WORKING PAPERS

N° 1461

August 2023

"Understanding World Economy Dynamics Based on Indicators and Events"

Yasser Abbas and Abdelaati Daouia

Toulouse School of Economics

Understanding World Economy Dynamics Based on Indicators and Events

Yasser Abbas* and Abdelaati Daouia†
Toulouse School of Economics

Abstract

Studying the content and impact of news articles has been a recurring interest in economics, finance, psychology, and political and media literature over the last 20 years. Most of these offerings focus on specific qualities or outcomes related to their textual data, which limits their applicability and scope. Instead, we use novel datasets that adopt a more holistic approach to data gathering and text mining, allowing texts to speak for themselves without shackling them with presupposed goals or biases. Our data consists of networks of nodes representing key performance indicators of companies, industries, countries, and events. These nodes are linked by edges weighted by the number of times the concepts were connected in media articles between January 2018 and January 2022. We study these networks through the lens of graph theory and use modularity-based clustering, in the form of the Leiden algorithm, to group nodes into information-filled communities. We showcase the potential of such data by exploring the evolution of our dynamic networks and their metrics over time, which highlights their ability to tell coherent and concise stories about the world economy.

Keywords: Dynamic clustering, graph theory metrics, influential economic actors, written media analysis, R, Gephi

1 Introduction

Studies of media output have been a mainstay feature of journals in economics, finance, psychology, and political and media literature for the last two decades. Regarded as the fourth estate, media outlets play a vital role in disseminating information and guiding public opinion and narratives. Of the many mediums permeating news media, the oldest by far is that of print, stretching as far back as the early 17$^{\text{th}}$ century and maintaining relevance to this day through newspapers, magazines, tabloids, and their more recently employed online counterparts. The proliferation of the latter made textual data available for mining and studying, inviting scores of researchers to dissect and explore news articles in hopes of uncovering something greater. For example, a set of studies attempted to establish that journalism not only reports on economic news, but also has tangible impacts on public perception (see Nadeau et al. [1999] and De Boef and Kellstedt [2004] to name a few), while another group discovered and examined a distinctly negative bias within media coverage (see Goidel and Langley [1995], Lamla and Lein [2008], and Soroka et al. [2015]). Others focused on the importance and influence of investigative journalism (see Hamilton [2016] and Mahone et al. [2019] for example), with Turkel et al. [2021] recently developing a method to measure the prevalence of said journalism in the space with the aid of text mining. Textual data has also seen recent use to develop a "crisis index" focusing on economic recessions [Le Mezo and Ferrari Minesso, 2020]. These are only a tip of the iceberg in terms of the depth and breadth of media-based research.

However, most of these approaches focus on their own goals or presupposed conclusions, limiting the way they extract, interact with, and represent the data. For human coding and keyword-search-based methods for example (like Hamilton [2016] and Mahone et al. [2019]),

*Corresponding author: yasser.abbas@tse-fr.eu
†Contributing author: abdelaati.daouia@tse-fr.eu. Orcid: 0000-0003-2621-8860

the application guides the selected keywords, which narrows the scope of the study to fit its objectives. In this paper, we use sprawling, holistic, and pliable datasets that can be molded to fit many a goal. This data comes in the form of networks of performance indicators of companies, industries, countries, and events, the edges between which weighted by the number of times they were linked in written media. The scale of the data allows using graph theory techniques to study, interpret, and visualize news-based data in new and innovative ways. Our work focuses on the dynamic case, where the evolution of the network can be observed and tracked rather smoothly, but the methods explained and applied in this work can be extrapolated to static snapshots of the data accumulated over long periods of time.

The article attempts to answer multiple questions, the first of which revolves around the topic of handling, manipulating, pruning, and aggregating data of this nature and scale into workable datasets that maintain the dynamic structure of the networks. To this end, we utilize multiple pruning techniques ranging from basic cutoffs of unimportant links and edges to a more advanced network connectivity-based approach. We also make use of the data's dynamic nature to create a series of weekly network snapshots between January 2018 and January 2022, and aggregate dynamically by implementing a fluid "memory" into the datasets. We then wish to examine community structure within the graphs, which is rather challenging given the dynamic nature of the data. Existing dynamic community detection algorithms tend to "oversmooth" in our case, as most algorithms emphasize efficient runtime through the enforced stability of clusters. However, the "memory" implemented into our series of networks creates smooth transitions and allows us to apply intuitive modularity-based clustering through the Leiden algorithm [Traag et al., 2019], a worthy successor to the popular Louvain algorithm [Blondel et al., 2008]. Using these datasets and clustering techniques, we then create compelling dynamic visualizations of the world economy from the perspective of (mostly Western) media. We utilize `R` [`R` Core Team, 2022] packages and the external software `Gephi` [Bastian et al., 2009] to generate interactive and elaborate yet legible figures that are adaptable to each economic actor's prominence within the network. The paper also attempts the ambitious task of quantifying the true influence of companies, industries, and countries in ways that do not directly rely on monetary measures like stock prices or GDP, but instead fully focus on the entities' media presence and interactions with other (non-)similar economic actors. We first do so through a study of nodes that employs popular graph theory metrics like the strength and betweenness centrality, but also through a novel measure of cluster dominance, which tries to assess the importance of an economic actor to the local community of nodes detected around it. We also extrapolate this exact same mindset to the study of links between nodes, where we employ our novel measure in similar fashion to reveal the most influential connections between economic actors. The paper also delves into a deeper analysis of community structure by studying the strongest clusters generated through the Leiden algorithm. We then apply all these methodologies to find the most influential companies from a media point of view and concisely explain their evolution over the course of the study period. All these approaches were discussed with the advent of Covid-19 in mind, which creates an extra layer of complexity to the analysis.

The remainder of the paper is organized as follows. In Section 2, we describe our data and its sources in detail. Section 3 briefly touches on the chosen modularity-based clustering algorithm and how we adapt it to our dynamic case in Subsections 3.1 and 3.2. It then features multiple graph theory-based applications to the data, including visualization in Subsection 3.3, studies of influential nodes, edges, clusters, and companies in Subections 3.4-3.7 respectively. Section 4 concludes.

2 Data description

Our data is directly supplied by a pioneering American firm known as Causality Link [1], which encompasses a tremendous amount of media-related information collected over the last 6 years. The company's servers use web-crawling algorithms to gather articles from 172 reputable western media/news related websites. They then utilize state-of-the-art text-mining algorithms to detect connections between *key performance indicators* (KPI), like supply, demand, profits, losses, and prices for different companies and industries. Laudy et al. [2022] detail these intricate methodologies that generate probabilistic causal models (also known as Bayesian networks) through natural language processing (NLP) techniques. For example, the statement "The increase in Apple demand leads to higher profits" would be interpreted as a positive connection between the KPIs "Apple demand" and "Apple profits", where the order of the KPIs indicates the direction of the connection. In this case, "Apple" would be the KPI's prefix or parent. These KPIs are not limited to companies and industries, as they can include macro-level performance indicators like GDP and exports/imports for countries and regions. To illustrate, the statement "The increase in population is creating stress on natural resources" would be parsed into a negative relationship between "world population" and "world resources", where "world" would be the prefix that encompasses general macro indicators. After some pruning (to be explained later), there are 97 companies and 69 industries within our data. Topical/seasonal KPIs such as trade deals, wars, natural disasters, and epidemics can also find their way into the transformed data, and can sometimes absolutely dominate its immediate neighborhoods, as our studies around Covid-19 will show. Each mention of a connection between two KPIs counts as one entry in the dataset and is referenced by date, which allows users to factor out older observations. The text-mining algorithms also attempt to determine whether the connection is discussing the past, present, or future, as well as the perceived strength of said connection. The datasets supplied by Causality Link should serve as a detailed and comprehensive summary of western media output, which ultimately reflects the media's view of the world and its economy.

On a more technical level, Causality Link defines its own ontology of countries, industries, and companies, as well as the KPIs for each category. The algorithms then search for mentions/synonyms for the key words and attempt to decode its meaning. This hard-coded approach allows the firm to focus on the most relevant and compelling parts of the studied texts, distilling an article's message to its purest form and generating a rough but concise summary of what it was trying to convey. The defined ontology is quite wide and exhaustive, and we rarely ever stumbled upon data that was not parsed or understood properly. That being said, it is worth noting that the firm only creates causal links between the KPIs of the parent countries, industries, and companies, and not between the parents themselves. While this aims to center the discussion around the connections between key performance indicators of the parents, it has the unintended side effect of limiting the influence of otherwise powerful parents whose KPIs are not explicitly examined very often in western media. For example, China, one of the word's most influential nations in terms of economic impact and dominance and currently the second largest economy, finds itself rather underrepresented in comparison to the USA in the processed data, as seen in the case study in Section 3. Since concrete data and statistics about the eastern superpower is not widely available, the text processing algorithms have a hard time parsing the vague mentions of China into meaningful causal links. While this complication might hold us back from modelling the true world economy, it can still give us a deeply insightful look into western media's view of the economy, exposing interesting quirks that can only be seen when examining textual data under the "macro-scope" employed in this work. In the sequel, we employ these weighted networks in their undirected form, as the inclusion of directed edges limits the set of tools we can use to analyze the data. Aggregated and processed versions of these datasets have been compiled into a database of weekly snapshots with a total of 1,108,866

[1] Causality Link's AI-powered research platform at https://causalitylink.com/

entries (81.7 MB). The processing mechanisms needed to create this database are laid out in the sequel, and the database as well as our implemented methods and code are available for download to the public at `https://github.com/ya-tls/world-economy-dynamics`.

3 Dynamic graph analysis and main results

3.1 Clustering and community detection

This section briefly introduces the preferred clustering approach for our data. First, it is worth noting that other well-known clustering methodologies like k-means [Lloyd, 1982] cannot handle the type of data we have at hand: k-means typically operates on points in a vector space and attempts to use available explanatory variables to group said points. It does not operate on adjacency matrices or edge lists directly. This is where modularity optimization shines. The framework attempts to maximize a suitable objective function that reflects the desired qualities of a partitioned graph [Newman and Girvan, 2004]. Multiple algorithms have attempted to optimize modularity through hierarchical agglomeration [Clauset et al., 2004], extremal optimisation [Duch and Arenas, 2005], spectral [Newman, 2006], and simulated annealing [Reichardt and Bornholdt, 2006] approaches. However, these approaches do not seem to scale well enough with our data. Such datasets call for fast greedy methods such as the Louvain algorithm [Blondel et al., 2008], or its more robust successor, the Leiden algorithm [Traag et al., 2019]. The latter will serve as our main clustering device going forward. For the sake of completeness, we provide brief explanations of modularity-based clustering and the Leiden algorithm. General modularity can be defined as

$$\mathcal{Q} = \frac{1}{2m} \sum_{i \neq j} \left(A_{ij} - \gamma \frac{a_i a_j}{2m} \right) \delta(c_i, c_j)$$

where A stands for the graph's adjacency matrix such that $A_{ij} = 1$ if nodes i and j are connected by an edge, and 0 otherwise, $a_i := \sum_j A_{ij}$ is the degree of node i (number of edges connected to node i), $m := \frac{1}{2} \sum_{i \neq j} A_{ij}$ is the total number of edges in the graph, and $c_i \in \{1, \ldots, k\}$ is the index of the cluster containing node i among the considered k clusters of the graph, with $\delta(c_i, c_j)$ being 1 if $c_i = c_j$ and 0 otherwise. This efficient formulation has the exemplary quality of weighing down links between nodes with high degrees, as such nodes have a high chance of being connected regardless of any underlying community structure. This prevents hubs, or nodes with high degrees, from dominating clusters. As for $\gamma \in \mathbb{R}_+$, it is known as the "resolution" parameter, which controls the granularity of clustering, *i.e.* larger γ values lead to bigger numbers of communities, and vice versa. As such, choices of γ are largely based on preference rather than optimality to some criterion. However, γ can also be interpreted as the coefficient weighing the contributions of links against non-links in a network. We thus select $\gamma = 1$ in our analysis to maintain parity between the two, and to open up avenues for comparison with our modularity clustering algorithms in the future, as most (like the popular Louvain algorithm) utilize this same choice. We use a weighted version of $\mathcal{Q}$ in our analysis, as implemented in the `find_partition` and `cluster-leiden` functions from the `R` packages `leidenAlg` [Kharchenko et al., 2021] and `igraph` [Csárdi et al., 2023] respectively, but we will restrict this explanation to the unweighted case to maintain brevity.

While the Leiden algorithm can be used to target different types of objective functions, we focus on optimizing $\mathcal{Q}$. The Leiden algorithm starts from a singleton partition (each node is its own cluster) and operates in three main stages:

- **Local Node Moving**: individual nodes are moved across clusters to generate the highest gains in modularity $\mathcal{Q}$.
- **Refinement**: each cluster created in the previous stage is treated as its own graph, and receives an additional stage of local node moving to refine each separate partition.

- **Network Aggregation**: each cluster of vertices is aggregated into a single node, and these nodes are then moved to increase the quality function further.

These steps allow the Leiden algorithm to improve on its Louvain counterpart in two aspects: the implementation of faster local node moving algorithms, and the inclusion of the additional refinement stage, which boosts intra-community connectivity. Ultimately, the Leiden algorithm offers an efficient and scalable approach to modularity optimization, two qualities that are of the utmost importance in the case of our large datasets.

However, all of the algorithms mentioned above, including the chosen Leiden algorithm, are predominantly used to cluster static data, whereas our case involves dynamically evolving data with varying numbers of nodes and edges over time. A few attempts have been made to handle such data in Held et al. [2016] and more recently in Seifikar et al. [2020] and Zhuang and Li [2019]. Nevertheless, these attempts emphasize runtime efficiency by promoting the stability of network partitions over time, which does not truly allow our data to speak for itself. To accommodate our unique database, we introduce an approach that strikes a balance between partition stability and sensitivity to changes by implementing a "memory" aspect into the data itself. At each point in time, data is collected and summarized from the last three preceding months, which yields smoothness in the evolution of the partitions but still permits shocks to fully manifest and impact the network and its structures. This data can then be fed directly into the Leiden algorithm at each time period. This novel framework is very intuitive and flexible, as the studied duration at each time period can be customized to fit the application at hand, and different clustering algorithms can be utilized without issues. In this paper, we use 3 months (a quarter) as the preferred period of study and the Leiden algorithm as our clustering method of choice because this combination ends up achieving the balance between stability and sensitivity that we strive for.

3.2 Data handling

Dynamic graphs offer a unique chance to study the evolution of our networks. In this analysis, we focus on the period between January 2018 and January 2022, which encompasses times of stability and relative peace, as well as times of great unrest with the advent of Covid-19. The data is organized into weekly snapshots observed every Monday. At each point of observation, we make use of data collected over the preceding 3 months. This data manipulation was performed directly on the company's proprietary AWS (Amazon Web Services) servers using `SQL`. This ingrained "memory" mechanism provides fluidity to the evolution of the network without hindering major shocks from manifesting. However, the massive scope of the data is challenging to deal with, as it creates unnecessary clutter and partially conceals interesting features of our dynamic networks. To account for this issue, we lightly prune the data by removing links with less than 25 mentions accumulated over each 3 month period. This filters out inconsequential links while maintaining the integrity of our network structure. We then employ more advanced screening by focusing on the largest connected subgraph within the network, *i.e.*, the biggest subset of the graph (by vertex number) where every node is connected, be it directly (direct edge between the nodes) or indirectly (multiple hops through edges). This stems from the data-based observation that the most influential actors in the economy are deeply connected and can form their own expressive graph. Figure 1 clearly demonstrates this observation. The figure showcases different metrics for information loss resulting from the reliance on the largest connected subgraph. Panel (a), which displays the number of nodes in both the full graph and the largest connected subgraph over the studied time period, shows that we are removing quite a bit of vertices. However, panels (b) and (c), which focus on the evolution of the number of edges and sum of edge weights respectively, reveal that the loss in either metric is barely noticeable. This implies that the dropped vertices, although numerous, were only connected by a few lightly-weighted edges, and that their impact on the overarching structure of the graph is rather minuscule. It is worth noting the huge shift in node count, edge count, and sum of edge weights around March 2020, which marks the point at which Covid-19 had begun to fully take over and reached pandemic status according to the World Health Organization. Since our pruning mechanism drops edges with less than 25 mentions over a 3 month "memory" period, these shifts imply that the increased activity attached to the virus affected a great number of nodes, allowing new entrants to pass through our pruning system. These jumps will be examined and expanded upon in the sequel.

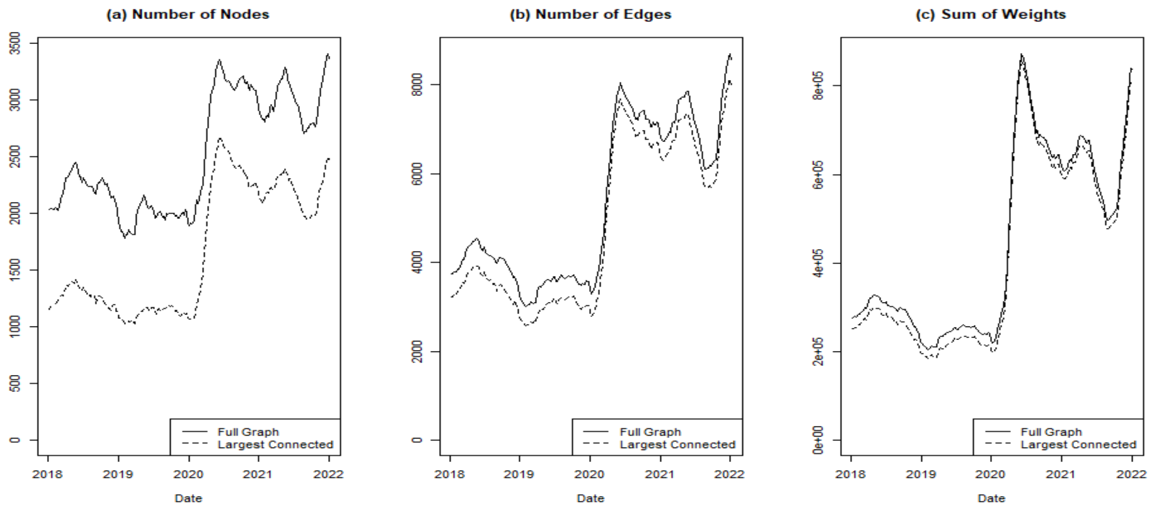


Figure 1: Studying information loss caused by using the largest connected subgraph through the weekly evolution of (a) the number of nodes, (b) the number of edges, and (c) the sum of edge weights.

After establishing the efficacy of our pruning, we can now begin delving into the data. We will first visualize our network at key points in time, then analyze the strongest nodes, links, communities, and companies within our dynamic network before and after Covid-19 using our weekly snapshots.

3.3 Visualization

While network data, particularly in its dynamic form, is not the easiest to unravel analytically, it makes up for it in spades in terms of visualization efficacy. To emphasize this dynamic quality of our data, we leverage `R` and the powerful `igraph` package, as well as `Gephi` and `SigmaJS`[2] to create an informative and interactive timeline of clustered networks, which can be found at https://ya-tls.github.io/world-economy-dynamics/networks/2018-01/. The timeline uses the aforementioned weekly data between January 2018 and January 2022, and utilizes the Force Atlas 2 placement algorithm [Jacomy et al., 2014] to create a suitable layout for each point in time. To ensure the legibility of our plots, we control the size of the nodes to match their degree, which is the number of links or edges connected to each vertex. This allows KPIs with high connectivity to shine in the midst of their lesser competition. We also tie each edge's width to its weight to give more prominence to links with large numbers of mentions. In addition to these features, we partition and color-code our graphs using the Leiden algorithm with the resolution parameter $\gamma = 1$, which facilitates tracking the evolution of community structures and sizes over time. This powerful timeline enables users to fully interact with the networks and their components, which presents a perfect gateway into the more advanced analysis we carry out in the subsequent sections. The timeline shows a rather consistent ensemble of nodes and clusters from January 2018 to February 2020, with nodes from

[2]The SigmaJS open-source library: https://www.sigmajs.org/

the world (macro indicators), USA, food products, oil and gas consumables, and capital markets prefixes forming their own respective clusters and dominating the graph. However, starting from around March 2020, we see massive shifts in our network structure with the "pandemic" nodes taking over the graph, which coincides with the WHO declaring Covid-19 as a pandemic. Since interactive timelines are not feasible visualization tools within the confines of this paper, and to give readers a taste of what the timelines can bring to the table, we visualize key points in time using `Gephi` and the Force Atlas 2 layout algorithm. The results can be found in Figure 2, where panels (a)-(f) respectively show our networks color-coded according to cluster size (number of nodes within the cluster) for the first week of January 2018, which is the first observation period, the last week of December 2019, which marks the last period of relative calm before Covid-19, the first weeks of both March and April 2020, which correspond with the WHO's pandemic announcement and the subsequent spread of the virus in Europe, the first week of June 2020, which is around the time the virus truly ravaged the USA (more than $100,000$ deaths and two million recorded cases; the period also coincides with the point at which the maximum sum of weights is achieved in panel (c) of Figure 1), and the final observation period in the first week of January 2022. We also present a small node dictionary to introduce the most influential nodes in Table 1. We will use this figure to give a summarized explanation of the network and community evolution seen in the interactive graph.

KPI	Meaning
world-gdp	A "macro" (world) indicator for production all around the world
capital-stock_market	An indicator centered around the stock market and its fluctuations
oil_gas-prices	A price indicator for oil and gas products and their many derivatives
food-production	A production indicator revolving around the worldwide supply chain for food products
world-holiday	A macro event indicator related to holidays celebrated around the world, especially Christmas
world-human_rights	A macro indicator that centers around human rights violations or discussions, specifically related to healthcare access and poverty
world-population	A macro indicator that focuses on birth, mortality, and immigration rates for populations, as well as casualties from disease/war
usa-project	An event indicator that relates to US government sponsored plans in the education, health-care, energy, capital markets, and agricultural sectors
usa-reform	An event indicator that is mostly concerned with US policy regulations and tax cuts/reforms
gold-prices	A price indicator for gold
education-usage	An indicator related to the usage and proliferation of new educational tools, specifically virtual and remote learning
pharmaceuticals-usage	An indicator centered around the usage and proliferation of pharmaceuticals, specifically the newly created Covid-19 vaccines

Table 1: Node dictionary for the most influential nodes in Figure 2.

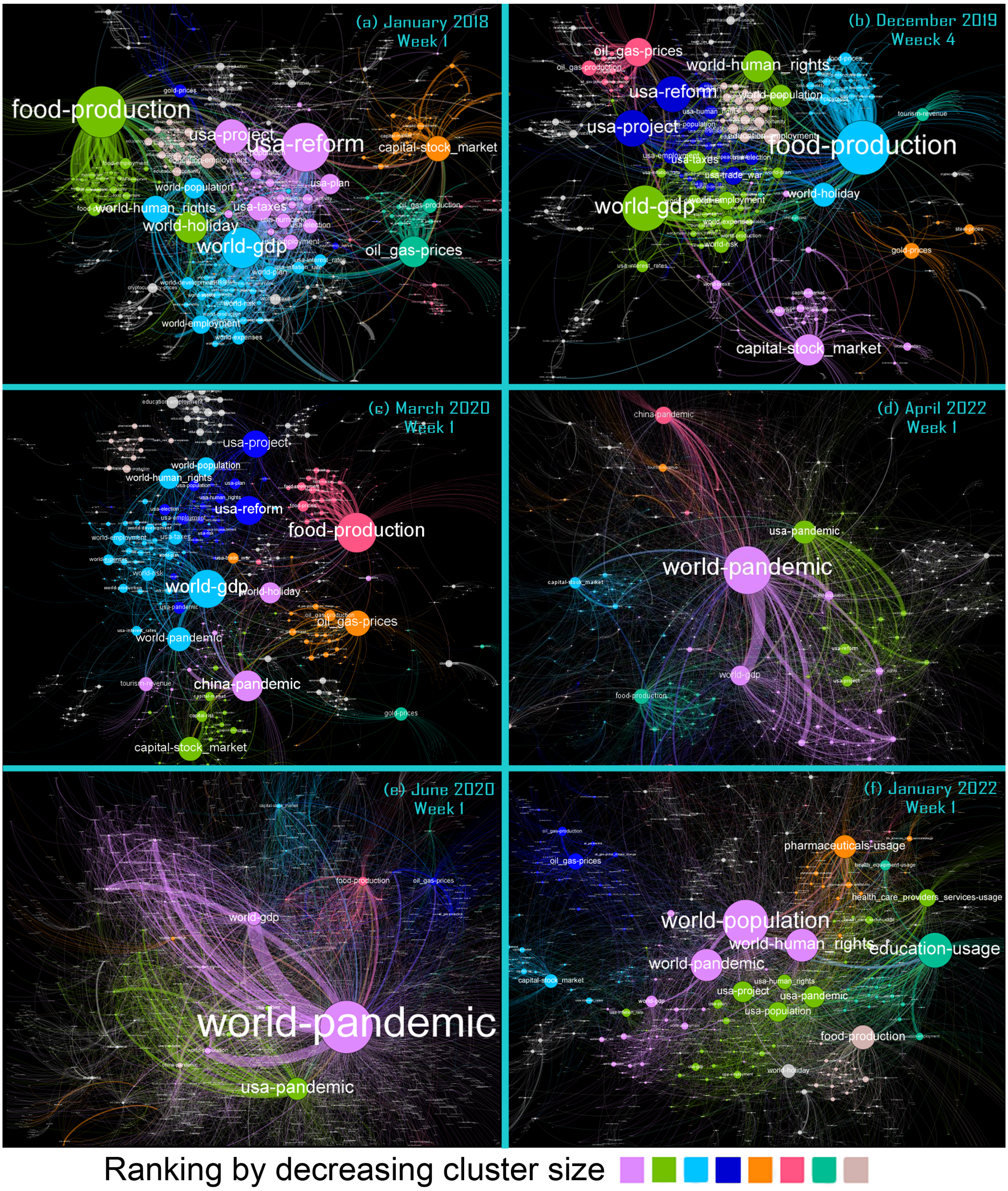


Figure 2: Clustered network visualization using weekly data for (a) the first week of January 2018 (first period), (b) the last week of December 2019, (c) the first week of March 2020, (d) the first week of April 2020, (e) the first week of June 2020, and (f) the first week of January 2022 (last period).

Panels (a) and (b) show a rather stable lineup of nodes and communities, even if those communities are trading spots in the cluster size rankings. Nodes from the food products prefix, specifically "food-production" seem to be highly influential, with the latter enjoying the highest degree (*i.e.* number of edges connected to each vertex), as represented by the node's size in both panels. Vertices related to macro indicators such "world_gdp" and "world_population", as well as those centered around projects, plans, reforms, and taxes in the USA also share center stage, with the latter exerting a powerful magnetic pull and creating the largest cluster in terms of node count in January 2018. This top spot is taken over by the capital markets cluster in December 2019, forming a high-powered and well-populated community around the stock market, specifically the New York Exchange in the form of "capital-stock_market". The oil and gas industry is also well represented, as it seems to pull a consistently large cluster in both panels, and the price indicator "oil_gas-price" enjoys respectable prominence. We start seeing a sizable shift in community structure as Covid-19 enters the picture in panel (c), with "china-pandemic" and "world-pandemic" suddenly jumping to the forefront of the network in March 2020. The relative stability of the graphs in panels (a) and (b) then violently implodes as Covid-19 spreads in April 2020, as panel (d) clearly shows the pandemic nodes "world-pandemic" and "usa-pandemic" dominating the graph and forming the two largest clusters in the network. Panel (e) provides an even more extreme visual of this dominance in June 2020, as the powerful nodes of the first two panels fade into the background, which further illustrates the severity of the pandemic's impact on every single aspect of the economy. Panel (f), on the other hand, shows the world economy in a state of recovery and adaptation, with nodes like "education-usage", "pharmaceuticals-usage", and "health_care_providers_services-usage" enjoying a huge push in the media, the first due to the seismic blow dealt to the education industry during the pandemic and the many efforts taken to re-establish some sense of the normality, and the other two reflecting the extensive media coverage of the pandemic's human toll and the numerous vaccines that attempted to counter and limit its spread. This short analysis of the panels reveals how much potential these types of datasets hold: a six-panel figure gave us enough information to write a concise yet insightful exposition about the evolution of the world economy. We strongly encourage the reader to experience this evolution through our interactive visualization in https://ya-tls.github.io/world-economy-dynamics/networks/2018-01/. We can push this type of analysis much further by employing the tools and metrics of graph theory to concretely identify nodes, clusters, and edges of interest, and track the evolution of companies and their dynamic interaction with the entire graph, as done in the next sections.

3.4 Influential nodes

A simple but intuitive way of identifying the main feature of a graph is to study its most powerful and influential nodes. However, it is not that simple to define what "powerful" and "influential" mean in the context of graph theory, as a plethora of node-related metrics attempt to quantify these concepts in different ways. We explore three such metrics in this section: two well-known devices, and one completely novel measure. The first is the strength, also known as the weighted degree, which can be expressed as

$$\text{Strength}_i = \sum_j w_{ij},$$

with i and j as nodes and $w_{ij} \in \mathbb{R}^+$ as the weight of the connection between the two nodes, or equivalently, the number of times they were linked in written media ($w_{ij} = 0$ implies the lack of an edge). In essence, this metric sums the weights of all connections linked to a certain node, which gives us a measure of its importance in the graph. Nodes with higher strength often feature as hubs of influence in weighted graphs, but the metric can be somewhat deceiving if a few links enjoy an abnormally large weight. To remedy this defect, we utilize a more refined way of quantifying the influence of a node while equally weighing the different edges connected to it. The metric is known as betweenness centrality (see p.47 of Kolaczyk and Csárdi [2014]), which relies on the number of "shortest paths" passing through each vertex. In essence, one can find a set of shortest paths between pairs of vertices, *i.e.*, the set of paths that connect two nodes (not necessarily with a direct edge) with the least required number of edges, or "jumps".

For nodes u, v, and i, the betweenness centrality of node i can be defined as follows:

$$\text{Betweenness Centrality}_i = \sum_{u \neq v \neq i} \frac{\sigma_{uv}(i)}{\sigma_{uv}},$$

where σ_{uv} is the number of shortest paths between nodes u and v, and $\sigma_{uv}(i)$ is the number of shortest paths between u and v that pass through node i. Vertices with high centrality are often regarded as highly influential, as they can exert control over the network by providing short traversal paths between otherwise disconnected nodes. Combined with strength, betweenness centrality can help identify truly powerful nodes within these huge, ever-evolving networks.

However, these two measures are predominantly global: they utilize all available edges with no consideration for whether a node can accumulate its own community. A novel way to examine a node's gravitational potential is by clustering the data, then checking whether said node "dominates" the cluster. To establish dominance, we use the Leiden algorithm explained in Section 3.1 with $\gamma = 1$ to partition networks, and then treat each community as its completely separate graph. Inside these subgraphs, we then re-compute betweenness centrality, and declare the node with the highest value of the metric as dominant. We choose betweenness instead of strength because the former is more concerned with centrality, which is exactly the quality we are looking for.

Using the above setup, we can now spot nodes with exceptionally high numbers of mentions using the strength, nodes that are highly central and pivotal to the overall connectivity of the graph using betweenness centrality, and nodes with significant gravitational pulls using the cluster dominance criterion. However, since we are dealing with dynamic graphs, we cannot closely examine each data snapshot separately. We opt for a more holistic approach by finding the top 5 nodes according to the first two metrics for each snapshot, and then counting the number of times a node was featured in the top 5 over time, divided by the length of the studied time period. This essentially generates the proportion or percentage of time a node managed to crack the top 5 for each metric. The third criterion, cluster dominance, is dealt with in similar fashion, except the top 5 rankings from each snapshot would correspond to the size of the clusters (number of nodes in a cluster) dominated by the node, so that nodes would be required to pull in large clusters to be considered for the ranking. We also choose to study the data in two portions: the pre-Covid-19 data, extending from January 2018 to February 2020, and the post-Covid-19 data, extending from March 2020 to January 2022. We choose March 2020 because it marks the pivotal point where the WHO declared Covid-19 as a pandemic. This allows us to analyze the state of the world before the advent of the infectious virus and contrast it with the times of its spread and aftermath. Figure 3 showcases a nested piechart of the top nodes for each metric, where the inner and outer charts present results for the pre and post-Covid-19 data respectively. On the charts, we only display node names and percentages for nodes that feature in the top 5 for each metric more than 10% of the time, while the legend shows all relevant information.

We first focus on the pre-Covid-19 era. The inner circles of Figure 3 reveal that "food-production" and "world-gdp" are extremely influential in both a global and local sense in the pre-virus era, as they successfully maintain high spots for strength (panel (a)), betweenness centrality (panel (b)), and cluster dominance (panel (c)). For example, "food-production" achieves 100s for both strength and betweenness, which implies that the node was in the top 5 vertices for each metric in every week before the advent of Covid-19. It also sits at a 91.1 for cluster dominance, which indicates that it managed to dominate one of the five biggest clusters for 91.1% of the weeks in the pre-virus period. Meanwhile, nodes like "usa-project" and "usa-reform" perform fairly well for the first 2 metrics but fall short in the cluster dominance department, which implies that these nodes do not enjoy the same gravitational pull as the two vertices discussed earlier. The oil and gas prices node follows a somewhat similar trend as it cements its position in the top 5 for node strength, but wavers in the other two rankings.

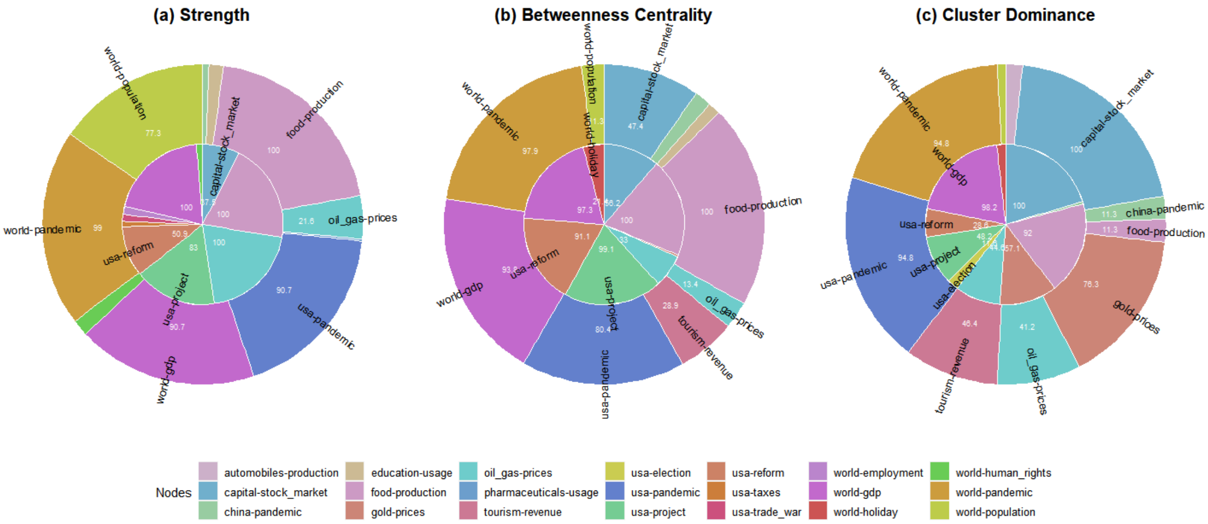


Figure 3: Top nodes by (a) strength, (b) betweenness centrality, and (c) cluster dominance. The inner chart represents 112 weeks between January 2018 and February 2020, while the outer chart represents 98 weeks between March 2020 and January 2022.

This suggests that while oil and gas prices receive extensive media attention, discussions of their impact seems to be rather isolated to the field and its immediate neighbors which does not translate into consistent placement for metrics that require high centrality and clustering potential. On the other hand, the stock market node falls behind in terms of strength but picks up some steam in the centrality rankings, and achieves a perfect 100% rate for the cluster dominance criterion. The node, which corresponds to stock markets and their fluctuations, seems to exert a powerful draw on a large number of nodes, which makes sense given the real world implications of this influential market. An even more extreme example of a node that falters globally but excels locally is the gold-prices vertex, which falls short of the top 5 for the first two metrics while performing fairly well for clustering potential (more on this in Section 3.6). Another interesting observation is that USA-based nodes are heavily featured at the top, while those of other economic superpowers, say China, are missing. Whether this is the result of an American leaning bias in western media, a lack of concrete data that complicates parsing relevant KPIs about China, or a combination of the two is unclear. This analysis, however, reveals very interesting quirks about the representative quality of data gathered from western media, and suggests that their view of the world is quite USA-centric.

We now turn our attention to the post Covid-19 era. The outer circles of Figure 3, tell a strikingly different story. The pandemic nodes "world-pandemic" and "usa-pandemic" reign supreme over nearly all three metrics, which is indicative of the severe impact of the virus on the world economy and the extensive coverage it subsequently received. It is worth noting that while the "china-pandemic" node does make it into the top 5, it again did not receive the same attention as its world and USA counterparts. The "world-population" node, which mainly focuses on access to healthcare service, makes its debut into the top 5 with an extremely strong showing in the strength criterion, which sheds light on the media's scrutiny of the accessibility to and performance of the healthcare system in face of crisis. Another newcomer is "tourism-revenue", which seems to find some footing in the betweenness centrality and cluster dominance metrics, highlighting the drastic toll of pandemic-related lockdowns on the industry. Returning faces like "food-production" and "world-gdp" continue to perform very well in the strength and centrality criteria, but their cluster dominance is much diminished by the overwhelming pandemic nodes. However, the stock market and gold prices nodes retain their clustering

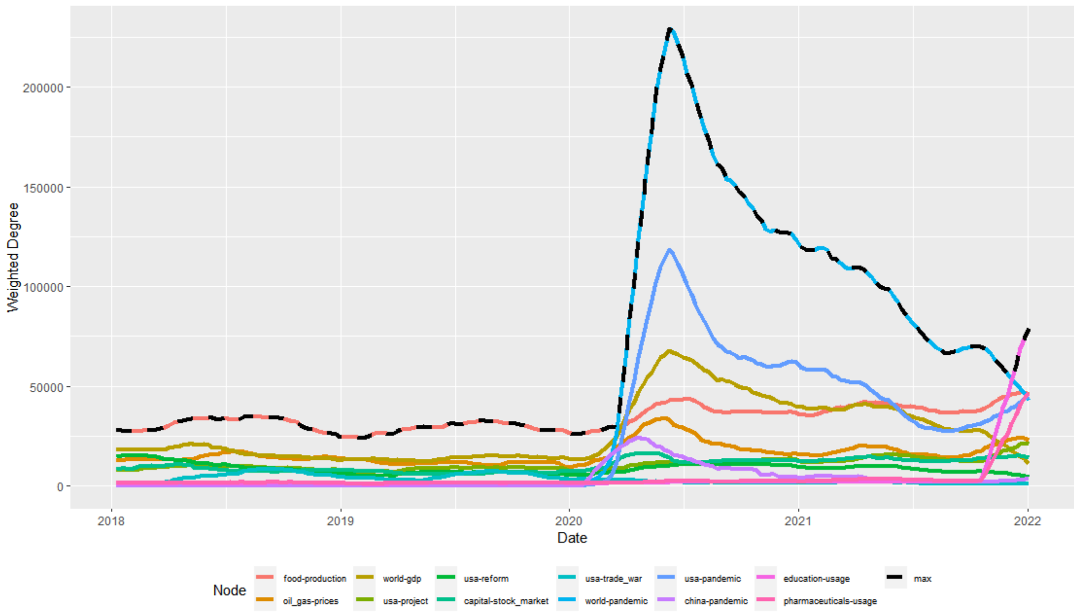


Figure 4: Weekly evolution of strengths for nodes of interest.

potential, which exemplifies their communal power. Examining this figure motivates studying the evolution of some of these metrics over time for our influential nodes. Figure 4 showcases the strength of some nodes of interest as solid lines, in addition to the maximum recorded metric as a dashed line, between January 2018 and January 2022, while Figure 5 displays their normalized betweenness centrality in a similar fashion; normalization is done through $2 \times (\text{Betweenness Centrality})/(n-1)(n-2)$, where n is the number of nodes in the network.

Figure 4 tells a very direct story: the three pandemic nodes accumulate negligible strengths between January 2018 and February 2020, where the food production node consistently dominates the metric. However, March 2020 sees the three nodes of interest begin a meteoric rise, with the USA and "world" pandemic nodes overtaking food production by a wide margin, and the "world"-based vertex establishing a comfortable lead over all competitors. It is worth noting that all three pandemic nodes peak rather early around May 2020, and then begin a systematic decline which signifies that the panic around the pandemic was slowly tapering off. However, even during this decline, the "world" pandemic node retains its position at the top, implying the devastating impact Covid-19 continues to have long after it first struck. Another interesting observation revolves around the "world-gdp" node, which also overtakes the dominant food production node around the beginning of the pandemic's expansion. This can be attributed to the catastrophic economic impact of the virus and its aftermath, as the potent food node seems to regain its lead on GDP as the pandemic slows downs. Figure 5 hits mostly similar narrative notes, as food production again dominates the graph between for the majority of the pre-Covid-19 era, save for a brief period in the beginning of 2018 where "usa-reform" took the lead, only to get completely crushed by the world pandemic node after March 2020. The latter's betweenness centrality spikes to unprecedented levels that it manages to maintain over the remaining duration of 2020 and the first three quarters of 2021. While the USA and China Covid-19 nodes fail to rise to the same heights as their "world" counterpart, their betweenness centrality still enjoys a sizeable bump around March 2020. World GDP again overtakes food production after the pandemic strikes, which is consistent with the behavior of the strengths seen earlier. These shifts in centrality, particularly that for the world pandemic node, is emblematic of the severe impact of Covid-19 and the chokehold it places on many different avenues of the world economy, bringing together vertices that might have been isolated before the emergence of the pandemic.

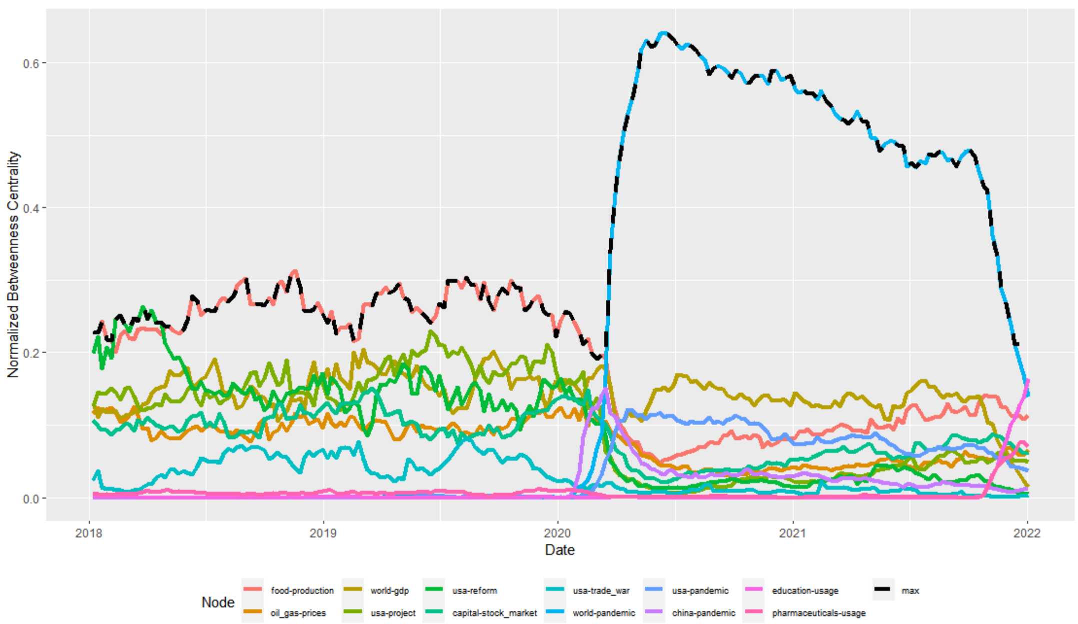
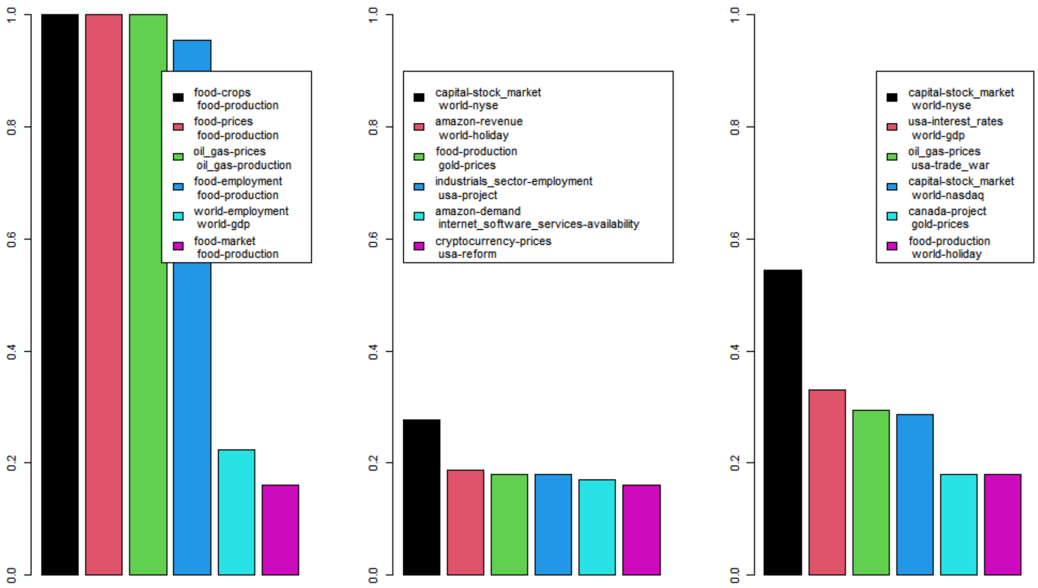


Figure 5: Weekly evolution of betweenness centrality for nodes of interest.

3.5 Influential links

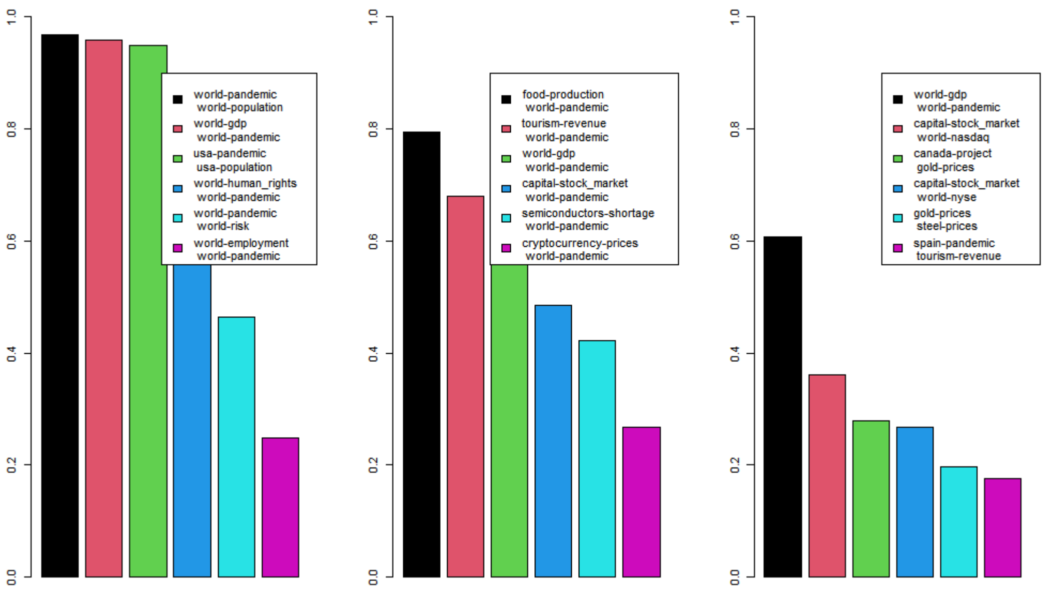
Having explored the most influential nodes in our network, the next logical step would be to study the links between them. We wish to unravel links that truly play a defining role within our network structures. One way to identify such links is to simply look at their weights, which reflects the number of mentions a specific link has received. While extremely straightforward and intuitive, the metric does not concern itself with the edge's contribution to the network's connectivity, *e.g.* it can end up focusing on intra-industry links that are indeed powerful but not as conducive to important inter-industry interactions. To this end, we can also evaluate links using the edge betweenness concept, which is simply a link analogue to the node betweenness centrality concept discussed earlier. Similarly to its node counterpart, this metric can reveal links that are inextricable to the overall connectivity of the graph, which should allow vital inter-industry connections to shine. However, since these two metrics are very global in nature, they tend to ignore the local impact of links that might be key to internal community structures within the graph. To capture this particular feature, we borrow the cluster dominance concept introduced in Section 3.4, but we apply it to edges by pinpointing links with the highest edge betweennes inside each subgraph. This allows us to highlight edges that are paramount to within-cluster connectivity. We look at the top 5 edges from each time period for each metric (and the cluster dominance ranking is determined by the size of the cluster "dominated" by the link) and then create a holistic ranking for the pre and post Covid-19 eras as seen in the last section. However, we choose to present the results as bar plots instead of piecharts, as tracking the top 5 edges from each time period produces an extremely large number of edges to keep an eye on, and bar plots allow us to limit the presentation to interesting edges. We plot the results in Figures 6 and 7 for the two respective periods.

Figure 6 focuses on the pre-Covid-19 era and sees a key link from the oil and gas industry and a plethora of powerful food chain-related edges dominate the edge weight category. Capital markets and stock exchanges receive better representation in the edge betweenness and cluster dominance categories, with the link between the New York Stock Exchange and the overall stock market topping both metrics. The connection between Amazon revenues and the "world-holiday" node takes second place in the edge betweenness ranking, which emphasizes Amazon's ever expanding role in satisfying consumer demand. The intricate connection between USA interest rates and the macro indicator "world-gdp" receives some well-deserved attention as it



(a) Edge Weights (b) Edge Betweenness (c) Cluster Dominance

Figure 6: Top six links by (a) edge weight, (b) edge betweenness, and (c) cluster dominance over the 112 weeks between January 2018 and February 2020.



(a) Edge Weights (b) Edge Betweenness (c) Cluster Dominance

Figure 7: Top six links by (a) edge weight, (b) edge betweenness, and (c) cluster dominance over the 98 weeks between March 2020 and January 2022.

contends for the top spots for cluster dominance. It is worth noting that the top links for the edge betweenness and cluster dominance metrics seem to struggle for consistency, with all but one of the strongest links from both categories featuring in the top 5 less than half the time. This high variability implies that links might be strongly tapping into the ever-changing news cycle, which focuses on transient scoops and stories while consistently making time for oil, gas, and food production. The post-Covid-19 results in Figure 7 are a lot more straightforward to interpret, with pandemic-related nodes making up 14 of the 18 links (top 6 links for each of the 3 metrics) on display. The pandemic's impact on population mortality rates, world GDP, food production, and capital markets clearly dominate all three metrics. Studying these links also highlights the pandemic's connections to the decline of the tourism and hospitality sectors, as well as the semiconductor supply shortage[3] and the cryptocurrency boom[4] witnessed around the beginning of the pandemic, which are two important events that we could not detect in Section 3.4. The post-Covid-19 era also sees the top links in the edge betweenness and cluster dominance categories achieve significantly higher consistency, which suggests that the news cycle was overwhelmed by the pandemic and its impact on the world economy.

3.6 Influential clusters

After taking a deep dive into the most influential nodes and links in our network in Sections 3.4 and 3.5, both before and after the advent of Covid-19, it would do us well to take a more holistic look at our data through the use of communities and clustering algorithms. Instead of identifying individual hubs of interest, clustering allows us to recognize powerful conglomerates of nodes that play an essential part in the economy. To this end, we again utilize the Leiden algorithm with $\gamma = 1$, thus fully allowing the data to speak for itself. Before delving into the structure of these clusters, we first track the number of generated communities over time between January 2018 and January 2022 in Figure 8. The figure shows a relatively consistent number of clusters between January 2018 and February 2020, followed by a sudden explosion around March 2020, which mirrors the jumps we saw in Figure 1, as well as the shifts in Figures 4 and 5. This is again indicative of the influx of new nodes attached to the increased activity around the rapid spread of the virus.

After analyzing the evolution of the number of generated clusters, we can now focus on understanding their structure and contents. However, one issue that arises with this novel type of analysis is that of naming the newly formed communities. We opt for an intuitive strategy that makes direct use of vertex names, particularly their prefixes. For each cluster, we track the number of vertices belonging to the same prefix inside the community and name the cluster by the prefix with the most vertices. This should give us a general idea about prefix hierarchy within each cluster. However, this method alone fails to account for highly heterogeneous clusters with many competing prefixes. To deal with these troublesome clusters, we add the caveat that the "dominant" prefix should account for at least 25% of the cluster's vertex count, and no other prefix should achieve that threshold within the same community. Clusters that violate these conditions will simply be referred to as "No Dominance". In the case of a single prefix dominating multiple clusters or the formation of many "No Dominance" clusters, which would yield duplicate cluster names, we order these communities by size (node count) and bestow the prefix name (or the aforementioned "No Dominance") on the most populated cluster. The remaining communities are then given numbered names (*e.g.* "No Dominance 1" for the second largest heterogeneous cluster). At each point in time, we rank clusters by size (number of vertices) and sum of edge weights within the cluster. The first ranking system rewards clusters with powerful hubs and strong magnetic pulls, while the second favors the existence of important

[3]Howley, D. (2021). These 169 industries are being hit by the global chip shortage, *Yahoo Finance*, April 25. Link.

[4]Locke, T. (2021). From bitcoin hitting $1 trillion in market value to Elon Musk's dogecoin tweets: 12 key crypto moments from 2021, *CNBC*, December 27. Link.

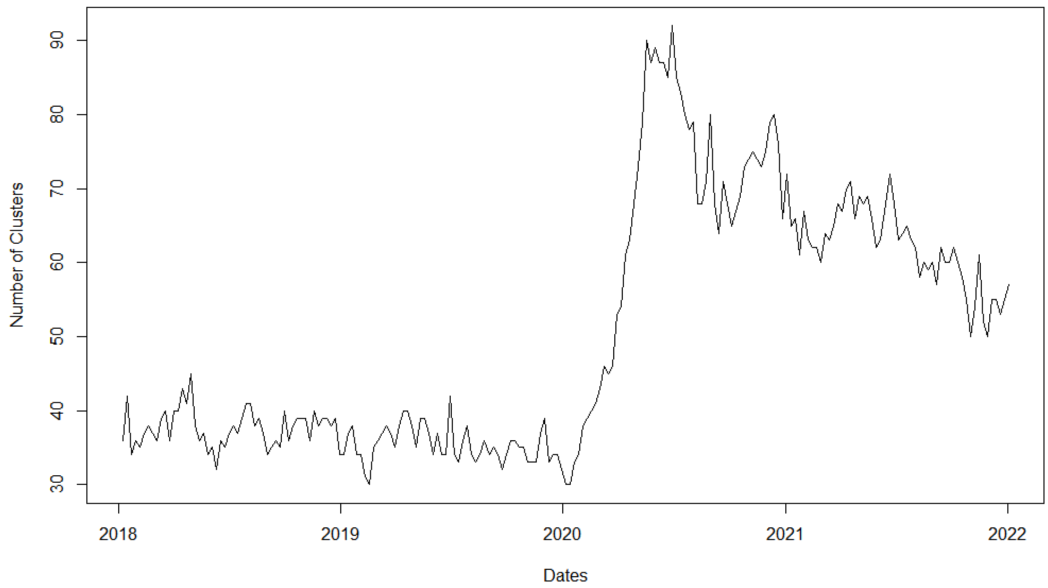


Figure 8: Evolution of the number of clusters obtained on a weekly basis.

edges within the community. For each ranking mechanism, we look at the top 5 clusters for each time period and then create a summarized leaderboard that tracks how many times a cluster featured in said top 5, divided by the maximum number of features, similarly to how we dealt with our influential nodes and links. We again split the data into the pre and post-Covid-19 datasets to address the potential heterogeneity in our datasets. The inner and outer charts of Figures 9 display the results for the two respective periods. Only clusters achieving 10% or higher are displayed on the piechart, while the legend showcases all relevant information.

It is clear from the figure that clusters dominated by the USA, world (macro indicators), food products, capital markets, and the oil industry prefixes dominate the largest clusters in terms of both vertex count and sum of internal edge weights. This list includes most of the usual suspects when talking about main players in the world economy. However, China is again a notable absentee, as one would expect the world's second biggest economy in terms of GDP to be at the top of such a list. This reflects the unbalanced nature of the collected data, which is mainly due to the predominance of the Western media's view of the world and its economy. We study the extent to which the Eastern superpower makes a tangible impact in our network later on in this section. It is also peculiar that "No Dominance" clusters were so prevalent at the top in terms of cluster size, especially in the post-virus era. To elaborate on this interesting observation, we first find the top node by strength and betweenness centrality inside the 2 most populated heterogeneous "No Dominance" clusters for each time period, then summarize these rankings in a similar fashion to what we have done earlier. We also track the top prefix by node count in each of the two largest "No Dominance" clusters in each time period and collate them into a concise ranking similarly to the first 2 metrics. We again present the results as proportions in the inner and outer circles of Figure 10 for the pre and post-pandemic eras respectively, and we restrict node names and prefixes on the chart to those that achieve at least 10%, while the legend shows everything.

The figure leaves very little room for doubt about which set of nodes overwhelms the "No Dominance" clusters. The gold prices vertex consistently tops the strength and betweenness centrality metrics, and the gold prefix outshines the rest in terms of node population within these clusters. In fact, it seems that these clusters are mostly centered around metals, as the "metals" and "steel" prefixes also feature quite heavily in the prefix rankings. These obser-

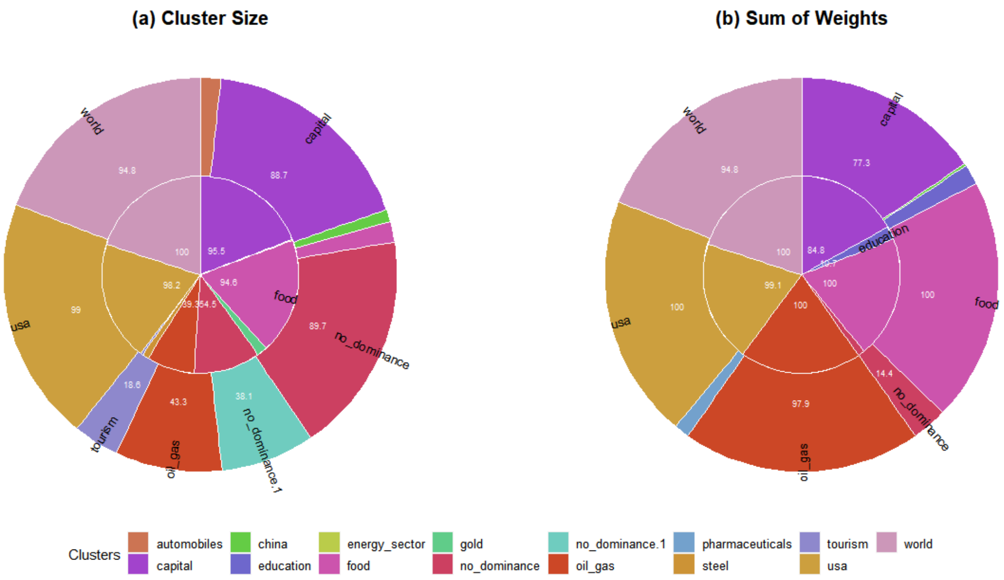


Figure 9: Top 5 clusters by (a) size and (b) sum of weights. The inner chart represents 112 weeks between January 2018 and February 2020, while the outer chart represents 98 weeks between March 2020 and January 2022.

vations make it quite clear that these expansive clusters assemble closely related industries to create heterogeneous yet logically consistent communities. On a different note, the tourism revenues vertex and tourism-related nodes in general establish a decent foothold in the post-virus era, which seems to be inline with the huge losses the industry suffered as the pandemic shut down travel. The "No Dominance" clusters, while vague in nature due to their name, seem to have their own convincing narratives to discover and explore.

Now that we have uncovered the mysteries of the "No Dominance" clusters, we can return to a more complete study of communities. To this end, we collapse each community into one node, and construct a new graph using these new aggregated vertices. These nodes will receive their respective cluster names. This yields nodes that are named after companies, industries, regions, and countries. To connect these community-representing vertices, it is sufficient to sum up the weights of connections between members of each community, *i.e.*, for two communities indexed by u and v,

$$\omega_{u,v} = \sum_{i \neq j} w_{ij}\, \delta(c_i, u)\delta(c_j, v)$$

will be the weight of the connection between collapsed vertices u and v, where w_{ij} is again the weight of the connection between the original non-collapsed vertices i and j, and c_i is the index of the community containing vertex i. We can then compute the strength and betweenness centrality of these collapsed vertices and track the top 5 nodes for each period, which leads to a summarized ranking in keeping with our ongoing trend. The results are collated in the inner and out circles of Figure 11 for the pre and post-pandemic eras respectively. We again only display cluster names on the chart when the 10% threshold is cleared, while the legend requires no such condition.

The results in the inner chart of Figure 11 are mostly consistent with those in Figure 9 for the pre-Covid-19 period, with collapsed nodes related to food products, USA, world, oil and gas, and capital markets dominating both the strength and betweenness centrality metrics. However, the post-pandemic results in the outer chart of Figure 11 slightly differ from those in Figure 9 in that they allow the pharmaceuticals and education "nodes" to shine, which reflects the massive influx of coverage related to these two sectors after the virus struck. China also barely makes it above the 10% threshold for the first time in our analysis, which emphasizes

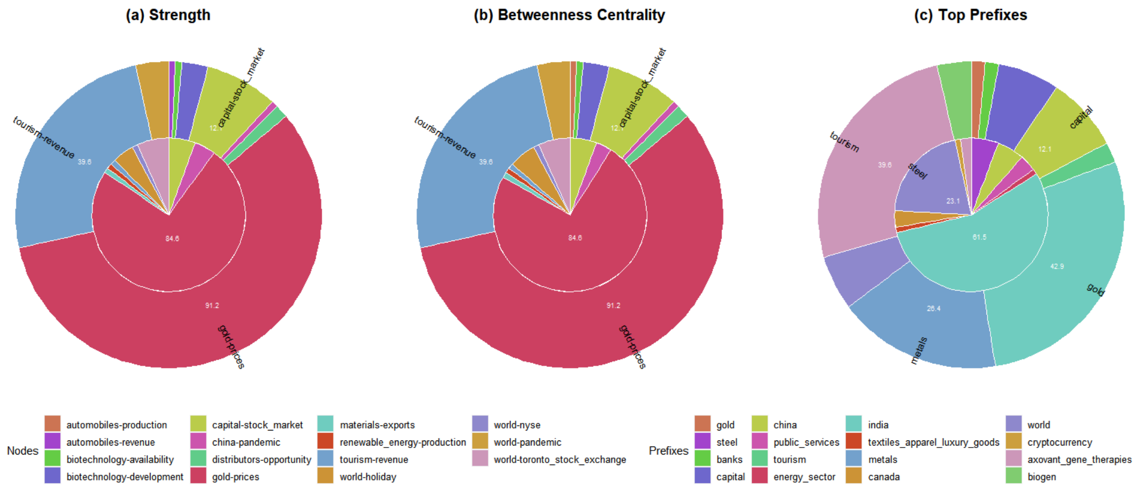


Figure 10: Understanding "No dominance" clusters through ranking their nodes by (a) strength and (b) betweenness centrality and (c) tracking their top prefixes. The inner chart represents 112 weeks between January 2018 and February 2020, while the outer chart represents 98 weeks between March 2020 and January 2022.

the severe understatement of the country's role in the world economy.

3.7 Influential companies

Our study of influential nodes, edges, and clusters is very informative, but also very holistic. It would be interesting to see how the methodologies established in the earlier subsections can be used to examine a specific subset of economic actors, or even a single actor. To this end, we choose to take a closer look at companies, as they are arguably the lifeblood of the world economy. A rather simple but telling approach would be to rank the top company-related nodes over time. Figure 12 shows the proportion of times a company-related node managed to break into the top 3 company-related nodes according to the strength and betweenness centrality criteria.

The figure sees Amazon demand, revenue, and employment nodes top the charts for both metrics and both time periods. Apple related nodes also consistently feature in the rankings for both strength and betweenness centrality, while Microsoft sees much better representation in the centrality metric, which highlights the company's pivotal role in the functioning of the world economy. Notable company mentions include Pfizer and Astrazeneca, two pharmaceutical companies that were instrumental in the post Covid-19 era. Studying company-nodes provides some interesting observations about which firms occupied much of public interest before and after Covid-19. We can supplement this analysis further by employing the clustering methodologies of Section 3.6 with a small twist. Instead of tracking a cluster's size/sum of weights, a very convincing measure of a company's magnetic pull would be whether the company managed to formulate its own cluster, *i.e.*, a company's nodes took up at least 25% of a cluster's constituents. In Figure 13, we showcase the proportion of times a company was able to attract and dominate its own cluster before and after Covid-19. Again, Amazon performs rather well before and after March 2020, forming its own cluster more than 50% of the time before the advent of the pandemic, and close to 70% of the time after it. Facebook, on the other hand, sees underwhelming performance in the first time period, and then skyrockets to the top after the virus struck. This type of graph also allows us to spot more seasonal stories like the Pacific Gas & Electric Company's ("pg_e" in Panel (a) of the figure) issues with electrical fires in

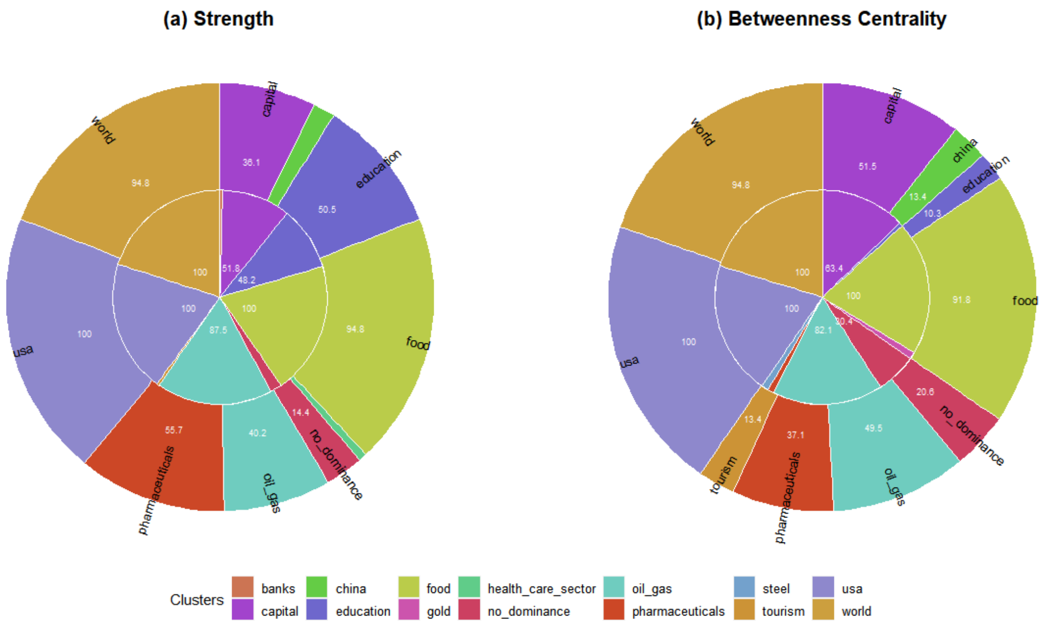


Figure 11: Cluster rankings by (a) strength and (b) betweenness centrality. The inner chart represents 112 weeks between January 2018 and February 2020, while the outer chart represents 98 weeks between March 2020 and January 2022.

California[5]. It is also worth noting that company cluster formation rates increase across the board in the post-pandemic era, which can be attributed to the influx of new nodes into the network (Figure 1) brought about by the virus' impact. However, this influx does not seem to help Pfizer and Astrazeneca, two of the most relevant companies in the post-pandemic era, to pull their own clusters enough times to make it into the top 10 companies for that time period. We study this particular peculiarity at the end of this section.

Examining Figures 12 and 13 separately yields a coherent view of companies in the world economy, but it does not tell the full story. In the case of Amazon for example, there is no guarantee that the top Amazon nodes (demand, revenue, and employment) from the first figure made it into the clusters showcased in the second figure. In other words, while an Amazon cluster might indeed exist, it will not necessarily contain the defining nodes of the global conglomerate, which can cast some doubt on the cluster's true influence. To get a more accurate measure of the cluster's importance, we instead focus on which clusters housed the top three Amazon nodes over time. To be more specific, the cluster needs to contain all 3 nodes to be considered. We use the "split" designation to denote a time period where the three nodes were spread across multiple clusters. We also choose to center the analysis around the post-pandemic period, where node and cluster population is more dense. Figure 14 shows the results, and sees the Amazon cluster on top again, as it manages to maintain the membership of the top company nodes around 70% of the time. Other clusters related to air freight logistics, IT, and not surprisingly, cryptocurrency, contribute in very diminished amounts. The latter can be attributed to the fact that multiple news reports came out tying Amazon to a crypto token release, but none of them have materialized thus far.

We can do a similar analysis for the notable abscentees from the clusters in Panel (b) of Figure 13, Astrazeneca and Pfizer. We again focus on the post-pandemic period, and target their most influential nodes: usage, production, and risk for Astrazeneca, and only the latter two for Pfizer, as seen in the Panel (a) of Figure 12. Figure 15 displays the results, and sees the

[5] Associate Press (2021). PG&E Is Charged With Manslaughter In A California Wildfire That Killed 4, *National Public Radio*, September 24. Link.

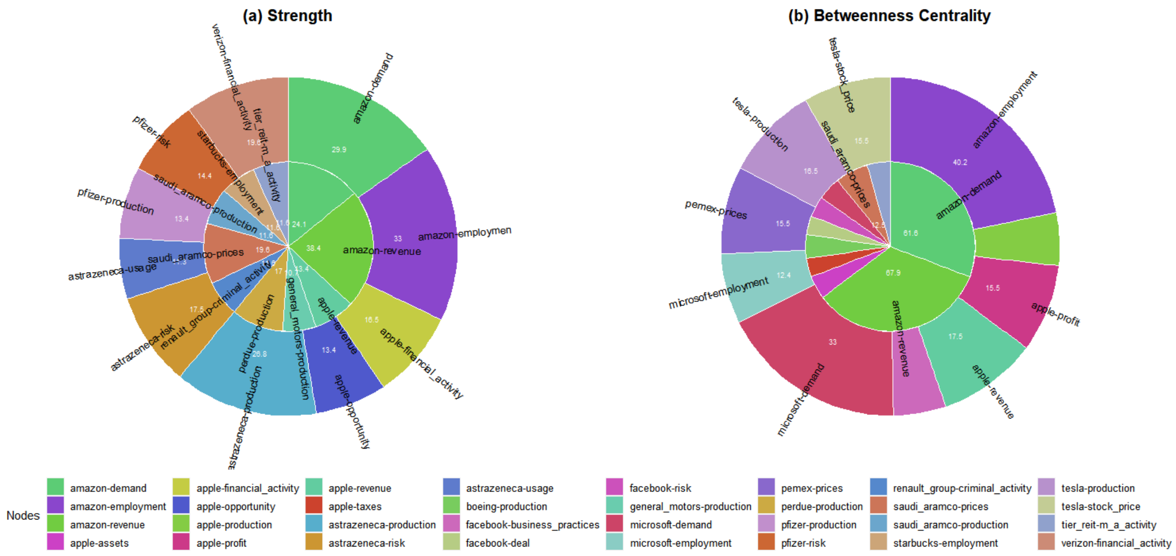


Figure 12: Company-related node rankings by (a) strength and (b) betweenness centrality. The inner chart represents 112 weeks between January 2018 and February 2020, while the outer chart represents 98 weeks between March 2020 and January 2022.

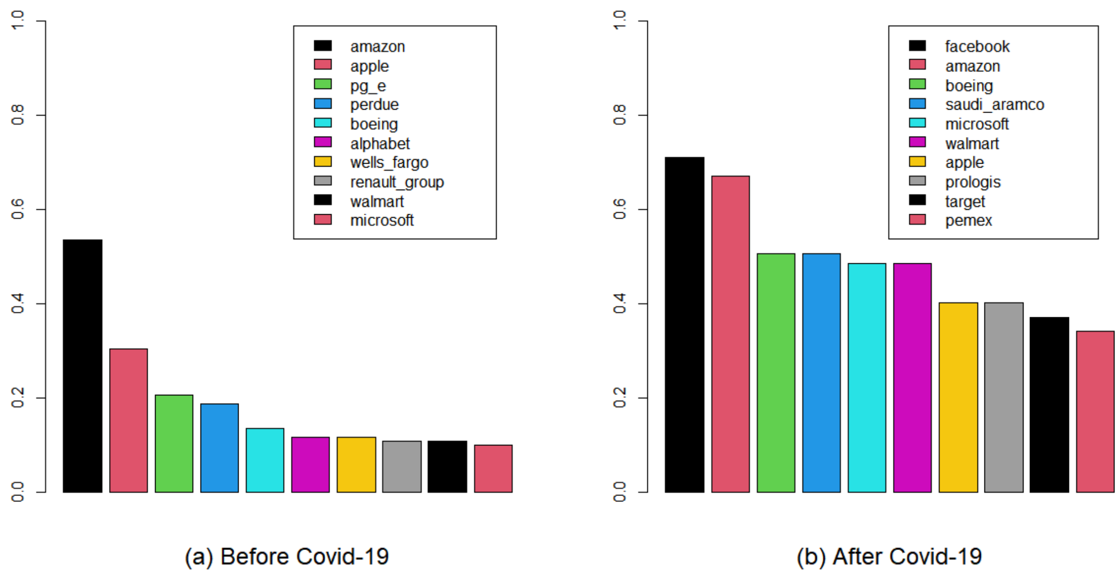


Figure 13: The proportion of times a company was able to dominate a cluster (a) between January 2018 and February 2020 and (b) between March 2020 and January 2022. The figure shows the top 10 companies for each period.

two companies' most relevant nodes get absorbed into the overarching pharmaceuticals cluster. This implies that even though the two companies were extremely influential during the struggle against Covid-19, they were still viewed by the media as a global collection of pharmaceutical companies, sometimes referred to as "big pharma" (in a derogatory manner).

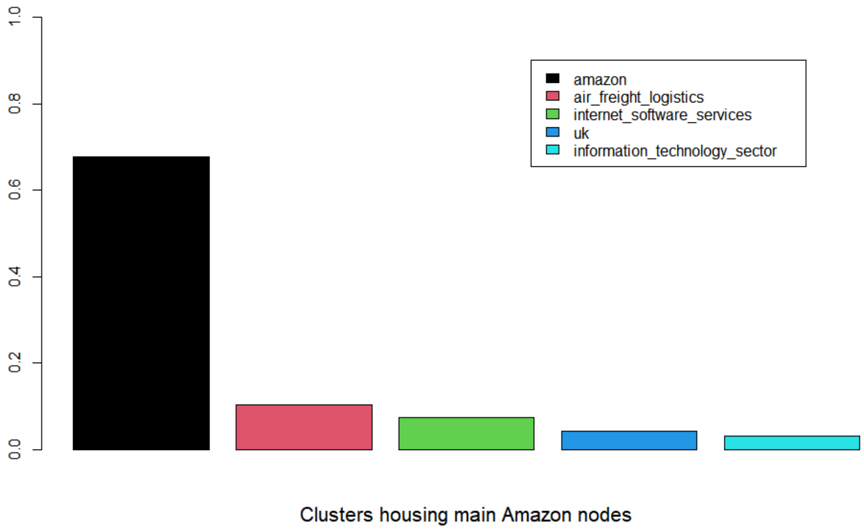


Figure 14: The proportion of times the three main Amazon nodes were placed into the same cluster between March 2020 and January 2022. The figure shows the top 5 clusters in terms of proportion.

Figure 15: The proportion of times the main (a) Astrazeneca and (b) Pfizer nodes were placed into the same cluster between March 2020 and January 2022. "Split" refers to situations where not all nodes were placed into the same cluster. The figure shows the top 5 clusters in terms of proportion for each company.

4 Discussion

In this article, we studied the Western media's view of the world economy through network data and graph theory techniques. Using key performance indicators of companies, industries, and countries as nodes, and the number of mentions in news articles as edge weights between said nodes, we managed to create rather comprehensive snapshots of the world economy through dynamic graph analysis, as these adaptive networks managed to weave a very realistic and concise narrative of the events that transpired around the emergence of the Covid-19 pandemic.

In Section 3.4, we studied the most influential nodes, or economic actors, in our networks over time through the use of the strength, betweenness centrality, and the novel cluster dominance metrics. The results clearly favor nodes like world GDP, food production, capital markets, and oil and gas prices. However, all these powerful nodes get eclipsed by the spectre of the world and USA pandemic nodes, as the latter two managed to establish an indomitable chokehold on the world economy at alarming speeds. Section 3.5 studied influential links within our dynamic networks, and found interesting connections between food, oil and gas, and world (macro) related nodes in the pre-pandemic era. However, these links again take a backseat to the connections formed around the pandemic nodes, specifically those that relate to the pandemic's severe repercussions on population mortality rates, world GDP, food production, capital markets, and the tourism industry, to name a few. Section 3.6 attempted to unravel the inner workings of community formation within our dynamic graphs through a study of the clustering outcomes of the Leiden algorithm. Using a fairly intuitive rule to name clusters, we found that communities dominated by USA, world, oil and gas, and capital markets consistently topped the charts in terms of cluster size and sum of internal edge weights. However, we also saw a decent chunk of data get pulled into "no dominance" clusters, the study of which revealed that one of the main culprits for this phenomenon is none other than gold prices. Section 3.7 focused the analysis to companies. A study of the most prominent company-related nodes saw Amazon and Apple nodes perform rather consistently before and after the advent of the pandemic. Production, usage, and risk nodes attributed to Astrazeneca and Pfizer experienced a rapid rise to prominence in the pandemic era, which was warranted given their influence on the evolution and containment of the virus. We then examined these same companies from a clustering standpoint, and quantified their magnetic pull by tracking the number of times each company managed to create its own cluster, and whether that cluster actually contained the company's most noteworthy nodes. In our brief analysis, we found that Amazon had no issues forming clusters around its mode powerful vertices, whereas Pfizer and Astrazeneca failed to differentiate themselves enough from their overarching pharmaceutical ecosystem to warrant the formation of their own clusters.

Such data has seen limited exploration in applied literature, and we believe that it hides plenty of untapped potential. While our application focuses on visualizing and describing the data, our foray into dynamic graphs suggests that temporal networks might have predictive power as an early warning system. Further research needs to be dedicated to news-based data of this nature, as these concise yet comprehensive summaries of expert opinions in Western written media can be key to new relevant variables and/or models. Other research directions include studying the unbalanced nature of the data at hand, especially in regards to China, as well as the impact of re-introducing directionality into our dynamic networks.

Computational Details

The results in this paper were obtained using `R` 4.2.1, `Gephi`, and `SigmaJS`. `R` itself and all packages used are available from the Comprehensive `R` Archive Network (CRAN) at https://CRAN.R-project.org/. `Gephi` can be downloaded from https://gephi.org/, and `SigmaJS` is available at https://www.sigmajs.org/. All relevant code and data can be found in the Github repository: https://github.com/ya-tls/world-economy-dynamics.

Acknowledgments

This research was financially supported by Causality Link in partnership with TSE-P, a research foundation under the aegis of the Toulouse School of Economics, the French National Research Agency under the grants ANR-19-CE40-0013 and ANR-17-EURE-0010, and the TSE-HEC ACPR Chair "Regulation and systemic risks".

References

Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1):361–362, Mar. 2009. doi: 10.1609/icwsm.v3i1.13937. URL https://ojs.aaai.org/index.php/ICWSM/article/view/13937.

Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008, 04 2008. doi: 10.1088/1742-5468/2008/10/P10008.

Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111, Dec 2004. doi: 10.1103/PhysRevE.70.066111. URL https://link.aps.org/doi/10.1103/PhysRevE.70.066111.

Gábor Csárdi, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller. *igraph: Network Analysis and Visualization in R*, 2023. URL https://CRAN.R-project.org/package=igraph. R package version 1.5.0.1.

Suzanna De Boef and Paul M. Kellstedt. The political (and economic) origins of consumer confidence. *American Journal of Political Science*, 48(4):633–649, 2004. doi: https://doi.org/10.1111/j.0092-5853.2004.00092.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0092-5853.2004.00092.x.

Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Phys. Rev. E*, 72:027104, Aug 2005. doi: 10.1103/PhysRevE.72.027104. URL https://link.aps.org/doi/10.1103/PhysRevE.72.027104.

Robert Kirby Goidel and Ronald E. Langley. Media coverage of the economy and aggregate economic evaluations: Uncovering evidence of indirect media effects. *Political Research Quarterly*, 48:313–328, 1995. URL https://api.semanticscholar.org/CorpusID:154602599.

James Hamilton. *Democracy's Detectives: The Economics of Investigative Journalism*. 10 2016. ISBN 9780674545502.

Pascal Held, Benjamin Krause, and Rudolf Kruse. Dynamic clustering in social networks using louvain and infomap method. pages 61–68, 09 2016. doi: 10.1109/ENIC.2016.017.

Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9:e98679, 06 2014. doi: 10.1371/journal.pone.0098679.

Peter Kharchenko, Viktor Petukhov, and Evan Biederstedt. *leidenAlg: Implements the Leiden Algorithm via an R Interface.*, 2021. URL https://CRAN.R-project.org/package=leidenAlg. R package version 1.1.0.

E.D. Kolaczyk and G. Csárdi. *Statistical Analysis of Network Data with R*. Use R! Springer New York, 2014. ISBN 9781493909834. URL https://books.google.fr/books?id=cNMhBAAAQBAJ.

Michael Lamla and Sarah Lein. The role of media for consumers' inflation expectation formation. *Journal of Economic Behavior & Organization*, 106, 07 2008. doi: 10.2139/ssrn.1150774.

Olav Laudy, Alexander Denev, and Allen Ginsberg. Building probabilistic causal models using collective intelligence. *The Journal of Financial Data Science*, 4(2), 2022. doi: http://dx.doi.org/10.2139/ssrn.3808233.

Helena Le Mezo and Massimo Ferrari Minesso. Using information in newspaper articles as an indicator of real economic activity. *Economic Bulletin Boxes*, 2:1, 2020. URL https://EconPapers.repec.org/RePEc:ecb:ecbbox:2020:0002:1.

S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28 (2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

Jessica Mahone, Qun Wang, Philip Napoli, Matthew Weber, and Katie McCollough. Who's producing local journalism? assessing journalistic output across different outlet types. *DeWitt Wallace Center for Media & Democracy*, 2019.

Richard Nadeau, Richard G. Niemi, and Timothy Amato. Elite economic forecasts, economic news, mass economic judgments, and presidential approval. *The Journal of Politics*, 61: 109–135, 1999. URL https://api.semanticscholar.org/CorpusID:154693439.

M Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 74:036104, 10 2006. doi: 10.1103/PhysRevE.74.036104.

Mark Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69:026113, 03 2004. doi: 10.1103/PhysRevE.69.026113.

Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Phys. Rev. E*, 74:016110, Jul 2006. doi: 10.1103/PhysRevE.74.016110. URL https://link.aps.org/doi/10.1103/PhysRevE.74.016110.

Mahsa Seifikar, Saeed Farzi, and Masoud Barati. C-blondel: An efficient louvain-based dynamic community detection algorithm. *IEEE Transactions on Computational Social Systems*, 7(2): 308–318, 2020. doi: 10.1109/TCSS.2020.2964197.

Stuart N. Soroka, Dominik A. Stecula, and Christopher Wlezien. It's (change in) the (future) economy, stupid: Economic indicators, the media, and public opinion. *American Journal of Political Science*, 59(2):457–474, 2015. doi: https://doi.org/10.1111/ajps.12145. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12145.

V. Traag, L. Waltman, and Nees Jan van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233, 03 2019. doi: 10.1038/s41598-019-41695-z.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL https://www.R-project.org/. R version 4.2.1.

Eray Turkel, Anish Saha, Rhett Carson Owen, Gregory J. Martin, and Shoshana Vasserman. A method for measuring investigative journalism in local newspapers. *Proceedings of the National Academy of Sciences*, 118(30):e2105155118, 2021. doi: 10.1073/pnas.2105155118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2105155118.

Di Zhuang and Mingchen Li. Dynamo: Dynamic community detection by incrementally maximizing modularity. *IEEE Transactions on Knowledge and Data Engineering*, PP:1–1, 11 2019. doi: 10.1109/TKDE.2019.2951419.