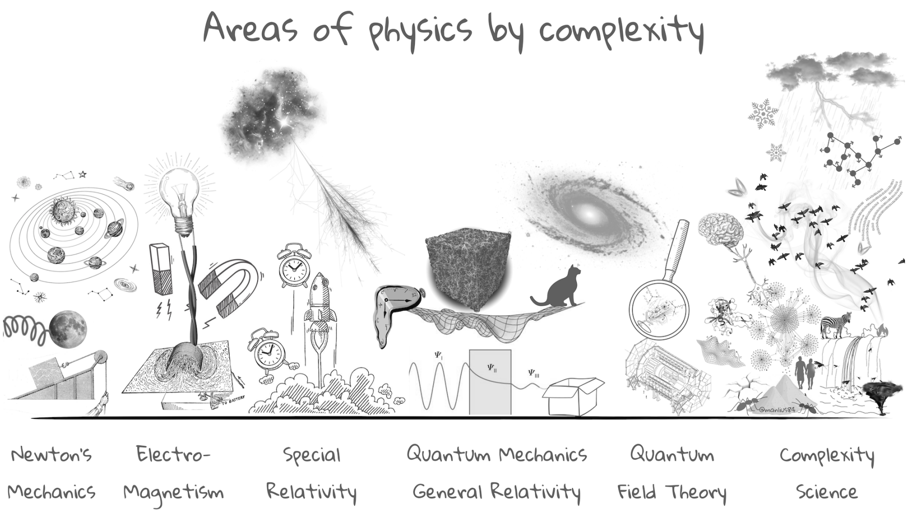
Università degli Studi di Padova



Project Climate Network

Gabriele Fiaschi

Last update: April 24, 2025

Contents

1 | Introduction

Earth system is one of the most impressive example of complex system, due the chaotic behaviour of its dynamics from mathematical and physical point of views. For this reason, the only analysis that we can do is a computational simulation, from which we can extract the time evolution of the variables of the system.

In this report I describe the pre-processing steps and the network analysis of climate data extracted from the Copernicus Datastore (a repository where are stored historical and projections data for different climatological variables) with the aim of establish a multilayer climate network between geographical areas (tipping elements) at different pressure levels; the network analysis is based on correlations between geographical sites.

Specifically, in this chapter I describe the dataset and the pre-processing steps, while the last part of the analysis is partial described in chapter 2 and 3 with special focus on the temperature variable.

1.1 | Earth system as complex system

One of the most difficult challenge for the analysis of climate change is the creation of physical Earth model [6]. To establish a realistic physical model we have to detect what is the main physical quantity to analyze. In this case this quantity is the fluid's time evolution in the Earth system, taking into account the rotational effect on it.

Scientists develop various *general circulation models* (GCM), to simulate the coupled dynamics of land, ocean and atmosphere. This is a mathematical model that allow us describing the climate's model; it is based on *Navier-Stokes equations* applied to the rotational system [2].

Mathematically, the model is a system of differential equations that are taking into account the fluid's movement. Finally, this quantity is numerically integrated over time, assuming that the whole Earth system is divided into a spatial grid. This assumption is needed for the numerical solution of the system. In particular, the grid consists in geographical points mapped through two coordinates (latitude and longitude), so as result we obtain a grid with area association.

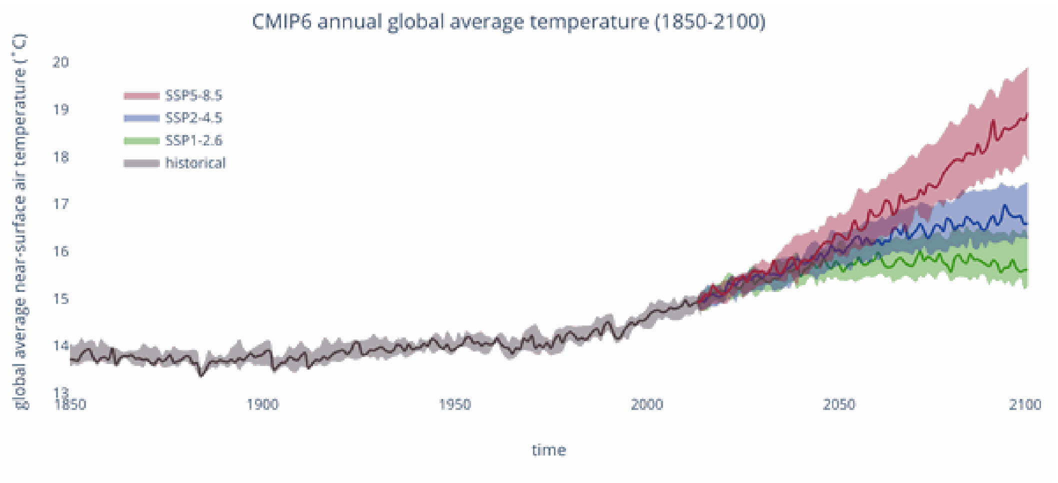


Figure 1.1: Evolution of historical average global temperatures (grey) and climatological projections (coloured) until the year 2100 for different SSP scenarios

.

1.1.1 CMIP6 projections

The previous model is employed for climate projections, particularly the main study is focused on the Couple Model Intercomparison Project (CMIP6) [5].

All that could be described as an ensemble of models (more than 100) and different scenarios develop from more than 50 research centers, and for each scenarios we can identify **Shared Socioeconomic Pathway** (SSP) [3], that are referring to different hypothetical situations with respect to the political choices about sustainability of the socio-economical system. These kinds of scenarios influence the projections allowing different results from the GCMs, that contributes to more or less dispersion in the evolution of the system with respect to the "mean model". This kind of dispersion in the variables' evolution are used as error bars on the model, as we can see in grey in 1.1. Finally, we can distinguish between different scenarios from the optimistic one, SSP1 to the catastrophic one, SSP5. In the first one, world shifts gradually towards ecological and sustainable path, this kind of scenario is the "green" path, visualization of results are displayed on 2.2 and 2.3, respectively mean temperature and precipitation flux. The SSP5 scenario, instead, is focused on the fossil-fueled development, so this is the "black" path, as we can see in Figures 2.4 and 2.1. To complete the description of the scenario with SSP experiment it is joined **Representative Concentration Pathway** (RCP), this quantity reflects the CO2 emission in the atmosphere, and it is the main parameter to evaluate how climate change influences life on Earth.

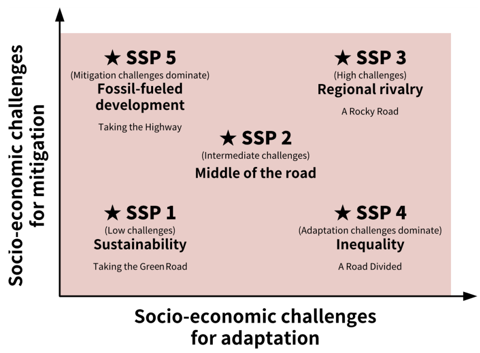


Figure 1.2: socioeconomic pathway experiments

2 | Data pre-processing

Before the analysis' phase, we have extracted data from a prior selection of models and scenarios, after that we have selected **precipitation**, **near surface air temperature** and **air temperature**, particularly for the last one we have selected difference pressure levels, from $850hPa$-$5hPa$, this choice is fundamental for the building of multilayer network based on temperature.
Specifically we have chosen **Cesm2**, **CMCC-ESM2**, **awi-cm-1-1-mr** as models and for each one we have analyzed SSP$1-2-6$ and SSP$5-8-5$ scenarios, temperatures and precipitations where SSP1 indicates the experiments and the remain quantity indicates the RCP associated values.
The first important step for the construction of the network is importing the raw data from copernicus Datastore. In particular, our analysis is focused on the time interval 2020-2100; so after this choice we have downloaded dataset in the zip format for each scenario and models.

After that the main procedure focuses on the unzipping files, here what we can see is that we have to manage a huge dataset size so to get information and extract anomalies (deviations from mean behaviour) we have to regrid all files.
For regrid procedure we have to remember that datas are organised in this way, for each geographical point identified by latitude and longitude (grid) we have specific value of precipitation/air temperature, the regridding procedure consists something similar to a coarse-grain, what we have to do is a sampling of a geographical point and after that we evaluate the mean of the selected variable taking into account the geographical neighbors point.
Before extracting anomalies we collect all data in a single netCDF4 file, that helps us for the final step of pre-processing analysis, indeed we will use this file as input for the evaluation of the anomalies with this formula:

$$A = \frac{S - C_{time}}{\sigma_{C_{time}}} \tag{2.1}$$

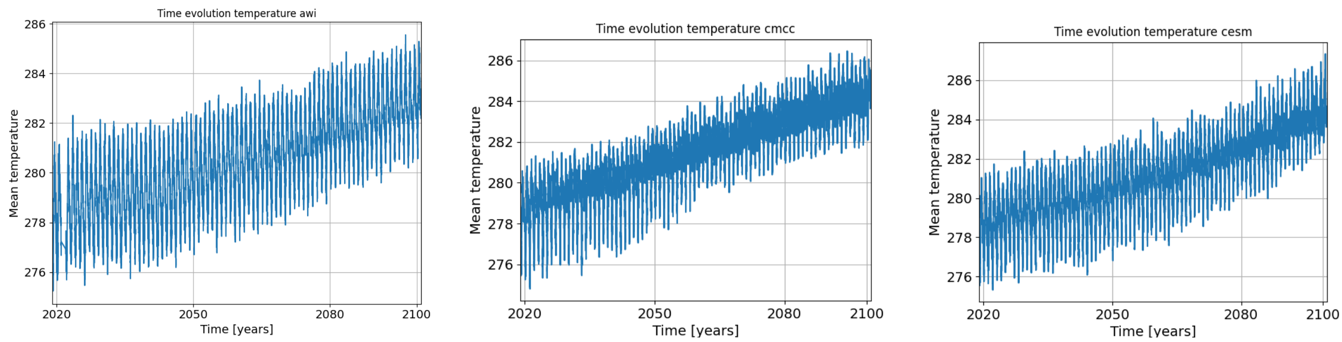
where:

- A are anomalies
- S data temperature
- C_{time} historical mean temperature
- $\sigma_{C_{time}}$ the standard deviation for the historical mean temperature

finally we evaluate anomalies as discrepancy between mean behaviour and single event variable, assuming that each of them are sampled from gaussian distribution; we divide for the standard deviation to detect how large is the discrepancy with respect to the mean behaviour.

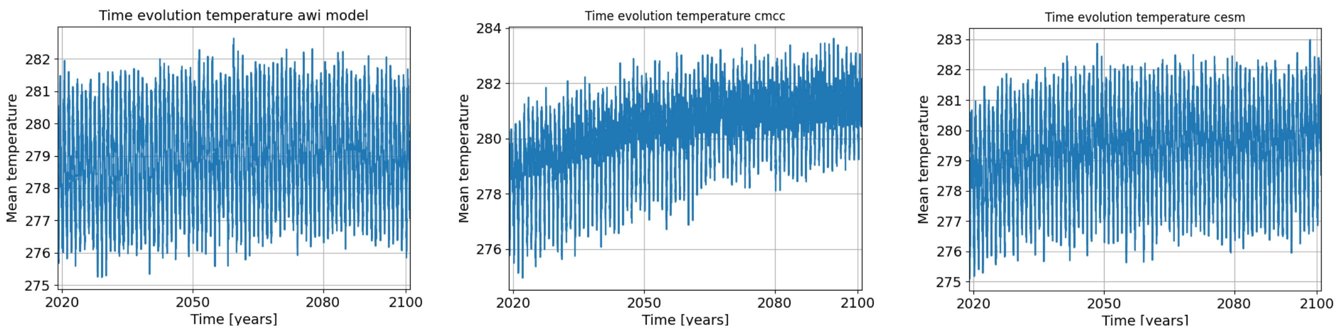
2.1 Results and Simulations

In the following we can see result of analysis that I implemented.



(a) awi model (b) cmcc model (c) cesm2 model

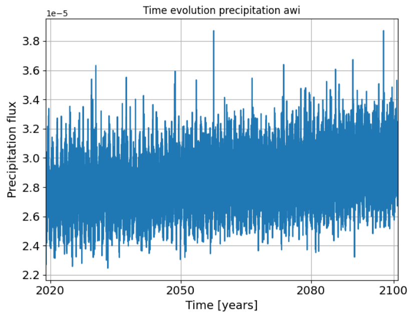
Figure 2.1: Analysis of mean temperature near surface for ssp5 scenario



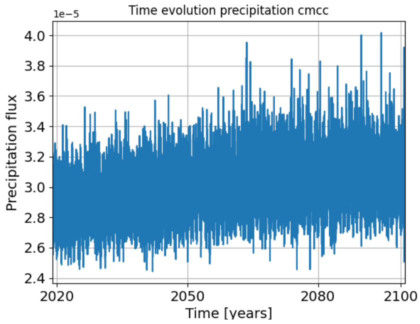
(a) awi model (b) cmcc model (c) cesm2 model

Figure 2.2: Analysis of mean temperature near surface for ssp1 scenario

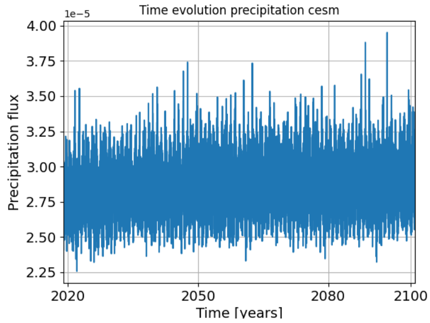
From analysis for optimistic scenario 2.2 we can deduce that mean temperature behaviour is not so high as we can see in the SSP5 scenario 2.1, more specifically we can evidence that trend in the pessimistic scenario is still growing in time, this highlights how extreme can be the mean temperature values with respect to SSP1 scenario, if there is no delimitation to the CO2 emission.



(a) awi model

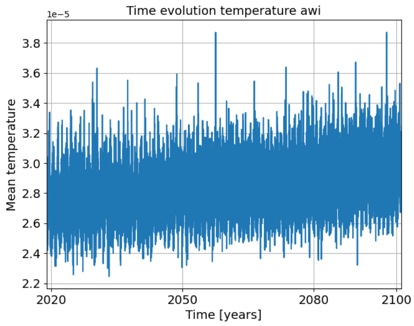


(b) cmcc model

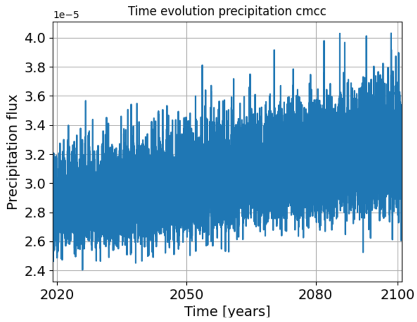


(c) cesm2 model

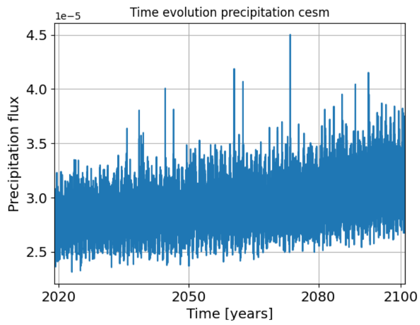
Figure 2.3: Analysis of mean precipitation for ssp1 scenario



(a) awi model



(b) cmcc model



(c) cesm2 model

Figure 2.4: Analysis of mean precipitation for ssp5 scenario

Finally, I show the temperature and the precipitation for cesm model, SSP1 and SSP5 scenario.

3 | Climate Network

The last part of the internship in climate network was focused on the creation of correlation network between different geographical regions, known as tipping points, which are interesting from climatological viewpoints [1]. The same procedure is repeated over all pressure level; to build up multilayer network.

Now the main steps are:

1. Compute the cross-correlation between all the pairs of "node"
2. Building a null model, using the IAAFT surrogates, to get a Z score
3. Compare with respect to null hypothesis to obtain a pvalue
4. Using a Bayesian procedure to obtain the posterior probability that a link exists between two sites.

With this procedure we evaluate a *functional network*, meaning that we are analyzing the activity of the network and not the "physical" links or the structural one.

The first step of the analysis is the evaluation of cross-correlation between two time series $x_i(t)$ and $x_j(t)$, representing monthly records of temperature anomalies at locations i and j, we compute the cross-correlation $C_{ij}(\tau)$ for all lags $\tau \in [-12, 12]$, which is defined as

$$C_{ij}(\tau) = \frac{\langle X_i(t)X_j(t+\tau)\rangle - \langle X_i(t)\rangle\langle X_j(t)\rangle}{\sigma^2} \tag{3.1}$$

Now the analysis to build networks is done on windows of ten years, from 2020 to 2100, with monthly frequency.
After this initial presentation we have to evaluate adjacency matrix, fundamental step to build a "physical" network, but in this case previous matrix is built from correlation one.
The further step after initial evaluation of cross-correlation is the **shuffle** procedure, we have to do that to destroy all information between the temporal series refer to node i and j [4].
All that is a necessary phase for the analysis of network, particularly for comparison with respect to **null hypotesis**. In this phase we have to compare the cross correlation value, after its maximization, between node i and j, not taking into account the autocorrelation, with respect to correlation from **Null hypotesis**, assuming that it is gaussian distributed.
To do that we extract Z_{score}, that tell us the anomalies of the data, for each couple of

nodes and we evaluate a Z_{score} matrix from which we can extract P_{value} matrix. All this kind of passages to finally get a probabilistic output, that is the entry of the **Fuzzy matrix**.
For the analysis of the network we have to sample multiple matrices, taking into account that we produce a casual number between 0 and 1, then we compare this number with respect to F_{ij}, if it is larger than the first one we can avoid it,so we replace the initial value of F_{ij} with 0 otherwise we can keep the F_{ij} associated value, finally we repeat this procedure for any different matrices; all that allows creating a density distribution probability. The last section describes the main steps to get the Fuzzy matrix, this kind of operations are based on the Bayesian' theorem.
The main effect to the use of Fuzzy network are based on :

- No arbitrary threshold
- No lose of information

that because generally to build a network we have to set a threshold and it causes the lack of information, due the elimination of links in network, so to avoid that and to guarantee universality future we need imposing Fuzzy network structure [4].

3.1 | Tipping elements connectivity

For the final analysis of the climate network we compute the correlations at different pressure levels between geographical areas. In particular, we focus on the time evolution of the correlation, as we can see in figure 3.1.

In the previous figure we can notice that across decades we have an increase of correlation between extreme north of Russia, Artic Pole and Greenland areas, with the formation of a triadic closure (the simplest cluster).

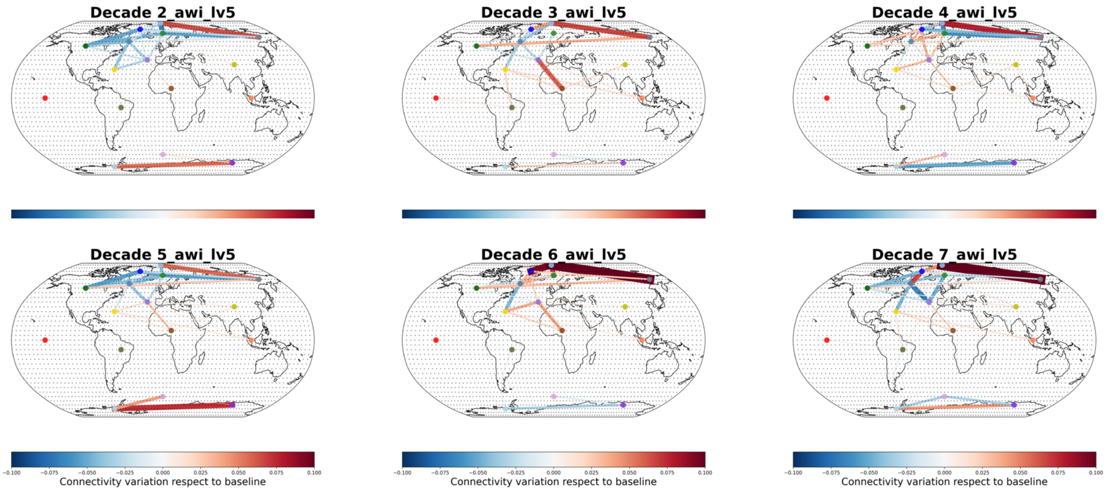
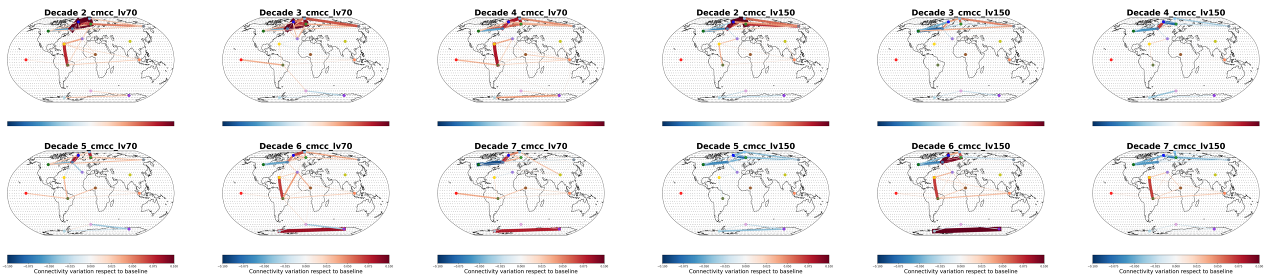


Figure 3.1: **Evolution of tipping elements connectivity over the past 40 years with respect to the** $2020-2030$ **baseline using the ERA5 temperature dataset using the awi model and SSP1 scenario.** The plot depicts the network of the changes in connectivity between the tipping elements compared to the $2020-2030$ baseline period. Edge thickness represents the connectivity C_{AB}^{base} of the baseline, while the color scale represents the variation $\Delta C = C_{AB} - C_{AB}^{\text{base}}$ between that decade and the baseline.

Very interesting is the pattern formation in the climate network varying the levels but fixing models and scenarios; indeed we can observe a pattern formation in the cmcc model for level 150hPa and level 70 hPa, as highlighted in 3.2.



(a) cmcc level 70hPa

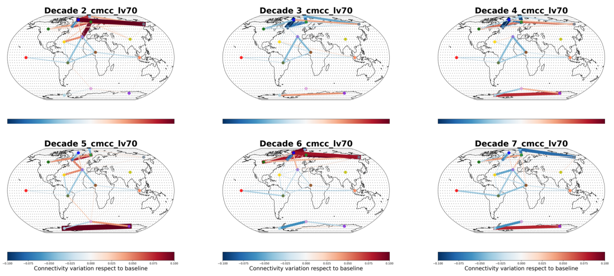
(b) cmcc level 150hPa

Figure 3.2: clustering formation across level of the cmcc model and scenario ssp1

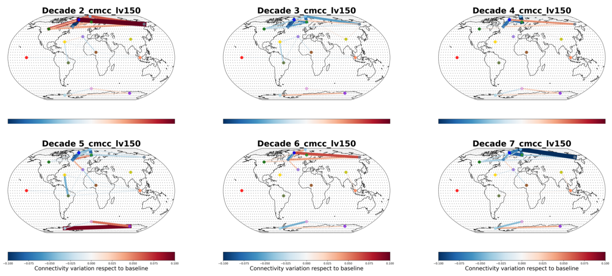
In this case we cannot deduce what is the physical effect of the cluster formation, but we can notice the initial increase of connectivity at level 70hPa with respect to the baseline, in our case fixed to the decade 2020-2030. The main emergent cluster in that pressure level takes into account areas like North Pole, North America, North Europe and Russia (Siberia region).

The same kind of cluster, but with different degree of the connectivity, emerges even for the 150 hPa level. This is an example of the analysis that I have done during my internship, that is repeated for all models and scenarios selected during the pre-processing phase.

Further figure is another example of clustering research in the scenario ssp5 at the same levels as done in 3.2. What we can observe is the emergence of different clustering structure with respect to the previous analysis, particularly in the Antartica's area.



(a) cmcc level 70hPa



(b) cmcc level 150hPa

Figure 3.3: clustering formation across level of the cmcc model and scenario ssp5

At the end to build up the network structure we have, firstly, selected air temperature at different pressure level (we did it during the pre-processing step) and in this phase we have applied the procedure previously depicted at the beginning of chapter 3. This operation is been repeated over all different pressure levels to establish, as we can see in figures: 3.2, 3.3, the presence of emergent phenomena in the single layer network, by changing models and scenarios.

All these procedures can be used for further multilayer structural analysis enlarging so the analysis done on the single layer network.

3.2 Conclusion

In summary my internship is focused on the pre-processing of the data and computational analysis of the single layer networks. Fundamental step for all the further analysis on the networks and clustering research is the data pre-processing . Particularly all steps are done using linux terminal and Python codes from the importing procedure of raw data to the visualization of the time evolution of climate variables. For selected variables as air temperature and flux precipitation I used Ncviewer software, that allow me to explore the evolution of the previous quantities.

Final steps is the network analysis, in this phase I looked for clustering or triadic closure into the single layer and I investigated if the same kind of the cluster structure survive across the pressure levels fixing the scenario and models.

4 | Bibliography

[1] Timothy M. Lenton, Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen, and Hans Joachim Schellnhuber. Climate tipping points — too risky to bet against. *Nature*, 575(7784):592–595, November 2019. doi: 10.1038/d41586-019-03595-0.

[2] T Lovato, D Peano, M Butenschön, S Materia, D Iovino, E Scoccimarro, PG Fogli, A Cherchi, A Bellucci, S Gualdi, et al. Cmip6 simulations with the cmcc earth system model (cmcc-esm2). *Journal of Advances in Modeling Earth Systems*, 14 (3):e2021MS002814, 2022.

[3] Malte Meinshausen, Zebedee RJ Nicholls, Jared Lewis, Matthew J Gidden, Elisabeth Vogel, Mandy Freund, Urs Beyerle, Claudia Gessner, Alexander Nauels, Nico Bauer, et al. The shared socio-economic pathway (ssp) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8): 3571–3605, 2020.

[4] Sebastian Raimondo and Manlio De Domenico. Measuring topological descriptors of complex networks under uncertainty. *Physical Review E*, 103(2):022311, 2021.

[5] Tido Semmler, Sergey Danilov, Paul Gierz, Helge F Goessling, Jan Hegewald, Claudia Hinrichs, Nikolay Koldunov, Narges Khosravi, Longjiang Mu, Thomas Rackow, et al. Simulations for cmip6 with the awi climate model awi-cm-1-1. *Journal of Advances in Modeling Earth Systems*, 12(9):e2019MS002009, 2020.

[6] Isla R Simpson, Julio Bacmeister, Richard B Neale, Cecile Hannay, Andrew Gettelman, Rolando R Garcia, Peter H Lauritzen, Daniel R Marsh, Michael J Mills, Brian Medeiros, et al. An evaluation of the large-scale atmospheric circulation and its variability in cesm2 and other cmip models. *Journal of Geophysical Research: Atmospheres*, 125(13):e2020JD032835, 2020.