

Identification of regions with a high responsiveness to anthropogenic carbon emissions

Tobias Hölzer

Master Student

University of Potsdam

Berlin, Germany

hoelzer@uni-potsdam.de

Abstract—The responsiveness of various Earth regions to anthropogenic carbon emissions was investigated using complex network methods. The approach utilizes mutual information to construct networks that reveal global teleconnections and with that regions sensitive to emissions. Following this approach, the atmospheric layers above oceans, particularly in the southern hemisphere, were marked to be more responsive to global CO² emissions. However, with the majority of found links were super-long teleconnections, the validity and robustness of the used approach is not clear.

I. INTRODUCTION

The presumably largest challenge humanity faces today is the global climate change. [1] Greenhouse Gases, such as carbon dioxide (CO²), hinders heat to escape earth and therefore warming the planet. Recent anthropogenic activities, particularly the burning of fossil fuels and deforestation, have led to unprecedented increases in atmospheric CO² levels. However, reducing CO² emissions is a challenging tasks, since most of the activities of human societies, like construction or commute, rely on the burning of fossil fuels. Knowing at which regions on earth emitting CO² has a higher effect in the short-term could help humans to move the necessary emittance away from these regions. This work applies complex network methods to search for such regions.

II. COMPLEX CLIMATE NETWORKS

Earth's climate is a complex, chaotic system with many components and actors with different roles. A common tool to analyse such complex systems are complex networks. [2] Complex network theory is a branch of graph theory that provides a robust mathematical framework for the analysis of complex systems. In the context of climate science, it offers a powerful tool to statistically investigate the topology of local and non-local statistical interrelationships, often referred to as teleconnections. [3]

Complex networks have gained significant attention in recent years. [4] They are essentially a set of elements V , known as vertices or nodes, with connections E between them, referred to as edges or links. Formally, such a network can be represented by an adjacency matrix

$$A_{i,j} = \begin{cases} 0 & \text{if } \{i,j\} \notin E \vee i = j \\ 1 & \text{if } \{i,j\} \in E \wedge i \neq j \end{cases} \quad (1)$$

for every $i \in V$ and $j \in V$. [3]

There are several measures used to describe complex networks, each providing unique insights into the structure and behavior of the network. [5] The **degree centrality** k_v is a fundamental metric that characterizes the connectivity of nodes. It represents the total number of connections per node. Nodes with a high degree-centrality tend to be well-connected and serve as network-hubs.

$$k_v = \sum_{i=1}^N A_{vi} \quad (2)$$

To quantify the typical distance between any pair of nodes in the network, one can use the **average link-length** \mathcal{L} :

$$\mathcal{L} = \frac{1}{\binom{N}{2}} \sum_{i < j} d_{i,j} \quad (3)$$

Another commonly used measure is the betweenness centrality. The betweenness centrality describes a node's importance, quantifying how often it acts as a bridge on the shortest paths between other nodes.

$$c_B(\nu) = \sum_{i,j \in V} \frac{\sigma(i,j|\nu)}{\sigma(i,j)} \quad (4)$$

where $\sigma(i,j)$ is the number of shortest (i,j) -paths, and $\sigma(i,j|\nu)$ is the number of those paths passing through some node ν other than i, j . [6]

In the context of climate research complex networks are used to represent teleconnections. Teleconnections in the climate system are essentially statistical correlations between meteorological or oceanographic phenomena occurring in widely separated regions. [7] For instance, the El Niño Southern Oscillation (ENSO) in the Pacific Ocean is known to influence weather patterns across the globe. Understanding these teleconnections is crucial for accurate climate modeling and prediction. [3]

Using complex networks one can represent these teleconnections as nodes and edges in a network, where nodes represent different regions or points in the climate system, and edges represent statistical relationships between them. This network representation can reveal important features of the climate system's structure, such as clusters of nodes that are strongly interconnected, or hubs that are connected to many other nodes. [8] Complex network measures can give insights into the overall architecture of the climate system and its teleconnections, and can help identify regions that play key roles in the system's dynamics.

There are several methods to obtain a adjacency matrix for a complex climate network. Early methods used the Pearson-Similarity to compare the time-series of each node. Edges are then defined as (Person-) similar pairs above a given threshold. To define the similarity matrix S , one use:

$$S_{i,j} = \frac{\text{cov}_{i,j}}{\sigma_i \sigma_j} \quad (5)$$

Based on this matrix a threshold θ is defined which then is used to define the adjacency matrix A :

$$A_{i,j} = \begin{cases} 0 & \text{if } S_{i,j} > t \\ 1 & \text{if } S_{i,j} \leq t \end{cases} \quad (6)$$

However, the Pearson-Similarity rely on a linear connection between two nodes which, in case of climate data, is often not the case. The climate could be described as a chaotic systems, because it is characterized by its seemingly random behavior, yet is governed by deterministic rules. These systems are very sensitive to initial conditions, leading to divergent trajectories over time. Hence, chaos is somewhat unpredictable in the mid-term even though the underlying dynamics follow well-defined rules. [9] However, to analyse the long-term predictability researchers use non-linear statistical models, which use a mix of physics-based approaches and data-driven techniques, such as the mutual information. [10] The mutual information is a measure from information theory which is able to detect linear and non-linear relationships. It is defined as follows:

$$\text{MI}_{i,j} = \sum_{\mu,\nu} p_{i,j}(\mu, \nu) \log \frac{p_{i,j}(\mu, \nu)}{p_i(\mu)p_j(\nu)} \quad (7)$$

with $p_i(\mu)$ and $p_j(\nu)$ are the probability density functions (PDF) of the time series at vertices v_i and v_j , and $p_{i,j}(\mu, \nu)$ is the joint PDF of a pair of vertices v_i and v_j .

III. CONSTRUCTION OF THE CARBON NETWORKS

To detect regions on earth which are strongly responsive to emissions two assets were constructed: (1) a network created based on mutual information of atmospheric carbon dioxide concentration data and (2) a network-layer of carbon emission data connected based on mutual information to previous concentration data. The first should reveal global teleconnections of carbon dioxide concentration in general, whereas latter should enable making assumptions about which regions of earth's atmosphere are more fraudand to short-term CO₂ emissions and which regions on earth's have a higher short-term impact on global CO₂ concentrations. The implementation was written largely in Python with Xarray, for some compute intense parts Rust was utilized and accessed from Python with the PyO3 framework. The code is available on GitHub with explainations on how to access the data. [11]

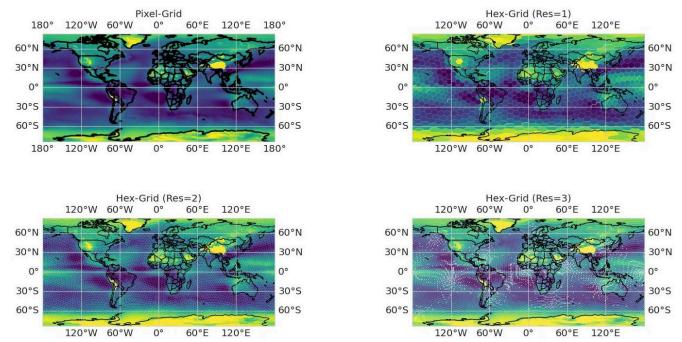


Fig. 1: Lat-Lon Grid vs. Hex-Grid with $r_{\text{hex}} = (1, 2, 3)$.

A. Data

For assessing atmospheric carbon dioxide concentration, the Sounder-SIPS dataset was utilized. The Sounder-SIPS dataset, derived from the Atmospheric Infrared Sounder (AIRS) instrument aboard the EOS Aqua satellite, provides insights into atmospheric composition, such as carbon dioxide (CO₂) and methane (CH₄) concentrations. Utilizing the CLIMCAPS (Community Long-term Infrared Microwave Coupled Product System) algorithm, temperature, water vapor, ozone, and trace gases were retrieved. The dataset offers a monthly resolution at one degree latitude by one degree longitude, ensuring global coverage from 2002 to 2023. [12]

The GRACED dataset, a near-real-time global gridded daily CO₂ emissions dataset, was utilized as carbon emission data. GRACED offers high spatial (0.1° × 0.1°) and temporal (1-day) resolution, covering emissions across seven sectors. The dataset is derived from Carbon Monitor's daily national CO₂ emissions estimates, multi-source spatial activity data, and satellite NO₂ data. Data is available from early 2019 to late 2023, spanning a relative short timerange. The dataset also includes a grid-level uncertainty analysis, ensuring its reliability despite the high spatiotemporal resolution. [13]

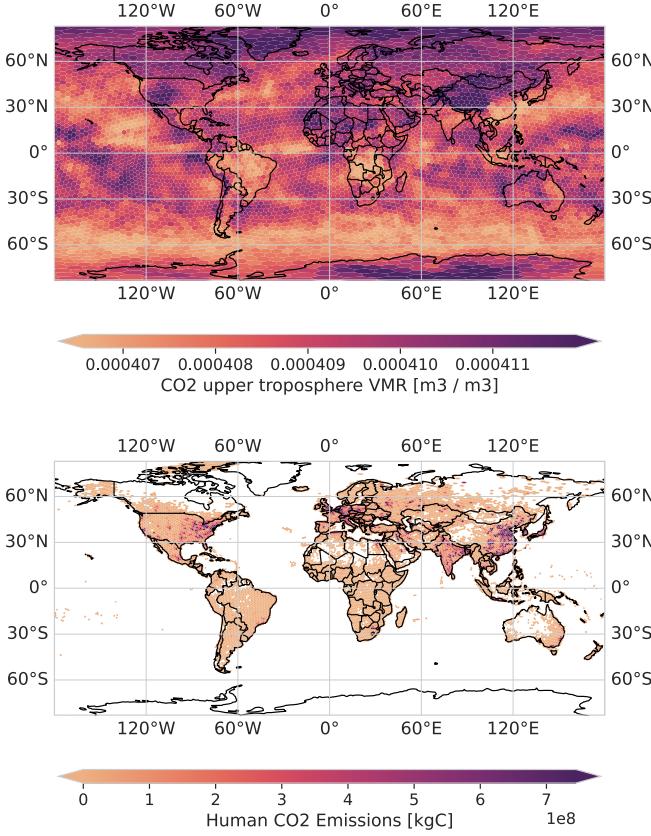


Fig. 2: AIRS CO₂ concentrations in the upper atmosphere in January 2020 (top) GRACED carbon emissions at the 1st January 2020 (bottom)

Both datasets were preprocessed and turned into hexagon-shaped data to remove the area-representation-bias towards the poles. For this Ubers h3 grid was used. [14] If a hexagon contained data from multiple pixels these values were aggregated by summing them up in case of the carbon emission data. For the concentration data the aggregation was done by meaning the data. The h3 library allows the use of different resolutions (r_{hex}). The resolution for the concentration data was chosen based on Fig. 1 which shows that with for $r_{\text{hex}} = 3$ some hexagons where empty. Therefore $r_{\text{hex}} = 2$ was chosen, which results of 5.882 vertices with an average area of 86802 km^2 . For the emissions-data $r_{\text{hex}} = 3$ with an average area of 12393 km^2 was chosen since the GRACED data has a higher original resolution. Because humans not emit carbon on every single point on the globe the resulting vertex count is with 5.228 hexagons lower than the total number of h3 hexagons for that resolution. Using hexagonal data also has the benefit of less data to be computed. With LatLon-Pixel data the adjacency matrices would span $(360 \cdot 180)^2 \approx 4.199M$ edges for the concentration data and $(360 \cdot 180) \cdot (3600 \cdot 1800) \approx 419.904M$ for a combination of the concentration data and the emission data. With a hexagonal grid both these numbers reduce to $5.882 \cdot 5.882 \approx 34M$ and $5.882 \cdot 5.228 \approx 31M$ edges. Fig. 2 shows the data from january 2020 of both processed datasets.

B. Memory-Constant Implementation of Mutual Information

Even with the drastical vertex-reduction through the use of hexagonal grids one of the challenging problems of this work was the limited access to comput power and memory. Since the mutual information measure is based on the PDF of both edges time series histograms needed to be calculated for the discrete case. Doing so in a vectorized way would have required approx. 200GB of memory. An naive non-vectorized version implemented in Python, which iterated over each edge seperately and therefore keeping the memory footprint low, would have taken more than 2h for a single adjacency matrix with $34M$ edges. Therefore another, optimized version was implemented in Rust. This version made use of various optimization tricks like fast stack-allocations, multi-threading and precomputation of histogram indicies. Through the compiled speed of Rust and mentioned optimizations a single adjacency matrix with $34M$ edges took 9s to compute with a constant memory usage.

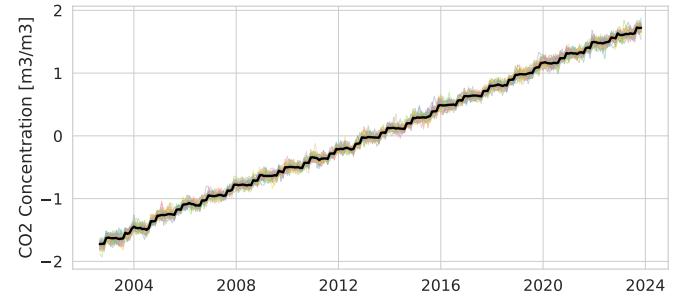


Fig. 3: Preprocessed data AIRS_{deseasoned} with it's mean in black.

C. Construction of the concentration network

Before the construction of the concentration network with the AIRS concentration data the data was first z-score normalized and then deseasonalized via months. Hence:

$$\text{AIRS}_{(\text{z-norm})_{i,t}} = \frac{\text{AIRS}_{i,t} - \overline{\text{AIRS}_i}}{\sigma(\text{AIRS}_i)} \quad (8)$$

$$\text{AIRS}_{(\text{deseasoned})_{i,t}} = \text{AIRS}_{(\text{z-norm})_{i,t}} - \overline{\text{AIRS}_{(\text{z-norm})_{i,t \in T}}}$$

where T describes all t of the same month as t. Fig. 3 shows the data over time.

Then the mutual information and adjacency matrix A_1 was calculated as described in Section II. Fig. 4 shows maps of Degree Centrality, Average Link-Length and Betweenness of the resulting network.

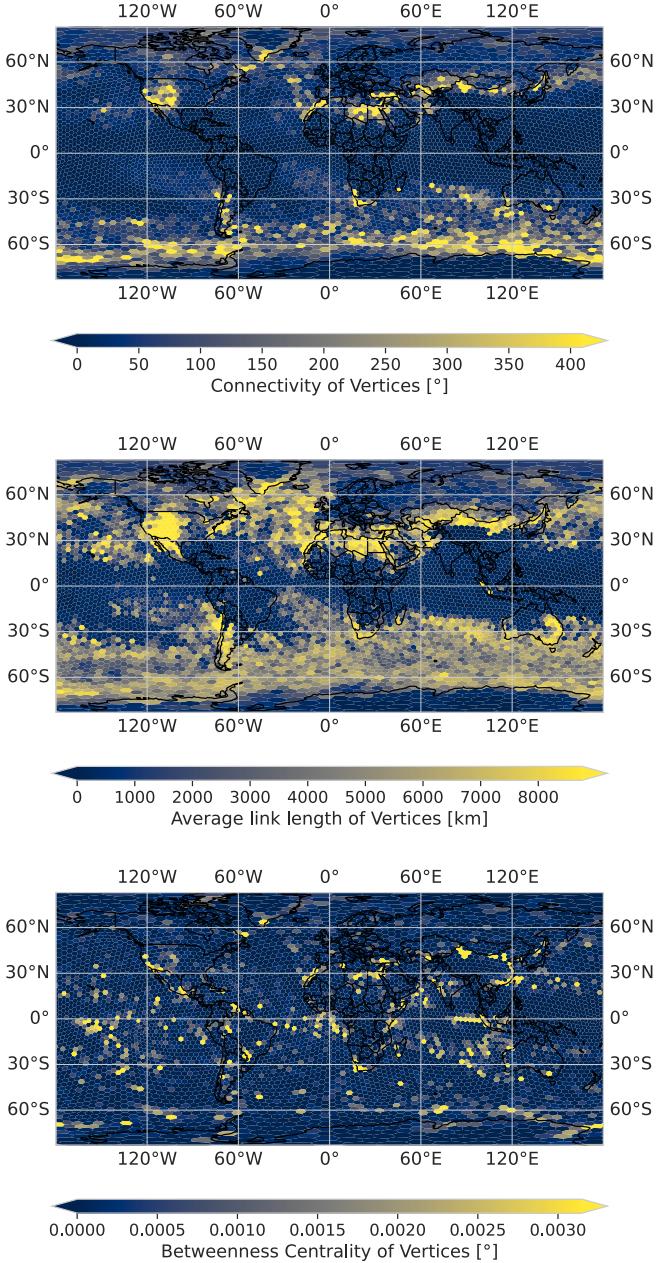


Fig. 4: from top to bottom: Degree Centrality, Average Link-Length and Betweenness of the concentration network

D. Construction of the inter-data network layer

Again, before the construction of the network-layer both datasets were further preprocessed. The emission data was monthly resampled via summation to fit the temporal resolution of the concentration data and then z-score normalized along the time axis to remove the spatial bias introduced by different population and economic power densities. Before the concentration data was z-normalized, the temporal discrete difference was calculated to fit the emission data's nature as data of changing masses instead of commulated mass. Fig. 5 shows the aligned data along the time-axis. This time, no deseasonalization was applied in both cases.

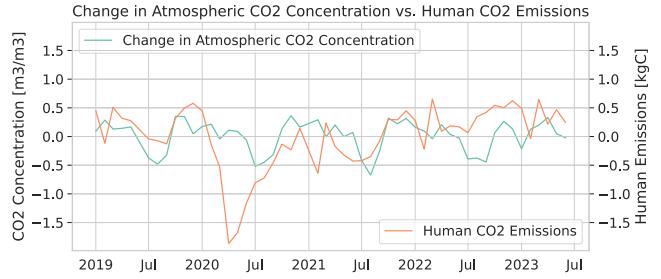


Fig. 5: total CO₂ emissions and average atmospheric concentration difference of CO₂ over time after alignment.

With the aligned data, the mutual information between every vertex v_e of the emission data and every vertex v_c of the concentration data was calculated, resulting in a 5.882×5.228 similarity matrix MI. A threshold of $\tau = 2.197$ was used to select 1% of edges for the computation of A_2 . The resulting adjacency matrix A_2 does not represent a complex network, since every edge is a bidirectional connection between two different types of vertices. However, it represents the relationship between CO₂ emission and concentration nodes. Therefore a degree centrality k and an average link-length \mathcal{L} can be computed for every vertex v_e and every vertex v_c , resulting in two world-graphs each. Fig. 6 shows these in detail.

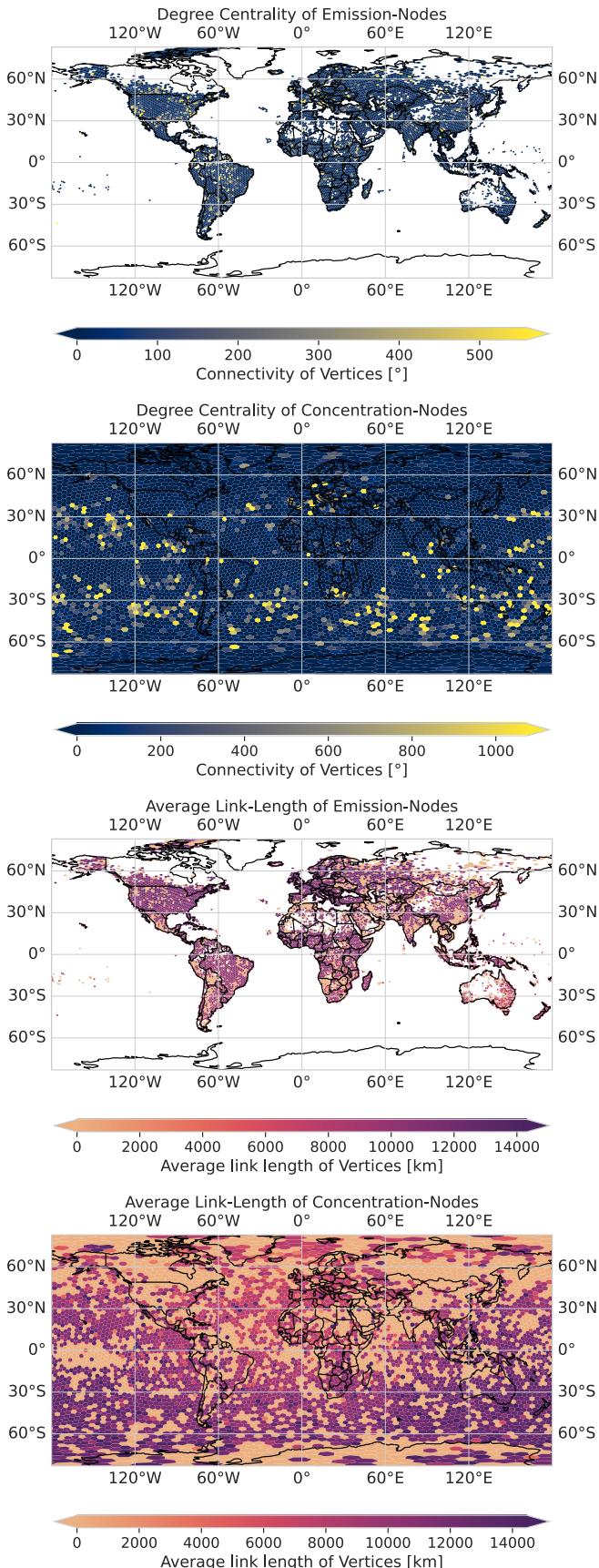


Fig. 6: Degree Centrality and Average Link-Length of emission and concentration nodes.

IV. RESULTS

Observing the degree centrality k and an average link-length \mathcal{L} of the concentration network in Fig. 4, a southern and a northern “stream” is visible. The degree centrality seems to be especially high around dry deserts, meaning the north Sahara, Gobi and the Colorado Plateau, as well near the east-antarctic land. The northern “stream” seems to have longer connections than the southern “stream”, the southern one is in return denser. Fig. 7 shows the distribution of link-lengths. Unfortunately, a statistical increase of long-ranging teleconnections, like in N. Boers, B. Goswami, and A. Rheinwald [15], could not be observed. It would be important to mention here, that the connections between the nodes are already very long, with most of them longer than 4000km. No significant pattern or cluster of hubs could be observed from the betweenness centrality.

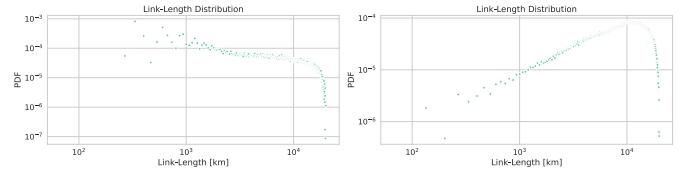


Fig. 7: Link Length Distribution of the concentration network (left) and of the network-layer (right) on a log-log scale.

The degree centrality and average link-lengths of the emission-concentration network layer are very noisy. There are very few nodes which are heavily connected to each other, but they seem to share a larger amount of connected neighbors. The heavily connected vertices on the concentration side are mostly found on oceans, not on land. Heavily connected vertices on the emission side seem to be randomly spread across the globe with no clear pattern visible. Because of the very even distribution towards the pacific of the connected concentration nodes, the average link-length of both node-types seem to be diverging from Europe away. A look on Fig. 7 also reveals that this network consists almost only of teleconnections, which seems to be a little bit odd.

V. DISCUSSION

Following the idea of the method, one may say that the atmosphere above the oceans, especially on the southern hemisphere, is more responsive to CO₂ emissions across the globe than other regions. The other way around, emissions at a lot of very specific regions on Earth, especially middle Europe, the Amazonas and the eastcoast of the U.S., tend to have stronger short-term effects on the CO₂ concentration of the atmosphere above mentioned oceans.

However, the resulting network-layer seems to be somewhat random, featuring a lot of very long teleconnections. Potential reason for this could be the short time-spanned data from which this layer was built. With only 52 datapoints in time, it could have been too less to draw meaningful information from it. Another reason could be the fast-circulation of CO₂ in the atmosphere, which potentially couldn't be tracked by the AIRS dataset. Therefore future research should be done with

long-term emission data and with a higher temporal resolution concentration data.

REFERENCES

- [1] Core Writing Team, H. Lee, and J. Romero, “IPCC, 2023: Summary for Policymakers,” *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1–34, 2023.
- [2] L. Amaral and J. Ottino, “Complex networks,” *The European Physical Journal B*, vol. 38, 2004.
- [3] J. Donges, Y. Zou, N. Marwan, and J. Kurths, “Complex networks in climate dynamics,” *The European Physical Journal Special Topics*, vol. 174, pp. 157–179, 2009.
- [4] F.-C. Ma, P.-H. Lyu, and X.-G. Wang, “Knowledge Discovery of Complex Networks Research Literatures,” in *Library and Information Sciences: Trends and Research*, C. Chen and R. Larsen, Eds., Springer Berlin Heidelberg, 2014, pp. 119–135. doi: 10.1007/978-3-642-54812-3_9.
- [5] M. G. Carneiro, B. C. Gama, and O. S. Ribeiro, “Complex Network Measures for Data Classification,” in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9533608.
- [6] U. Brandes, “A faster algorithm for betweenness centrality* ,” *The Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001, doi: 10.1080/0022250X.2001.9990249.
- [7] S. Feldstein, “The Timescale, Power Spectra, and Climate Noise Properties of Teleconnection Patterns,” *Journal of Climate*, vol. 13, pp. 4430–4440, 2000.
- [8] H. A. Dijkstra, E. Hernández-García, C. Masoller, and M. Barreiro, *Networks in Climate*. Cambridge University Press, 2019.
- [9] K. Sun, S. He, and H. Wang, “Introduction,” in *Solution and Characteristic Analysis of Fractional-Order Chaotic Systems*, Springer Nature Singapore, 2022, pp. 1–26. doi: 10.1007/978-981-19-3273-1_1.
- [10] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, “The backbone of the climate network,” *EPL (Europhysics Letters)*, vol. 87, no. 4, p. 48007, Aug. 2009, doi: 10.1209/0295-5075/87/48007.
- [11] T. Höller, *Chaotic Carbon Networks*. (Mar. 2024). [Online]. Available: <https://github.com/relativityhd/chaotic-carbon-networks>
- [12] C. Barnet, *Sounder SIPS: AQUA AIRS IR-only Level 3 CLIMCAPS: Comprehensive Quality Control Gridded Monthly V2*. (2019). Accessed: Mar. 15, 2024. [Online]. Available: https://disc.gsfc.nasa.gov/datasets/SNDRAQIL3CMCCP_2/summary
- [13] X. Dou *et al.*, “Near-real-time global gridded daily CO₂ emissions 2021,” *Scientific Data*, vol. 10, Feb. 2023.
- [14] I. Brodsky, D. Ellis, Z. Knudsen, N. Rabinowitz, and A. Friend, *h3*. (2024). Accessed: Mar. 15, 2024. [Online]. Available: <https://github.com/uber/h3>
- [15] N. Boers, B. Goswami, and A. Rheinwalt, “Complex networks reveal global pattern of extreme-rainfall teleconnections,” *Nature*, vol. 566, pp. 373–377, 2019.