GRAPH NEURAL NETWORK MODEL AS A SERVICE

ARBËR CENI, LEDIA BOZO, ALDA KIKA

Department of Informatics, Faculty of Natural Sciences, University of Tirana

e-mail: arber.ceni@fshn.edu.al

Abstract

Nowadays, the use of Social Media networks is growing rapidly, providing a substantial pool for unstructured data. The use of social media data is a growing area of interest, which has opened a new path for social science research concerning human behaviours. Having a tool to analyse and understand the dissemination of fake news in a social media provides a contribution to this area. This paper aims to present a graph neural network model as a service. We have implemented a service model to analyse the large amount of data in an understandable format for not professional users. First, a "Graph Generation" service is implemented based on real-time requests from an interface. Second, the "Prediction" service takes the generated graph as a parameter and according to the specified requirements, communicates real-time with a trained model to calculate high probability users of being the spreader of fake news. Finally, the results are displayed in a visualised graph format. The use of data visualisation is essential in patterns identification and future trends forecast.

Key words**:** *Model as e Service, Social network data, Graph, Twitter.*

Përmbledhje:

Në ditët e sotme, përdorimi i rrjeteve të mediave sociale po rritet me shpejtësi, duke ofruar një burim të konsiderueshëm për të dhënat e pastrukturuara. Përdorimi i të dhënave të mediave sociale është një fushë interesi në rritje, e cila ka hapur një rrugë të re për kërkimin e shkencave sociale në lidhje me sjelljet njerëzore. Të kesh një mjet për të analizuar dhe kuptuar përhapjen e lajmeve të rreme në një media sociale jep një kontribut në këtë fushë.

Ky botim ka për qëllim të paraqesë një model rrjeti neural në grafe si një shërbim. Kemi implementuar një model shërbimi për të analizuar sasinë e madhe të të dhënave në një format të kuptueshëm për përdoruesit jo

profesionistë. Së pari, është implementuar një shërbim "Gjenerimi i Grafit" bazuar në kërkesat në kohë reale nga një ndërfaqe. Së dyti, shërbimi "Parashikim" merr si parametër grafin e gjeneruar dhe sipas kërkesave të specifikuara komunikon në kohë reale me një model të trajnuar për të llogaritur përdoruesit me probabilitet të lartë të të qenit përhapës i lajmeve të rreme. Së fundi, rezultatet shfaqen si një graf i vizualizuar. Përdorimi i vizualizimit të të dhënave është thelbësor në identifikimin e modeleve dhe parashikimin e prirjeve të ardhshme.

Fjalë kyçe: *Model si Shërbim, Të dhënat e rrjetit social, Graf, Twitter.*

Introduction

The use of social media networks has become increasingly prevalent, with a significant portion of the global population actively engaging in these platforms (Hansen et al., 2011; Kaur & Sandhu, 2021). It has a significant impact on communication, information sharing, and social interaction (Sood et al., 2020). The rise in usage can be attributed to the enjoyment derived from these networks, as well as the ability to connect with others and businesses (Curran & Lennon, 2011). However, the impact of social media on individuals can be both positive and negative, with studies revealing therapeutic effects such as increased self-esteem, as well as negative consequences like the activation of a narcissistic state (Kaur & Sandhu, 2021).

The acceptance and spread of fake news in social media is a significant issue. This problem is exacerbated by the ease of access and low cost of distribution on social media, which also acts as a source of fake news (Narwal, 2018). This makes it extremely important to combat the spread of fake news and to aware the society. Analysing social media data and providing a tool to understand and predict the dissemination of fake news has a great interest.

This paper presents an approach for modelling and analysing social media data as a service and the need for tools and methodologies to effectively analyse and utilise this data. This modelling provides structuring social media data to address challenges such as fake news and manipulation.

The paper is divided into five sections. The first section introduces motivating scenarios and core ideas of the paper. The second section highlights the related work. Section “Data and Methodology” describes data manipulation and processes from data collection to services implementation and data

visualisation. The results and analysis are interpreted and explained in Section IV. The paper is finally concluded in Section V.

Literature Review

A range of studies have explored the use of social data analysis to understand user behaviour. Benevenuto et al. (2009) used clickstream data to identify patterns in social network workloads, revealing that browsing accounted for the majority of user activities. Tang et al. (2011) applied social network analysis to a local social media platform, uncovering correlations between social networks and user behaviours. Bhagat et al. (2015) proposed a cut-based classification approach for analysing user behaviour on social networking sites, with a focus on sentiment analysis. These studies demonstrate the value of social data analysis in understanding user behaviour, with implications for interface design, content distribution, and decision-making.

There are various detection and mitigation approaches that are proposed to deal with the challenge of fake news on social media. Hakak et. al (2020) highlights the need for a data sharing platform to facilitate machine learning, while Reddi and Eswar (2021) present a Modified Long Short-Term Memory network as an effective method for fake news recognition. Mahid et al. (2018) categorises detection techniques into content-based, social context-based, and hybrid-based methods, and identifies open research challenges. These authors underscore the importance of ongoing research and awareness in addressing the threat of fake news in social media.

The use of social network data as an e-service is a growing area of interest, with a focus on real-time data analysis (Jang & Kwak, 2014). This data can be leveraged for service innovation and transformation, particularly in community-based e-services, through the synthesis of semantic social network and attraction theory (Yuan & Fei, 2010). Tang and Terzi (2010) highlight the rich content and linkage data available in online social networks, providing opportunities for data analytics. Krishna (2021) further emphasizes the use of data mining techniques, machine learning, and statistical modelling for the analysis of social networking data.

Data and Methodology

In this section we show data collection and the applied methodology in order to perform experiments and achieve results on collected data.

Data Collection

First, data is collected from Twitter, through Twitter API 1.1, using free access mode. As we were interested about communities that comment on Covid-19 and vaccines, a range of related words are specified for this topic. The set of searching terms can be generalized in categories and subcategories, which are different approaches for the specified topic.

Every day a scheduler was executed in order to collect tweets for a period of nine months from November 2021 to July 2022. The information is stored in a database stamping the whole data in a JSON format. (Ceni et al., 2023) This information has to be processed in other stages in order to analyse data properly.

Data Modelling

We need to define a model to present the large amount of data in an understandable format. As authors have studied in (Ceni et al., 2023), a number of possible graph file formats were considered to be used for storing the generated graphs. In particular they investigated GEXF, GDF, GraphML and adjacency matrix. Their study resulted in choosing GraphML as an appropriate format for graph generation.

The GraphML file format results from the joint effort of the graph drawing community to define a common format for exchanging graph structure data. It uses an XML-based syntax and supports the entire range of possible graph structure constellations including directed, undirected, mixed graphs, hypergraphs, and application-specific attributes. (Brandes et al., 2014)

As authors have explained in Ceni et al. (2023), the general idea of the graph generation algorithm is to get from the database all the records matching the model requested by the user. The author's posts are extracted and stored in a dictionary in order to prevent duplicated users. For each of the records, the algorithm extracts users that have been mentioned or replied to and adds them to the dictionary as well if they are not present. Finally, for each record, the algorithm creates the appropriate edges of the graph if one of the actions in the list is applied: "ReplyTo", "Retweet", "Mention", "Tweet" (self-loop) or "MentionInRetweet".

Service Modelling

One major challenge in this paper is to develop data services to efficiently process the collected data from specified social information services and display this information in a web application. The need for agility and innovation in today's world has led to the emergence of service orientations and transformations, which can support organizational agility (Demirkan, 2016). In this context, we have implemented several modules, which enable receiving the information defined by the user and presenting this information in the form of a visual graph more suitable for a non-professional user to understand.

First, we have created a model, which processes the raw data, performs data cleaning and preprocessing in order to increase data quality and performs some general filters to get the requested information. The data is processed to make the users unique, and the important information is cleaned and stored in an indexed table. We have considered the following filters displayed in Table 1 in modelling our model.

Table 1. Service Model Parameters

Service Parameters	**Description and Limitations**
Start Date	Gets the data from the start date. Required parameter.
End Date	Gets the data until the end date. Required parameter.
Topic (category/subcategory)	Lists of relevant words related to the specified topic are considered.
Language	All languages are considered. Special interest in Albanian

	language posts.
Graph Size	Several Graph sizes are considered from 100 nodes to 1000 nodes.

Second, the "Graph Generation" module is provided as a service. It is integrated with the web application enabling to receive real-time requests from the user and generate the graph in GraphML format. It is implemented in .NET by considering each of the relationships as an edge with equal weight in our graph. The users build the nodes of the graph and other important information is included as attributes for nodes and edges. This format is flexible and can be customised according to the selected attributes.

Third, the "Prediction" service has been i

mplemented, which takes the graph generated as a memory stream parameter according to the specified requirements and feeds it to the trained model studied by the authors in (Kika et al., 2023) to enable the calculation of the probability of being inactive for all users of the graph. This service will return a dictionary with users (nodes) and the prediction if that user account is going to be inactive according to the GNN model.

The service is conceived and designed consisting of several modules, which increase flexibility by enabling us to use libraries that better address our needs. The component layers of the service are:

- Data
- The model
- Infrastructure
- API

The implemented modules are presented in Figure 1.

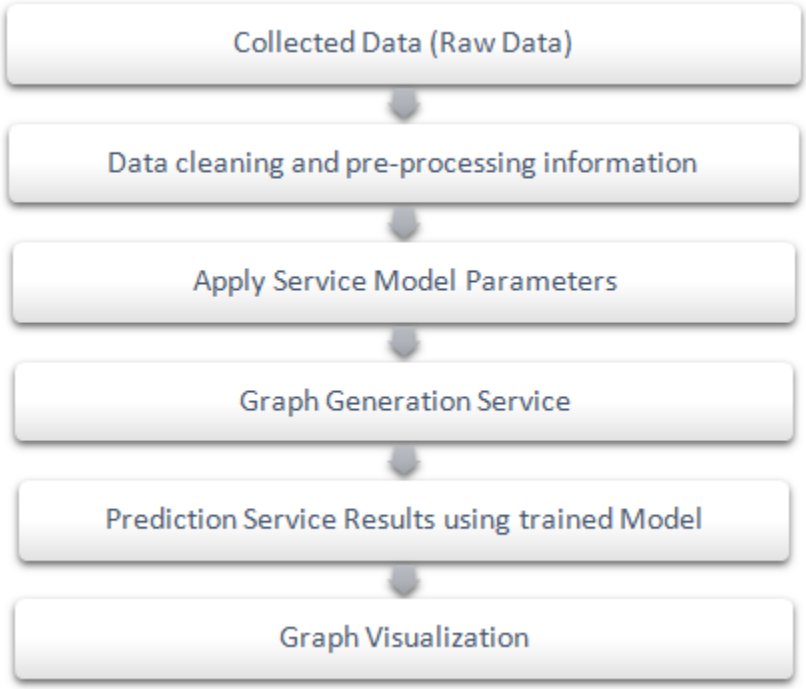

Figure 1. The implemented modules

Finally, the results have to be displayed in a visualized graph format. We have used Sigma.js and graphology to visualize the final graph. Sigma.js is a modern JavaScript library for rendering and interacting with network graphs in the browser. It works in symbiosis with graphology, a multipurpose graph manipulation library, which is used to handle graph data models & algorithms. Sigma.js library is aimed at visualizing graphs of thousands of nodes and edges and provides features to personalize the way your graphs are rendered. (SigmajsOrg, 2023)

Analysis & Results

The services implementation enables the model data to be easily accessed by different users. The main aim of this paper is to integrate all the features

optimized and to offer this functionality to non-professional users for further analysis.

An interesting statistic that users need is the most active users filtered by the period of time, user language, selected topic, and graph size. This information is displayed in real-time in a graphic form. Figure 2 presents the most active users discussing “moderna”.

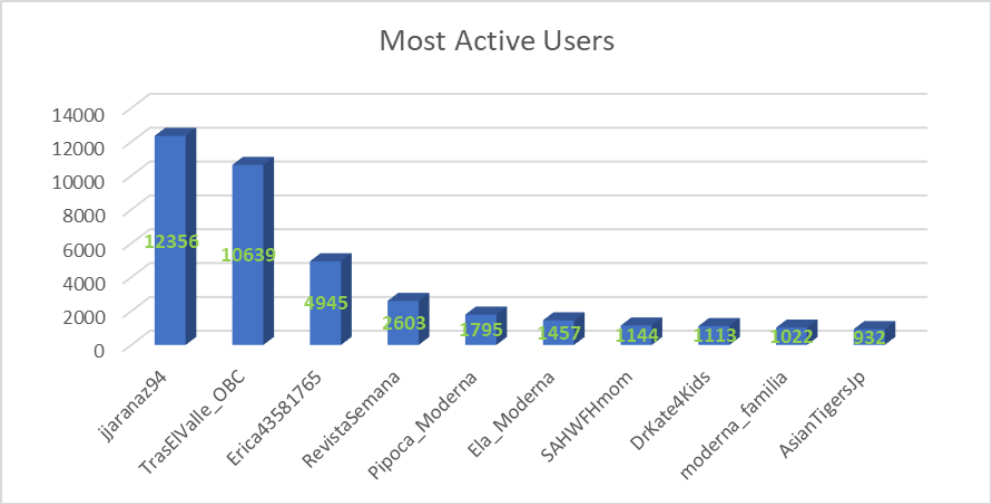

Figure 2. Most active users discussing “moderna”

A big challenge of this implementation is model performance. Graph generation, manipulation and visualization reduce performance if the model is not optimised. In order to achieve better results in performance we have used modular design, new components and optimised parallel algorithms.

Finally, the trained model is provided to the user through a web application where it is possible to define some criteria and present the results in real-time in graphic form with 1000 nodes as presented in Figure 3.

The graph can be analysed in real-time considering any suspicious joints. The application provides rich information for every single user, such as name, profile, page, description, image, followers, date etc.

The model has predicted the central node as a fake news spreader. This information is displayed in the visualised graph with red colour. The model provides the information that the “red” user represents a potential risk to the network and it is necessary for the analysers to make further verifications and take the necessary measures.

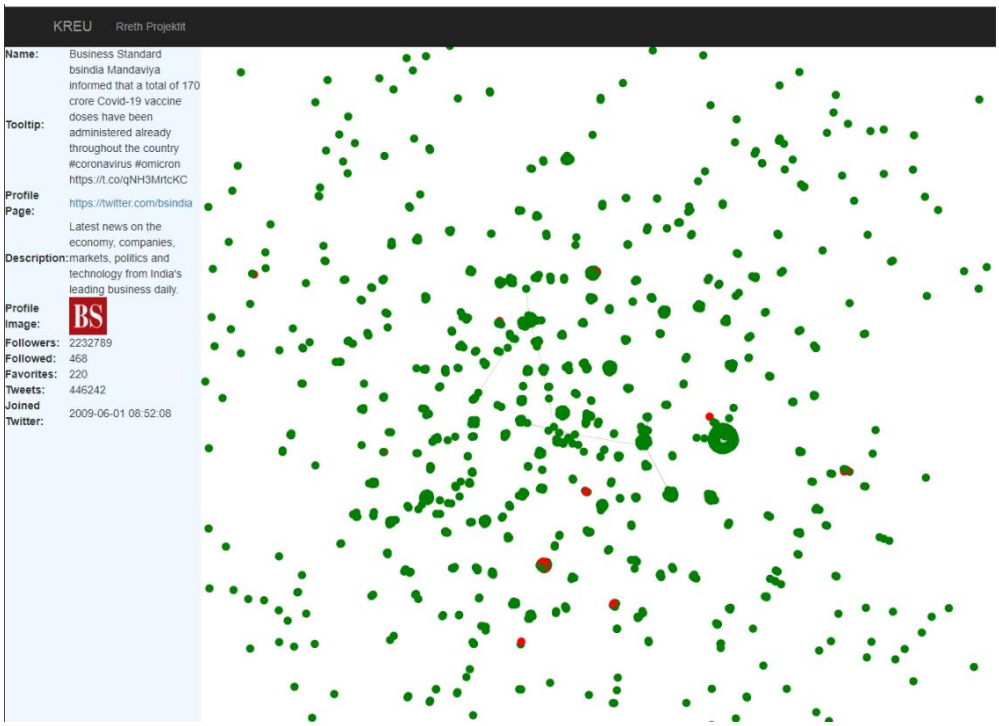

Figure 3. Graph visualization

The service implemented for the calculation of the probability of active users has enabled the generation of the graph in Figure 4. In this visualization, only one group we are interested in is considered. The service has taken as a parameter only the data collected at the beginning of November 2021 for the term "Covid-19" in the English language with a maximum of 500 users. The Prediction service takes as a parameter the graph of the specified users and based on the model data, returns the probability of whether these users are

active (green color) or not (red color). The service processes data asynchronously, significantly improving the performance of returning the result in real-time.

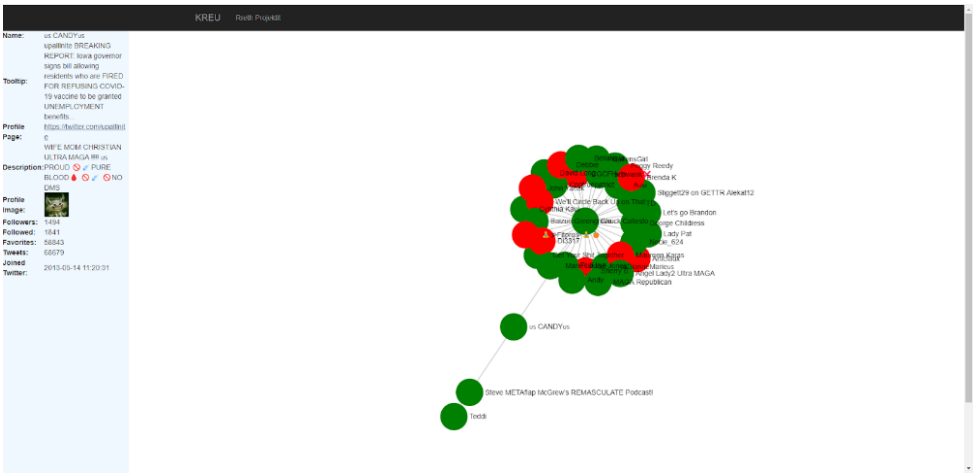

Figure 4. Graph visualization according to specified criteria

Conclusions

In this paper we present an architecture to implement graph neural networks as a service. The neural network chosen to be implemented as a service is the model presented in (Kika et al., 2023). This graph neural network is based on the GraphSMOTE model and is trained on Twitter data. The model takes a graph of Twitter users as input and tries to predict which user account is going to be inactive (suspended) based on the data it has been trained on. Our architecture consists of three components: graph generation, prediction and visualization.

The graph generation component will generate a graph based on the criteria specified by the user. The prediction component will pass this generated graph to the trained model and store the output of the model. The visualization component will visualize the generated graph and the model's output where

each “active” prediction will be visualized as a green node in the graph and each “inactive” prediction will be visualized as a red node in the graph. Implementing such a service has a huge impact on the usability of machine learning models. Usually, these models are difficult to be used and understood by non-professional users, so a way to ease these difficulties needs to be implemented. Another positive aspect of this work is that with the visualization offered by the service we have a better understanding of the results, a better interpretability, and a better general sense of the validity of the results.

In this aspect, we notice that the model does not always perform as it should regarding accuracy where we can distinguish some false negatives. A possible improvement to the model can be to use active learning to improve the model's accuracy. Given the vast amount of raw data, querying the database takes most of the processing time during graph generation. A possible improvement can be to use subpartitioning by date. We leave these proposed improvements as future work.

Acknowledgements

This publication is made possible with the financial support of the National Agency for Scientific Research and Innovation (AKKSHI). Its content is the responsibility of the authors, the opinion expressed in it is not necessarily the opinion of AKKSHI.

References

A. Ceni, A. Kika and D. Çollaku, A Social Network Analysis of COVID-19 Vaccines Tweets, 5th International Conference on Recent Trends and Applications in Computer Science and Information Technology, April 26-27, 2023.

A. Kika, A Ceni, D. Collaku, E. Loka, L. Bozo, K. Hoxha, Imbalance node classification with GNN: A study on a Twitter dataset, (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 14, No. 11, 2023

A. P. Bhagat, K. A. Dongre and P. A. Khodke, "Cut-based classification for user behavioral analysis on social websites," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Greater Noida, India, 2015, pp. 53-59, doi: 10.1109/ICGCIoT.2015.7380427.

B. Narwal, "Fake News in Digital Media," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, pp. 977-981, 2018.

D. L. Hansen, B. Shneiderman, M. A. Smith, “Analyzing Social Media Networks with NodeXL”, Morgan Kaufmann, pp 3-9, 2011.

F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM conference on Internet measurement (IMC '09). Association for Computing Machinery, New York, NY, USA, pp. 49–62. 2009. https://doi.org/10.1145/1644893.1644900

G. Krishna, Social Networking Data Analysis Tools and Services. In H. Bansal, G. Shrivastava, G. Nguyen, & L. Stanciu (Eds.), Social Network Analytics for Contemporary Business Organizations pp. 19-34. IGI Global, 2018.

H. Demirkan, J.C. Spohrer, Emerging service orientations and transformations (SOT). Inf Syst Front 18, pp. 407–411, 2016. https://doi.org/10.1007/s10796-016-9656-8

J. Curran & R. Lennon, Participating in the Conversation: Exploring Usage of Social Media Networking Sites. Academy of Marketing Studies Journal, vol. 15, pp. 21, 2011.

J. Tang & E. Terzi, Social Network Data Analytics, Computer Science, 2010.

N. Kaur and A. Sandhu, “The Use of Web-Based Social Networks: Trends, Antecedents, and Consequences”, Special issue on Management and Social Media, vol. 18, pp. 233-245, 2021.

S. Hakak, W.Z. Khan, S. Bhattacharya, G.T. Reddy, KK.R. Choo, Propagation of Fake News on Social Media: Challenges and Opportunities. In: Chellappan, S., Choo, K K.R., Phan, N. (eds) Computational Data and Social Networks. CSoNet 2020. vol 12575. Springer, Cham. 2020.

S. Sood, R. Verma, & A.K. Pal, A Review of Use of Social Media and its Significance, 2020.

S. Reddi, G.V. Eswar, Chapter 9 - Fake news in social media recognition using Modified Long Short-Term Memory network, Fadi Al-Turjman, B.D. Deebak, In Intelligent Data-Centric Systems, Security in IoT Social Networks, Academic Press, pp. 205-227, ISBN 9780128215999, 2021.

S.-T. Yuan, Y.-L. Fei, A synthesis of semantic social network and attraction theory for innovating community-based e-service, Expert Systems with Applications, vol 37, Issue 5, pp. 3588-3597, ISSN 0957-4174, 2010.

SigmajsOrg. (2023), Retrived from

https://github.com/jacomyal/sigma.js/blob/main/README.md

T. Tang, M. Hämäläinen, A. Virolainen, and J. Makkonen. Understanding user behavior in a local social media platform by social network analysis. 15th International Academic

MindTrek Conference: Envisioning Future Media Environments (MindTrek '11). NY, USA, pp. 183–188. 2011. https://doi.org/10.1145/2181037.2181067

U. Brandes, M. Eiglsperger; J. Lerner and C. Pich, "Graph Markup Language (GraphML)". In Tamassia, Roberto (ed.). Handbook of Graph Drawing and Visualization (PDF). CRC Press. pp. 517–541, 2014.

YJ. Jang, J. Kwak, Social Network Service Real Time Data Analysis Process Research. In: Park, J., Zomaya, A., Jeong, HY., Obaidat, M. (eds) Frontier and Innovation in Future Computing and Communications. vol 301. Springer, Dordrecht. 2014.

Z. I. Mahid, S. Manickam and S. Karuppayah, "Fake News on Social Media: Brief Review on Detection Techniques," 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, 2018, pp. 1-5, doi: 10.1109/ICACCAF.2018.8776689.