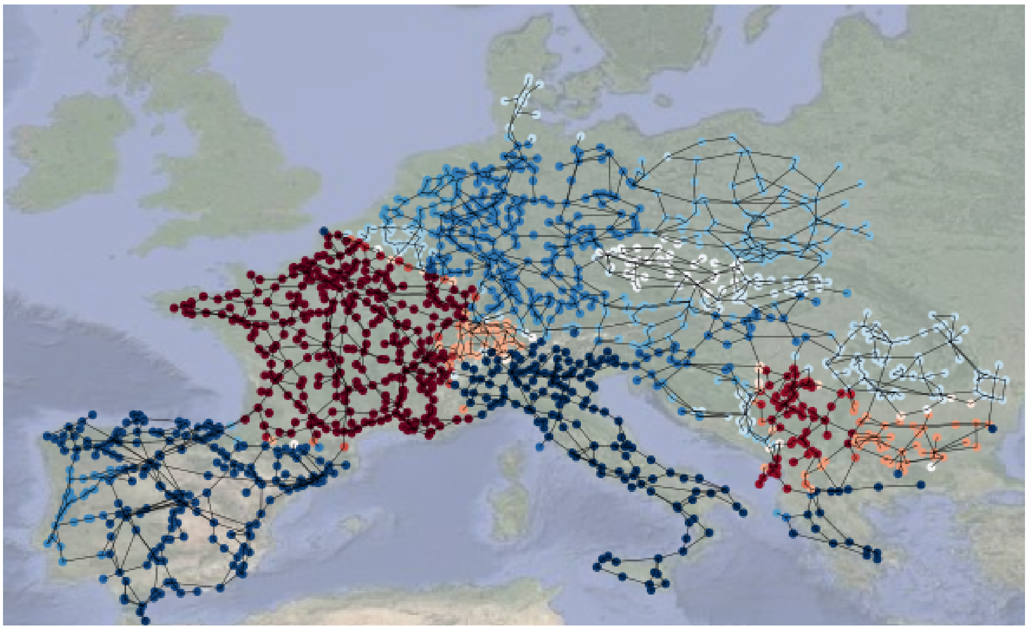
EE558 - A Network Tour of Data Science
Final Project Report



Francesco Gallo, Kay Lächler,
Roberto Chedraui Abud, Viktor Crettenand

January 2020

1 Introduction

The power grid is the backbone of modern society. Analyzing its components and structure allows to tackle one of the greatest problems our society is confronted with: climate change. Modern power grids were designed for centralized generation, but in the near-future they will rely on decentralized and non-dispatchable power production due to renewable-energy sources, such as solar panels and wind turbines. Those energy sources have a fluctuating power output unlike traditional energy sources such as hydroelectric plants, nuclear plants and thermal plants that have stable and adjustable outputs.

This motivates the search for new and innovative ways to model large scale power grids and to make reliable predictions of renewable energy production. This project investigates the use of network analysis tools to model the power grid and draw conclusions from it. Another objective is the development of a tool which can make renewable power output predictions more reliable.

Since Gustav Kirchhoff published a paper in 1847 where concepts of graph theory were used to model electrical networks, graph theory has been used for the analysis and optimization of power systems [1]. Power networks consist of nodes, which are points of interconnection between generators and consumers, as well as edges between nodes, which represent transmission lines. Additional information about power production and consumption, energy mix as well as production and demand forecast can be added to the nodes.

The RE-Europe dataset [2] has all the mentioned elements and was chosen to perform the desired network analysis in this project. It includes the location of European nodes, information about the maximal output, and type of power plants all over the continent, as well as information about the transmission lines that connect the nodes to form the power grid. In addition, it contains a time series with the hourly power consumption of each node, as well as the forecast and effective solar and wind output power for each node for a period of three years. Overall, this data set allows for both a spatial and a temporal evaluation of the power grid. The graph is weighted using the inverse of the length of the transmission lines.

A visualization of the network and the information of the connected generators can be seen on Figures 1 and 2. This was plotted using the coordinates given in the data set and the tools used in the course.

Figure 1: RE-Europe dataset graph

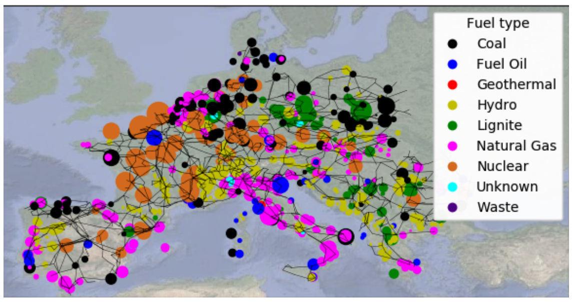


Figure 2: Generators with fuel type and capacity

2 Exploration

To begin the exploration of the data set, some basic parameters were calculated. The results are shown on Table 1. The network is connected, which makes sense since generators all over the continent are linked through transmission lines to allow power flow between countries. The degree distribution is shown on Figure 3. Looking at this distribution, the graph appears to be of power law type, which makes sense since having a lot of transmission lines connected to the same node renders the regulation of each line highly complex. This conclusion is also supported by the very small clustering coefficient, which suggests that having clusters and complex graph structures is not beneficial for a stable power grid. Finally, most of the nodes have a degree of two, so the graph consists of several paths containing intermediate nodes, but does not branch often. The graph is quite sparse, as shown by the adjacency matrix in 8 in the Appendix.

The eigenvalues of the graph Laplacian are shown on Figure 4. Only the first eigenvalue is zero (to the limit of floating point precision), which is consistent with the number of connected components. All other eigenvalues are low, which should be the case because nodes are usually connected to close-by nodes. This is because it would not make sense to connect nodes that are far away, as such transmission line is not viable.

Table 1: Some network parameters

Connected components	1
Diameter	48
Clustering coefficient	0.105975

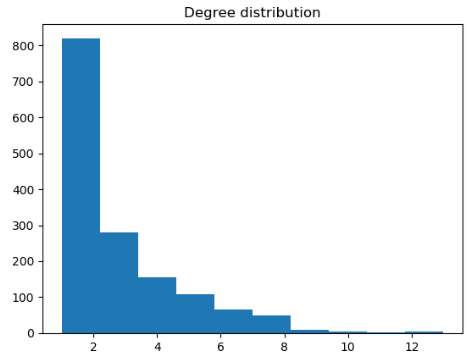


Figure 3: Degree distribution of the graph

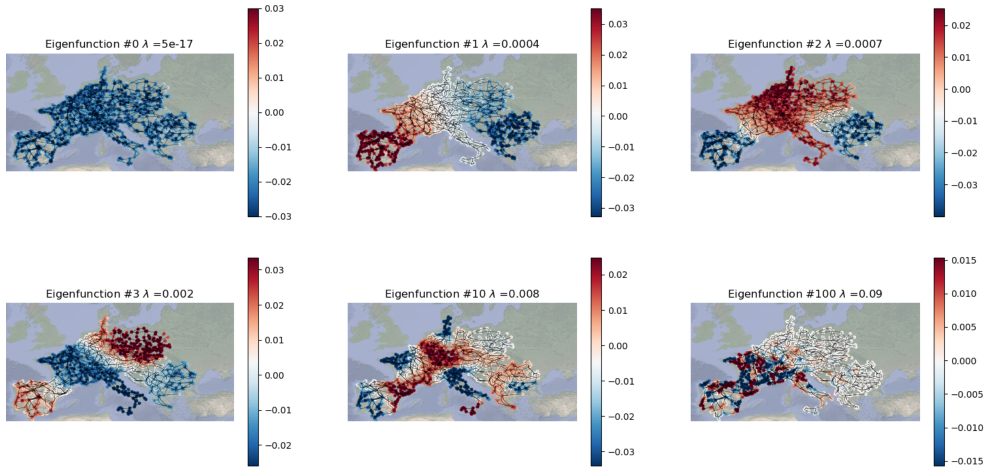


Figure 4: Eigenfunctions of the graph laplacian

For a more elaborate analysis of the data, the Fourier transform of the power consumption signals was done. The result for four specific frequencies is shown on Figure 5. The chosen frequencies correspond to periods of 12 hours, 24 hours, one week and one year. The Fourier components corresponding to these frequencies are relevant because they carry information about the periodicity at which power is used in different regions. It is surprising to see how some country boarders are clearly visible with this analysis, suggesting that the difference in power use is more related to the country than the geographical region.

One striking example is France, which can be clearly seen in red on the plot in the lower right corner, which corresponds to a frequency of once a year. The fact that this Fourier component is much larger for France than for its neighbouring countries may be due to the extensive usage of electric heating in France [4]. Since heating consumes a lot of energy and since it is mostly used in the winter (frequency of once a year), this may explain why that Fourier component is so large. In other countries houses are heated in large part with non-electric boilers, which explains why their Fourier component for a frequency of once a year is lower.

On the upper plots on Figure 5, one can see the Fourier components for periods of 12 and 24 hours. What stands out in these plots is that France and Switzerland have lower Fourier components than the rest of Europe. This may be due to the adaptive price of electricity in France and Switzerland, which encourages people and companies to use electricity evenly throughout the day and the night. In Switzerland the price of electricity is set hour by hour [6] which is an incentive for a uniform consumption, which is desirable in power systems. In France there are two prices [7] : one for peak times and one for off-peak times. The peak-time lasts for 16 hours each day. During that time span the price is constant which means the power consumption is still prone to fluctuate. This explains why France has low Fourier coefficient for a frequency of once every 24 hours and why Switzerland has low Fourier coefficients for frequencies of both once every 12 and 24 hours. In Italy, for instance, this policy was just starting to be implemented at the time this data was taken [3].

Finally in the lower left corner, on the plot for a period of one week, the Balkans have low Fourier components. This could be explained by low economic activity relative to other regions, since industrialized countries consume more power on working days [5].

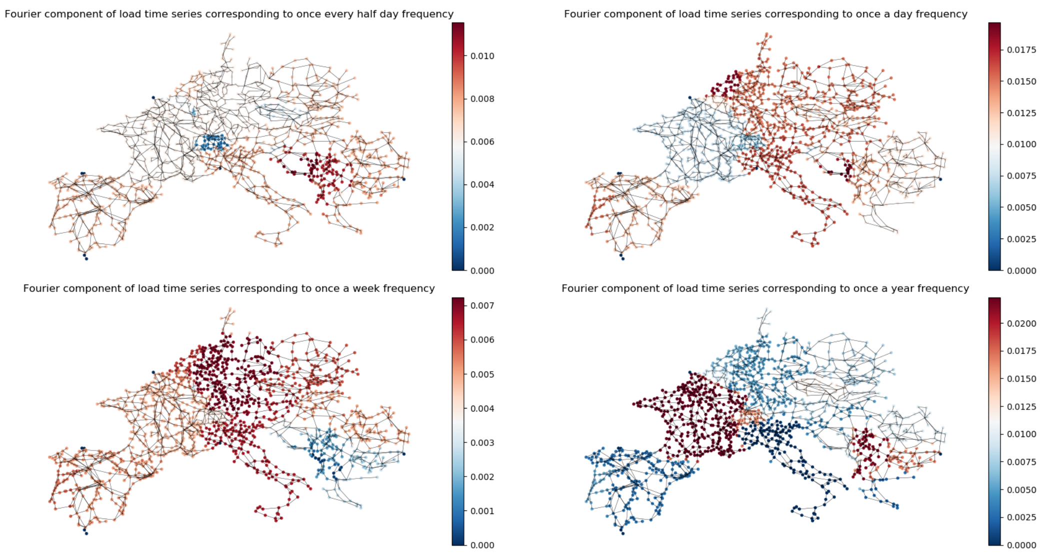


Figure 5: Fourier components of power demand time series

3 Exploitation: Graph Convolutional Network (GCN)

3.1 Model Description

In this section, the goal was to predict the renewable energy forecast of the nodes from their neighboring nodes. To achieve this, a GCN was used, which takes the structure of the graph and the edge weights into account. It consists of 3 similar layers which are calculated as follows:

$$X^{(t+1)} = D^{-1}AWX^{(t)}$$

where $D \in \mathbb{R}^{D\times D}$ is the diagonal degree matrix, $A \in \mathbb{R}^{D\times D+1}$ is the augmented adjacency matrix with the first column set to 1's representing the bias node, $W \in \mathbb{R}^{D\times D+1}$ is the weight matrix where row i corresponds to the weight vector of node i, $X \in \mathbb{R}^{D+1\times N}$ is the feature matrix where row i corresponds to node i and column j corresponds to the time sample j of the forecast time series and AW is the element-wise multiplication of A and W. Concerning the dimensions: D is the number of nodes and N is the number of time samples.
By multiplying the weight matrix with the adjacency matrix, the structure of the graph is integrated into the layer, in fact, each layer only calculates the weighted sum of the neighboring nodes. Since the diagonal of A is 0, the predicted node itself is not taken into account in the calculation. (This also means the power production forecast of a node is of course not needed to predict its forecast). By consecutively applying the above equation 3 times (3 layers), each node is predicted from the nodes that are no more than 3 edges away from it. This turns out to deliver a good compromise between accuracy and complexity of the model. Two separate models with identical base structure were trained to predict the solar and wind energy forecasts. The training, using the entire data set of 1494 nodes and 21'043 hours of forecasting data (2GB) was performed on EPFL's SCITAS computing cluster. The code for the GCN using Pytorch is given in the Appendix.

3.2 Results

With the training complete, the GCN was tested by predicting the forecast of solar and wind energy production over the course of 5261 hours on the entire network. An example of such a prediction over 5 days is given in Figure 6. The prediction had good results, especially for solar which is perhaps more predictable than wind.

It is interesting to visualize what the average error is for the entire graph. For solar, 75% of the nodes have an error inferior to 10% and no pattern is noticeable. For wind, however, there are certain regions where the error is relatively low and others with errors above 70-80%. A hypothesis is that these regions correspond to ones with many mountains, e.g. the alps, which are among these high error areas. Since the prediction for a node is done with respect to adjacent nodes, perhaps the wind varies a lot in short distances due to the mountains cutting it off.

Overall this can be a useful tool to predict the renewable energy output for the entire power grid based on forecasts for only certain nodes. It could be used in situations where some weather station are out of order to still have a forecast for every node. The model clearly has limitations, especially for wind power predictions, which appear to have a correlation with the region where the node is located. Thus, for this renewable resource, the model could be useful in certain areas only. It would be interesting to incorporate information about the location of mountain ranges and relief.

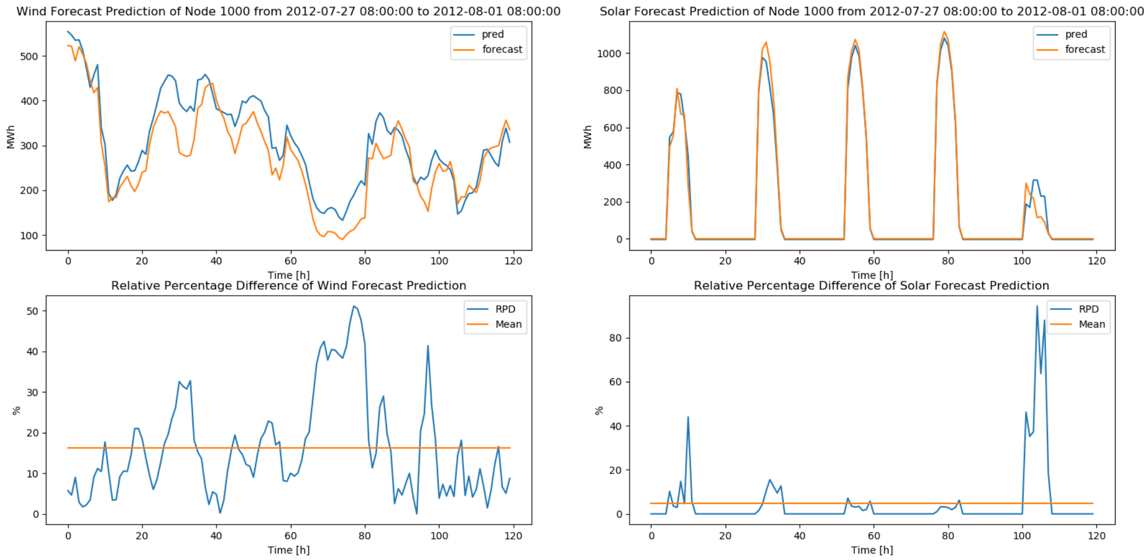


Figure 6: GCN prediction on one of the nodes during a five day period

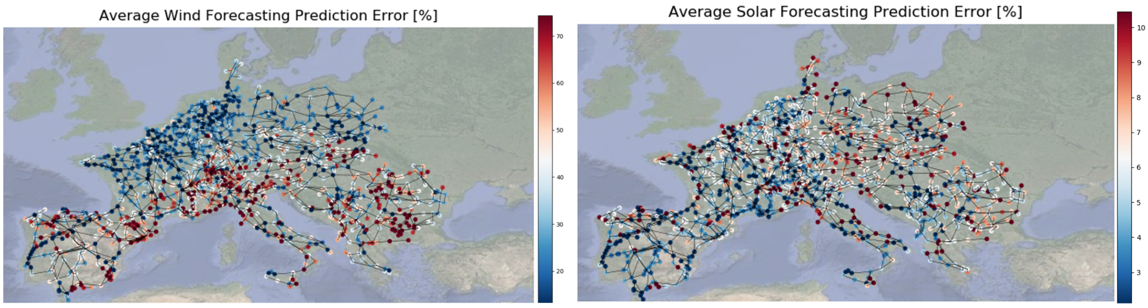


Figure 7: Average prediction error for the entire graph.

4 Conclusion

With the results obtained in this project, it can be seen that the use of network science on power grids can be useful. The GCN predictions are an excellent example of the potential these tools have. The impact of energy policies and regulations could be seen by performing a Fourier analysis, which clearly draws the boarders between some countries, suggesting an underlying social or political reasons. The even distribution of power consumption over time is very desirable, and it was seen that a Fourier analysis of the power demand signal (who's components were then plotted on the graph) was a useful measure.

On the other hand, a Graph Convolutional Network was used, which can be a useful product to obtain a full renewable energy forecasting given limited data. The prediction has some limitations for wind power in some regions, but it delivered very promising results for the prediction of solar energy output.

Appendix

```
class GCN(nn.Module):
    def __init__(self, A, D):
        super(GCN, self).__init__()
        self.A = torch.cat((torch.Tensor(A), torch.ones(A.shape[0], 1)), dim=1)
        self.D = torch.Tensor(np.diag(np.diag(D)**-1))
        self.W1 = nn.Parameter(torch.rand(self.A.shape))
        self.W2 = nn.Parameter(torch.rand(self.A.shape))
        self.W3 = nn.Parameter(torch.rand(self.A.shape))

    def forward(self, X):
        X = X.T
        X = torch.cat((torch.Tensor(X), torch.ones(1, X.shape[1])), dim = 0) #bias node
        X = self.D.mm(self.A*self.W1).mm(X)
        X = torch.cat((torch.Tensor(X), torch.ones(1, X.shape[1])), dim = 0) #bias node
        X = self.D.mm(self.A*self.W2).mm(X)
        X = torch.cat((torch.Tensor(X), torch.ones(1, X.shape[1])), dim = 0) #bias node
        X = self.D.mm(self.A*self.W3).mm(X)
        return X.T
    def reset_parameters(self):
        self.W1 = nn.Parameter(torch.rand(self.A.shape))
        self.W2 = nn.Parameter(torch.rand(self.A.shape))
        self.W3 = nn.Parameter(torch.rand(self.A.shape))
```

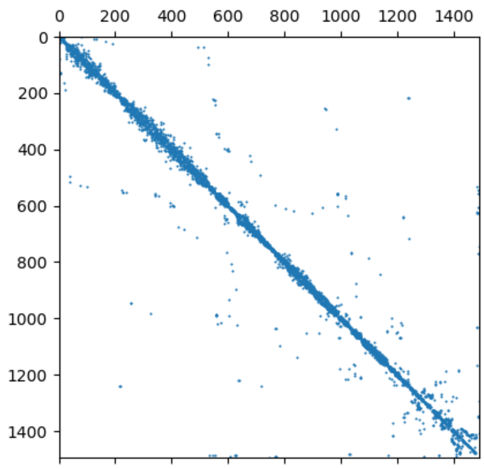


Figure 8: Adjacency matrix of the graph

References

[1] T. Santanoe Oepomo, “Graph Theory and Topology for 3 Phase Power System Under Faulted Studies”, IJRAS, vol. 10, pp 219-246, Feb. 2012.

[2] T.V. Jensen & P. Pinson, “RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system”, SCIENTIFIC DATA, vol. 4, Nov. 2017.

[3] S. Maggiore, M. Gallanti, M. Benini & W. Grattieri, “Evaluation of the effects of a tariff change on the Italian residential customers subject to a mandatory time-of-use tariff”, ECEE Summer Study Proceedings, pp 1909-1918, June 2013.

[4] F. Blin, “La repartition des differents types de chauffages en 2017”, QuelleEnergie.fr, May 2018, https://www.quelleenergie.fr/magazine/nouvelles-energie/quels-types-de-chauffage-energie-2017.

[5] H.R. Khosravani, M. Castilla, M. Berenguel, A. Ruano & P. Ferreira, “A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building”, Energies, vol. 9 (1), article n. 57, Jan. 2016.

[6] J. Abrell, “The Swiss Wholesale Electricity Market”, SCCER CREST, May 2016.

[7] X. Pinon, “Tarif Jour Nuit: Quel est le prix de l’électricité?”, Kelwatt, Nov. 2019, https://www.kelwatt.fr/guide/tarif-jour-nuit comprendre.