменён на частное стоимости экспорта (импорта) и средней цены за взятый период. Альтернативный способ замены пропусков нулями или статистиками средних исказили бы данные ещё сильнее, а поиск истинных значений очень трудоёмкий процесс.

Были посчитаны основные характеристики данных, где PrimaryValue – объём импорта в долларах США, а QTY – объём импорта в кг:

Import in 2001-2005

| | PrimaryValue | Qty |
|-------|--------------|--------------|
| count | 3.158000e+03 | 3.158000e+03 |
| mean | 1.648562e+08 | 1.483628e+09 |
| std | 2.017955e+09 | 2.963325e+10 |
| min | 1.000000e+00 | 0.000000e+00 |
| 25% | 4.397854e+03 | 1.837028e+03 |
| 50% | 1.561085e+05 | 2.231150e+05 |
| 75% | 4.171023e+06 | 1.165874e+07 |
| max | 9.970456e+10 | 1.357616e+12 |

Рис. 1

Import in 2006-2010

| | PrimaryValue | Qty |
|-------|--------------|--------------|
| count | 3.761000e+03 | 3.761000e+03 |
| mean | 3.032655e+08 | 8.048553e+08 |
| std | 2.735147e+09 | 7.630945e+09 |
| min | 1.000000e+00 | 0.000000e+00 |
| 25% | 6.683000e+03 | 1.794000e+03 |
| 50% | 3.293010e+05 | 2.545450e+05 |
| 75% | 1.077464e+07 | 1.552785e+07 |
| max | 1.143193e+11 | 2.956420e+11 |

Рис. 2

Import in 2011-2015

| | PrimaryValue | Qty |
|-------|--------------|--------------|
| count | 3.988000e+03 | 3.988000e+03 |
| mean | 4.354452e+08 | 7.903938e+08 |
| std | 3.399199e+09 | 6.360170e+09 |
| min | 1.000000e+00 | 0.000000e+00 |
| 25% | 6.306822e+03 | 1.031750e+03 |
| 50% | 3.162814e+05 | 1.987670e+05 |
| 75% | 1.272082e+07 | 1.558088e+07 |
| max | 7.418018e+10 | 1.346474e+11 |

Рис. 3

Import in 2016-2020

| | PrimaryValue | Qty |
|-------|--------------|--------------|
| count | 3.998000e+03 | 3.998000e+03 |
| mean | 3.151319e+08 | 8.445560e+08 |
| std | 2.393007e+09 | 6.500350e+09 |
| min | 1.370000e+00 | 0.000000e+00 |
| 25% | 8.312748e+03 | 1.358343e+03 |
| 50% | 3.120317e+05 | 2.172777e+05 |
| 75% | 1.247360e+07 | 2.244283e+07 |
| max | 6.425158e+10 | 1.397167e+11 |

Рис. 4

Как можно видеть, данные распределены очень неоднородно, поскольку коэффициент вариации много больше 33%, а также очень асимметрично, поскольку среднее больше медианы в несколько раз.

Визуализация

Далее был построен небольшой ряд визуализации полученных данных.

Круговые диаграммы

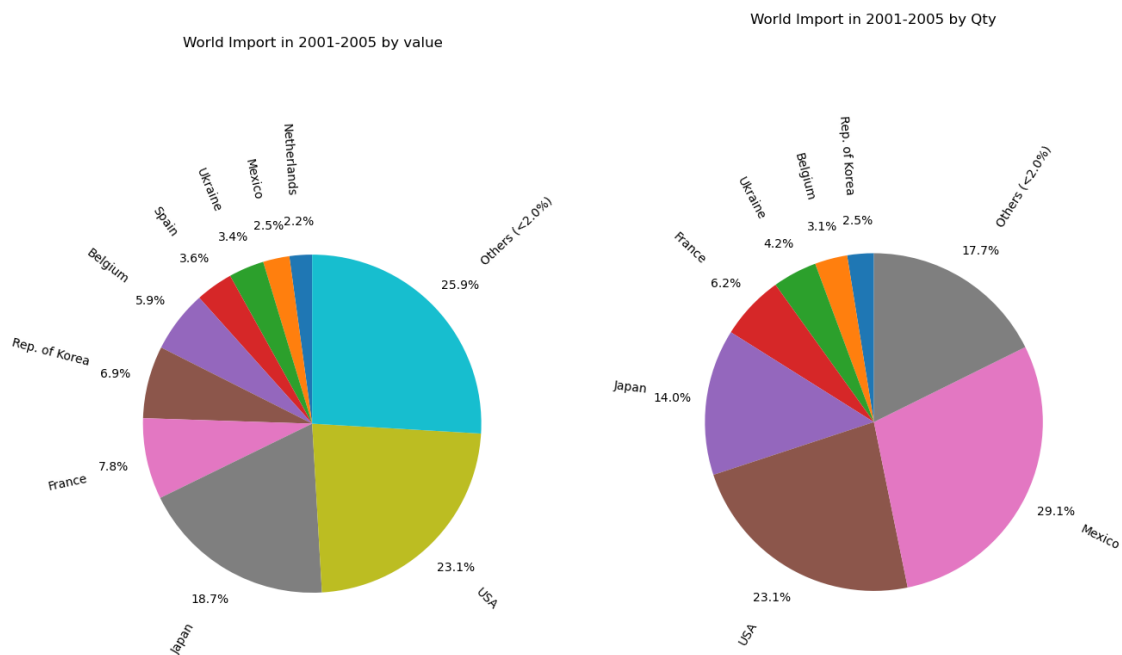


Рис. 5 Круговые диаграммы импорта по стоимости и объёму в период 2001-2005

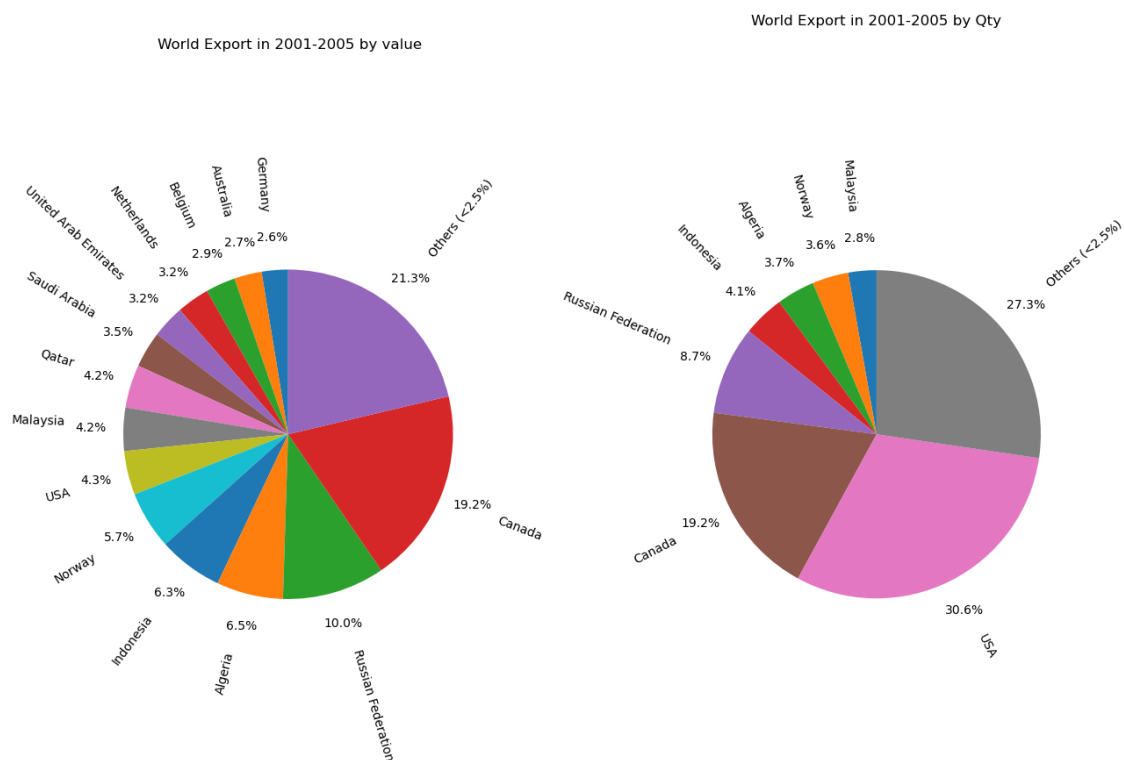
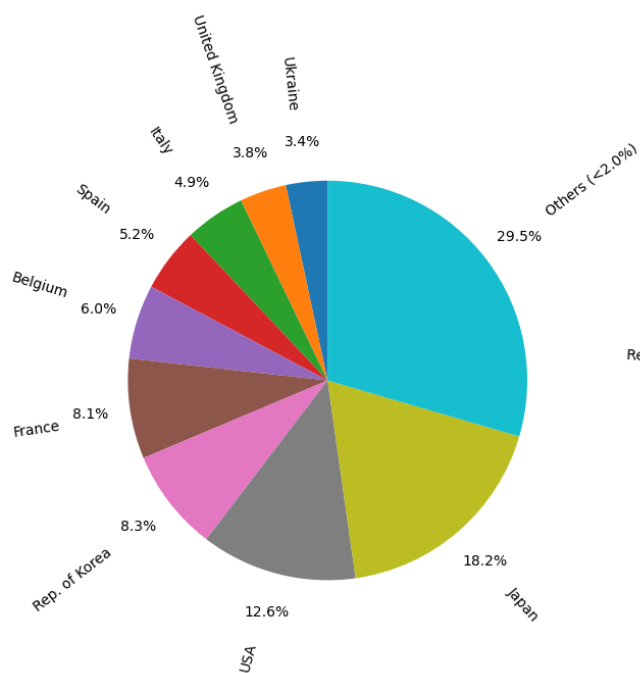


Рис. 6 Круговые диаграммы экспорта по стоимости и объёму в период 2001-2005

World Import in 2006-2010 by value



World Import in 2006-2010 by Qty

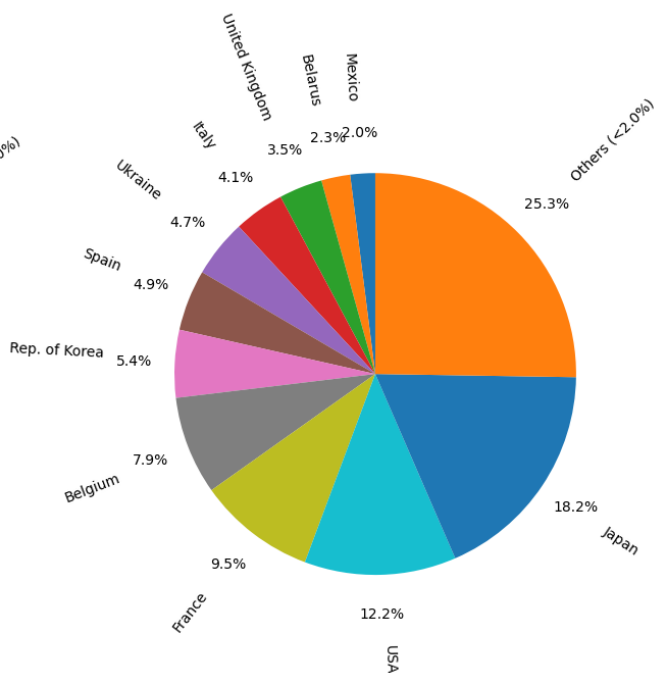
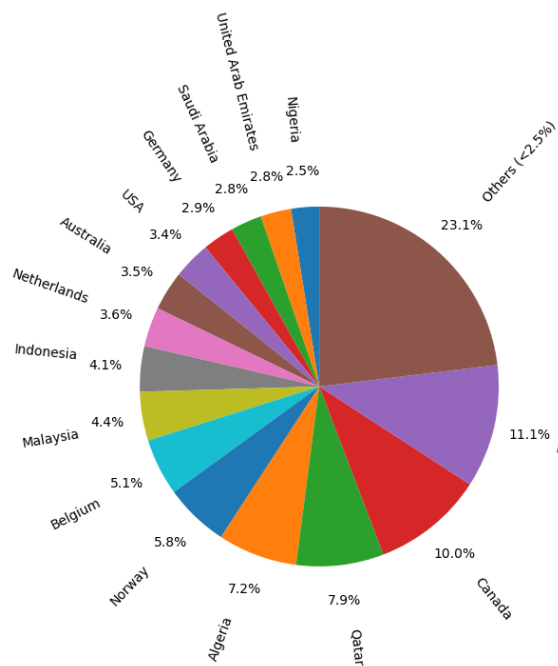


Рис. 7 Круговые диаграммы импорта по стоимости и объёму в период 2006-2010

World Export in 2006-2010 by value



World Export in 2006-2010 by Qty

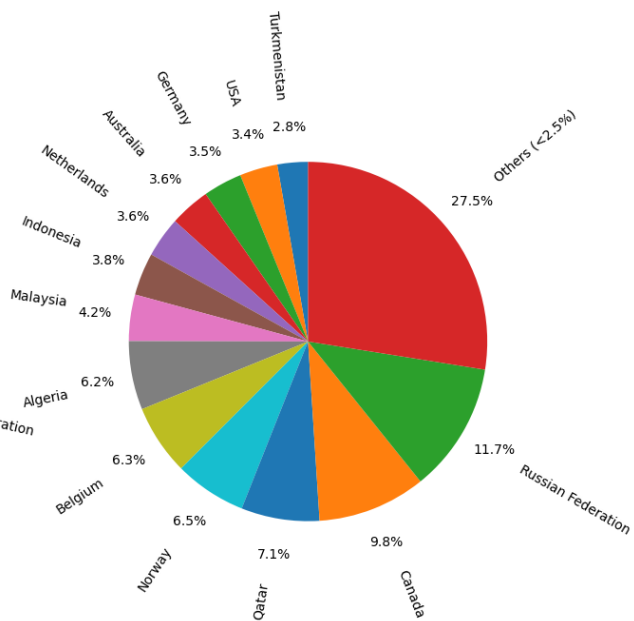


Рис. 8 Круговые диаграммы экспорта по стоимости и объёму в период 2006-2010

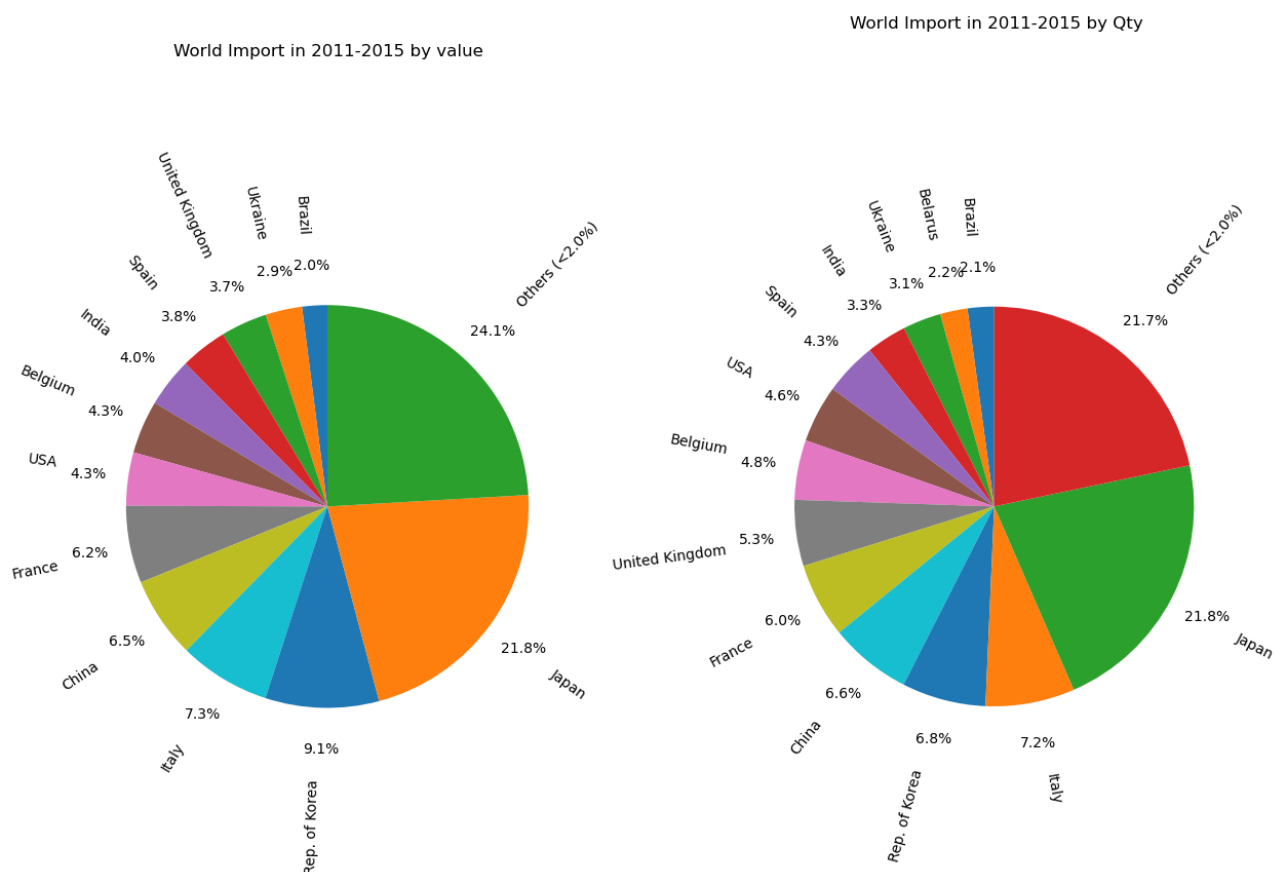


Рис. 9 Круговые диаграммы импорта по стоимости и объёму в период 2006-2010

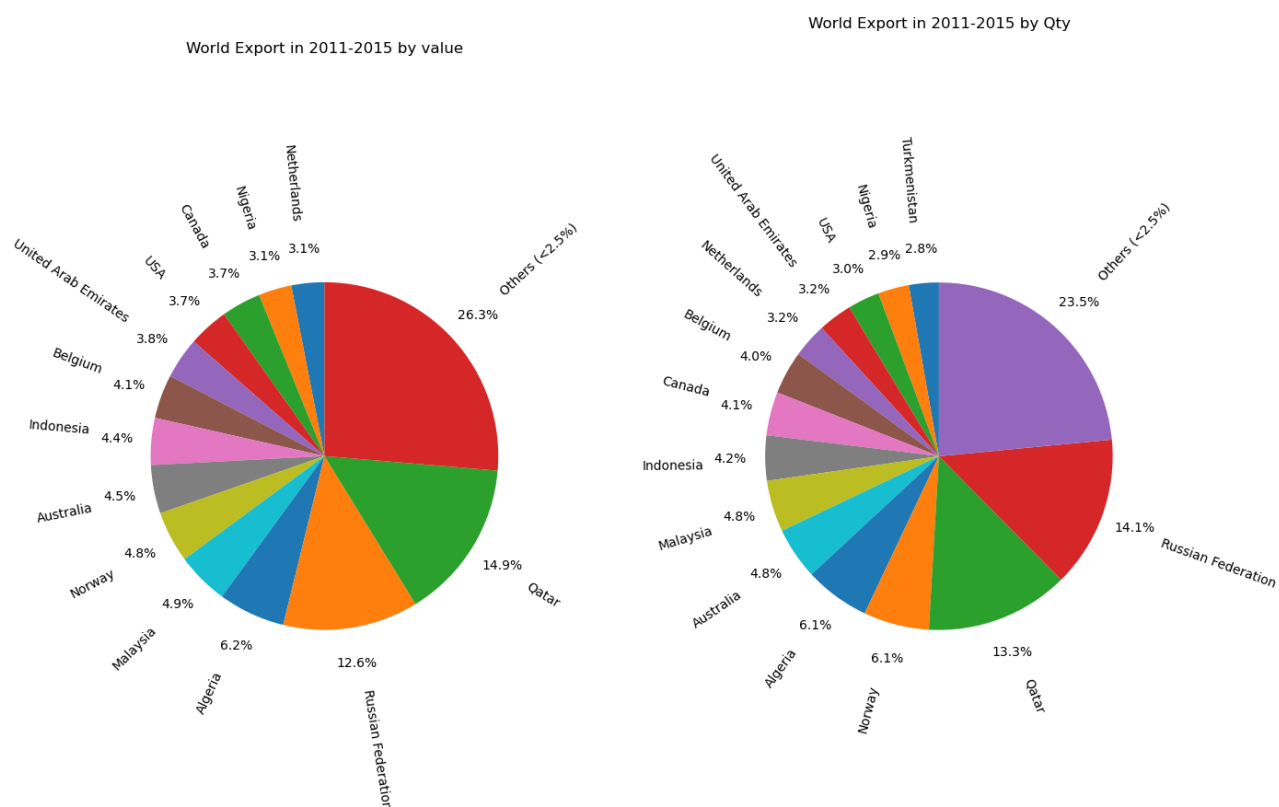


Рис. 10 Круговые диаграммы экспорта по стоимости и объёму в период 2011-2015

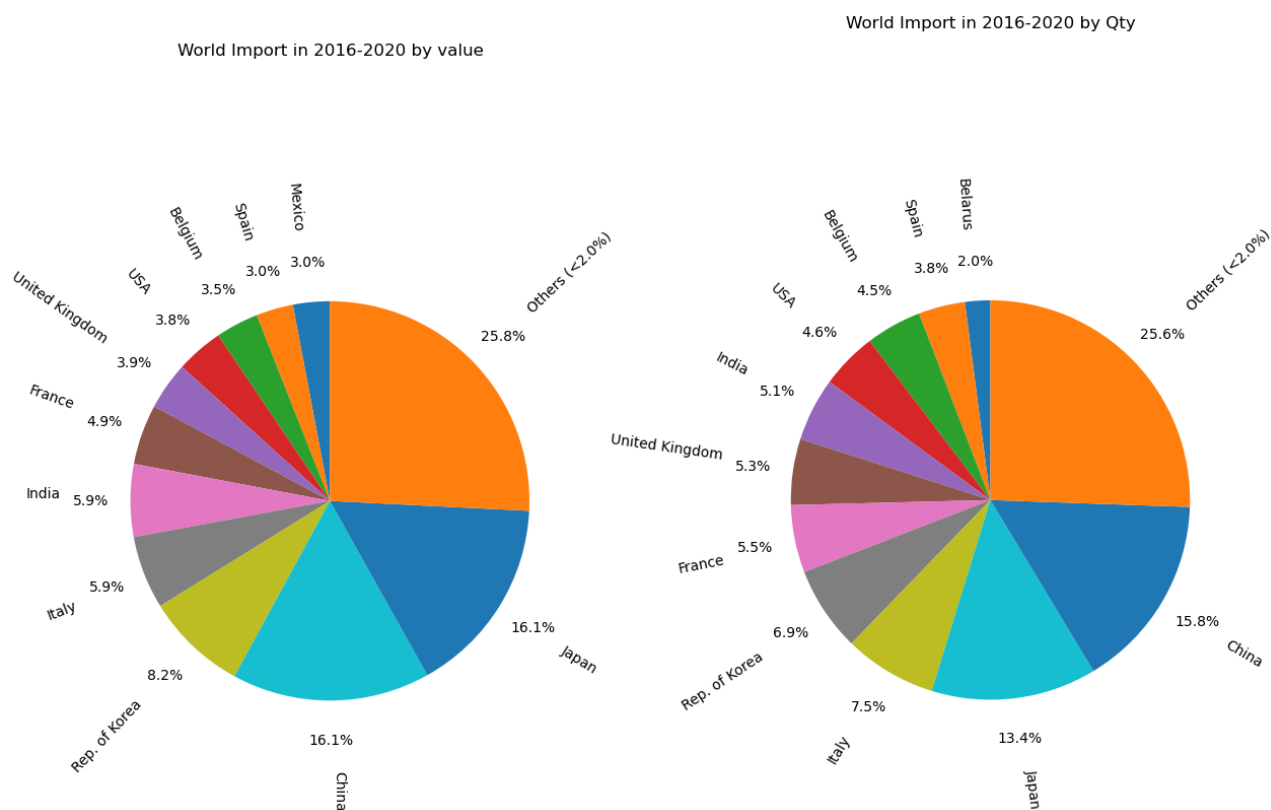


Рис. 11 Круговые диаграммы импорта по стоимости и объёму в период 2016-2020

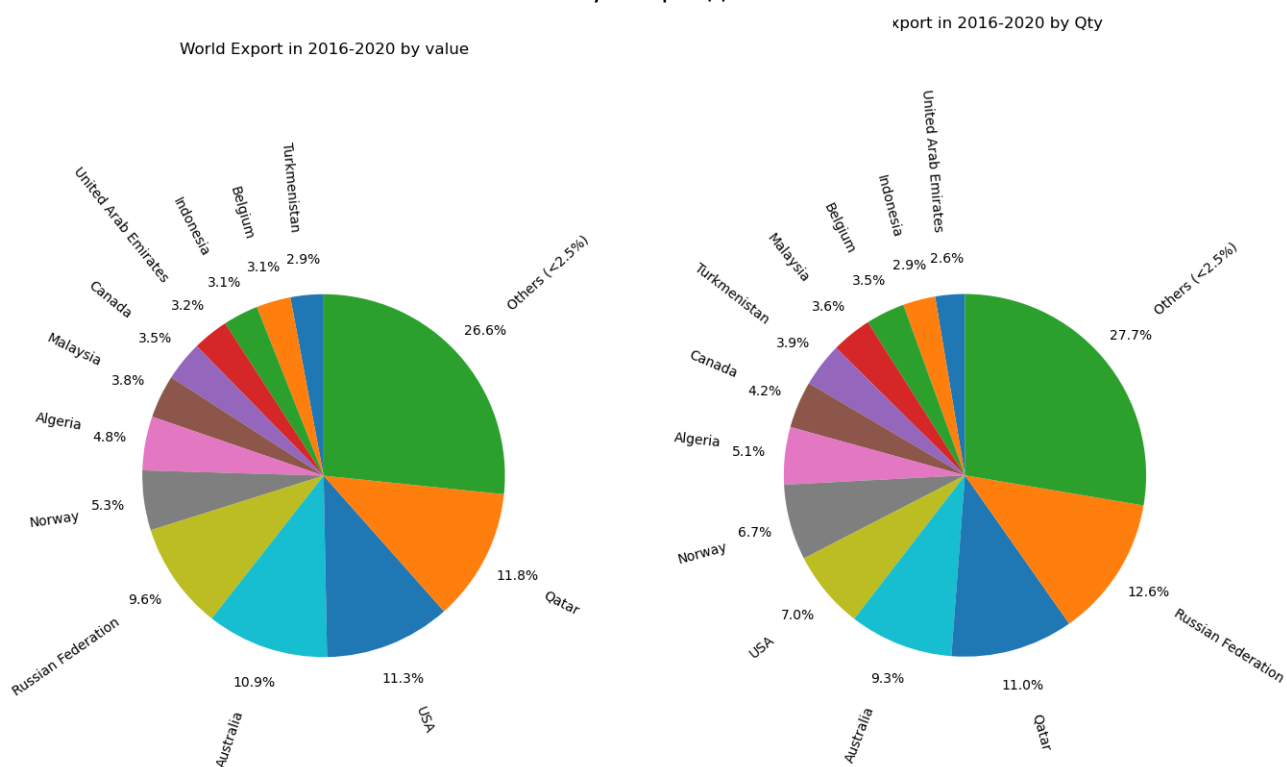


Рис. 12 Круговые диаграммы экспорта по стоимости и объёму в период 2016-2020

Кривые экспорта/импорта

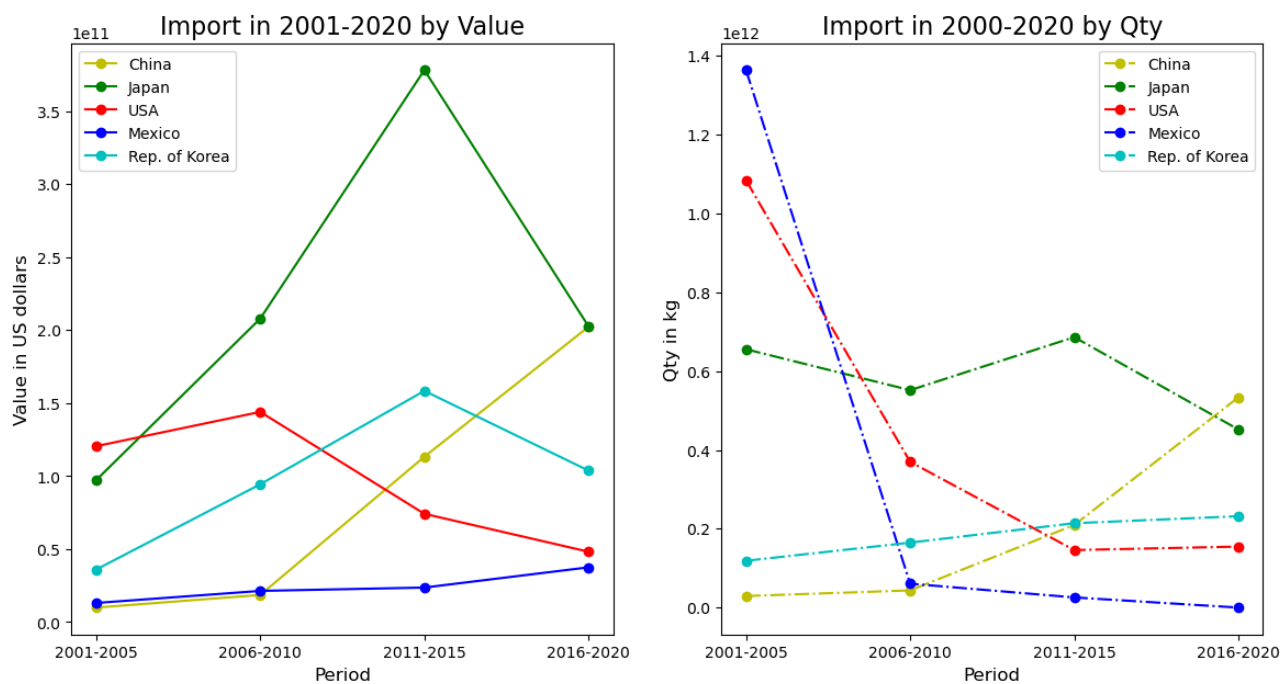


Рис. 12 Кривые импорта по стоимости и объёму в период 2000-2020

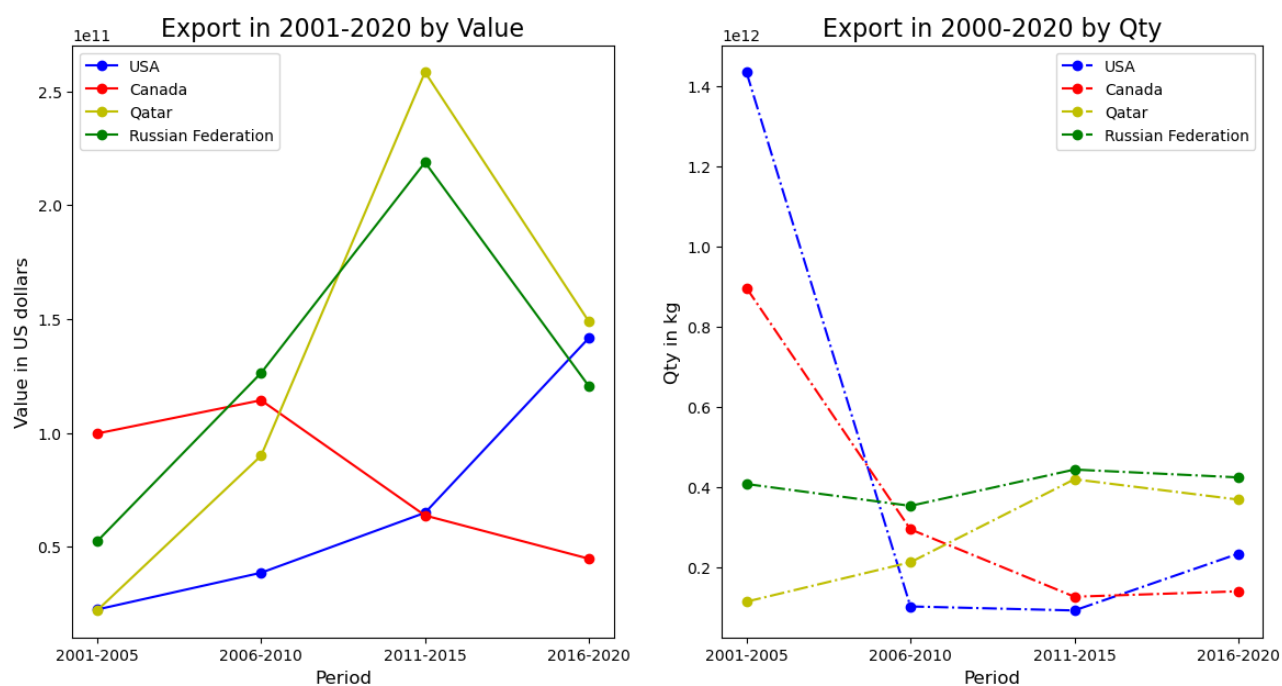


Рис. 12 Кривые экспорта по стоимости и объёму в период 2000-2020

Проблема данных

Как можно видеть на графиках, очень сильно выбивается пара Мексика - США. Если обычно разница между стоимостью и количеством (кг) долей от общего импорта (экспорта) составляет не более 1%, то для этой пары отклонение очень сильное. В блоке 2001-2005гг в мировом импорте по стоимости Мексика составляла долю в 2,5%, то по объёму 29.1%. Аналогичная ситуация у Америки только по экспорту. Это связано с тем, что в этой строке Мексика – США есть пропуск объёма, который был заполнен на основе средней цены. А истинная цена, вероятно, была выше средней. Таких несоответствий может быть несколько, поэтому проводить анализ по объёму будет некорректно. Из-за неоднородности экспортируемого газа, достаточно сложно будет более грамотно заполнить пропуски, поэтому я буду делать анализ по стоимости экспорта/импорта.

Обзор литературы

Ниже описаны основные индексы центральности.

Пусть мы имеем взвешенный направленный граф $G = \{V, E, W\}$, где $V = \{1, 2, \dots, n\}$ множество узлов, $|E| = |N|$, $E \subseteq V \times V$ множество рёбер, и $W = \{w_{ij}\}$ множество весов – вещественные числа, предписанные каждому ребру $(i, j) \in E$. Тогда для него можно определить следующие индексы центральности:

- 1) Weighted in-degree centrality (Freeman, 1979)

$$C_i^{w \text{ in-degree}} = \sum_{j=1}^n w_{ji}$$

Говоря иначе, сумма весов рёбер, входящих в вершину.

- 2) Weighted out-degree centrality (Freeman, 1979)

$$C_i^{w \text{ out-degree}} = \sum_{j=1}^n w_{ij}$$

Говоря иначе, сумма весов рёбер, исходящих из вершины.

- 3) Weighted degree centrality (Freeman, 1979)

$$C_i^{w \text{ degree}} = C_i^{w \text{ in-degree}} + C_i^{w \text{ out-degree}}$$

Говоря иначе, сумма весов рёбер, входящих и исходящих из вершины.

- 4) Betweenness centrality¹ (Freeman, 1979)- центральность по посредничеству

Отношение числа кратчайших путей, проходящих через данный узел и соединяющих два узла, к общему числу кратчайших путей, соединяющих эти узлы.

$$C_i^{\text{betweenness}} = \sum_{jk} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

Где, σ_{jk} – количество кратчайших путей, соединяющие вершины j и k .

- 5) Closeness centralities (Rochat, 2009) – центральность по близости

$$C_i^{\text{closeness}} = \frac{1}{\sum d_{ij}}$$

Где d_{ij} длина пути от вершины i до вершины j .

¹ В работе (Fuad Aleskerov, 2016) этот индекс называется relative betweenness.

- 6) Индекс PageRank является одним из наиболее известных и широко используемых алгоритмов центральности в сетевом анализе. Он был разработан Ларри Пейджем и Сергеем Брином, сооснователями поисковой системы Google, и был первоначально применен для оценки важности веб-страниц.

Индекс PageRank рассчитывается на основе идеи, что важность узла в сети определяется в основном количеством и качеством связей, которые он имеет с другими важными узлами. Этот индекс основывается на концепции "важность по ссылке". Суть состоит в том, что если узел А имеет ссылку на узел В, то это свидетельствует о том, что узел А считает узел В важным.

Расчет индекса PageRank происходит по итеративному алгоритму. На первом шаге каждый узел получает одинаковое начальное значение PageRank. Затем на каждой итерации PageRank каждого узла пересчитывается путем учета вклада соседних узлов, которые ссылаются на данный узел. Чем больше ссылок имеет узел и чем важнее ссылающиеся на него узлы, тем выше будет его PageRank. Процесс повторяется до тех пор, пока значения PageRank не стабилизируются.

Значение индекса PageRank для узла можно интерпретировать как вероятность случайного перехода по сети и остановки на данном узле. Узлы с более высокими значениями PageRank считаются более важными и центральными в сети, так как они имеют большее количество и качество входящих связей. (Sergey Brin, 1998)

- 7) Short-Range Interactions Centrality (Fuad Aleskerov, 2016)

Группа соседей вершины i $\Omega(i) \subseteq \overline{N_i}$ критическая, если $\sum_{j \in \Omega(i)} w_{ij} \geq q_i$, где q_i заранее заданный порог. Ключевой называется вершина в критической группе, при удалении которой группа становится некритической.

Интенсивность прямого влияния p_i^j вершины j на вершину i вычисляется как $p_j^i = \frac{w_{ji}}{\sum_k w_{ki}}$, а интенсивность непрямого влияния как

$$p_{jy}^i = \begin{cases} \frac{w_{jy}}{\sum_k w_{ki}}, & \text{если } w_{yi} > 0, w_{jy} < w_{yi} \text{ и } y \neq j \\ \frac{w_{yi}}{\sum_k w_{ki}}, & \text{если } w_{yi} > 0, w_{jy} > w_{yi} \text{ и } y \neq j \\ 0, & \text{иначе} \end{cases}$$

Итоговая интенсивность влияния узла на узел агрегируется по интенсивности всех групп, где узел является ключевым относительно к размеру группы. Влияние каждого узла на узел равно нормированному значению конечной меры интенсивности. После этого, индекс агрегируется по всем соседним вершинам.

У этого индекса есть недостаток – во внимание берётся только прямое влияние на первом уровне, в то время как могут быть дальние взаимодействия, оказывающие влияние на вершину. Этот недостаток устраняет индекс LRIC.

8) LRIC – Long-Range Interactions Centrality. (Fuad Aleskerov, 2016)

Пусть $\Omega_p(i)$ – набор ключевых вершин в группе $\Omega(i)$. После вычисляется матрица интенсивности C по отношению к весам w_{ij} , порогу q_i и критической группы $\Omega(i)$ как:

$$\begin{cases} c_{ij} = \frac{w_{ij}}{\max_{\Omega(j) \leq N_j | i \in \Omega_p(j)} \sum_{i \in \Omega_p(j)} w_{ij}}, & \text{если } i \in \Omega_p(j) \leq \bar{N}_j \\ 0, & i \in \Omega_p(j) \in \bar{N}_i \end{cases}$$

После этого по определенной функции вычисляется влияние вершины i на вершину j для пути от i до j . После того как вычислено влияние вершины i на вершину j по всем путям длиной меньше s , вычисляется агрегированное влияние i на j с помощью определённой функции. Мной был выбран алгоритм MaxMin, для которого функция влияния влияния вершины i на вершину j для пути от i до j имеет вид:

$$f(P_k^{ij}) = \min(c_{il_1^k}, c_{l_1^k l_2^k}, \dots, c_{l_{n(k)-1}^k j})$$

А функция агрегированного влияния i на j по всем путям:

$$c_{ij}^*(s) = \max_{k: |P_k^{ij}| \leq s} f(P_k^{ij})$$

Для каждой вершины считаются относительные влияния $u_i = \frac{\sum_j w_{ij}}{\sum_k \sum_j w_{kj}}$,

поскольку чем выше влияние узла на другие узлы, тем выше его вес.

Значение LRIC для каждой вершины будет равно произведению матрицы C на вектор u .

Подсчёт индексов

На языке Python с помощью библиотек Numpy, Pandas, NetworkX, SLRIC (взятой с <https://github.com/SergSHV/slric>)² были подсчитаны индексы центральности, указанные в литературе. А с помощью пакета Matplotlib и программы Gephi был построен ряд визуализации итоговых расчётов. Исходный код работы (файл Gas_analysis) и исходные данные доступны по [ссылке](#).

Weighted In-degree centrality

В нашем случае, Weighted In-degree центральность не что иное, как мировой импорт, т.е. сколько всего страна импортировала в данный период. Чтобы не загромождать работу таблицами со значениями индексов³, я буду строить визуализацию по странам с наибольшими значениями индексов.

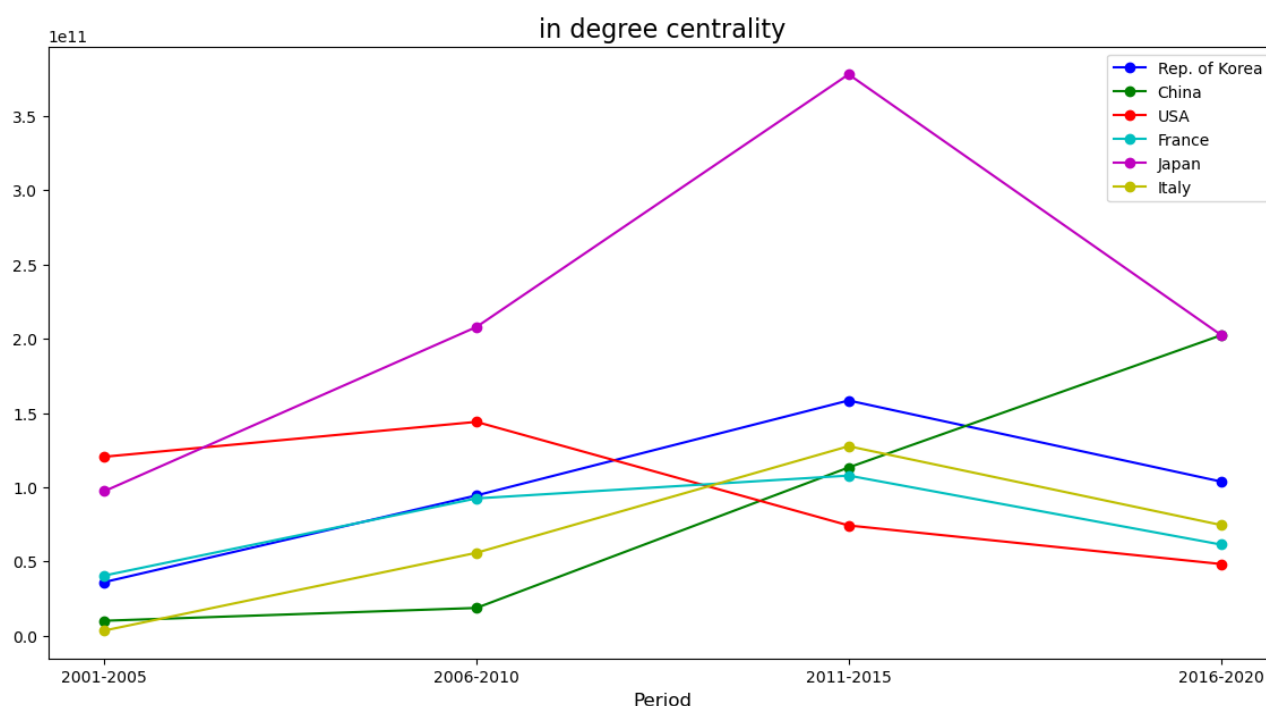


Рис. 13 Значение Weighted In-degree центральности в период 2001-2020

Таким образом, самыми большими импортёрами являются следующие страны: Южная Корея, Китай, США, Франция, Япония и Италия.

² Эта библиотека не работала должным образом «из коробки», поэтому потребовалось немалое количество времени, чтобы привести её в рабочее состояние.

³ Значение всех индексов могу предоставить в виде таблицы или продемонстрировать в среде разработки

Weighted Out-degree centrality

В нашем случае, Weighted Out-degree центральность не что иное, как мировой экспорт, т.е. сколько всего страна экспортировала в данный период.

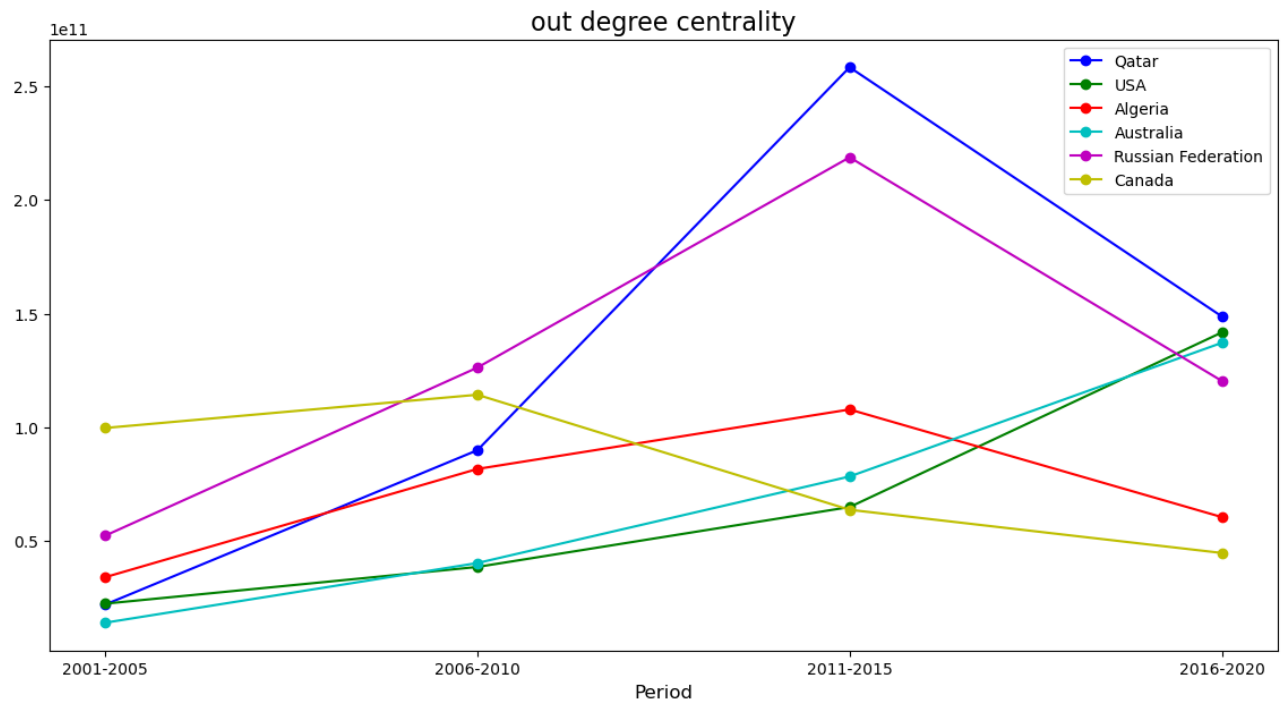


Рис. 14 Значение Weighted Out-degree центральности в период

Как можно видеть, самыми большими экспортёрами являются следующие страны: Катар, США, Алжир, Австралия, Россия и Канада.

Weighted Degree centrality

Степенная центральность представляет собой сумму in и out-degree центральность, то есть сумма экспорта и импорта для каждой страны.

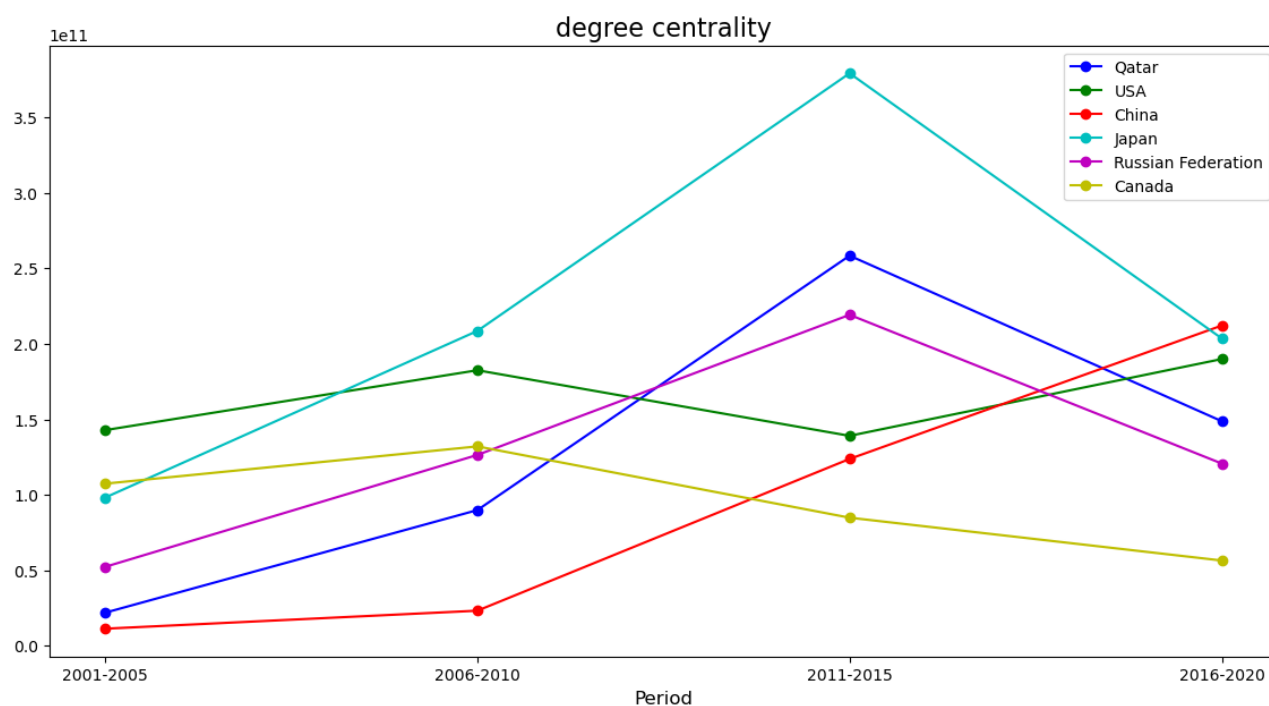


Рис. 15 Значение Weighted Degree центральности в период 2001-2020

В топ стран по индексу Degree вошли Катар, США, Китай, Япония, Россия и Канада.

Betweenness centrality

Центральность по посредничеству показывает, насколько часто определенный узел выступает в качестве посредника или промежуточного элемента на кратчайших путях между другими парами узлов в сети. Узлы с высокой центральностью по посредничеству играют важную роль в связи и передаче информации между другими узлами в сети.

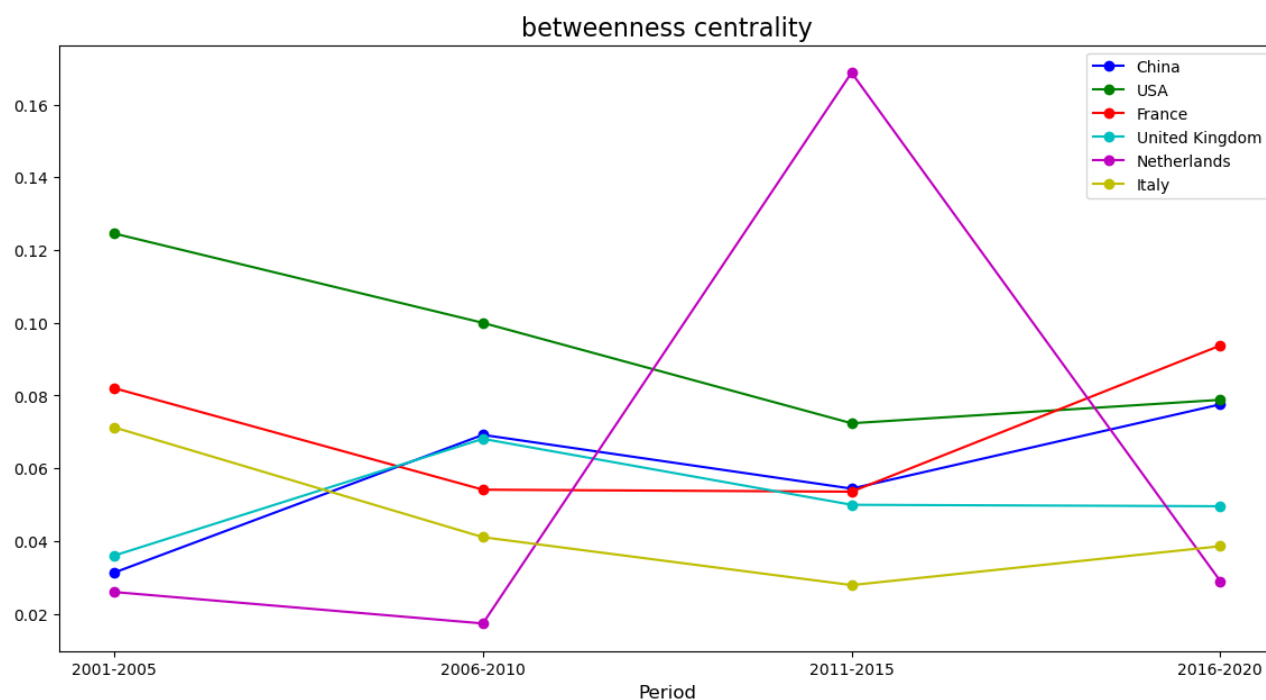


Рис. 16 Значение центральности по посредничеству в период 2001-2020

Наиболее центральными по посредничеству стали страны Китай, США, Франция, Великобритания, Нидерланды и Италия

Closeness centrality

Центральность по близости показывает, насколько близко (или насколько быстро) узел может достичь всех других узлов в сети. Узлы с более высоким значением центральности по близости считаются более "близкими" к другим узлам и имеют более быстрый доступ к информации или ресурсам в сети.

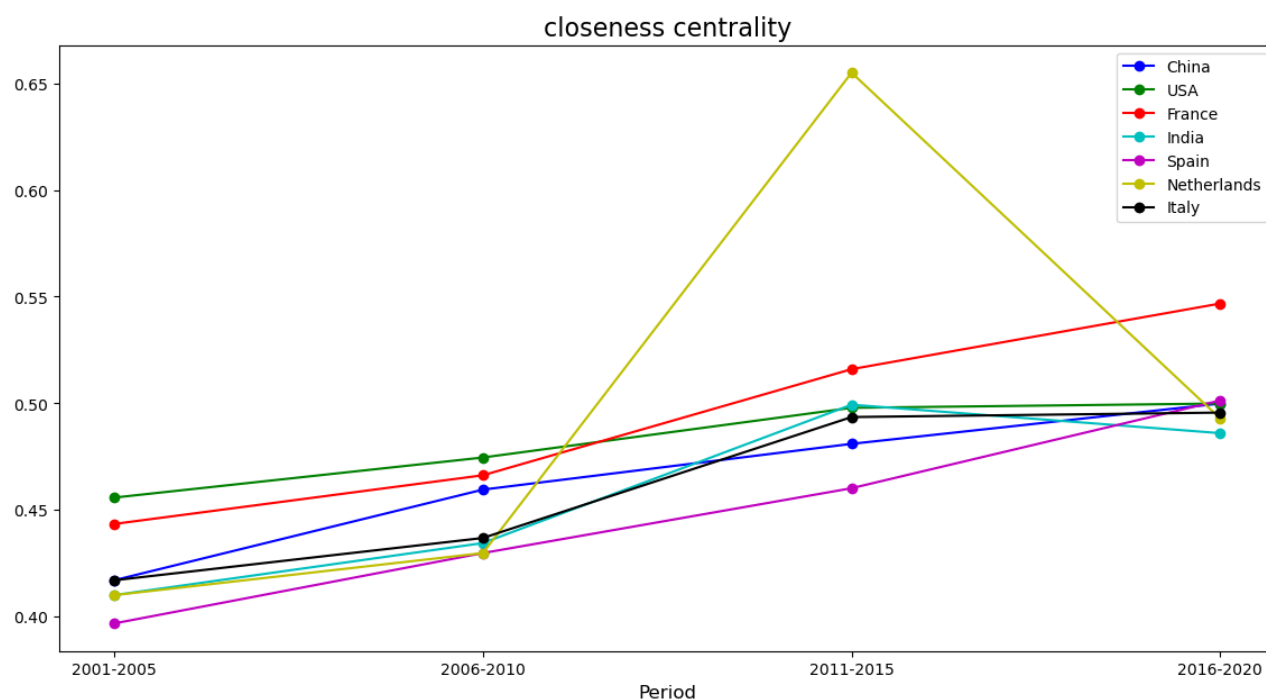


Рис. 17 Значение центральности по близости в период 2001-2020

По графику видно, что самыми центральным по близости являются страны Китай, США, Франция, Индия, Испания, Нидерланды и Италия.

PageRank

Также был посчитан индекс PageRank для данных сетей.

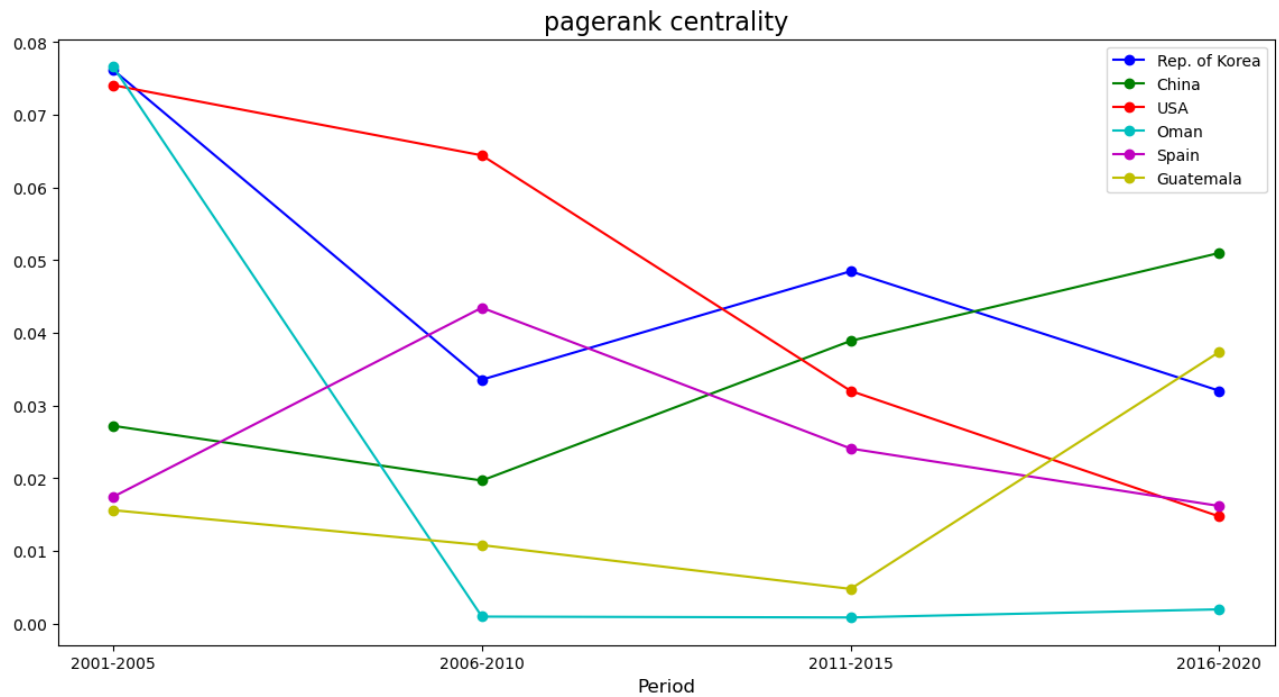


Рис. 17 Значение индекса PageRank в период 2001-2020

По индексу PageRank наиболее важными оказались страны: Южная Корея, Китай, США, Оман, Испания и Гватемала. Дополнительно, в Gephi была построена сеть на основе PageRank. Чем больше значение PageRank, тем больше название страны.

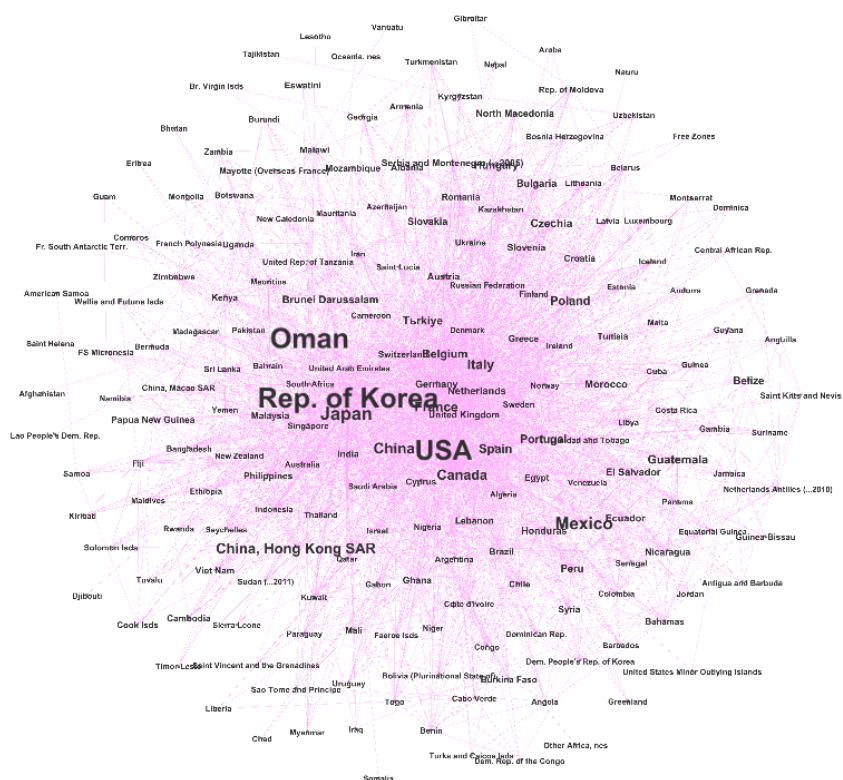


Рис. 18 Сеть 2001-2005гг

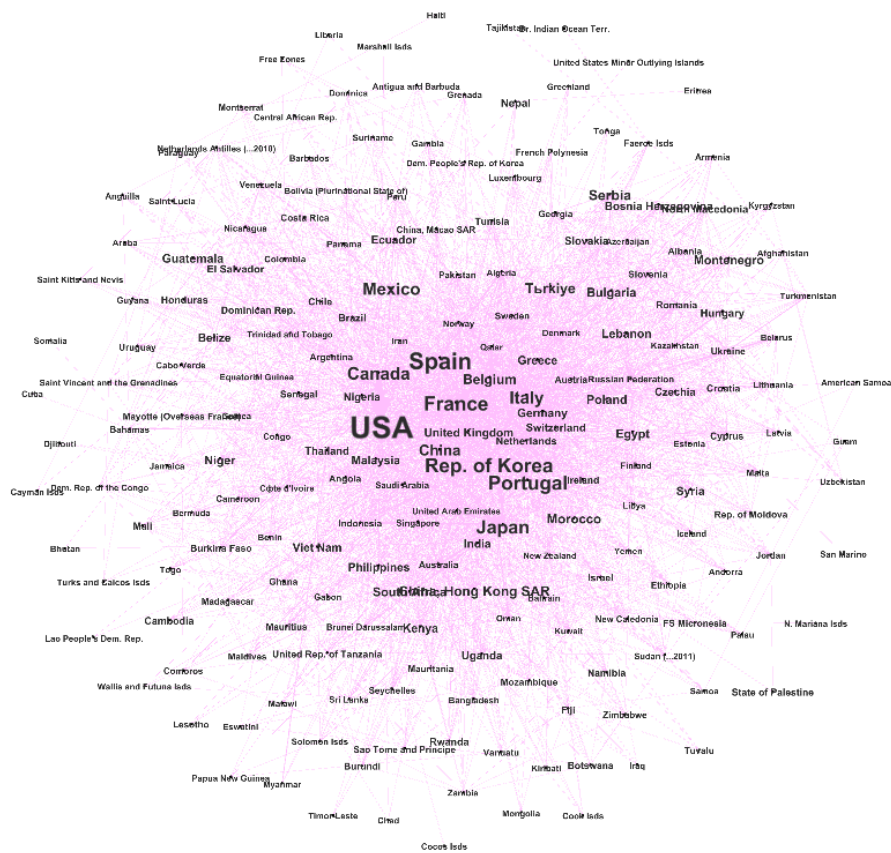


Рис. 18 Сеть 2006-2010гг

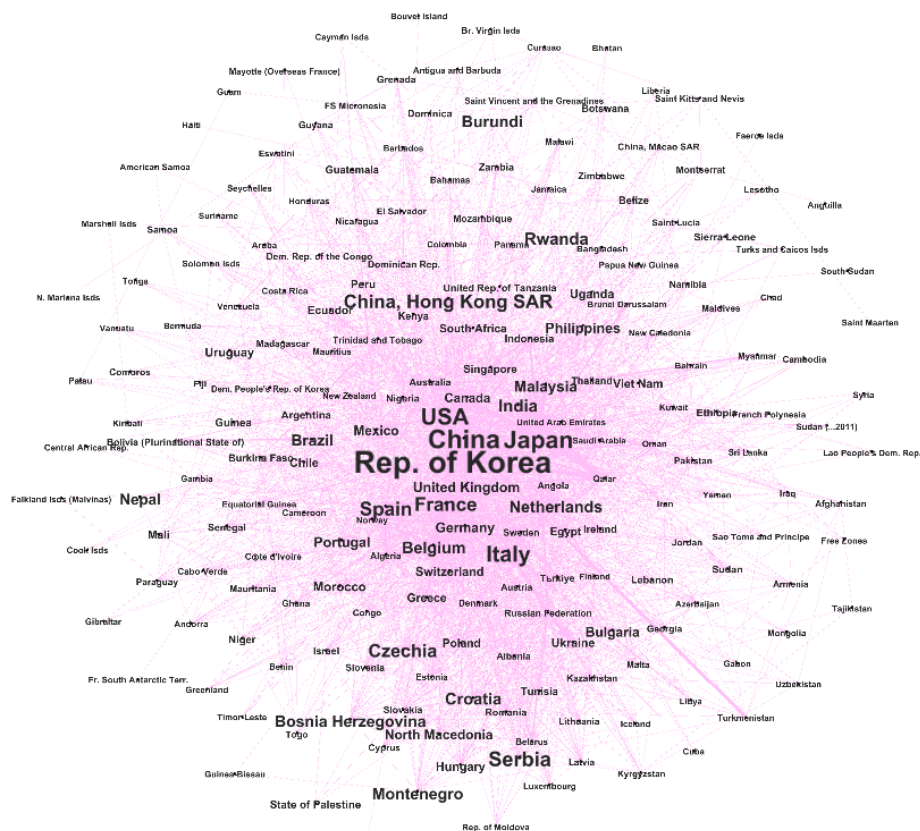


Рис. 19 Сеть 2011-2015гг

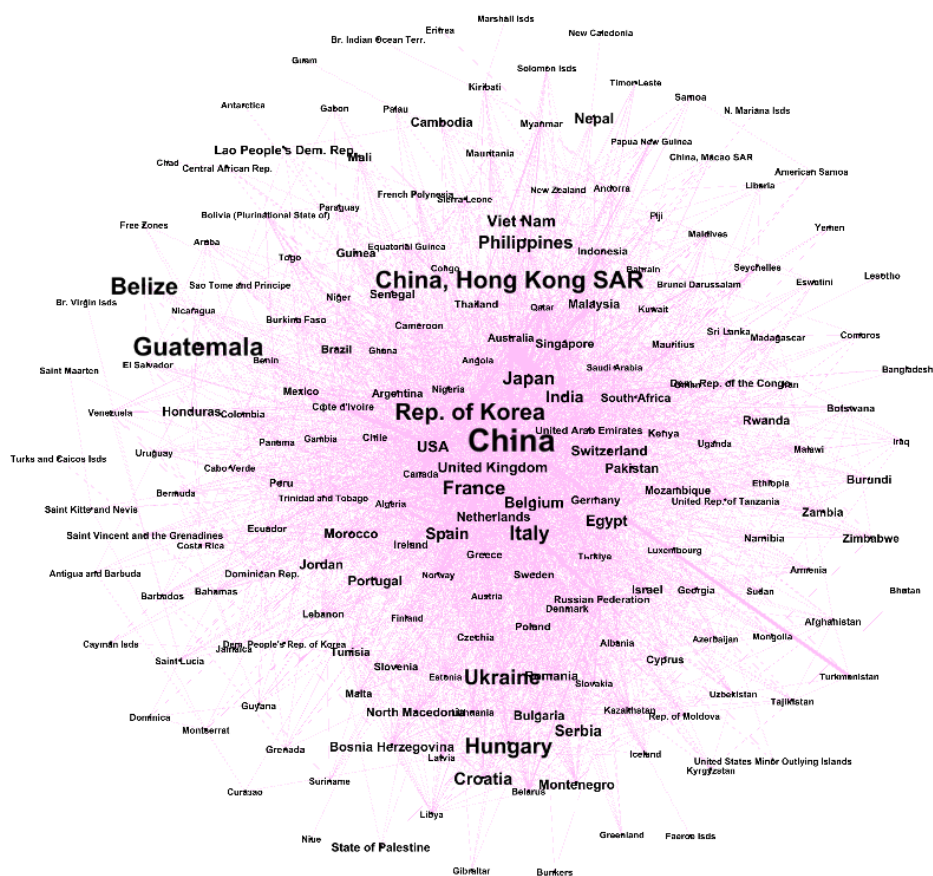


Рис. 20 Сеть 2016-2020гг

Short-range interaction centrality

SRIC индекс отражающий важность вершины на основе влияния на соседние вершины, и, в отличие от других индексов, более точно показывает скрытые влияния в сети, но действует лишь на первом уровне. В качестве квоты было взято $q = 50\%$, поскольку эту долю будет достаточно сложно скомпенсировать с помощью других партнеров.

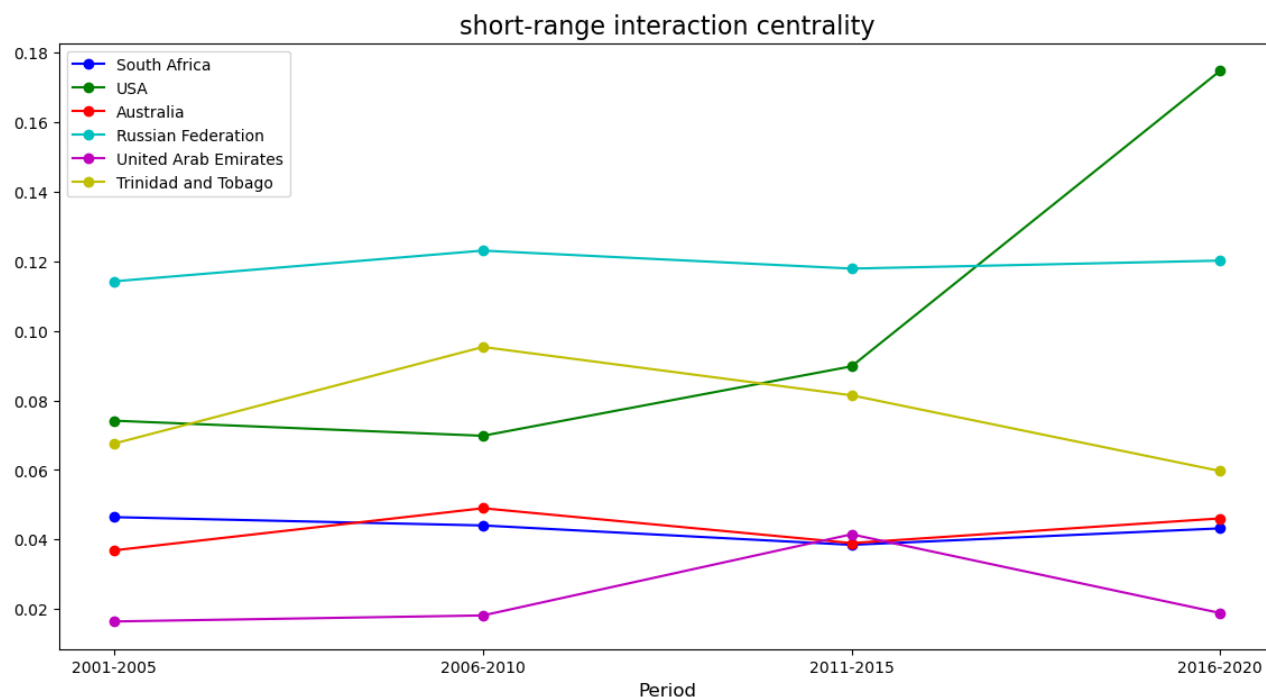


Рис. 21 Значение индекса SRIC в период 2001-2020

Как можно видеть, в топе стран по индексу SRIC оказались страны, которых не было в предыдущих индексах. Наиболее центральными странами на основе индекса SRIC оказались Южная Африка, США, Австралия, Тринидад и Табаго, Россия и ОАЭ

Long-range interaction centrality

LRIC индекс является усовершенствованной версией SRIC, но также учитывает дальние влияния. В качестве квоты было взято $q = 50\%$, модель MaxMin и максимальная длина пути равна 3.

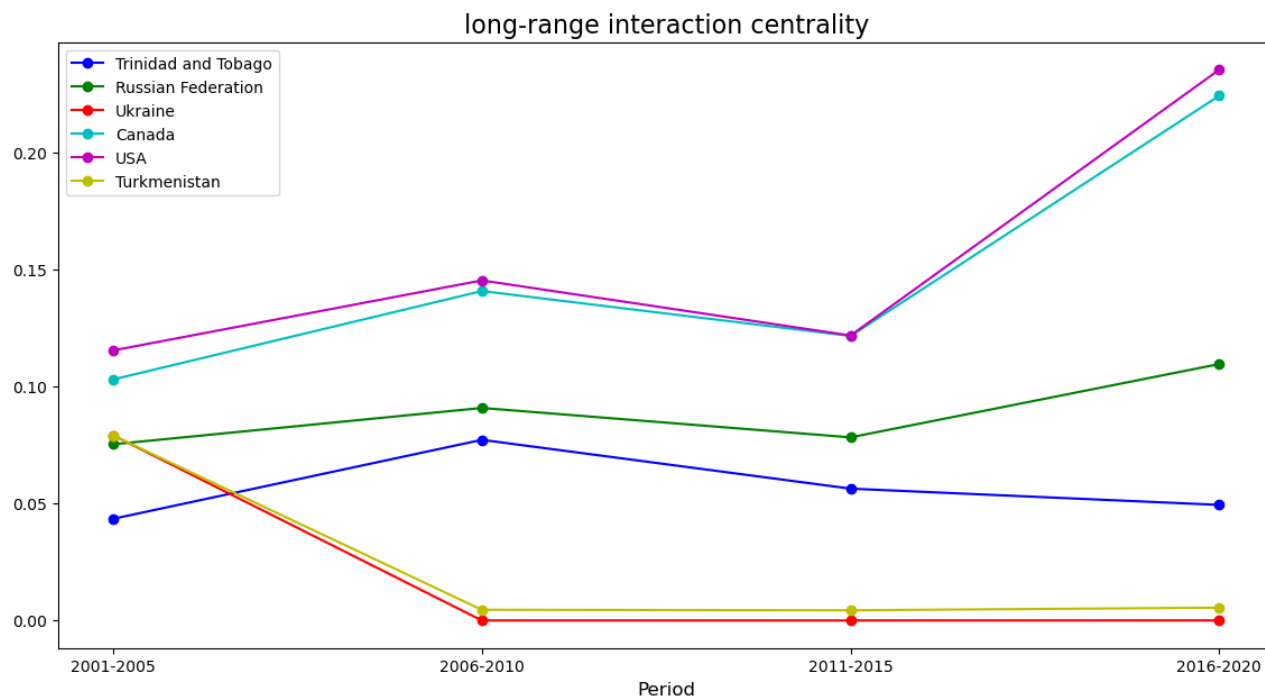


Рис. 22 Значение индекса LRIC в период 2001-2020

Страны оказывающие наибольшее влияние оказались США, Канада, Россия, Тринидад и Тобаго, Туркменистан и Украина.

Заключение

Исходя из значения индекса LRIC, как наиболее подходящего для анализа, США занимает первое место по значимости на протяжении 20 лет (2001-2020гг), а Канада второе на протяжении 20 лет (2001-2020гг). Россия занимает 3 место по значимости на основе индекса LRIC с 2006г по 2020г. 4 место с 2006г по 2020г является страна Тринидад и Тобаго, что является очень неожиданным результатом. Таблица значения индекса LRIC выглядит следующим образом:

| | country | 2001-2005 | 2006-2010 | 2011-2015 | 2016-2020 |
|-----|---------------------|--------------|-----------|-----------|-----------|
| 130 | USA | 1.154315e-01 | 0.145357 | 0.121806 | 0.235437 |
| 167 | Canada | 1.030218e-01 | 0.140809 | 0.121608 | 0.224325 |
| 1 | Russian Federation | 7.531335e-02 | 0.090847 | 0.078267 | 0.109604 |
| 70 | Trinidad and Tobago | 4.344928e-02 | 0.077211 | 0.056345 | 0.049395 |
| 152 | Mozambique | 7.552327e-07 | 0.040874 | 0.039008 | 0.043743 |
| 115 | Australia | 3.951179e-02 | 0.040876 | 0.026400 | 0.038289 |
| 77 | South Africa | 3.160158e-02 | 0.040875 | 0.034674 | 0.038275 |
| 223 | Germany | 1.668541e-02 | 0.018202 | 0.017786 | 0.016616 |
| 22 | China | 4.789399e-03 | 0.022741 | 0.013021 | 0.016607 |
| 67 | Belgium | 4.776252e-03 | 0.040896 | 0.034926 | 0.016559 |

Рис. 23 Значение индекса LRIC в период 2001-2020гг
(отсортированы по 2016-2020гг)

Список литературы

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. В L. C. Freeman.

Fuad Aleskerov, N. M. (2016). CENTRALITY MEASURES IN NETWORKS BASED ON NODES ATTRIBUTES, LONG-RANGE INTERACTIONS AND GROUP INFLUENCE. В N. M. Fuad Aleskerov. препринт WP7/2016/04.

Rochat, Y. (2009). *Closeness Centrality Extended To Unconnected Graphs*.

Sergey Brin, L. P. (1998). The anatomy of a large-scale hypertextual Web search engine. В L. P. Sergey Brin.