Rapid prediction of ground shaking intensity with Graph Neural Networks

Stefan Bloemheuvel- Tilburg University, Department of Cognitive Science and Artificial Intelligence, Tilburg, 5037 AB, The Netherlands, and Jheronimus Academy of Data Science (JADS), 's-Hertogenbosch, 5211 DA, The Netherlands, e-mail: s.d.bloemheuvel@jads.nl

Jurgen van den Hoogen– Tilburg University, Department of Cognitive Science and Artificial Intelligence, Tilburg, 5037 AB, The Netherlands, and Jheronimus Academy of Data Science (JADS), 's-Hertogenbosch, 5211 DA, The Netherlands, e-mail: j.o.d.hoogen@jads.nl

Dario Jozinović- Istituto Nazionale di Geofisica e Vulcanologia, via di Vigna Murata 605, 00143 Roma, Italy, and Swiss Seismological Service, ETH Zurich, Sonneggstrasse 5, 8092, Zurich, Switzerland e-mail: dario.jozinovic@sed.ethz.ch)

Alberto Michelini– Istituto Nazionale di Geofisica e Vulcanologia, via di Vigna Murata 605, 00143 Roma, Italy, e-mail: alberto.michelini@ingv.it

Martin Atzmueller– Osnabrück University, Semantic Information Systems Group, Wachsbleiche 27, 49090, Osnabrück, Germany, and German Research Center for Artificial Intelligence (DFKI), Berghoffstraße 11, 49090, Osnabrück, Germany, e-mail: martin.atzmueller@uni-osnabrueck.de

Abstract: Rapid accurate prediction of strong ground shaking can be crucial for earthquake early warning. Recently, machine learning (ML), with its advances in Deep Learning (DL), has shown great potential in analysing seismic waveforms. More specifically, when using the data acquired by a seismic network, the incorporation of additional information consisting of the network station positioning into the DL model has been found beneficial to improve the accuracy of the ground motion predictions (Jozinović et al., 2022). Such spatial information can be exploited thoroughly by adopting graph structures, along with the seismic waveforms. Recent advances in adapting DL to graphs have shown promising potential in various graph-related tasks. However, these methods have not been completely adapted for seismological tasks. In this work, we advance an architecture capable of processing a set of seismic time series acquired by a network of stations using the benefits of Graph Neural Networks (GNNs) (see Fig. 1). The objective of the study is the rapid determination of the ground motion (PGA, PGV, and SA 0.3s, 1s and 3s) at farther stations that have not been yet reached by the strong ground shaking by availing of the first signals recorded at the stations close to the epicentre. The work builds upon the GNN approach proposed in Bloemheuvel et al. (2022) and incorporates transfer learning, see Jozinović et al. (2022). We apply the methodology to two datasets having very different source-receiver geometries sited in central Italy (CI, Jozinović et al., 2020, Jozinović et al., 2022) and in north-western central Italy (CW), respectively (Fig. 2). The two datasets have already been the object of similar studies using convolutional neural networks which serve as baselines for comparison. We find that the GNNs are highly suited for the analysis of seismic data from a set of stations and show improvement when compared to the previous work (Bloemheuvel et al., 2022 and Jozinović et al., 2022). We exemplify the early warning capabilities of the proposed approach.

Keywords: Graph Neural Networks, Seismogram, Convolutional Neural Networks, Sensors, Regression, Earthquake Ground Motion, Seismic Network, EEW

1. Introduction

In seismology, the adoption of graph neural networks is at the very beginning. The problem belongs to what in the machine learning community is called multivariate time series regression using graph-based methods (Figure 1). Initial attempts in seismology have been made by van den Ende and Ampuero (2020), Yano et al. (2021), Kim et al. (2021), yet each have some shortcomings. van den Ende and Ampuero (2020) mention that they designed a GNN for the localization of earthquakes from waveform data. However, they only append the (latitude, longitude) information to the time series being handled by a CNN. Therefore, while prediction scores improved, no actual GNN layers were used. Second, Yano et al. (2021) proposed a graph partitioning algorithm that works together with a CNN. However, they make use of classical graph theory techniques and a GNN method is not applied. Lastly, Kim et al. (2021) recently suggested a method that uses CNNs and GNNs for seismic event classification. However, (1) no spatial information is used at all, i.e., each edge has a weight of 1, nor (2) meta information about the stations is added, and (3) only three nodes are examined for each observation, which could be difficult to interpret as a full-fledged/complex network. Finally, another recently proposed method for handling seismic data with GNNs is from McBrearty and Beroza (2022). Here, the location and magnitude of earthquakes is predicted. However, their input are pre- calculated characteristics of the earthquakes for each station, which differs from our goal to use raw waveform data.

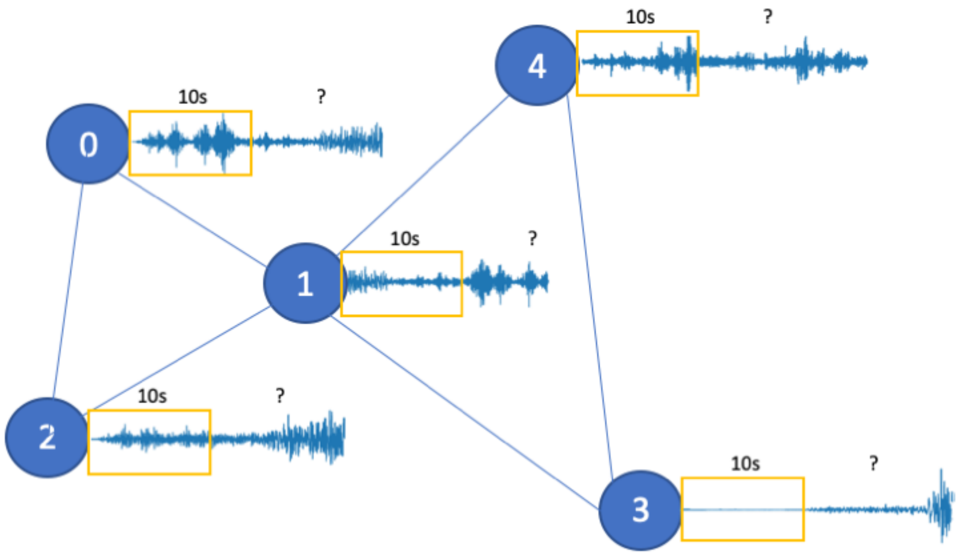


Fig. 1 – Simplified diagram showing the Graph Neural Network (GNN) approach. Stations represent the nodes and the connections between them represent the edges of the GNN. The objective is to determine the maximum ground shaking at the nodes by exploiting the pattern of the recorded seismograms in the e.g., first 10 s. The node features of the GNN are the calculated features from the convolutional layers in the model.

In this work, we extend the work by Bloemheuvel et al. (2022) that present a technique that takes full advantage of the GNN architecture to process multivariate time series. In practice, the technique combines the capabilities of convolutional layers (feature extraction) and graph convolutional layers (spatial information) to manage the feature sizes that are common in high-frequency time series data arising from multiple sensors. This work is enriched with the adoption of transfer learning (Bozinovski, 2020; Pan and Yang 2009 for a review) to alleviate the paucity of the datasets by taking advantage of the developments by Jozinović et al. (2022). The seismological problem that is addressed consists of predicting the peak ground motion at a set of farther located seismic stations from the epicenter using the very first seconds of the first recoding stations. This approach does not require the provision of the earthquake location and magnitude and it can reveal useful for earthquake early warning purposes (Jozinović et al., 2020, 2022).

2. Data

In this work, we have used two datasets – the Central Italy (CI) and the Central Western Italy (CW) ones. They consist of three-component earthquake waveforms data. The CI dataset includes the 2016 Central Italy sequence (Chiaraluce et al. 2017) recorded by 39 stations in the epicentral area and its surroundings. We use earthquakes in the study area bounded by latitude [42◦, 43.75◦] and longitude [12.3◦,14◦] which occurred from 1 January 2016 to 29 November 2016. All the events occur within crustal depths in the range 1.6 km < D < 28.9 km. Using these criteria, 915 earthquakes with magnitude M ≥ 3 have been used (Figure 2). The CW dataset includes three-component waveforms of 256 earthquakes with magnitude 2.9 ≤ M ≤ 5.1 (Fig. 1a), from 39 stations. The earthquakes occurred between 2013 January 1 and 2017 November 20. The earthquake depths range from 3.3 to 64.7 km (Fig. 1b). The stations and the earthquakes are located in the area bounded by latitude [41.13◦, 46.13◦] and longitude [8.5◦, 13.1◦](Fig.2). The dataset are available at https://doi.org/10.5281/zenodo.3669969 and https://doi.org/10.5281/zenodo.5541083 for CI and CW, respectively.

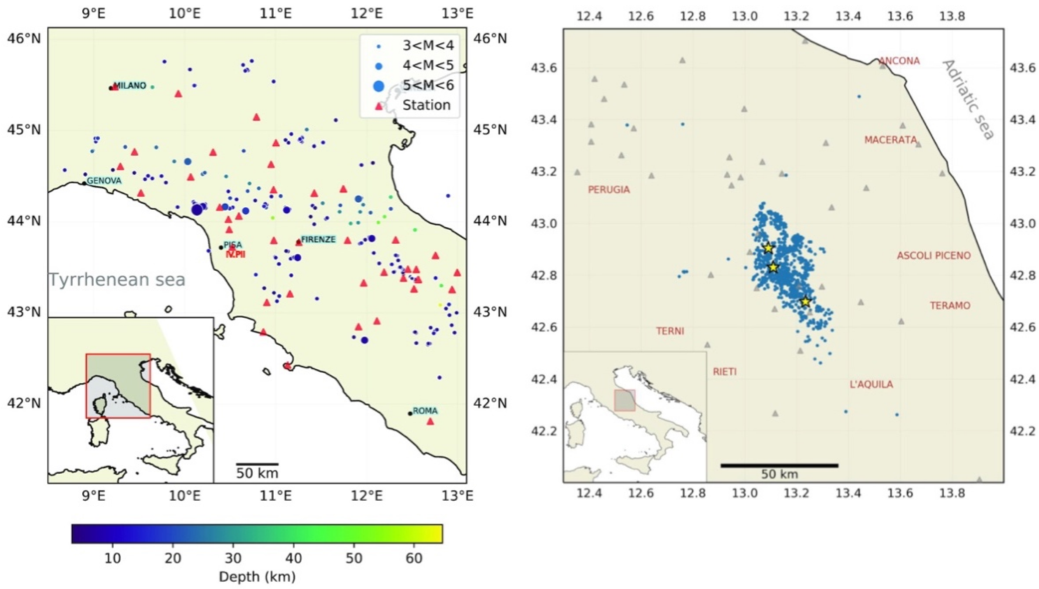


Fig. 2 – Maps showing the source-receiver geometries of the CW (left) and CI (right) datasets analysed by Jozinović et al. (2020, 2022) and used in this study. The CW dataset includes earthquakes and stations extending on a wide area in north-west central Italy with earthquakes occurring at depths between the near surface and about 65 km whereas the CI dataset spans a much smaller area and earthquakes at shallow depths from the 2016 central Italy earthquake sequence.

3. Method

The method adopts the model developed by Bloemheuvel et al. (2022) based on the CNN model developed by Jozinović et al. (2020). In practice, the model replaces the spatial convolutional layer with a graph convolutional layer. In Figure 3, we show the simplified diagram of the model. The first two layers perform single station feature extraction through convolution. In addition, transfer learning is inserted in these first two layers using features extracted from the STEAD (Mousavi et al., 2019) dataset as in Jozinović et al. (2022). After tensor re-shaping, graph convolutional (GCN) layers are applied. These are followed by flattening plus a dense layer to estimate the regression values and terminate the model. The stations represent the nodes and the connections between them represent the edges of the GNN. The objective is to determine the maximum ground shaking at the nodes by exploiting the pattern of the recorded seismograms in the e.g., first 10 s.

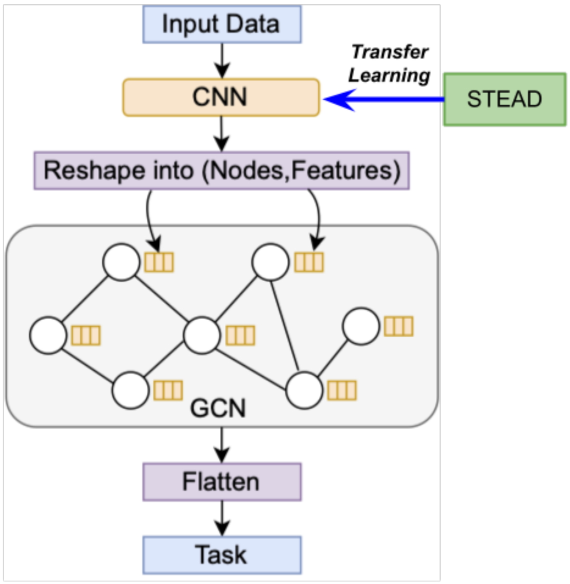


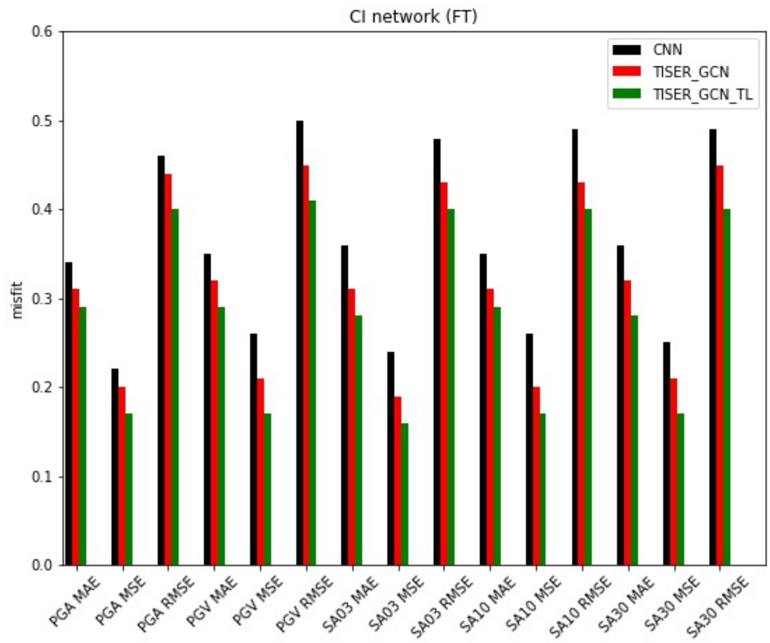
Fig. 3 – Diagram showing the adopted GCN model. CNN refers to the first two layers in which the convolution is carried out for each station 3C waveforms independently. Coefficients obtained from the STEAD dataset training are inserted in these two layers. Reshaping is applied before adopting GCN layer. Flattening and estimation of the regression values terminate the model. The stations represent the nodes and the connections between them represent the edges of the GNN. The objective is to determine the maximum ground shaking at the nodes by exploiting the pattern of the recorded seismograms in the e.g., first 10 s.

4. Application

The method has been applied to both the CI and the CW datasets. The results are summarized in Figure 4 by plotting the error for the five IMs according to three different error functions - maximum absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). The results shown include the TISER-GCN model fully described in Bloemheuvel et al. (2022), the TISER-GCN-TL developed mainly in this study and the CNN model of Jozinović et al. (2020) for comparison.

We see that for the CI network dataset, the adoption of the graph formalism to replace the spatial convolution of the CNN model leads to significant error reductions for all the IMs and the addition of the transfer learning improves in some cases the performance. For the CW network dataset, that features a very different source receiver geometry, range of event magnitudes and hypocentral depth, the TISER-GCN model results also in significant improvements when compared to the CNN.

We included transfer learning by inserting a model trained on the STEAD dataset (Mousavi et al., 2019) for magnitude determination (cf. Jozinović et al., 2022). The transfer learning in the first two layers was applied in two different ways. In both cases, we made the first layer not trainable. That is, we fixed the layer and used the coefficients provided by the model pre-trained on STEAD. The second layer was made either trainable or fixed. The results shown below in Fig. 4 are for the case in which the second layer is made trainable while using the STEAD coefficients as prior.



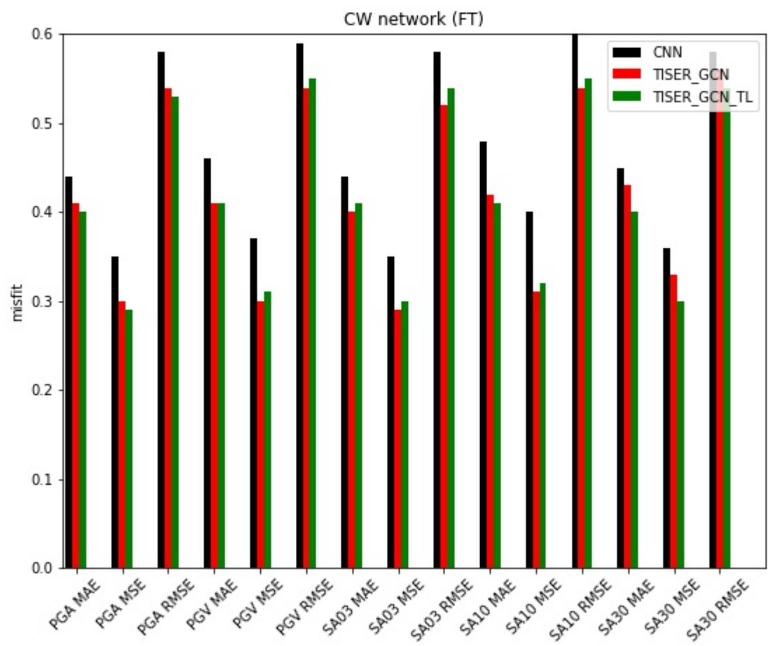


Fig. 4 – Bar plots showing the performance of the CNN (Jozinović et al., 2020), the graph convolutional network (TISER-GCN, Bloemheuvel et al., 2022) and the latest development that include the transfer learning in the previous model (TISER-GCN-TL). CI network (top) and CW network (bottom) are shown for MAE, MSE and RMSE for all the five IMs (PGA, PGV, SA(0.3), SA(1.0) and SA(3.0)).

In detail, inclusion of transfer learning for the CI dataset leads to significant improvements in all the error metrics adopted. In contrast, the results obtained using the CW dataset do not show the same behaviour and the results are either similar to or slightly worse than those obtained using the GCN. However, we note that some improvements were obtained when also the second layer was fixed (not shown) in the analysis of the CW dataset. This different behaviour could be explained by the limited resolving power of the CW dataset and the inclusion of the coefficients provided by the model pre-trained on STEAD becomes very relevant.

5. Conclusions

The results of the analysis presented in this work indicate that adoption of the graph neural network formalism can be extremely effective in the analysis of seismological data acquired by seismic networks spread unevenly across a territory. Overall, the GCN formalism performs always consistently better than the CNN approach.

Inclusion of transfer learning benefits the results of the GCN approach even further. Care must be observed, however, on whether to insert the external information through transfer learning fully or not. Our preliminary results suggest that for poor source receiver geometries it is preferable to use fully the external information (e.g., for the CW dataset in this study) whereas for datasets that have a good resolving power (e.g., the CI dataset), inclusion of transfer learning is still effective through a balanced insertion of the amount of information.

Acknowledgements

This work has been funded by the Interreg North-West Europe program (Interreg NWE), project Di-Plast - Digital Circular Economy for the Plastics Industry (NWE729), and partially funded by the INGV project Pianeta Dinamico 2021 Tema 8 SOME (grant no. CUP D53J1900017001) funded by the Italian Ministry of University and Research "Fondo finalizzato al rilancio degli investimenti delle amministrazioni centrali dello Stato e allo sviluppo del Paese, legge 145/2018" and by the European Union's Horizon 2020 research and innovation program (grant no. 821115), Real-time Earthquake Risk Reduction for a Resilient Europe (RISE).

References

Bloemheuvel, Stefan, Jurgen van den Hoogen, Dario Jozinović, Alberto Michelini, and Martin Atzmueller. "Multivariate Time Series Regression with Graph Neural Networks." *ArXiv:2201.00818 [Cs]*, January 3, 2022. http://arxiv.org/abs/2201.00818.

Bozinovski, S., 2020. Reminder of the first paper on transfer learning in neural networks, 1976, Informatica, 44(3), 291–302

Jozinović, Dario, Anthony Lomax, Ivan Štajduhar, and Alberto Michelini. "Rapid Prediction of Earthquake Ground Shaking Intensity Using Raw Waveform Data and a Convolutional Neural Network." Geophysical Journal International 222, no. 2 (August 1, 2020): 1379–89. https://doi.org/10.1093/gji/ggaa233.

———. "Transfer Learning: Improving Neural Network Based Prediction of Earthquake Ground Shaking for an Area with Insufficient Training Data." Geophysical Journal International 229, no. 1 (April 1, 2022): 704–18. https://doi.org/10.1093/gji/ggab488.

Kim, G., Ku, B., Ahn, J.-k., Ko, H.: Graph convolution networks for seismic events classification using raw waveform data from mul- tiple stations. IEEE Geoscience and Remote Sensing Letters (2021)

McBrearty, I.W., Beroza, G.C.: Earthquake location and magnitude estimation with graph neural networks. arXiv preprint arXiv:2203.05144 (2022)

Mousavi, S. Mostafa, Yixiao Sheng, Weiqiang Zhu, and Gregory C. Beroza. "STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI." *IEEE Access* PP, no. 99 (October 15, 2019): 1–1. https://doi.org/10.1109/ACCESS.2019.2947848.

Pan, S. J. & Yang, Q., 2009. A survey on transfer learning, IEEE Trans. Knowl. Data Eng., 22(10), 1345–1359.

van den Ende M.P. and Ampuero, J.-P.: Auto- mated seismic source characterization using deep graph neural networks. Geophysical Research Letters 47(17), 2020–088690 (2020)

Yano, K., Shiina, T., Kurata, S., Kato, A., Komaki, F., Sakai, S., Hirata, N.: Graph- partitioning based convolutional neural net- work for earthquake detection using a seismic array. Journal of Geophysical Research: Solid Earth 126(5), 2020–020269 (2021)