International Journal on Advances in Networks and Services

2023 vol. 16 nr. 1&2

The *International Journal on Advances in Networks and Services* is published by IARIA.
ISSN: 1942-2644
journals site: http://www.iariajournals.org
contact: petre@iaria.org

Sponsored by IARIA
www.iaria.org

International Journal on Advances in Networks and Services
Volume 16, Number 1 & 2, 2023

Christos Bouras, University of Patras, Greece
Mahmoud Brahimi, University of Msila, Algeria
Marco Bruti, Telecom Italia Sparkle S.p.A., Italy
Dumitru Burdescu, University of Craiova, Romania
Diletta Romana Cacciagrano, University of Camerino, Italy
Maria-Dolores Cano, Universidad Politécnica de Cartagena, Spain
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain
Eduardo Cerqueira, Federal University of Para, Brazil
Bruno Chatras, Orange Labs, France
Marc Cheboldaeff, Deloitte Consulting GmbH, Germany
Kong Cheng, Vencore Labs, USA
Dickson Chiu, Dickson Computer Systems, Hong Kong
Andrzej Chydzinski, Silesian University of Technology, Poland
Hugo Coll Ferri, Polytechnic University of Valencia, Spain
Noelia Correia, University of the Algarve, Portugal
Noël Crespi, Institut Telecom, Telecom SudParis, France
Paulo da Fonseca Pinto, Universidade Nova de Lisboa, Portugal
Orhan Dagdeviren, International Computer Institute/Ege University, Turkey
Philip Davies, Bournemouth and Poole College / Bournemouth University, UK
Carlton Davis, École Polytechnique de Montréal, Canada
Claudio de Castro Monteiro, Federal Institute of Education, Science and Technology of Tocantins, Brazil
João Henrique de Souza Pereira, University of São Paulo, Brazil
Javier Del Ser, Tecnalia Research & Innovation, Spain
Behnam Dezfouli, Universiti Teknologi Malaysia (UTM), Malaysia
Daniela Dragomirescu, LAAS-CNRS, University of Toulouse, France
Jean-Michel Dricot, Université Libre de Bruxelles, Belgium
Wan Du, Nanyang Technological University (NTU), Singapore
Matthias Ehmann, Universität Bayreuth, Germany
Wael M El-Medany, University Of Bahrain, Bahrain
Imad H. Elhajj, American University of Beirut, Lebanon
Gledson Elias, Federal University of Paraíba, Brazil
Rainer Falk, Siemens AG - Corporate Technology, Germany
Károly Farkas, Budapest University of Technology and Economics, Hungary
Huei-Wen Ferng, National Taiwan University of Science and Technology - Taipei, Taiwan
Gianluigi Ferrari, University of Parma, Italy
Mário F. S. Ferreira, University of Aveiro, Portugal
Bruno Filipe Marques, Polytechnic Institute of Viseu, Portugal
Ulrich Flegel, HFT Stuttgart, Germany
Juan J. Flores, Universidad Michoacana, Mexico
Ingo Friese, Deutsche Telekom AG - Berlin, Germany
Sebastian Fudickar, University of Potsdam, Germany
Stefania Galizia, Innova S.p.A., Italy
Ivan Ganchev, University of Limerick, Ireland / University of Plovdiv “Paisii Hilendarski”, Bulgaria
Miguel Garcia, Universitat Politecnica de Valencia, Spain
Emiliano Garcia-Palacios, Queens University Belfast, UK
Marc Gilg, University of Haute-Alsace, France

Debasis Giri, Haldia Institute of Technology, India
Markus Goldstein, Kyushu University, Japan
Luis Gomes, Universidade Nova Lisboa, Portugal
Anahita Gouya, Solution Architect, France
Mohamed Graiet, Institut Supérieur d'Informatique et de Mathématique de Monastir, Tunisie
Christos Grecos, University of West of Scotland, UK
Vic Grout, Glyndwr University, UK
Yi Gu, Middle Tennessee State University, USA
Angela Guercio, Kent State University, USA
Xiang Gui, Massey University, New Zealand
Mina S. Guirguis, Texas State University - San Marcos, USA
Tibor Gyires, School of Information Technology, Illinois State University, USA
Keijo Haataja, University of Eastern Finland, Finland
Gerhard Hancke, Royal Holloway / University of London, UK
R. Hariprakash, Arulmigu Meenakshi Amman College of Engineering, Chennai, India
Eva Hladká, CESNET & Masaryk University, Czech Republic
Hans-Joachim Hof, Munich University of Applied Sciences, Germany
Razib Iqbal, Amdocs, Canada
Abhaya Induruwa, Canterbury Christ Church University, UK
Muhammad Ismail, University of Waterloo, Canada
Vasanth Iyer, Florida International University, Miami, USA
Imad Jawhar, United Arab Emirates University, UAE
Aravind Kailas, University of North Carolina at Charlotte, USA
Mohamed Abd rabou Ahmed Kalil, Ilmenau University of Technology, Germany
Kyoung-Don Kang, State University of New York at Binghamton, USA
Sarfraz Khokhar, Cisco Systems Inc., USA
Vitaly Klyuev, University of Aizu, Japan
Jarkko Kneckt, Nokia Research Center, Finland
Dan Komosny, Brno University of Technology, Czech Republic
Ilker Korkmaz, Izmir University of Economics, Turkey
Tomas Koutny, University of West Bohemia, Czech Republic
Evangelos Kranakis, Carleton University - Ottawa, Canada
Lars Krueger, T-Systems International GmbH, Germany
Kae Hsiang Kwong, MIMOS Berhad, Malaysia
KP Lam, University of Keele, UK
Birger Lantow, University of Rostock, Germany
Hadi Larijani, Glasgow Caledonian Univ., UK
Annett Laube-Rosenpflanzer, Bern University of Applied Sciences, Switzerland
Gyu Myoung Lee, Institut Telecom, Telecom SudParis, France
Shiguo Lian, Orange Labs Beijing, China
Chiu-Kuo Liang, Chung Hua University, Hsinchu, Taiwan
Wei-Ming Lin, University of Texas at San Antonio, USA
David Lizcano, Universidad a Distancia de Madrid, Spain
Chengnian Long, Shanghai Jiao Tong University, China
Jonathan Loo, Middlesex University, UK
Pascal Lorenz, University of Haute Alsace, France

Albert A. Lysko, Council for Scientific and Industrial Research (CSIR), South Africa
Pavel Mach, Czech Technical University in Prague, Czech Republic
Elsa María Macías López, University of Las Palmas de Gran Canaria, Spain
Damien Magoni, University of Bordeaux, France
Ahmed Mahdy, Texas A&M University-Corpus Christi, USA
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France
Gianfranco Manes, University of Florence, Italy
Sathiamoorthy Manoharan, University of Auckland, New Zealand
Moshe Timothy Masonta, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa
Hamid Menouar, QU Wireless Innovations Center - Doha, Qatar
Guowang Miao, KTH, The Royal Institute of Technology, Sweden
Mohssen Mohammed, University of Cape Town, South Africa
Miklos Molnar, University Montpellier 2, France
Lorenzo Mossucca, Istituto Superiore Mario Boella, Italy
Jogesh K. Muppala, The Hong Kong University of Science and Technology, Hong Kong
Katsuhiro Naito, Mie University, Japan
Deok Hee Nam, Wilberforce University, USA
Sarmistha Neogy, Jadavpur University- Kolkata, India
Rui Neto Marinheiro, Instituto Universitário de Lisboa (ISCTE-IUL), Instituto de Telecomunicações, Portugal
David Newell, Bournemouth University - Bournemouth, UK
Ngoc Tu Nguyen, Missouri University of Science and Technology - Rolla, USA
Armando Nolasco Pinto, Universidade de Aveiro / Instituto de Telecomunicações, Portugal
Jason R.C. Nurse, University of Oxford, UK
Kazuya Odagiri, Sugiyama Jyogakuen University, Japan
Máirtín O'Droma, University of Limerick, Ireland
Jose Oscar Fajardo, University of the Basque Country, Spain
Constantin Paleologu, University Politehnica of Bucharest, Romania
Eleni Patouni, National & Kapodistrian University of Athens, Greece
Harry Perros, NC State University, USA
Miodrag Potkonjak, University of California - Los Angeles, USA
Yusnita Rahayu, Universiti Malaysia Pahang (UMP), Malaysia
Yenumula B. Reddy, Grambling State University, USA
Oliviero Riganelli, University of Milano Bicocca, Italy
Antonio Ruiz Martinez, University of Murcia, Spain
George S. Oreku, TIRDO / North West University, Tanzania/ South Africa
Sattar B. Sadkhan, Chairman of IEEE IRAQ Section, Iraq
Husnain Saeed, National University of Sciences & Technology (NUST), Pakistan
Addisson Salazar, Universidad Politecnica de Valencia, Spain
Sébastien Salva, University of Auvergne, France
Ioakeim Samaras, Aristotle University of Thessaloniki, Greece
Luz A. Sánchez-Gálvez, Benemérita Universidad Autónoma de Puebla, México
Teerapat Sanguankotchakorn, Asian Institute of Technology, Thailand
José Santa, University Centre of Defence at the Spanish Air Force Academy, Spain
Rajarshi Sanyal, Belgacom International Carrier Services, Belgium
Mohamad Sayed Hassan, Orange Labs, France
Thomas C. Schmidt, HAW Hamburg, Germany

Véronique Sebastien, University of Reunion Island, France
Jean-Pierre Seifert, Technische Universität Berlin & Telekom Innovation Laboratories, Germany
Dimitrios Serpanos, Univ. of Patras and ISI/RC ATHENA, Greece
Roman Y. Shtykh, Rakuten, Inc., Japan
Salman Ijaz Institute of Systems and Robotics, University of Algarve, Portugal
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal
Florian Skopik, AIT Austrian Institute of Technology, Austria
Karel Slavicek, Masaryk University, Czech Republic
Vahid Solouk, Urmia University of Technology, Iran
Peter Soreanu, ORT Braude College, Israel
Pedro Sousa, University of Minho, Portugal
Cristian Stanciu, University Politehnica of Bucharest, Romania
Vladimir Stantchev, SRH University Berlin, Germany
Radu Stoleru, Texas A&M University - College Station, USA
Lars Strand, Nofas, Norway
Stefan Strauβ, Austrian Academy of Sciences, Austria
Álvaro Suárez Sarmiento, University of Las Palmas de Gran Canaria, Spain
Masashi Sugano, School of Knowledge and Information Systems, Osaka Prefecture University, Japan
Young-Joo Suh, POSTECH (Pohang University of Science and Technology), Korea
Junzhao Sun, University of Oulu, Finland
David R. Surma, Indiana University South Bend, USA
Yongning Tang, School of Information Technology, Illinois State University, USA
Yoshiaki Taniguchi, Kindai University, Japan
Anel Tanovic, BH Telecom d.d. Sarajevo, Bosnia and Herzegovina
Rui Teng, Advanced Telecommunications Research Institute International, Japan
Olivier Terzo, Istituto Superiore Mario Boella - Torino, Italy
Tzu-Chieh Tsai, National Chengchi University, Taiwan
Samyr Vale, Federal University of Maranhão - UFMA, Brazil
Dario Vieira, EFREI, France
Lukas Vojtech, Czech Technical University in Prague, Czech Republic
Michael von Riegen, University of Hamburg, Germany
You-Chiun Wang, National Sun Yat-Sen University, Taiwan
Gary R. Weckman, Ohio University, USA
Chih-Yu Wen, National Chung Hsing University, Taichung, Taiwan
Michelle Wetterwald, HeNetBot, France
Feng Xia, Dalian University of Technology, China
Kaiping Xue, USTC - Hefei, China
Mark Yampolskiy, Vanderbilt University, USA
Dongfang Yang, National Research Council, Canada
Qimin Yang, Harvey Mudd College, USA
Beytullah Yildiz, TOBB Economics and Technology University, Turkey
Anastasiya Yurchyshyna, University of Geneva, Switzerland
Sergey Y. Yurish, IFSA, Spain
Jelena Zdravkovic, Stockholm University, Sweden
Yuanyuan Zeng, Wuhan University, China
Weiliang Zhao, Macquarie University, Australia

Wenbing Zhao, Cleveland State University, USA
Zibin Zheng, The Chinese University of Hong Kong, China
Yongxin Zhu, Shanghai Jiao Tong University, China
Zuqing Zhu, University of Science and Technology of China, China
Martin Zimmermann, University of Applied Sciences Offenburg, Germany

International Journal on Advances in Networks and Services
Volume 16, Numbers 1 & 2, 2023

CONTENTS

Analysis of Politicians' Tweets to Explore Political Communication with Social Network Analysis

Heidi Schuhbauer, Sebastian Schötteler, Johannes Niu, Bernhard Schiffer,
David Wolfarth, Tim Groß, Thomas Kreidl, and Fabian Model
Computer Science Department
Nuremberg Institute of Technology
Nuremberg, Germany
heidi.schuhbauer@th-nuernberg.de; sebastian.schoetteler@th-nuernberg.de; niujo64943@th-nuernberg.de; schifferbe69185@th-nuernberg.de; wolfarthda82341@th-nuernberg.de; grossti74853@th-nuernberg.de; kreidlth74662@th-nuernberg.de; modelfa74656@th-nuernberg.de

Abstract—**This paper illustrates the practical application of cluster analysis, social network analysis, sentiment analysis, and topic analysis in a case study on Twitter data. These techniques provide insights into the public communication patterns between German Members of Parliament (MPs) on Twitter around the time of the 2021 federal election. The question of this work was to determine whether a potential shift in communication towards the inaugurated "Ampel" coalition, made up of the parties SPD, Greens, and FDP, can be derived from Twitter interactions. Twitter data were collected and separated into two time slots: before and after the election. In distinct scenarios, mention, retweet, and reply interactions are first considered together and then separately. In these scenarios, the Girvan-Newman Algorithm detects clusters of MPs dependent on the interactions observed. Then, the average inbreeding homophily and other network metrics of the pre- and post-election area are compared. An additional scenario focuses on intra- and inter-party sentiments conveyed within tweet texts. In a fourth scenario, MPs are grouped according to their party affiliation, the average inbreeding homophily values of parties, and potential coalitions. A topic analysis handled relevant discussion topics between the successful coalition parties. Changes in communication behavior at these two different time slots are visible. The communication clusters of those MPs differ mostly before and after the election. The average sentiment of the parties towards each other changed positively, although no significant tendency could be derived regarding later coalition formations. A determination of the density of the network for the topics also indicates that the importance of the individual topics before and after the formation of the government has a consistent relevance.**

Keywords-Cluster Analysis; Microblog; Network Metrics; Sentiment Analysis; Social Network Analysis; Topic Analysis.

I. INTRODUCTION

This paper is an extended version of the one presented in SOTICS 2022 [1].

For political communication between parties, politicians, and their constituents, social media platforms play an important role. By communicating through platforms, such as Facebook, Twitter, and Instagram, political actors reach wide audiences within a short period. On these platforms, politicians publicly communicate with each other. As one consequence of this trend, social media platforms are gradually outclassing traditional media outlets – such as television and print media – as the main source of news and information. Case in point, a survey conducted in 2018 determined that social media platforms are the most popular information and news sources for American adults aged between 18 and 29 [2]. Corroborating, another survey conducted in 2021 determined that up to 76% of adults rely on social media as a news and information source, depending on the country [3].

Among those services, Twitter is a very popular microblog that is used by many persons also for professional use. Twitter promotes the dialogue between politicians and between politicians and their constituents via mention and reply interactions, which allow users to engage in direct communication. Consequently, social media have become a central component of political communication.

Relations between individual MPs can be examined in more detail using social network analysis. "A social network consists of a finite set or sets of actors (depicted as nodes) and the relation or relations defined on them (depicted as ties)" [4]. Translated to microblogs, two types of social networks can be derived from such platforms – follower and interaction networks. A follower network can be defined as a set of microblog users connected via follower ties regarding who follows whom. An interaction network can be defined as a set of microblog users connected via interaction ties that diffuse content between said users. The application of social network analysis on Twitter data offers a variety of research possibilities. Interactions can be derived from public tweets referring to other people, i.e., retweets of or replies to another user's tweet, or mentions of a user. Analysis of interaction networks explores these relations, as well as their textual contents, which can be examined through sentiment analysis.

This article applies methods of social network analysis to explore changes to the communication of German MPs from selected political parties around the 2021 federal election. After 16 years of a government led by the Union, made up of the Christian Democratic Union (CDU) and Christian Social Union (CSU), – in coalition with the Social Democratic Party of Germany (SPD) for three and the Free Democratic Party (FDP) for one legislative period – Germany had a change in government after the federal elections in 2021. The new government consisting of the SPD, Alliance 90/The Greens (Greens), and FDP was formed on December 8, 2021. For the first time in its history, a government formed by a coalition of more than two parliamentary groups thus governs the Federal Republic of Germany.

In the past, federal government alliances consisted of coalitions between the Union/SPD, Union/FDP, SPD/FDP, and SPD/Greens. A coalition between the Greens and FDP within a tripartite coalition currently governs in the states of Schleswig-Holstein and Rhineland-Palatinate, in a so-called Jamaica coalition (together with the CDU) and an Ampel ('traffic light') coalition with the SPD, respectively. On a federal level, however, these two parties had yet to form a coalition.

This article applies social network analysis, cluster analysis, sentiment analysis, and topic analysis to explore changes to the communication of German MPs from select political parties around the 2021 federal election.

Section II of this paper presents related works, formulates the research gap, and specifies the hypotheses. Section III introduces the methodology used to aggregate and analyze the data for the network scenarios. Section IV presents the results of each network scenario and discusses them. Section V describes the methods for the topic analysis of relevant subjects discussed by the "Ampel" coalition parties who won the election. Section VI contains the result of the topic analysis. Section VII illustrates the limitations of this research, as well as starting points for possible future work.

II. Research Gap

Virk [5] compares different Social Network Services (SNS) as a type of social media and explores the special role of Twitter in public communication. The author examines the communication patterns between Twitter users and applies the tie strength theory postulated by Granovetter [6] to conclude that interactions on Twitter – unlike other SNS – focus on content rather than user relationships, and thus can reach wider audiences.

Lassen and Brown [7] examine Twitter use by members of congress in the United States of America. They state that SNSs enable politicians to communicate more directly and personally with peers and supporters by eliminating limits on message visibility, allowing content to be redistributed beyond one's followers. The application of social network analysis to political networks shows the fragmentation and clustering of politicians, parties, or political systems.

Boireau [8] identifies communities among Belgian MPs along party and linguistic lines. For this purpose, the Girvan-Newman Algorithm (GNA) was applied on a network generated from the MPs' connections to followers, and retweet interactions to find hidden communities and homogeneous clusters by calculating their homophily indices, which express the degree of similarity of members within a cluster.

Caetano et al. [9] analyze social networks among Twitter users during the 2016 American presidential election by analyzing tweets about the candidates. Users were clustered based on their sentiment towards a candidate with their mentioning behavior and hashtag use. By obtaining homophily indices of these clusters, the authors could identify users with high degrees of relative similarity.

Sentiment analysis attempts to quantify attitudes conveyed in a text. Giachanou and Crestani [10] discuss common procedures for sentiment analysis, as well as their respective limitations, e.g., the detection of irony or emotions. The work explicitly focuses on methods suitable to retrieve sentiments from tweets.

Boras and Singh [11] show that Twitter is the social network most used for political discussions. However, the intensity of the discussion is strongly dependent on the topic. In Meier et al. [12], among other things, the issue of engagement of German politicians on certain topics was examined in the election year 2017. To determine which topics were intensively discussed on Twitter, a hashtag network was created, with which each tweet with a hashtag could then be directly assigned to a specific issue. Using additional metrics, such as the Z-score or the Q-modularity, it was found that the communication behavior of the parties changed significantly before the election, especially in terms of interaction and discussion with other parties. A topic analysis of political discussion topics on Twitter based on hashtags was also done by Boras and Singh [11] for the Indian political landscape. To examine whether discussion on Twitter about the identified topics leads to political polarisation, a mention and retweet network was formed. These networks can indicate discussion and advocacy. Content and sentiment analysis can be used to identify shifts in opinion on specific topics as well as internal network relationships. Garcia-Sanchez et al. [13] used cluster analysis to identify different ideologies concerning various political issues of Brazilian parliamentarians in 2019. Using the degree of centrality, it was possible to identify particularly active users and thus the central key figures. A comparison between possible techniques for topic modeling can be found in Egger et al. [14]. The techniques investigated here were latent Dirichlet allocation (LDA), non-negative matrix factorization (NMF), Top2Vec, and BERTopic. Tweets are used as the data basis of the comparison, and the advantages and disadvantages in the different application areas are discussed.

Until now, literature does not describe possible changes in Twitter communication behavior between MPs before and after an election. An exploration of the change in tone by analyzing the sentiment of tweets before and after an event has also not yet been described. An analysis of the topics discussed by the parties is missing. Interesting aspects of political communication behavior on social media are expected results of this analysis.

Consequently, this article examines how Twitter interactions (mentions, retweets, replies) between MPs of possible coalition partners (CDU, CSU, SPD, Greens, FDP) changed before and after the 2021 German federal election. It furthermore explores potential differences in intra- and inter-party communication and attempts to show whether the political shift towards the inaugurated "Ampel" coalition could be derived from the observed changes. Changes in intra- and inter-party communication are additionally presented in relation to the various topics.

The following hypotheses form the basis for the communication behavior analysis: The article hypothesizes that different interactions between MPs can be observed during the pre- and post-election period (H1) and that the resulting interaction networks for each period show a difference in intra- and inter-party communication (H2). The article further assumes that "Ampel" MPs' mutual sentiment changed positively (H3). By analyzing the sentiment between parties, as well as the average homogeneity within parties and party groups, political tendencies towards an "Ampel" coalition can be observed (H4). In the topic networks, intra-party homophily changes to inter-party homophily as a function of the observation time points (H5). For the change in communication, it is also important to consider the density of the social network. The communication intensity of the topic networks depends on the observation times (H6).

Thus, this article attempts to describe the change in communication between MPs by analyzing their Twitter interactions before and after the federal election 2021. It aims to understand whether changing interaction intensities between MPs of potential coalition partners yield conclusions about the emerging "Ampel" coalition. This would be of relevance for future research into the interdependencies of political communication on Social Network Services, such as Twitter.

III. Methods for Scenarios

Mention, retweet, and reply interactions between MPs from the SPD, Greens, FDP, CDU, and CSU were collected to explore changes in communication on Twitter. One MP using another MP's handle denotes a mention interaction. A Twitter handle, which is commonly known as a username**,** is the name with which a user has registered on Twitter. Since it serves as an account's identifier, no two usernames on the social network are the same [16]. Retweets refer to the redistribution of another user's tweet and can contain commentary by the retweeter. A reply is defined as a comment posted under another MP's tweet. The resulting social networks of MPs connected by their interactions are analyzed in four separate scenarios.

A. Network Scenarios

Scenario 1 considers all interaction types, while in scenario 2, a) mention, b) retweet, and c) reply interactions were examined separately. For each scenario, MPs were grouped using automated cluster detection and examined for modularity and homophily. Modularity measures the strength of division of a network into clusters. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules [17]. Homophily in network structures means the principle that nodes in a network tend to have links to other nodes with similar attributes [14].

In scenarios 3 and 4, MPs were grouped based on their party affiliation. In scenario 3, interactions were examined for the tweet author's sentiment towards the addressed MP using sentiment analysis. The sentiment for every interaction was evaluated based on the tweet's text. To determine changes to the inter-party relations, each party's average sentiment toward all other parties was then calculated and compared between the pre- and post-election networks. Scenario 4 examined the average homophily within each party and party group. Party groups were based on politically and numerically possible coalition compositions ("Ampel", "Jamaica") and for the Union parties.

B. Data Aggregation

Publicly available Twitter data can be divided into three categories: (1) User information, such as the username, the Twitter handle (identified by @), or account description; (2) following and liking behavior of a user, and the user's followers; (3) the user's tweet timeline, in which all self-published or retweeted tweets appear, as well as the user's replies to others' tweets.

As a basis for this study, publicly available tweets from MPs of the 19th (2017-2021) and 20th (2021-2025) legislative sessions were collected for the period from July 26, 2021, 0:00 a.m. to November 26, 2021, 12:00 p.m. The end date was chosen to serve as cut-off due to the official presentation of the coalition agreement between the SPD, Greens, and FDP on November 24, 2021. To collect reactions to this announcement, two more days were added. The period between the closing of polls on September 26, 2021, at 6 p.m., and the end date covers 60 days and is considered as the post-election period. An equally long time before the closing of polls was considered for the pre-election period.

Twitter accounts were selected from all MPs with a public Twitter timeline who are members of the parties SPD, CDU, CSU, Greens, and FDP. Members of the parties "The Left" and AfD were not included in this analysis, as neither party was relevant for coalition negotiations after the election. The timelines of all selected accounts were then scraped from Twitter's website.

Data Collection. Scraping of timelines was done using the Python package Scweet [15]. Scweet uses the Chrome plugin Selenium [19] to access the desired Twitter page, extract the information of the tweet from the page, and save it to a CSV file.

Data Processing. A custom Java application was developed to generate uniformly formatted and sanitized datasets. The data originally scraped from Twitter included the timelines of all MPs, i.e., all their tweets, retweets of, and replies to other tweets within the time frame. The information generated for each of these messages included the time of publication, the author's username and handle, the textual contents of the tweet, as well as information on whether it was posted as a retweet of, or reply to another tweet. If other users were mentioned within the tweet, they could be identified through their handles.

Additionally, the application enriches the data with information on party affiliation and membership of the 19th or 20th legislative period. It produced output data in the GEXF format [20], which is limited by specified procedures. First, all tweets that did not represent a connection between two MPs were removed. The dataset was then divided into a pre- and a post-election partition. For this purpose, all tweets that were created before the time of the closing of polls on September 26, 2021, 6:00 p.m. were assigned to a first partition. The elements from the timeline after this date were assigned to a second partition. Additionally, the output is restricted to specific interaction types. This allowed the creation of one pre-election and one post-election dataset for each of the scenarios defined.

Data Description. The data set collected from Twitter consisted of 26,888 German-language tweets from 736 Twitter accounts. 15,770 of these tweets were posted before and 11,118 after election day. 1,030 MPs were elected for the 19th and 20th legislative periods. 71.5% of them maintained a Twitter account. Once filtered, the dataset consisted of 622 accounts and 9,582 tweets. After removing all tweets that did not connect two MPs, 5,766 tweets from 466 MPs remained in the pre-election dataset, and 3,816 from 476 MPs in the post-election dataset. Figure 1 shows the percentage distribution of all tweets among the parties before and after the election.

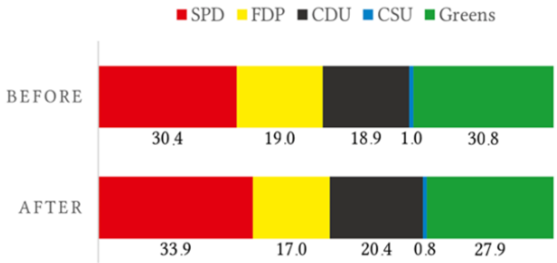


Figure 1. Percentage distribution of MPs' tweets by party

The pre- and post-election data contain nodes and edges depending on the interaction types selected during the data processing step. Scenarios 1 and 4 thus contained all MPs, while scenario 2 contained three separate data sets, differentiated by interaction types. Scenario 3 handled only those interaction types whose tweet text fields were not empty. The aggregated data and source code can be accessed at [14].

C. Cluster Detection

Cluster detection extracts groups of individuals from a network based on the similarity of one or more attributes. This work used connectivity-based clustering, which identifies clusters based on the connections between nodes in the network, as well as the weights of connections. For this purpose, the Girvan-Newman Algorithm [22] was used. This algorithm assumes that members of a cluster have more connections to other members of the same cluster, and fewer connections to other nodes in the remaining network. By iteratively removing connections whose Edge Betweenness Centrality (EBC) is the highest, clusters are separated from each other. The EBC is defined as "the number of [the] shortest paths between pairs of vertices that run along it" [10]. In each step, the edge with the highest EBC is removed from the network and its modularity is calculated. The modularity of a network denotes how well clusters are separated from each other. The iteration continues until every connection between nodes has been eliminated. The intermediate step with the highest modularity is the result of the algorithm.

To guarantee that an MP's allocation to a cluster is based on their interactions and not their party affiliation, a χ^2 test is performed on the network. The test's p-value denotes the probability p of MPs' party affiliation determining the results of the cluster detection.

D. Sentiment Analysis

The textual contexts of MPs' tweets were examined to analyze the sentiment for which the Python package TextBlob [15] was used. The package uses a lexicon-based approach to compute the sentiment. For the analysis of German language texts, the plugin TextBlobDE [24] was used. A predefined dictionary of words associated with positive or negative emotions is used to weigh a text's sentiment. An individual score is assigned to each word in the examined text. The overall sentiment is defined by the average sentiment across all words in the text. The algorithm generates a polarity score from –1.0 to +1.0 for each tweet, which classified the tweet as either positive, neutral, or negative. Each tweet in the data set is then enriched with the polarity value, as well as the polarity class as additional attributes.

E. Homophily

The homophily index H measures a cluster's relative homogeneity. To determine H for a cluster i, the connections of all nodes of the cluster are examined. Caetano et al. [9] calculate $H_i = \frac{s_i}{si+d_i}$ where s_i denotes homogeneous links, i.e., those that connect a node of class i to other nodes of the same class, while d_i denotes heterogeneous connections, i.e., those that connect a node of class i to nodes of another class. By normalizing H_i over the whole network, H can be compared across different clusters. This inbreeding homophily index IH is determined by $IH_i = \frac{H_i - w_i}{1-w_i}$, where w_i denotes the relation of nodes between cluster i and the total number of nodes in the network. Clusters whose IH_i is greater than 0 are considered homogeneous. The average of IH across all clusters in a network is used to compare the clusters detected in the pre- and post-election networks.

F. Evaluation

The procedure resulted in a set of network pairs, each consisting of a pre- and a post-election network. The two networks created for scenario 1 contained all MPs that have interacted via mentions, retweets, or replies within the respective timeframe. The number of connections between two nodes weighted the edges.

Scenario 2 generated one network pair for each of the three interaction types. Thus, one pre- and one post-election network each were generated, which included all those MPs that a) mentioned each other, b) replied to one another, and c) retweeted each other. Edges represent the connections. They are weighted by the interaction count. These scenarios were examined separately. For each network, automated cluster detection was applied. The H and IH indices were calculated to determine the homogeneity of each cluster. Additionally, the number of nodes and edges in the network, the number of clusters identified by the GNA, as well as their networks' average homophily and inbreeding homophily indices, and the maximum modularity were determined. Statistical significance was ensured using the χ^2 test. The results of these analyses were then compared for the pre- and post-election network pair. To illustrate the results of the automated cluster detection, each pre- and post-network pair is visualized as a cluster graph.

In scenario 3, each party's average sentiment towards all other parties was examined. For this purpose, MPs were clustered according to their party affiliations.

Scenario 4 looked at the inbreeding homophily of each party, as well as the coalition options before and after the election. The *IH* values for the coalitions were also checked for statistical significance using the χ^2 test and its p-value.

IV. Results of The Scenarios

A. Scenario 1: Multiple Interactions

In scenario 1, automated cluster detection included all interaction types. An overview of the collected metrics can be found in Table I.

Table I. Network and cluster metrics considering all interactions

Metric	Value (pre)	Value (post)	Difference
Number of nodes	466	476	10
Number of edges	5766	3816	-1950
Number of clusters	256	188	-68
Maximum modularity	0.026	0.356	0.330
Average *IH*	0.0212	0.0571	0.0359
p-value from χ^2-Test	< 0.001	< 0.001	

The number of MPs (nodes) tweeting after the election did not vary much from that before the election. However, the number of connections (edges) was reduced by 33%, which suggests that tweeting activity was distributed more equally among MPs after the election. The GNA identified 256 clusters of the pre-election network with 466 MPs, and very low modularity, homophily, and inbreeding homophily indices. After the election, 476 MPs could be assigned to 188 clusters. The maximum cluster size was reduced by 54.5% to 97. The modularity increased by 1369%, from 0.026 to 0.356, and homophily and inbreeding homophily also increased significantly. Figure 2 shows a visualization of these clusters. Node colors represent each MP's party affiliation. The size of a node depicts the sum of all incoming and outgoing edges, i.e., the node's degree. Edges were omitted from these figures for improved visibility.

Pre-election, the visualization shows a distinctive, large cluster that unites MPs across all parties. Outside of this cluster, many MPs are scattered into tiny groups or unassigned to any notable cluster. Post-election, four large clusters separated along party affiliation can be identified. A heterogeneous group of MPs was not assigned to any notable cluster.

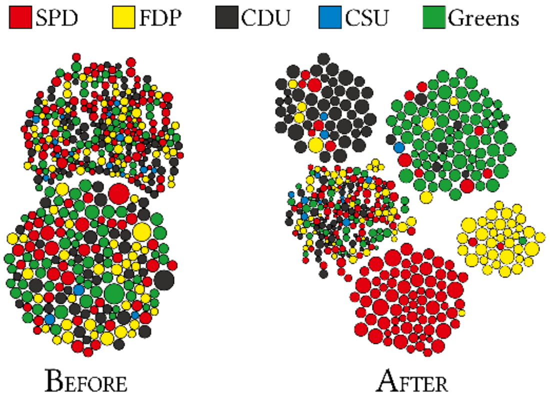


Figure 2. Clusters found by GNA before and after the election considering all interactions

The pre-election results of scenario 1 show that MPs were likely allocated to the dominant cluster based on their general activity on Twitter. Nodes with higher degrees were allocated to the dominant cluster. Post-election, distinct clusters are clearly separable, which consist mainly of MPs of either the SPD, CDU, Greens, or FDP. The number of nodes that could not be allocated to any major cluster decreased. This indicates that post-election, MPs predominantly communicated within their parties, while they communicated much more openly before the election. The overall count of interactions decreased significantly.

B. Scenario 2: Single Interactions

When interaction types are considered separately, these findings can be analyzed in more detail.

Mentions. In this particular scenario, clusters were determined based on mentions only. Table II shows the collected metrics.

Table II. Metrics of network and clusters derived from mentions

Metric	Value (pre)	Value (post)	Difference
Number of nodes	433	428	-5
Number of edges	3247	1758	-1489
Number of clusters	95	38	-57
Maximum modularity	0.237	0.441	0.204
Average *IH*	0.1158	0.4550	0.3292
p-value from χ^2-Test	< 0.001	< 0.001	

Almost as many (433 vs 428) MPs mentioned one another in the pre- and post-election period. Interactions decreased by 54%, and the number of detected clusters decreased by 40%. After the election, 38 clusters with a modularity of 0.441 could be identified, compared to 95 clusters with a modularity of 0.237 before the election. The average IH across all clusters in both networks increased by more than 300%. Figure 3 visualizes the detected clusters.

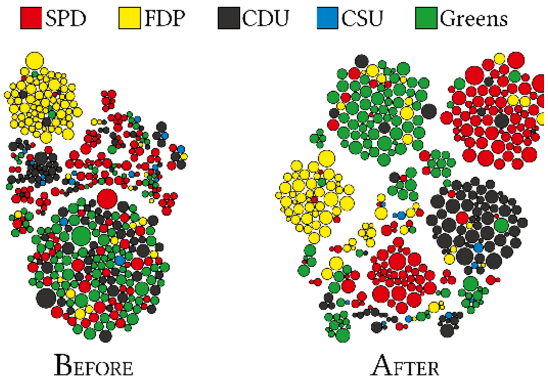


Figure 3. Clusters found by GNA before and after the election considering only mentions

Pre-election, three distinct clusters can be identified, one portraying a large cluster mainly dominated by Greens but including MPs across all parties, one dominated by FDP MPs, and a smaller one dominated by CDU MPs. The large, heterogeneous cluster dominated by Green MPs could be caused by many mentions of the Greens' chancellor candidate, Annalena Baerbock.

Distinct clusters are detected in the post-election network separated along party lines. Two SPD clusters are found, as well as several smaller but still homogeneous clusters. The number of mentions increased. Subsequent analysis revealed that the distinct party clusters might be caused by MPs congratulating their party peers.

In the next analysis, only retweet connections were examined. The metrics on this can be found in Table III.

TABLE III. METRICS OF NETWORK AND CLUSTERS DERIVED FROM *RETWEETS*

Metric	Value (pre)	Value (post)	Difference
Number of nodes	178	144	-34
Number of edges	253	151	-102
Number of clusters	19	26	7
Maximum modularity	0.713	0.825	0.112
Average *IH*	0.8809	0.9567	0.0758
p-value from χ^2-Test	< 0.001	< 0.001	

The number of nodes was reduced by 19.1% from 178 to 144, and the number of edges by 40% (253 to 151). Nineteen clusters could be identified before the election, and twenty-six after, which explains the reduction of the median cluster size from 6 to 2.5. Conversely, the modularity increased slightly, just like the average IH. Figure 4 shows a visual representation of both networks' clusters.

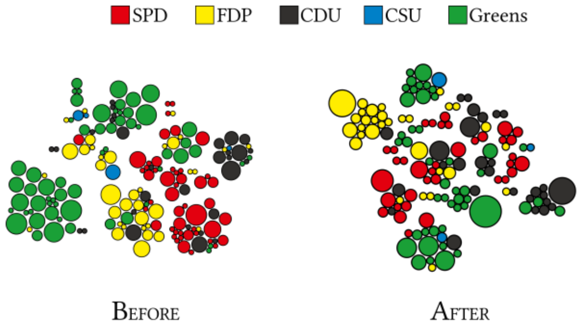


Figure 4. Clusters found by GNA before and after the election considering only retweets

Clustering detected several well-separated clusters with relatively high homogeneity both before and after the election, with smaller clusters found in the latter. A possible explanation is that MPs attempted to promote the tweets of party peers. The clusters in the post-election network were smaller. Retweets play a smaller role in communication among MPs.

The last analysis of this section considered reply connections only. The metrics can be found in Table IV.

TABLE IV. METRICS OF NETWORK AND CLUSTERS DERIVED FROM *REPLIES*

Metric	Value (pre)	Value (post)	Difference
Number of nodes	351	388	37
Number of edges	2266	1907	-359
Number of clusters	233	24	-209
Maximum modularity	0.113	0.401	0.288
Average *IH*	0.0636	0.7112	0.6476
p-value from χ^2-Test	< 0.05	< 0.001	

More nodes could be identified in the post-election network, but fewer connections between them. The algorithm identified 233 clusters before and 24 after the election, a reduction of 89.7%. Modularity and IH increased significantly. A visualization of identified clusters can be found in Figure 5.

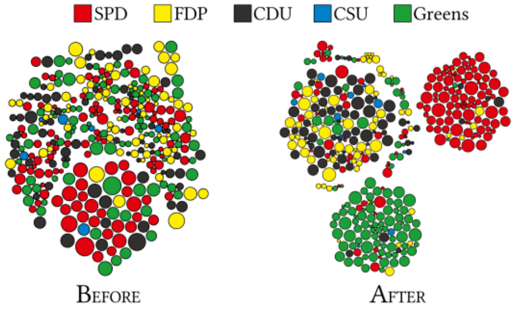


Figure 5. Clusters found by GNA before and after the election considering only replies

Pre-election, one large and many small clusters were found, with the main cluster containing many nodes with a high in– and out-degree.

Post-election, two main clusters were identified, notably consisting mainly of SPD and Green party members respectively. Another large cluster contained a heterogeneous mix of MPs.

Solely considering reply interactions, one large and many small clusters were found in the pre-election network. The main cluster contains many nodes with a high in– and out-degree. In the post-election network, more nodes are identified but fewer connections between them are found. Two main clusters were identified, notably consisting mainly of SPD and Green party members. One cluster of CDU and FDP MPs indicates active conversations between these two parties, potentially on the FDP's willingness to enter coalition negotiations with the SPD and Greens shortly after the election, which supports hypothesis H1.

C. *Scenario 3: Sentiment Analysis*

Each interaction's textual content was analyzed to retrieve the parties' mutual sentiments. The average sentiment of interactions from MPs of one party towards MPs of the other parties was calculated. The results are shown in Table V. Notably, polarity does not score very highly overall, except for the sentiment from MPs of the CSU towards MPs from the CDU. FDP MPs communicated neutrally in general. The SPD scores positively towards the "Ampel" parties. On average, Green party MPs showed positive polarities only towards other MPs of their party.

TABLE V. AVERAGE SENTIMENT BETWEEN PARTIES BEFORE THE ELECTION

Target Source	SPD	FDP	CDU	CSU	Greens
SPD	0.25001	0.21293	0.00002	-0.11499	0.35683
FDP	0.05095	-0.01008	0.06981	0.09734	0.01032
CDU	0.00070	0.02997	0.13179	-0.12469	0.00483
CSU	0.04297	-0.01875	0.70728	0.10625	0.11405
Greens	-0.16582	0.03257	-0.16458	0.00053	0.35588

Table VI shows the average sentiment between parties after the election. The post-election sentiments between parties notably tend towards an overall positive sentiment. The SPD received overall positive interactions, especially from the CDU. The SPD communicated relatively neutrally, both internally, as well as towards their subsequent coalition partners. The polarity of the interactions among MPs of the Greens and interactions from MPs of the CSU towards CDU MPs did not change significantly from their pre-election scores. The overall sentiment across all parties after the election was on average more positive than before the election. The FDP especially shows notable increases in positive sentiments towards the SPD and the Greens, considering that the FDP moved towards the "Ampel". This strongly hints at successful coalition negotiations, which ended with the signing of the coalition contract.

TABLE VI. AVERAGE SENTIMENT BETWEEN PARTIES AFTER THE ELECTION

Target Source	SPD	FDP	CDU	CSU	Greens
SPD	0.00166	0.25408	0.31106	0.84063	0.06433
FDP	0.33102	0.54495	-0.11953	0.00391	0.27281
CDU	0.79865	-0.00598	0.09291	-0.08487	0.67012
CSU	-0.16250	0.24688	0.59688	0.12500	0.39146
Greens	0.43225	0.09978	0.62791	-0.06024	0.38109

D. *Scenario 4: Party and Group-Dependent Clustering*

In this scenario, MPs were clustered along party affiliation. Additionally, the two potential government coalitions, "Ampel" (SPD, Greens, FDP) and Jamaica (CDU, CSU, Greens, FDP), as well as the Union (CDU, CSU), were clustered. To compare the homogeneity within each cluster, the average *IH* before and after the election was calculated and compared. Table VII displays the average *IH* values of each party, as well as the coalition and union clusters for the pre- and post-election networks.

TABLE VII. RELATIVE IH IN PARTIES AND PARTY GROUPS

	Before	After	Difference
CDU	0.4749	0.5596	0.0848
CSU	0.0721	0.0516	-0.0205
SPD	0.5272	0.6392	0.1120
Greens	0.5682	0.5729	0.0047
FDP	0.5397	0.4618	-0.0779
"Ampel" Coalition	0.4519	0.6272	0.1754
Jamaica Coalition	0.5209	0.5307	0.0098
Union Group	0.4632	0.5422	0.0791
p-value from χ^2-Test	0.057764	0.106983	

The biggest differences are between the SPD and CDU. Their relative homophily increased. CSU and FDP decreased in *IH*. SPD received the biggest increase in homogeneity. This could be explained by their win of the election, and the positive feedback MPs received from their peers, as well as the election of SPD MPs Olaf Scholz as chancellor and Bärbel Bas as president of the parliament. The biggest positive change among grouped MPs took place in the "Ampel" coalition, but *IH* increased for the Jamaica and Union clusters as well. However, a significant statistical independence of these findings is not reliably provable, as the χ^2-test results in relatively high p-values for the pre- and post-election homophily.

V. METHODS FOR TOPIC ANALYSIS

The change of government was preceded by a volatile election campaign, which was interspersed with multifaceted debates on various topics [25]. The change of working communities, which also include public decision-making bodies, is often accompanied by a change in the working and communication climate because the inner diversity changes [26]. The analysis of these changes helps to understand how the patterns of interaction within and across parties change over this period. In the context of this study, a qualitative analysis of the interactions between the members of the factions of the governing parties is done based on exploratory research. Changes in the social network for certain topics are examined. Changes in intra- and inter-party communication are presented in relation to relevant topics.

A. Data Work

The topic analysis focuses on the communication between the traffic light parties because this was the successful coalition. For this purpose, a new data acquisition had to be done:

The tweets of all members of the traffic light coalition are used as the data basis for further network analyses. In the first step, the Twitter handles of the parliamentarians are fetched from a list of all members of the Bundestag, which is dynamically managed and made available by Twitter [27]. This data scratch is implemented using the Twitter API and the Python library Tweepy [28]. The data of the first period refers to the time from 26 September 2020 00:00 to 25 September 2021 23:59. The comparison period is defined similarly, offset by one year, from 26 September 2021 00:00 to 25 September 2022 23:59. This period is longer than the period of the previous analysis to get more data. The data contains information on the date, the text, the user ID, and the tweet ID. In addition, information on retweets, replies, and mentions are included, which are important for the formation of the networks in the further process of the study. Since party affiliation plays an important role, this information is subsequently added to the individual tweets with the help of master data from the German Parliament [29]. With this information, only tweets from members of the Greens, the SPD, or the FDP are kept to sort out the tweets from the opposition parties. In total, the remaining dataset contains tweets from 118 members of the three parties with approximately 120,000 tweets. No tweets could be found for the remaining 298 members of the government parties.

To carry out a topic analysis, the data is first pre-processed. In this process, only tweets with their conversation ID are retained and tweets with a line length of less than 100 are sorted out. In addition, interfering words and characters such as links or Twitter-specific characters are removed from the tweets. Pre-processing reduces the size of the data to approximately 80,000 tweets, which are analyzed with the help of the topic modeling technique BERTopic [30] and classified into 10 topics. Only 10 topics are determined to prevent fragmentation into several smaller topics. This is necessary so that representative networks of an appropriate size can be formed. Clustering into many topics greatly reduces the number of tweets on a topic.

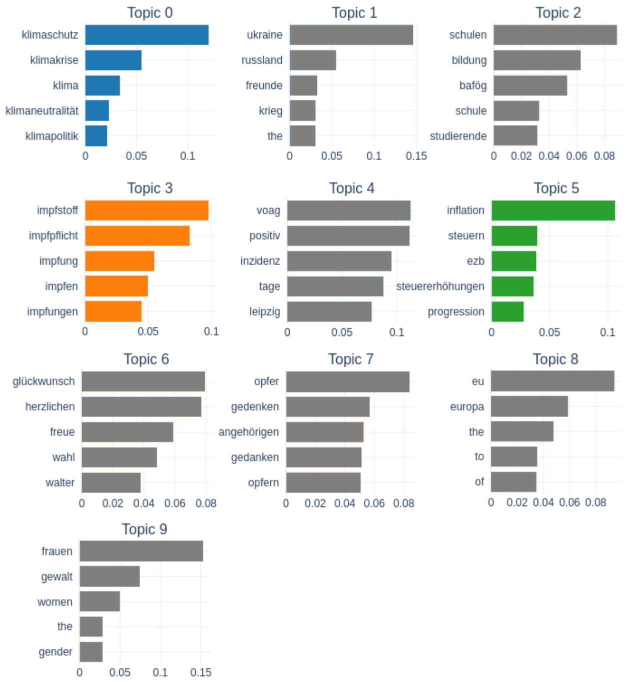


Figure 6. Word Scores per Topic

The Topic Word Scores in modelfa74656@th-nuernberg.de 6 show the distribution of the detected topics. The most frequent five words are listed for each of the 10 topics. The division of the topics results in approximately 65,000 tweets that can be assigned to a topic. This results in approximately 55,000 tweets that are present in the original data set of 120,000 tweets but could not be assigned to a topic. In the next step, connections are detected between the assigned and unassigned tweets to be able to assign these tweets to a topic. With the help of retweets and replies, about 10,000 unassigned tweets can be assigned to a topic. It is assumed that a reply and a retweet mean that the tweet is about the same topic as the original tweet. This procedure results in the topics, which are shown in Table VIII and the respective number of tweets they contain.

TABLE VIII. TOPIC EVALUATION

Topic	Amount	IRR	Precision	Topic Assessment
0	**6500**	**92 %**	**95 %**	**✓ (Climate)**
1	6500	96 %	93 %	✗ (Ukraine-Conflict)
2	3000	75 %	81 %	✗ (Education)
3	**2000**	**98 %**	**98 %**	**✓ (Vaccination)**
4	4500	89 %	20 %	✗ (Covid)
5	**5000**	**92 %**	**100 %**	**✓ (Finance)**
6	30000	88 %	7 %	✗ (Elections)
7	5500	73 %	27 %	✗ (Disasters)
8	5000	80 %	8 %	✗ (Europe)
9	3000	68 %	14 %	✗ (Conflicts)

To select suitable topics, it is necessary to determine their quality in advance. For this purpose, suitable generic terms for the respective topics are sought at the beginning. In cooperation with a political scientist, suitable terms are defined for the keywords in Figure 6.

The evaluation of the assessment is done using the interrater reliability (IRR) value. This involves verifying whether the tweets thematically match the keywords and the generic terms. The procedure is based on Newman et *al.* [31]. The IRR value is used to check whether all observers agree on the assessment. A representative sample of 50 tweets is reviewed for each topic and the IRR value is calculated from this. Only topics whose IRR value is high enough are of sufficient data quality and are therefore taken into account in the subsequent analyses. Values in the range of at least 80 percent are considered high enough.

To determine the accuracy of the assignment of the tweets to the topics, all tweets of a topic are included for which all observers agree that the generic term and the keywords match the respective tweet or not. The IRR for the tweet under investigation must therefore be 100 percent. The values for calculating the accuracy are 0 or 1, depending on whether the topic matches the tweet (1) or not (0). The mean value is determined from these binary values, which corresponds to the accuracy in percent.

The result of the evaluation indicates that 7 from 10 topics have a sufficiently high IRR value. However, the accuracy is only high enough for topics climate, the Ukraine conflict, education, vaccination, and finance. Topic 1 (Ukraine-Conflict) is omitted, as this topic is almost exclusively discussed after election day, and thus no comparison is possible concerning the two observation periods.

Finally, this leads to the topics of climate, vaccination, and finance, which are analyzed because their IRR values and accuracies are high enough.

B. Data Analysis

First, graphs are generated for visualization and analysis with the software Gephi [14a] using the Python package NetworkX [32]. These contain all nodes and edges of a respective topic. A node corresponds to a participant in the communication network of the topic, thus a Twitter account. An edge corresponds to a connection between two participants in the topic network. The connection can take the form of a mention, a retweet, or a reply. During the creation process, self-loops are removed from the graph, as they have no relevance in the study of interactions between politicians. In addition, the attribute "party" with the value FDP, SPD, or Greens is added to each node for the subsequent analyses. All remaining nodes that cannot be assigned to an “Ampel” party are removed. The resulting graphs are visualized in Gephi and are exemplarily shown for Topic 0 (Climate) in Figure 7 before (on the left side) and after (on the right side) the election. The colors of the points visible in the following images represent the party colors of the SPD (red), FDP (yellow), and Greens (green).

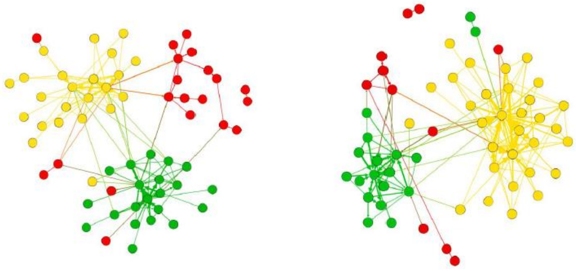


Figure 7. Network Visualisation Topic 0 (Climate)

NetworkX offers a variety of functions for calculating metrics, which are used to determine basic network properties for each topic before and after the election. These include the number of nodes and edges of the respective topics, which provide information about the change in the number of nodes and edges over the observation period and thus the size of the network. Based on the nodes and edges, the connectivity of each topic area is analyzed using the node density metric. This corresponds to the ratio between the number of edges in each network and the theoretical number of edges if the network were fully connected and is expressed in *Stegbauer et al.* [33] using $p = \frac{Actual\ number\ edges}{Expected\ number\ edges}$. The expected number of edges in a loop-free directed network is defined as *N(N-1)*, where *N* is the number of nodes. A node density of one would correspond to a fully interconnected network. The node density is determined across parties at the topic level as well as at the party level for each topic. This allows the analysis of the change in network node density for the entire topic as well as a party-specific breakdown of the change. Node density is a metric that enables the measurement of the connectedness of a network. The density can be used to compare networks with each other and to determine the intensity of social dynamics within a network.

Subsequently, the dyadicity at the party level is determined for each topic network. Dyadicity represents the connectedness between nodes of the same party relative to the standard connectedness of the network. In *Wang et al.* [34] to calculate dyadicity, first, the average connectedness in the network of two nodes is determined with $p = \frac{2M}{N(N-1)}$, where *N* is the number of nodes and *M* is the number of edges. Within a party *i*, the dyadicity is expressed by $D_i = \frac{Actual\ number\ edges}{Expected\ number\ edges}$. The actual number of edges is the number of edges between nodes of the same party. The expected number of edges is determined by $\frac{N_i(N_i-1)}{2}p$, where N_i represents the nodes of a party. A network is dyadic if the nodes belonging to the same group are more connected to each other than in a random network. This is the case for $D > 1$. In addition to the intra-party determination of the dyadicity for each topic, the average of the intra-party dyadicities for each topic is calculated. This provides information about the dyadicity of the topic as a whole.

The determination of the homophily of the network is based on the calculations in *Currarini et al.* [35]. First, the ratio of nodes of a party to the total number of nodes is represented as $w_i = \frac{N_i}{N}$, where i represents the party under consideration. The homophily within a party is calculated with $H_i = \frac{s_i}{s_i + d_i}$, where s_i represents the edges to nodes of the same party and d_i the edges to nodes of a different party. Since this measurement is susceptible to bias and different group sizes, the result is normalized to the internal homophily with $IH_i = \frac{H_i - w_i}{1 - w_i}$. Here the distortion is placed in relation to the maximum possible distortion $1 - w_i$. If $IH_i < 0$, it is a heterophilic network. A homophilic network is defined with $IH_i > 0$. With $\overline{H}_{Total} = \frac{1}{n}\sum_{i=1}^{n} IH_i$, the homophily is then built over the entire network, where n corresponds to the number of parties.

VI. Results of Topic Analysis

Figure 8 shows the number of nodes for each topic and both periods under consideration. It can be observed that the number of nodes for the topic areas is almost similar in both periods. Only for topic 3 (Vaccination), there is a slightly higher deviation in the number of nodes.

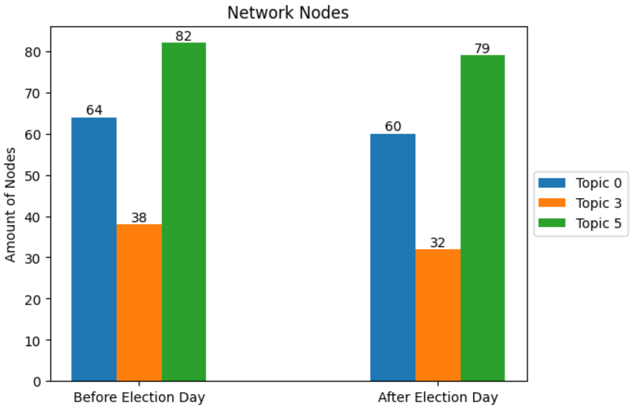


Figure 8. Network Nodes Distribution

Figure 9 shows the number of edges of the selected topics for both periods. The number of edges refers to the actual sum of interactions between accounts. Multiple interactions between the same nodes are also listed as multiple edges. It should be emphasized that the number of edges has only increased for topic 0 (Climate) over the course of the observation period. For all other topic areas, there is a reduction in the number of edges in the second period compared to the first period. The popularity of topics 3 (Vaccination) and 5 (Finance) decreased over time, whereby a lower need for discussion has led to a lower number of edges. For topic 3 (Vaccination), this development can be explained by the course of the pandemic countermeasures.

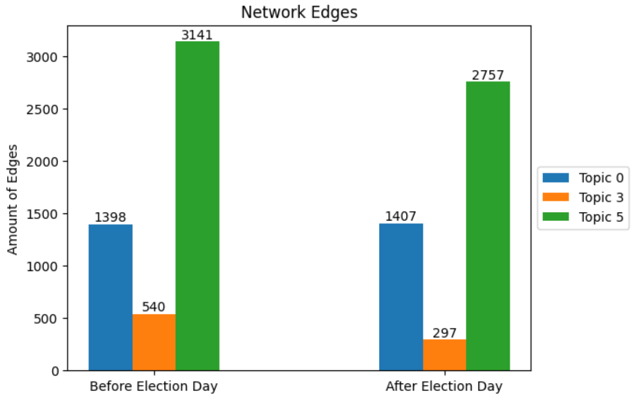


Figure 9. Network Edges Distribution

The results of the network node density analysis are shown in Figure 10. The node density is determined based on a directed network; it is the directed node density of the entire network. For topic 0 (Climate), the network node density increases slightly over time, this topic is examined in more detail. Figure 11 serves as an example of a breakdown of the party node density based on a topic.

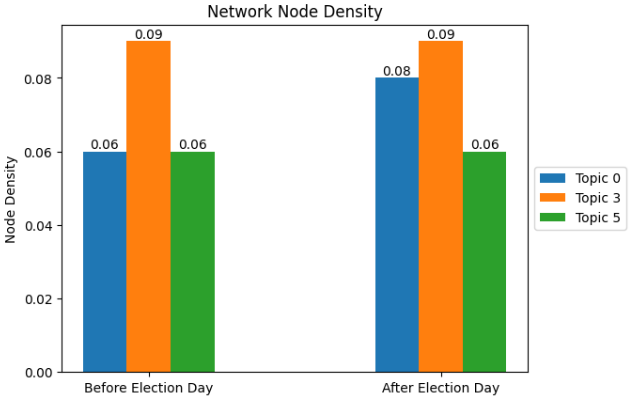


Figure 10. Network Node Density

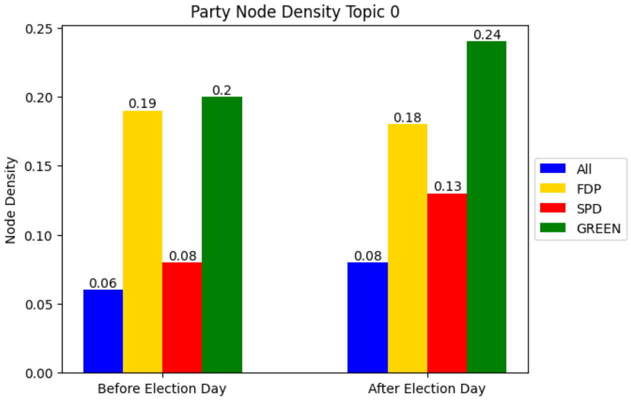


Figure 11. Party Node Density Topic 0 (Climate)

Table IX shows the average value of the node density for each party as well as for the entire network over the period for all topics.

TABLE IX. PARTY NODE DENSITY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	0.20	0.19	-0.01
SPD	0.12	0.17	+0.05
Greens	0.20	0.22	+0.02
All	0.07	0.08	+0.01

The results of the analysis of the dyadicity of the parties are explained through the lens of the individual topics. Figure 12 shows the dyadicities of the networks of the individual topics. For topic 3 (Vaccination), the dyadicity increases over the period under observation. For topics 0 (Climate) and 5 (Finance), the dyadicity value decreases slightly. An increase in dyadicity indicates an increase in intra-party communication for the respective topic, while a decrease in dyadicity indicates an increase in inter-party communication.

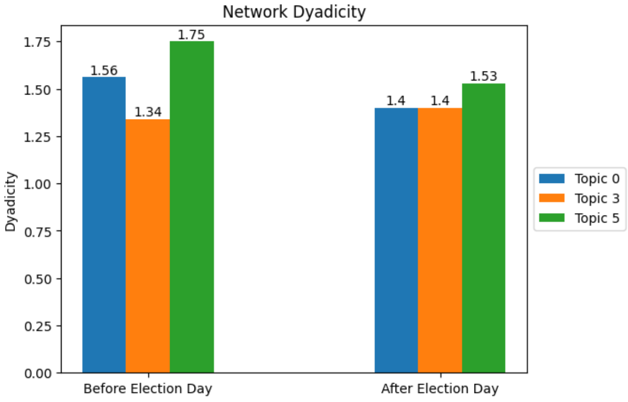


Figure 12. Network Dyadicity Distribution

Table X shows the development of the average dyadicity of the parties across all topics. The dyadicity values of the individual parties lead to the conclusion that the Greens and the FDP cultivate strong intra-party communication, while the SPD prefers inter-party communication. The developments shown in Table X suggest that the SPD has slightly increased its intra-party communication over time, while the FPD and the Greens have developed a slight trend towards inter-party communication.

TABLE X. PARTY DYADICITY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	1.85	1.54	-0.31
SPD	0.80	1.08	+0.28
Greens	1.99	1.70	-0.29

Figure 13 shows the analysis of party dyadicity for topic 0 (Climate). Here it can be seen that the dyadicity decreases at the FPD and the Greens. The dyadicity of the SPD increases for the same topic. This suggests that, compared to the previous period, the FDP and the Greens have stronger inter-party communication than the SPD. However, the SPD has strengthened its intra-party communication.

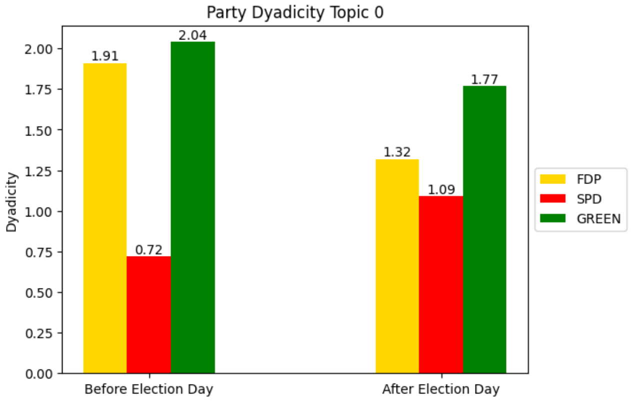


Figure 13. Party Dyadicity Topic 0 (Climate)

Table XI shows the change in the average homophily of the parties for all topics. A decrease in homophily can be observed for all parties. It indicates a reduction in intra-party debate in relation to inter-party communication. This means that the parties developed a trend towards communication between each other rather than communication within their party boundaries.

TABLE XI. INBREED HOMOPHILY PARTY AVERAGES

Party	Before Election	After Election	Difference
FDP	0.85	0.83	-0.02
SPD	0.20	0.07	-0.13
Greens	0.75	0.64	-0.11

Overall, there have been changes in the parties' communication behavior concerning the analyzed topics. The SPD’s homophily declines. At both other parties, the changes are small and the homophily remains at a high level, which does not align with the initial expectations. An answer to the research question can therefore be formulated as follows:

The communication behavior of the traffic light parties for the period under consideration concerning the selected topics has not changed as a result of the formation of the government in 2021. From the perspective of a political scientist, this can be attributed to the way compromises are reached in government circles. While compromises are discussed and found in the respective committees, social media platforms such as Twitter usually serve to profile successes, at least among coalition partners. In general, public criticism is avoided among coalition partners and rather expressed towards the opposition.

VII. CONCLUSION

This paper illustrates the application of techniques from social network analysis, sentiment analysis, cluster analysis, and topic analysis in combination to explore communication on social media, especially on microblogs.

H1 is proven, as differences are found for mention and reply interactions. The networks for each interaction type yield differences in both intra- and inter-party interactions, which is shown by the results of the GNA. These findings are statistically significant due to the low p-values. H2 can therefore be considered true. The p-value of the χ^2 test indicates a low likelihood that party affiliation influences the assigned cluster.

H3 cannot be answered clearly. MPs' mutual sentiment changed positively. The FDP's positive change towards the coalition partners SPD and Greens can be considered a sign of a generally improved attitude towards these parties. However, the notable overall increase in positivity across most parties could indicate that the findings of the FDP are not unique. The generally positive attitude between parties after the election can be caused by MPs congratulating one another. A lack of German language sentiment analysis models for short text fragments limits this research. Improved models utilize machine learning techniques and so can comprehend sentiments on a broader level and can also recognize nuances.

Statements about H4 are not reliable. However, while positive tendencies towards an "Ampel" coalition can be shown from both the sentiment analysis and the inter-party and intra-coalition homogeneity, neither can be proven as statistically significant.

Concerning hypothesis H5, the specific metrics do not indicate a change in homophily. No significant change in homophily can be detected in the topic networks depending on the observation period.

Concerning hypothesis H6, the intensity of communication has decreased across all topics. The communication intensity is determined by the number of edges per topic.

Different interactions between MPs can be observed during the pre- and post-election periods and the resulting interaction networks for each period show a difference in intra- and inter-party communication. However, this paper handles political communication only via Twitter. Results are partially transferable to other countries.

Future work may include "The Left" and AfD in these considerations to produce more information. Expanding the evaluated timeframes or continuous monitoring would produce more data. Analyzing follower and friend networks and MPs' liking behavior in combination with the findings of this article would yield insights into differences in parties' mutual relationships around elections.

The change of communication networks within German parties based on political issues could be of interest to several actors. The results of this study could contribute to a better understanding and analysis of the political climate within the "traffic light parties" in the context of political analysis. The results of this study can help to understand the behavior of people in groups and their communication within political parties. This could be of interest to social scientists working on questions of group dynamics and political socialization. Findings from this study could also be useful for communication practitioners by providing insights into the way communication networks develop and change within political parties. This could be of interest to communication professionals involved in the design and management of communication strategies.

References

[1] H. Schuhbauer, S. Schötteler, J. Niu, B. Schiffer, and D. Wolfarth, "A Quantitative Social Network Analysis of Politicians' Tweets to Explore Political Communication". Proceedings of The Twelfth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 22). Lisbon, Portugal, Oct. 2022.

[2] E. Shearer, "Social Media Outpaces Print Newspapers in the U.S. as a News Source", Pew Research Center, [Online]. Available from https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/, 2018.

[3] A. Watson, "Usage of Social Media as a News Source Worldwide 2021", Statista, [Online]. Available from https://www.statista.com/statistics/718019/social-media-news-source/, 2021.

[4] S. Wasserman and K. Faust, "Social Network Analysis – Methods and Applications", 1st edn. Cambridge University Press, Cambridge, USA, 1994.

[5] A. Virk, "Twitter: The Strength of Weak Ties", University of Auckland Business Review, Vol. 13, No. 1. University of Auckland, Auckland, AUK, NZL, pp. 19-21, Jan. 2011.

[6] M. S. Granovetter, "The Strength of Weak Ties", American Journal of Sociology, Vol. 78, No. 6. The University of Chicago Press, Chicago, IL, USA, pp. 1360-1380, May 1973.

[7] D. S. Lassen and A. R. Brown, "Twitter: The Electoral Connection?", Social Science Computer Review, Vol. 29, No. 4. SAGE Publications, Thousand Oaks, CA, USA, pp. 419-436, Nov. 2011.

DOI: https://doi.org/10.1177%2F0894439310382749

[8] M. Boireau, "Uncovering Online Political Communities of Belgian MPs through Social Network Clustering Analysis", Proceedings of the 2015 2nd International Conference on Electronic Governance and Open Society: Challenges in Eurasia (EGOSE '15). Association for Computing Machinery, New York, NY, USA, pp. 150-163, Nov. 2015. DOI: https://doi.org/10.1145/2846012.2846049.

[9] J. A. Caetano, H. S. Lima, M. F. Santos, and H. T. Marques-Neto, "Using sentiment analysis to define Twitter political users' classes and their homophily during the 2016 American presidential election", Journal of Internet Services and Applications, Vol. 9, Article 18, Sep. 2018. Springer Open, DOI: https://doi.org/10.1186/s13174-018-0089-0.

[10] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods", ACM Computing Surveys, Vol 49, No. 28. Association for Computing Machinery, New York, NY, USA, pp. 1-41, Jun. 2017. DOI: https://doi.org/10.1145/2938640.

[11] A. Boras and S. R. Singh, "Investigating political polarization in India through the lens of Twitter", Springer, 2022. DOI: https://doi.org/10.1007/s13278-022-00939-z.

[12] F. Meier, A. Bazo, and D. Elsweiler, "Using Social Media Data to Analyse Issue Engagement During the 2017 German Federal Election", Commun. ACM 22, 1, 25, pp. 1-25, Feb. 2021. DOI: https://doi.org/10.1145/3467020.

[13] E. Garcia-Sanchez, P. R. Benetti, G. L. Higa, M. C. Alvarez, and E. Gomez-Nieto, "Political discourses, ideologies, and online coalitions in the Brazilian Congress on Twitter during 2019". New Media & Society, 25(5), p.1130-1152, 2021. DOI: https://doi.org/10.1177/14614448211017920.

[14] R. Egger and J.A. Yu, “Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts”. Front. Sociol. 2022, 7, 886498. https://doi.org/10.3389/fsoc.2022.886498

[15] Y. A. Jeddi, Scweet (Version 1.6), [Online]. Available from: https://github.com/Altimis/Scweet, Dec. 2021.

[16] Twitter Handle, [Online]. Available from: https://influencermarketinghub.com/glossary/twitter-handle/#:~:text=What%20is%20Twitter%20Handle%3F%20A%20Twitter%20handle%2C%20which,usernames%20on%20the%20social%20network%20are%20the%20same, Dec. 2022.

[17] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner, "On Modularity Clustering". IEEE Transactions on Knowledge and Data Engineering. 20 (2): 172–188, Feb. 2008. doi:10.1109/TKDE.2007.190689 . S2CID 150684.

[18] A. Evtushenko and J. Kleinberg, “The paradox of second-order homophily in networks”, Sci Rep 11, 13360, Jun. 2021. https://doi.org/10.1038/s41598-021-92719-6

[19] The Selenium Project, Selenium (Version 4.1.0), [Online]. Available from: https://github.com/seleniumhq/selenium, Dec. 2021.

[20] GEXF Working Group, GEXF File Format (Version 1.2), The Gephi Community Project, 2009. [Online]. Available from: http://gexf.net, Dec. 2021.

[21] Data work, [Online]. Available from: https://git.informatik.fh-nuernberg.de/wolfarthda82341/sna-germanys-members-of-parliament-on-twitter, Oct. 2022.

[22] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks”, Proceedings of the National Academy, Vol. 99, No. 12. National Academy of Sciences of the United States of America, Washington DC, USA, pp. 7821-7826, Jun. 2002.

DOI: https://doi.org/10.1073/pnas.122653799.

[23] S. Loria, TextBlob (Release v0.16.0) Documentation. TextBlob: Simplified Text Processing. [Online]. Available from: https://textblob.readthedocs.io/en/dev/index.html#, Dec. 2021.

[24] M. Killer, textblob-de. (Version 0.4.3), German language support for TextBlob by Steven Loria. [Online]. Available from: https://github.com/markuskiller/textblob-de, Dec. 2019.

[25] E. Quadbeck, "Früh und schmutzig – der Bundestagswahlkampf 2021," RedaktionsNetzwerk Deutschland, [Online]. Available from:

https://www.rnd.de/politik/bundestagswahlkampf-2021-haerter-und-schmutziger-MAAJZPIDTFAD7OMUJXUXE2OE5I.html, May 2023.

[26] S. K. Horwitz and I. B. Horwitz, “The Effects of Team Diversity on Team Outcomes: A Meta-Analytic Review of Team Demography”, Journal of Management, 33(6), pp. 987–1015, 2007. https://doi.org/10.1177/0149206307308587

[27] Twitter Account Deutscher Bundestag, List members [Online]. Available from: https://mobile.twitter.com/i/lists/912241909002833921/members, Nov. 2022

[28] J. Roesslein, “Tweepy Documentation” [Online]. Available from: https://docs.tweepy.org/en/stable/index.html, Jan. 2023.

[29] Deutscher Bundestag, “The Open Graph Viz Platform” [Online]. Available from: https://www.bundestag.de/services/opendata, Jan. 2023

[30] M. Grootendorst, “Bertopic: Neural topic modeling with a class-
based tf-idf procedure”. arXiv preprint arXiv:2203.05794, 2022.

[31] D. Newman, Y. Noh, E. Talley, S. Karimi, and T. Baldwin, “Evaluating topic models for digital libraries”, Proceedings of the 10th annual joint conference on Digital libraries (JCDL '10). Association for Computing Machinery, New York, NY, USA, pp. 215–224, 2010. https://doi.org/10.1145/1816123.1816156

[32] NetworkX Developers, Software for Complex Networks [Online]. Available from: https://networkx.org/documentation/stable/index.html, Jan. 2023

[33] C. Stegbauer and R. Häußling. Handbuch Netzwerkforschung. Verlag für Sozialwissenschaften, 2010. ISBN: 9783531158082.

[34] X. Wang, O. Varol, and T. Eliassi-Rad, “Information access equality on generative models of complex networks”, Appl. Netw. Sci. 7, 54, 2022. https://doi.org/10.1007/s41109-022-00494-8

[35] S. Currarini, M.O. Jackson, and P. Pin, “An Economic Model of Friendship: Homophily, Minorities, and Segregation”. Econometrica, 77: pp. 1003-1045, 2009. https://doi.org/10.3982/ECTA7528

Towards Design and Implementation of the Breakthrough Web

Santipong Thaiprayoon
Chair of Communication Networks
FernUniversität in Hagen
Hagen, Germany
santipong.thaiprayoon@fernuni-hagen.de

Herwig Unger
Chair of Communication Networks
FernUniversität in Hagen
Hagen, Germany
herwig.unger@fernuni-hagen.de

Abstract—**With rising demands for accessibility, security, and privacy, the future of the Web has attracted significant attention from the digital economy, focusing on improving data protection and user experience. This article proposes a conceptual framework enabling local users to directly and safely access and share web content and services on the local and global web through their mobile devices. The local web is configured to run on a sandbox server within a specific area over a local network. In addition, the proposed framework incorporates Web 3.0, which makes the Web better understand contextual data and automatically provides personalized responses that match users. In contrast, local users retain private control over their data. The results of the experiments revealed that the proposed framework is secure, scalable, and reliable enough to be used in real-world environments. This framework could also be highly valuable in evolving the power of a decentralized Web.**

***Index Terms*—web 3.0; privacy protection; user personalization, proximity authentication, network communication.**

I. Introduction

This article is an extension of research work that was originally proposed in the Fourteenth International Conference on Advances in Future Internet (AFIN 2022) [1], which provided a personalized context-aware recommendation for the development of an autonomous intelligent agent for an individual user on online social networks.

The World Wide Web, commonly referred to as WWW, W3, or the Web, is a network system of interconnected documents and other web resources to be accessed through the Internet, linked by hyperlinks and Uniform Resource Locators (URLs) [2]. Since the Web was developed by Tim Berners-Lee, a British computer scientist, while working at CERN, the European physics research organization, it is a free "wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents." However, due to the lack of strict regulations [3], any user with a server connected to the Internet could freely and easily access and share any type of information, including texts, multimedia, and user-generated content.

As there is no connection between information and web addresses where it is stored, search engines [4] are developed to support users in an inestimable information space, and rapidly expanding social network platforms are established to connect not only content but also users in a single, worldwide-spanning system, which leads to the commercialization of the Web. Consequently, the Web has become a place that overfloods its users with plenty of useful, useless, and sometimes even dangerous or criminal information and services [5]. As a result, government institutions in all countries attempted to regulate the Internet use with many more or less valuable regulations, making Internet use sometimes more stressful than helpful [6].

In this process, users have gradually lost more and more control over what information about themselves is collected, what part of the overall available information is offered to them, and how much outdated advertising they have to tolerate. In particular, the following drawbacks can be identified:

- The Web architecture today is a centralized system of information control, which means that all personal data and user behavior are stored in a single place and controlled by one corporation or organization. This makes it easy for governments or hackers to censor content and for users to be tracked and access all of their personal information without their consent.
- Web 2.0 enables any user to publish their content in large social network systems, which is a significant improvement over Web 1.0, which only allows static linking of web pages with potentially dynamic content. However, real interaction between previously unknown users in a given context or location is still impossible. Additionally, centralized platforms can cause numerous issues for users. For instance, users are unable to quickly transfer data between platforms or switch between applications that would reuse their data.
- Information is not connected to any location or context. Typically, a web search engine returns results from any site worldwide, although that information may be senseless out of its location and context.
- The amount of information on the Web has made it difficult and time-consuming for users to find specific information. Any search engine is still a single event; previous search results cannot be refined with the help of the system. In particular, defining locations and addressing groups of people, information sources, and contexts are tedious tasks in today's search engines. Generally, the large number of search results overwhelms users, and as a result, only the top 10 to 30 results are considered, which may not be the result they prefer. The search engines should return results that are tailored to the needs of a particular user.
- Manifold data security and protection laws like the European DS-GVO [7], permanent advertisements, and push-up requests annoy users and prevent the Web from being efficient in information support. Nevertheless, it is unclear to most users what happens with their data, what meta-information is collected, and how to suppress or remove such information from third-party servers.

In conclusion, the Web of today is controlled by large corporations and governments such as Amazon, Meta (formerly Facebook), and Google. At the same time, users have to follow their guidelines, are limited in the design of the information they offer, and need help to address it to the right, intended group of users. Moreover, the existing information flood discriminates against small information providers and merchants, which disappear in the sponsored offers of tech giants. Following Matthew Hodgson [8], “A decentralized web would give power back to the people online,” also regarding their privacy, data portability, and security. However, those architectures are still not available. An additional barrier to their construction is the fact that those systems cannot be built in competition with existing structures but must be integrated into cooperating with them.

To deal with the drawbacks, a trusted conceptual framework is proposed. The proposed framework could focus on the development of personalized services and the purpose of re-decentralizing the Web by giving users complete control over personal data while managing privacy, security, transparency, and the Internet experience [9]. Additionally, users can grant or revoke access to their personal data as needed. This way enables direct interaction between users and external services hosted on local servers without intermediaries.

This article is structured as follows: In Section II, a literature review is conducted. Section III explains the architectural framework. In Section IV, communication designs are described. The experimental details are presented in Section V. Section VI contains a discussion of the results. Section VII gives examples of use cases. Finally, the conclusion and suggestions for future works are presented.

II. Literature Review

In this section, the fundamental concepts for designing and implementing the proposed framework are introduced in detail, including the evolution of the Web, Wi-Fi technology, Bluetooth technology, context-aware recommender systems, proximity authentication, user matchmaking, and user personalization. Then, related research works are discussed.

A. *The Evaluation of the Web*

In the early 1980s, the Internet evolved into a global communication network infrastructure that allows computer networks worldwide to connect to one another. The Internet has become an essential part of human interactions and connectivity. It enables every individual to gain access to digital information through various applications, particularly the Web. The Web is a collection of web pages containing documents and other web resources. Users can access web content via the Internet on their devices using web browsers or web-based applications [10], [11].

The development of the Web, known as Web 1.0, began in the 1990s. This is the first stage in the evolution of the Web. All Internet users are content consumers in the Web 1.0 era, where content creators provide content in web pages that are stored on web servers in the HyperText Markup Language (HTML) format. These web pages are represented as the read-only web, which consists primarily of static content and allows users to only search for and read information [12].

The lack of active interaction between users and the web pages resulted in Web 2.0. Web 2.0, the current age of the Web, is the second stage of the Web revolution. It is an improved version of Web 1.0 due to the transition from static to dynamic content that responds to user input. Web 2.0 is defined as the read-write web that emphasizes the importance of user-generated content. In the current era, any user can be both a content producer and a content consumer. With the growth of mobile technologies, they can also contribute information and communicate with other users via websites using smart devices. Meta, Twitter, YouTube, and Instagram are well-known Web 2.0 commercial platforms that allow users to contribute content, share information, and interact with other Internet users in a virtual community [13].

A large amount of new content has been currently being created and shared on Web 2.0 applications. This content information is under the control of some of the giant tech companies. This means that all of that data, including personal and sensitive data, is exploited for business purposes, such as targeted advertisements and marketing campaigns. Web 3.0 is, therefore, a concept for a new iteration of the Web that aims to make the Web more context-aware and intelligent in decentralized infrastructures. Essentially, it can understand the meaning of words and emotions through data analysis in order to automatically provide the user with highly personalized and appropriate suggestions of items by leveraging emerging technologies that heavily rely on blockchain technology, Artificial Intelligence (AI), Natural Language Processing (NLP), Machine Learning (ML), Internet of Things (IoT), Augmented Reality (AR), and Virtual Reality (VR). Users will then have a better experience driven by enhanced data connectivity. This will be achieved by empowering each user to become the owner of their data and enhancing the overall user experience through the implementation of numerous innovations. In Web 3.0, users have ownership and control over their data and can choose to share or monetize their data on their own terms. This gives users more privacy and control, addressing the concerns of data centralization and lack of privacy that are prevalent in Web 2.0. In addition, Web 3.0 enables participants to interact freely, publicly, and privately with others without the need for permission or central authorities, thus avoiding scalability and single-point-of-failure issues [14]–[16].

In summary, Web 3.0 is still a concept that is being developed. However, some businesses attempt to develop products that can be transformed into Web 3.0 applications. Some of the most widely used Web 3.0 technology can be seen in virtual assistants like Siri and Alexa and connected smart homes.

B. *Wi-Fi Technology*

Wi-Fi is a wireless communication technology that allows devices such as computers, mobile devices, and other equipment to interface with wireless networks [17], [18]. It is commonly used for a Wireless Local Area Network (WLAN) of devices and Internet access, allowing nearby digital devices to exchange data over radio waves. Wi-Fi technology uses radio waves to transmit and receive data between Wi-Fi devices, such as laptops, smartphones, and Wireless Access Points (APs). The radio waves used in Wi-Fi technology operate in the 2.4 GHz and 5 GHz frequency bands. These frequencies can be divided into multiple channels, and APs and Wi-Fi devices communicate over a specific channel. The channel can affect the speed and stability of the Wi-Fi connection. One of the key advantages of Wi-Fi technology is that it can support multiple devices simultaneously without needing physical cables or wires. Wi-Fi also enables multiple devices to connect to the same network, allowing users to share resources such as printers and files.

The data sent over a Wi-Fi connection is encoded using a protocol called the Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard. The protocol defines rules and procedures governing how Wi-Fi devices transmit, receive, and manage each other over a Wi-Fi network. The Wi-Fi protocol also includes a range of security measures to protect against unauthorized access and ensure the privacy of data transmitted over the network. Some of the security measures built into the Wi-Fi protocol include the Wired Equivalent Privacy (WEP) protocol, Wi-Fi Protected Access (WPA), and Wi-Fi Protected Access II (WPA2) [19].

When a user needs to connect to a Wi-Fi network, the user with a Wi-Fi-enabled device begins to scan the environment for available Wi-Fi networks. It looks for APs broadcasting their network name, the SSID and determines whether the network is secured or open. If the network is open, the device can immediately connect to it, but the user will need to enter a password to gain access if it is secured. Once the device is connected to a Wi-Fi network, it starts communicating with other devices on the network [20].

Wi-Fi technology is used in a wide range of applications, from home networking to large-scale enterprise networks. In homes, Wi-Fi connects devices such as smartphones, laptops, and smart home devices. Wi-Fi is also utilized in public areas such as airports, cafes, and hotels to provide customers with free or paid Wi-Fi access [21].

C. *Bluetooth Technology*

Bluetooth technology is a wireless communication standard for data exchange over short distances between different Bluetooth devices, such as smartphones, laptops, and IoT devices, which attempt to build personal area networks (PANs). It allows users to simultaneously send or receive data between devices or neighboring active devices within the range of Bluetooth signals [22].

The history of Bluetooth [23]–[26] started with the first generation developed by the Bluetooth Special Interest Group (SIG) in the 1990s, which provided the Basic Rate (BR) for basic functionalities. Each new version of Bluetooth usually comes with a new mode that adds new features or makes certain things work better. In 2 and 3 generations, the SIG introduced Enhanced Data Rate (EDR) and High Speed (HS) to boost throughput in different manners, which comprised one fundamental part of Bluetooth, BR/EDR. Bluetooth Low Energy (BLE) was introduced in the four generations to increase the application fields of Bluetooth for low-power devices. Indeed, the fifth generation represents a significantly improved new generation. According to the SIG, the latest version of Bluetooth, version 5.2, which was introduced in 2019, achieved two times the transmission speed, four times the transmission range, and eight times the broadcasting capacity of Bluetooth 4.2. With Bluetooth 5.2, upgraded devices are able to transfer data more quickly and establish connections with richer content that are more stable. Currently, Bluetooth is officially successful in several areas, including speed, coverage, advertising capacity, robustness, and network capacity.

There are three primary variants of Bluetooth technology: (1) Bluetooth Classic (BC), also known as Bluetooth Basic Rate or Enhanced Data Rate (BR/EDR); (2) Bluetooth Low Energy (BLE); and (3) the recently released Bluetooth Mesh (Mesh). BC and BLE devices form a piconet in a central-peripheral manner, whereas mesh enables devices to create a mesh network based on BLE advertising.

For transferring data between two Bluetooth devices, they first establish a communication channel using a pairing process. A discoverable device should accept incoming connection requests. Generally, a device finds the discoverable device using a service discovery process. After the discoverable device agrees with the pairing request, the two devices exchange security keys to perform authentication and encryption to complete the bonding process. After the pairing and bonding processes are complete, the two devices are ready to transmit data. The security keys generated during the bonding process are reused if devices disconnect and reconnect. The two devices must always be paired and connected, with each of the two devices trusting the other and being able to exchange data in a secure manner using encryption to provide confidentiality and authenticity guarantees for communication against attackers.

In Bluetooth technology, Received Signal Strength Indication (RSSI) is a metric used to estimate the distance between two Bluetooth devices in close proximity [27]. The basic principle is that the strength of the signal decreases as the distance between the two Bluetooth devices increases. The RSSI value is measured in decibels (dBm) and indicates the power level of the received signal that reaches a device when a Bluetooth receiver is detected. Since Bluetooth receivers can broadcast their advertising packets with varying transmission power (TX) values, a combination of the RSSI and TX power values is used to estimate the distance to the device. The TX power value is the strength of the signal measured at 1 m from the device. The precision of the measured TX power value is crucial for calculating the distance between the device and the Bluetooth receiver, as the signal strength varies with the device's distance. The distance between two Bluetooth devices can be calculated using the equation (1).

$$d =^{10^{\frac{TX-RSSI}{10n}}} \tag{1}$$

where d is the distance in meters, and *TX* is an RSSI value known as the signal strength when a device and a Bluetooth receiver are at 1 meter. *RSSI* is the device's RSSI value. The constant n depends on the location of the Bluetooth receiver or the environmental factor, ranging in value between 2 and 4.

D. *Context-Aware Recommender Systems*

Traditional Recommendation Systems (RS) can be modeled as a two-dimensional (2D): $Users{\times}Items$ space. However, considering only information about users and items is not enough in applications. Therefore, additional contextual information should be considered in the recommendation process. Contextual information includes the location, the time, the weather (e.g., the current temperature), the user's mood, the user's current activity, the user's current goals, the presence of other individuals accompanying the user, and the user's communication capabilities.

With advances in ubiquitous and mobile computing, the lack of analysis of contextual information in recommendation systems has been strongly attacked. Thus, researchers and developers have mainly focused on solving classic problems of recommendation systems, such as the cold start problem, spam vulnerability, high dimensionality, and many others [28], [29]. Recently, researchers working on recommendation systems have recognized the need to investigate them in domains where context information is relevant. In order to improve the recommendations based on contextual information, the authors extend the classical 2D paradigm to a multidimensional recommendation model that provides recommendations based on multiple dimensions: $Users \times Items \times Contexts$ space. RS incorporating context information in the recommendation process are expressed as Context-Aware Recommender Systems (CARS). In other words, CARS attempt to accommodate user preferences in various contexts. Since user preferences may vary depending on the context, it is necessary to consider context information when generating the most relevant recommendations [30]. For example, a user may prefer a different type of restaurant for a business lunch compared to a casual dinner with friends. CARS can also consider other factors, such as time of day, location, and weather conditions to provide more personalized recommendations.

E. Proximity Authentication

Proximity authentication [31]–[33] is a method of authentication that uses the proximity of a physical device, such as a smartphone or smart card, to verify the identity of a user. This method is commonly employed to enhance the security of physical access control systems, such as buildings or secure areas, or to grant access to digital resources, such as online accounts or computer systems. The basic principle of proximity authentication is that a device, such as a smartphone or smart card, is associated with a specific user and can be used to verify their identity. When a user approaches a secure area or attempts to access a digital resource, the device is brought close to a reader or sensor that can communicate with it. The reader or sensor may use a variety of technologies to communicate with the device, such as RFID, NFC, Bluetooth, or Wi-Fi. The device then sends a signal to the reader or sensor, which verifies the identity of the user and grants or denies access appropriately.

There are various research studies on proximity authentication. Zhang et al. [34] proposed a novel proximity-based authentication mechanism for IoT devices called Move2Auth. Move2Auth detects proximity by comparing the RSS trace and smartphone sensor trace during two user gestures with large RSS variations. Kalamandeen et al. [35] introduced a system that determines if two devices are in close physical proximity by taking advantage of the similarity of the channel between these devices and a third observing device. The system leverages the many devices that users already possess to aid in this process.

F. User Matchmaking

User matchmaking [36], [37] is the process of connecting users and suggesting potential matches to users with similar user profiles, interests, preferences, or goals in a digital environment. This way is commonly used in online games, dating apps, social media platforms, and other online communities where users can interact with one another. The process of user matchmaking typically involves collecting information about each user, including their age, gender, location, interests, and past behaviors. This information is then used to identify other users who are most compatible with them based on shared interests, similar activity patterns, and geographical proximity. Moreover, user matchmaking is essential for creating engaging and personalized experiences for online community users.

There are several different approaches to user matchmaking. The first approach is content-based filtering [38]–[40]. This approach involves analyzing the content of user profiles to identify shared interests or preferences. The second approach is collaborative filtering. This approach examines the behavior patterns of a large number of users to identify similarities and differences. Hybrid approaches are the third approach in that user matchmaking systems incorporate content-based and collaborative filtering to improve precision and efficiency [41]. Building a user matchmaking system requires a deep understanding of user data, such as profiles, activity histories, and behavioral patterns. This data is then analyzed using one or more of the abovementioned approaches based on machine learning algorithms to identify potential matches.

G. User Personalization

User personalization [42], [43] is the process of filtering information systems that use machine learning algorithms and data analysis techniques to identify patterns and make predictions about the user. The goal of user personalization is to suggest items or content that are tailored to the specific needs, preferences, and interests of individual users. It involves analyzing data about certain contextual situations and past behaviors of each user, such as their location, time of day, device type, search history, browsing history, purchase history, and other factors, to make relevant recommendations. In recent years, research on user personalization has been used in several fields, including e-commerce sites, social media platforms, and other online services, to suggest products, services, or content that users are likely interested in. Liu et al. [44] proposed a hierarchical framework for personalized movie recommendations. The weekly ranking of a movie is used for association and recommendation. Moreover, movie content and user preferences are integrated to generate dynamic movie synopses for personalized navigation. Yan et al. [45] proposed a personalization framework for complementary product recommendation. The model encodes user purchase history into a personalized embedding and learns product features with graph-attention networks. It is then trained jointly via a re-ranking module. Xin and Wan [46] proposed a POI recommendation model based on an improved factorization machine and BERT to extract the social, user, POI, and sequence characteristics of users.

In conclusion, user personalization is an effective way of providing users with information that is customized to their specific preferences and interests. It can help businesses to create a more personalized and engaging user experience, increasing customer satisfaction and revenue.

H. Existing Research

In the past decade, the rapid development of intelligent Web systems has been accelerated by the emergence of new Web technologies and innovative Web usage concepts. Several research studies have been reported on various application areas. Chen [47] proposed a personalized learning path generation scheme that simultaneously considers the courseware difficulty level and the concept continuity of successive courseware based on incorrect pre-testing responses while implementing personalized curriculum sequencing during learning processes. Sharma et al. [48] proposed the application of semantic web mining, focusing on web personalization. In addition to providing the user with personalized web pages, the web personalization system offers the user a list of domains in which the user may be interested. Thus, users can switch interests while searching the Internet for information. Essam et al. [49] presented a decentralized platform for social Web called Solid. Solid is a decentralized platform for the social web that ensures data independence and simple yet powerful data management. Users store their data in Personal Online Datastores (PODs), which can be hosted on personal servers or public servers by POD providers. Users can have multiple PODs and choose from various providers based on privacy, reliability, and legal protection. Solid applications are client-side web or mobile applications that directly access and manipulate data from PODs. The platform facilitates the development and use of social features, allowing applications to aggregate data from different sources and enabling multiple applications to reuse the same data on a POD. Users can switch between applications without losing access to their existing data, as applications are decoupled from the data they use.

However, most of the existing studies ignore concerns about privacy and data protection issues for individuals who use the Web. In addition, they rarely use AI techniques to improve the performance of the Web based on the specific needs and interests of individual users in their context.

III. Architecture Framework

The Web has dramatically impacted the way people communicate, interact, and collaborate through various web-based platforms. Online users can contribute information, such as text posts, documents, videos, and photos, using computer desktops, laptops, and mobile devices. As a consequence, individuals are now able to stay connected to their local web as well as the broader global web that runs on top of the Internet using communication protocols like Hypertext Transfer Protocol (HTTP) and Hypertext Transfer Protocol Secure (HTTPS).

The main difference between the global and local webs is their scope and accessibility. The local web is a network of websites and services that are located within a specific geographic area or physical location, like a home, office, or specific network. These websites and services typically provide information and resources to individuals within a local community. It is typically smaller and more localized than the global web, and it can be created using a variety of communication technologies such as Wi-Fi, Bluetooth, Zigbee, and others. These technologies allow devices on the local web to communicate with each other directly. This allows for faster and more reliable communication between devices and increases privacy and security. In contrast, the global web refers to the network of websites and services accessible by different individuals over a network through specific communication protocols.

One major weakness of the Web is the lack of privacy protections and accessibility for users in specific environments. This issue can be resolved by utilizing locally hosted servers, also called local servers, in a specific geographical area. It is suitable for restricted environments, such as campus networks and company intranets. The use of local servers can increase accessibility for these users by providing them faster access to information, more control over the data stored on them, and less reliance on costly and unreliable Internet connections. Similarly, local servers can keep local networks safe, which are still invisible from the Internet. Local servers can also improve security because they are less likely to be attacked or have their data stolen than commercial or global servers.

In addition, with the emergence of Web 3.0 technology, the Web is shifting towards decentralized structures that provide enhanced security, privacy, and trust. This means that users can locally store their data on their own devices or on an independent server rather than in a centralized location like Google or Facebook. The idea behind this concept is that users will have greater control over their personal information. Individual users have the ability to own and manage their personal information to preserve their privacy. Local web refers to the concept of creating and maintaining web content and services on a local level rather than relying on centralized servers. Similarly, local servers refer to using small, decentralized servers to store and distribute information rather than relying on a few large, centralized servers. Figure 1 illustrates an overview of the local and global webs for local users to access information and services.

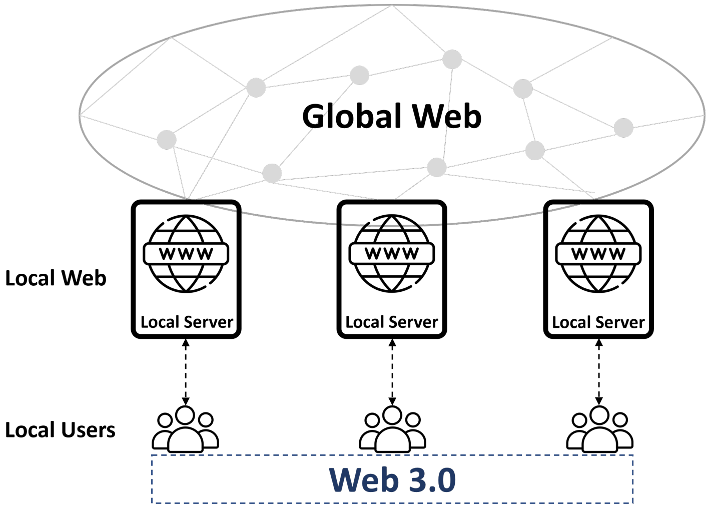


Fig. 1. An overview of the local and global web

From Figure 1, the Web can be expanded from the independent local level to the global level. Web resources are stored on a standalone local server, while they are accessed from other local servers on the Internet or public networks simultaneously. This means that each local server can securely communicate and synchronize based on HTTPS, which acts as a tunneling network. In the tunneling network, data is encrypted and encapsulated within a tunnel or secure communication channel, which is then transmitted across the network. HTTPS also uses encryption to protect data transmissions between local servers over a secure connection. Consequently, each local server enables its local users to securely and remotely participate in and access web pages or services hosted on other or neighboring local servers. The Web can be divided into two levels: (1) local web; and (2) global web. Web resources and services at the level of the local web are stored on a single local server that enables local users to access and share information about local events, news, or emergency services within a specific area. In contrast, the level of the global web is designed to be a global network system that is not controlled by any centralized servers. Web resources on several local servers are accessed by each other across that network. To describe the conceptual framework of a sandbox server in detail, the process overview is explained and illustrated in Figure 2.

A sandbox server, also called a local server, can be a physical machine or a virtual machine running on a computer connected to a local network within a sandbox environment to run applications in isolation safely. The framework is a proposed alternative approach to the current Web architecture to create a more secure

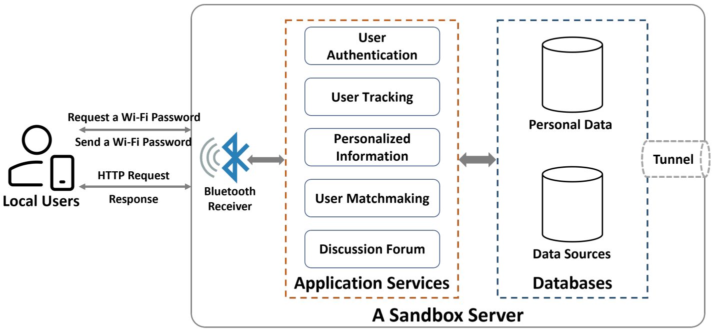


Fig. 2. The architecture framework of a sandbox server

and open Internet that prioritizes individual privacy. The proposed framework also provides a solution to privacy and data protection issues on the Web by independently separating personal data stored on mobile devices from services on local servers. Each user controls their own data through their mobile device, enabling direct interaction between users and their local server without the need for central servers. The personal data relates to an identified or identifiable natural person, including information that specifically identifies an individual, such as a name, address, telephone number, mobile number, e-mail address, credit card number, bank information, identification number, location data, or an online identifier, and information about that individual or their activities that is directly linked to the individual. The proposed framework also incorporates advanced technologies such as AI, ML, and NLP under Web 3.0 to generate personalized intelligent recommendations based on their contextual information and current vicinity, and increased privacy benefits. The proposed framework also helps prevent copyright infringement on the original works of users, such as written texts, music, images, videos, and software codes, when users publish them on the local web.

Each local user who holds a mobile device first connects to a sandbox server through a Bluetooth receiver to obtain a Wi-Fi password before accessing a local Wi-Fi network. This is a way to verify the identity of local users and protect against unauthorized access. Once the authentication process is complete, the local users are granted access to the local Wi-Fi network. Meanwhile, some of the user context stored on their mobile device is also shared and synchronized with different services on the sandbox server. These user contexts are used to identify patterns, make predictions about what the users may like, and recommend relevant information to them. The local users can then access and contribute information to the sandbox server. Furthermore, the local web can be integrated into the global web to improve accessibility and inclusivity for local users. On the level of the global web, several sandbox servers are connected across a public network, such as the Internet, to distribute access to web resources using the tunneling network that allows for the secure movement of data from one network to another. As a result, all connected sandbox servers can communicate securely and remotely with other sandbox servers and share information, services, and resources without centralized servers. Sandbox servers are autonomous and can freely join and leave the global network system at any time, which makes the system continuously and highly dynamic. For example, when a sandbox server wants to read a web page, it will look for other sandbox servers in the neighborhood that have the file and establish a direct connection between them. The file is transferred directly between the sandbox servers. One of the main benefits is that it removes bottlenecks or central points of failure from a system, as files are transferred directly over the network between two sandbox servers rather than through a central server. In addition, the proposed framework could help local users, who interact with the local and global web, identify opportunities to enhance their experiences and make the most of the resources available through the Web.

The proposed framework is designed and implemented to enable local users to access information and services on both the local and global web. Local users have the ability to have more control over their data, increase privacy, and reduce the power and authority of central servers. To explain the importance of research work, the main contributions are summarized as follows:

- A conceptual framework for the local and global web that focuses on providing information and services to local users is proposed. It enables local users to access and share information through a sandbox server and multiple sandbox servers connected together in a network, making the network more resilient and resistant to failure.
- The proposed framework can be incorporated into the global web in order to increase accessibility and inclusivity for local users.
- The local web provides more personalized and relevant results to local users based on their profiles and locations. They can be utilized to discover local information about restaurants, organizations, and opportunities for community engagement. This way can also assist in avoiding the problem of information overload.
- The combination of the local and global web is an alternative way for local users to access information from both levels of global and local information resources in different locations.
- Multiple sandbox servers operating within a network can provide benefits, such as enhanced scalability, security, and privacy, as well as the ability for local users to share resources and data directly.

In summary, the proposed framework expects to move towards becoming a web of highly intelligent interactions in the near future. It aims to rebuild the technical architecture of the Web based on the principles of decentralization. This will be accomplished through the use of artificial intelligence and communication technologies to provide Internet experiences with greater stability, security, and freedom for users, depending on their current contextual data. Moreover, the proposed framework is designed to bring power back into the hands of individuals by allowing them to own and control their data on mobile devices by utilizing the idea of peer-to-peer networks rather than companies or governments, which may use that information for their own purposes or sell it to advertisers, marketers, or others who might want access to it. This provides greater privacy and security for users, as no single entity can control what data is stored or how it can be used. It will also allow individuals to decide how much information they want to share with third parties or other people rather than having all their information stored in one place.

IV. Communication Designs

This section describes the details of the communication designs of the proposed framework. The idea behind this framework is to provide web content and services to local users. The sandbox server is designed as an isolated environment with restricted access within a specific geographical area. Moreover, each sandbox server has the possibility to collaborate with each other via efficient ad-hoc communication. The proposed framework for a sandbox server consists of four main components: (1) application services; (2) databases; (3) Bluetooth receivers; and (4) tunneling network.

1) Application Services: These application services run on a sandbox server and provide access over a local Wi-Fi network, allowing local users to interact with the application services from a specific area and providing secure communication channels. The application services can offer specific services to local users, such as local weather and news services, which can provide dynamic content and helpful information, such as information that changes frequently or is appropriate for local users. This can help reduce the spread of misinformation and disinformation, as local users can access credible and relevant information specific to their locations and contexts.

2) Databases: This acts as a database server, providing a centralized data management platform for storing, managing, and accessing data in a database. The main role of the database server is to receive requests from the application services, search for the requested data, and return the results. In addition, this database server can hold data both temporarily and permanently. Especially, sensitive data is automatically removed after 24 hours, according to a timestamp. This provides extra security for the database, ensuring that unauthorized users cannot access sensitive data for an extended period of time. There are two databases stored on the database server, including personal data and data sources. Personal data is the collection of user data, such as name, address, date of birth, age, gender, weight, height, education, user interests, areas of expertise, and short texts from the biography. In addition, the handling of personal data by application services is governed by privacy laws and regulations, and individuals have the right to access and control their data. On the other hand, data sources refer to general raw data that is widely available through various sources, such as books, websites, and news media, typically in the form of numbers, words, images, or other forms of data. It is used to analyze data to gain a better understanding and provide users with insights and information. It can also affect privacy and security, so organizations need to be careful when collecting and using data to ensure they follow all the relevant rules and laws.

3) Bluetooth Receivers: The Bluetooth receiver acts as a terminal device for Bluetooth connections, receiving the signal from source devices, including smartphones and laptops. Bluetooth-enabled devices communicate with each other using a process called pairing. To establish a secure wireless connection, the pairing process involves exchanging information between two Bluetooth-enabled devices, such as their unique identification numbers and encryption keys.

4) Tunneling Network: This is utilized in communication networks governed by the HTTPS protocol to provide a direct connection between two sandbox servers for the security and privacy of data transmission across a public network. The HTTPS protocol operates on the basis of a client-server model. A client, such as a web browser, sends an HTTPS request to a local server, such as a web server. The local server processes the request and returns an HTTPS response to the client. The response contains data such as HTML, images, videos, and other files that can be displayed via the web browser. HTTPS uses SSL/TLS (Secure Sockets Layer/Transport Layer Security) to encrypt the transmitted data. This provides a secure channel that protects against eavesdropping, tampering, and data theft. When a client requests a web page or resource using HTTPS, the server sends its SSL/TLS certificate to the client, which verifies the certificate's authenticity and establishes a secure connection.

The main tasks of the application services consist of five services: (1) user authentication; (2) user tracking; (3) user personalization; (4) user matchmaking; and (5) discussion forum. Each service is explained in the following subsection.

A. User Authentication Service

When a local Wi-Fi network is deployed, a robust authentication mechanism is the first layer of defense for the local Wi-Fi network. Therefore, strong authentication helps to protect network access and user-sensitive data and provides security for data communications between devices on the local Wi-Fi network so that authorized individuals can only use the network. This application service proposes a scheme for Wi-Fi authentication based on Bluetooth proximity, offering an additional layer of security and functionality. When users log in to the local Wi-Fi network with Bluetooth authentication, they can access and manage data through a series of secure, encrypted connections. Proximity-based authentication is a method of authentication that relies on the physical proximity of an object or device to verify the identity of the user. The main idea of this application service is to verify the identity of local users before granting them access to the local Wi-Fi network and the resources they need based on their presence or proximity. Suppose the local users are in close enough proximity to a sandbox server. In that case, they can perform the authentication process by connecting with a Bluetooth receiver to get Wi-Fi passwords sent to their registered mobile numbers via SMS messages. In the meantime, traffic data transmitted over the local Wi-Fi network is encrypted and decrypted using the WPA2-PSK encryption standard to prevent sensitive data from leaking or being compromised. Moreover, it is a convenient way for local users with confidence to connect to the Internet through secure communications on local Wi-Fi networks effectively. The process of Wi-Fi authentication is shown in Figure 3.

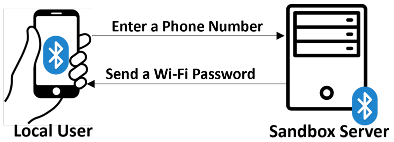


Fig. 3. The process of Wi-Fi authentication

The process begins with a local user holding a Bluetooth-enabled mobile device that attempts to connect to web content and services on a local Wi-Fi network. Before accessing the local Wi-Fi network, the local user needs to use the Bluetooth-enabled mobile device to scan for advertising signals broadcast from available Bluetooth receivers to request a Wi-Fi password. Bluetooth receivers act as devices to advertise and wait for connections, which accept an incoming connection request after advertising. When the mobile device is within the range of the Bluetooth receiver signal, it can connect and communicate with each other. The mobile device triggers an action on a mobile application to automatically redirect to an authentication page for network access verification. Then, the local user is required to provide a phone number on the authentication page via a Bluetooth connection. Then, the sandbox server generates a Wi-Fi password and sends it directly to the registered mobile phone number of the user via SMS message. SMS is a short message containing a Wi-Fi password that is sent to the mobile phone of the local user who initiated the request. To gain local Wi-Fi network access, the local user must enter the Wi-Fi password into a captive portal authentication page on their mobile device to prove their identity. This way, it ensures that the local user accessing the local Wi-Fi network has been verified by the owner of the phone. This scheme indicates that it is extremely convenient, safe, and smooth for users, thereby enhancing their trust and confidence in using resources and services on the local Wi-Fi network.

The Wi-Fi authentication protocol relies on a combination of authentication and encryption processes to provide maximum protection for local Wi-Fi networks. The Wi-Fi authentication protocol is illustrated in Figure 4.

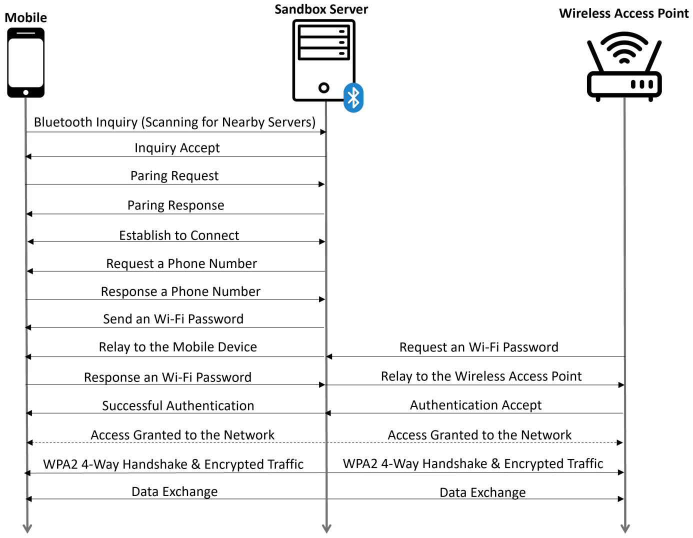

Mobile
Sandbox Server
Wireless Access Point
Bluetooth Inquiry (Scanning for Nearby Servers)
Inquiry Accept
Paring Request
Paring Response
Establish to Connect
Request a Phone Number
Response a Phone Number
Send an Wi-Fi Password
Relay to the Mobile Device
Request an Wi-Fi Password
Response an Wi-Fi Password
Relay to the Wireless Access Point
Successful Authentication
Authentication Accept
Access Granted to the Network
Access Granted to the Network
WPA2 4-Way Handshake & Encrypted Traffic
WPA2 4-Way Handshake & Encrypted Traffic
Data Exchange
Data Exchange

Fig. 4. The Wi-Fi authentication protocol

The user authentication process on a local Wi-Fi network performs a multi-step process involving three progressive states.

a) Registration State: A local user with a Bluetooth-enabled mobile device scans for nearby Bluetooth receivers to make a connection. After a Bluetooth receiver accepts the pairing request, the two devices complete a bonding process in which they exchange security keys. After the pairing and bonding processes are complete, the two devices exchange information. The local user then enters a phone number into an authentication web page before accessing wireless network resources. A sandbox server generates a Wi-Fi password and sends it back to the mobile phone number specified by the mobile user via an SMS message.

b) Authentication State: After the registration state is complete with a Bluetooth connection, the local user must enter the Wi-Fi password to access resources and services on the local Wi-Fi network. After authorization is complete, network access is granted to the local user.

c) Encryption State: When the authentication process is successful, the process of a WPA2 4-way handshake is performed to encrypt the data being transmitted between wireless access points and mobile devices. The local Wi-Fi network enables seamless data exchange through a single wireless access point at a time.

The Wi-Fi authentication and encryption are used in pairs to primarily prevent local Wi-Fi networks from unauthorized and malicious access attempts and secure wireless transmissions. The Wi-Fi authentication based on Bluetooth proximity acts as an interface in the middle between wireless access points and mobile devices. It helps to block all traffic except for authentication traffic. When the authentication server verifies the credentials of the user, it unblocks and permits all wireless traffic. This part could enhance the security and privacy capabilities of local Wi-Fi networks and improve the user experience.

The Wi-Fi authentication enables only the phone number owner to receive a Wi-Fi password, allowing them to log in to a local Wi-Fi network and verify their identity using a password sent via an SMS message. This makes it difficult for attackers to obtain unauthorized access to data and resources or to steal user credentials. It differs from traditional password authentication, which may continue to be useful for attackers with stolen credentials. This authentication method is an efficient way for businesses and organizations to integrate it into their authentication strategies because it has the potential to be implemented at a low total cost and directly reaches all users' existing mobile devices.

B. User Tracking Service

This application service presents a user tracking mechanism to recognize local users if they are inside a certain area based on Bluetooth technology, which is suitable for an RSSI algorithm. RSSI is a well-known location method that uses a known mathematical model that describes signal path loss with distance. The aim of this application service is to automatically determine and track local users who hold Bluetooth-enabled mobile devices and enter within the range of a Bluetooth receiver signal in real time. A sandbox server also displays that local users are present at a particular location. A Bluetooth receiver is responsible for detecting the location information of local users, tracking whether local users are still in signal coverage, and estimating the spatial distance between the Bluetooth receiver and a mobile device. The process of Bluetooth user tracking is shown in Figure 5.

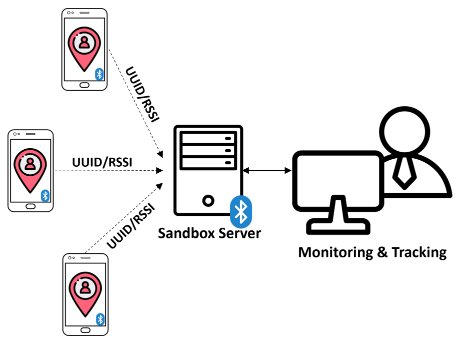


Fig. 5. The overview of user tracking

The user tracking mechanism relying on Bluetooth technology utilizes a range-based method involving the measurement of the RSSI of the Bluetooth signal from a Bluetooth receiver fixed at a prominent location. Alternatively, a fixed Bluetooth receiver is capable of calculating RSSI values transmitted by

Bluetooth-enabled mobile devices. Typically, the range-based method achieves an accuracy of a few meters and is used to determine whether an asset or a person is within a predefined room. Figure 5 demonstrates that the local users holding Bluetooth-enabled mobile devices are automatically located inside a room with a sandbox server that receives advertising messages from the mobile device. Administrators can also observe and track people staying in a specific area at a specific time to gain insight into behavioral patterns and person counts. In this case, with only one Bluetooth receiver, the local users can be roughly detected within a certain distance of the sandbox server.

On the basis of Bluetooth technology, the subsequent steps for detecting local users in a particular environment can be described.

- A fixed Bluetooth receiver starts periodically scanning advertising signals broadcasted from nearby mobile devices within the range of the fixed Bluetooth receiver signal. At the same time, each mobile device constantly advertises a Bluetooth signal containing identifying information within its range, enabling other Bluetooth devices to monitor and connect to them. When the fixed Bluetooth receiver receives Bluetooth signals from a mobile device, the mobile device is recognized as being in range. Typically, a Bluetooth broadcast signal contains a Universally Unique Identifier (UUID) and raw RSSI values. The scanning interval is set at 1000 milliseconds (1 second) to produce positioning results every second. The RSS value is measured in the unit decibel-milliwatts (dBm) and typically ranges from near 0 dBm (excellent signal) to less than -100 dBm (poor signal). A pretty value, for instance, is below -50 dBm; a reasonable value is between -70 dBm and -80 dBm, and a value of -100 dBm indicates no signal at all.
- If a mobile device appears within a specific proximity range of the fixed Bluetooth receiver signal, a Bluetooth connection between the fixed Bluetooth receiver and the mobile device is created. The mobile device is then located to determine whether it stays in a target area by measuring the signal strength with RSSI values. If the mobile device is not found, the fixed Bluetooth receiver tries to scan regularly to see whether mobile devices are present in their coverage radius. The signal strength between the fixed Bluetooth receiver and the mobile device is compared with a threshold value set at 60 dBm. If the signal strength exceeds the threshold value (<60 dBm), the fixed Bluetooth receiver identifies that the mobile device is in the range. If the signal strength is greater than the threshold value (>60 dBm), the fixed Bluetooth receiver identifies that the mobile device has moved far from the range of the fixed Bluetooth receiver. The fixed Bluetooth receiver captures a Bluetooth signal containing a raw RSSI value broadcasted from the mobile device to estimate the distance and recognize whether the mobile device is still inside the signal coverage or the targeted area.
- The fixed Bluetooth receiver roughly determines the geographical location of the mobile device by obtaining the known fixed location of the Bluetooth receiver. Then, the fixed Bluetooth receiver shows the location of the mobile device on a simple map.
- The fixed Bluetooth receiver processes the raw RSSI value from the Bluetooth network to calculate the approximate distance between the mobile device and the fixed Bluetooth receiver. Both Bluetooth devices need to be within Bluetooth range to estimate the distance. The distance between two Bluetooth devices is formulated using the equation (1).
- Administrators can continuously monitor and track the position and distance of users carrying Bluetooth-enabled mobile devices in real-time through a web-based monitoring tool.

Finally, this application service could be used to implement proximity solutions that provide location data and detect whether users enter a target area, usually within a closed environment, such as a building, shopping center, home, office, hospital, airport, conference room, or museum. Once users are identified, the system can track their location and identify them among the group of users. In addition, most smartphones are equipped with Bluetooth signal-receiving modules. Therefore, smartphones enable new Location-Based Service (LBS) capabilities. LBS provides advanced services to customers and managers, assisting individuals in determining their position. These include asset tracking, location-based advertising, business intelligence and analytics, social networking entertainment, and retail experiences.

C. User Personalization Service

The goal of this application service is to automatically provide users with the relevant information within their current situations by analyzing and understanding contextual data. A local user holding a mobile device requests and receives information and services from a sandbox server through the interface of the mobile device. The sandbox server is setup as host applications running on a web server designed to receive requests via the HTTP protocol in order to deliver static content from a website or other resources stored in a database, such as HTML, text, images, video, and other media files, to local users within a vicinity area. The content is then displayed via a web browser or mobile application. This application service of personalized information could also refer to the user tracking to deliver personalized and context-dependent recommendations of a list of relevant items based on their current location, such as nearby restaurants or shopping centers. Moreover, it allows local users to connect to the sandbox server via a local Wi-Fi network and automatically offers interesting services, such as private messages, games or puzzles, local point-of-interest, and advertising campaigns, such as promotions or discounts. This application service helps businesses enhance the user experience and improve their engagement with the product or service. This application service provides individual customers with a unique and personalized experience, allowing them to receive services that are tailored to their specific needs and preferences.

D. User Matchmaking Service

The main objective of this application service is to produce a list of potential friends with similar interests, ranked according to a similarity score based on their personal information. The matchmaking algorithm compares a user profile with other profiles using a text similarity technique and suggests a suitable list of similar users. User interests, expertise, and biographies are combined to improve the accuracy of recommending similar users. The text similarity technique measures the similarity score between two pieces of personal information based on lexical and semantic similarity, covering both word level and context level using NLP techniques, word embeddings, and cosine similarity. Each user profile is cleaned up and transformed from unstructured textual data into an appreciable format. A word embedding technique encodes and converts textual data into a numeric format as a vector representation. Two vectors are compared using cosine similarity to extract semantically similar text from user profiles and return a similarity score.

E. Discussion Forum Service

The discussion forum is a virtual platform where individuals can come together to discuss a specific topic or set of topics. This application service is designed to facilitate online conversations and exchanges between users and can be used for a wide range of purposes. The key features of discussion forum tools include the following:

a) User Profiles: The ability for users to create personal profiles that include information such as their name, location, and interests, as well as the ability to upload profile pictures and other media.

b) Threaded Discussions: The ability for users to start new discussions or add to existing ones, creating a threaded conversation.

c) Chat: The ability for users to communicate with other users in real-time.

d) Search Functionality: The ability for users to search the discussion forum for specific topics, keywords, or posts.

e) Notifications: The ability for users to receive notifications when new content is added to the discussion forum, such as new posts, comments, or replies.

These application services have the potential to greatly enhance the customer experience, increase customer loyalty, and drive business growth.

V. Experimental Details

This section describes the experimental design for the proposed framework. The experiments are performed to evaluate the performance of local communication networks and application services to ensure the proposed framework can run in real-world use cases efficiently. The detailed experimental procedures are shown in the following subsections.

A. Experimental Strategies

The experimental strategies are divided into three strategies, including (1) Bluetooth signal testing; (2) network latency testing; and (3) application service testing. The following strategies are explained below.

1) Bluetooth Signal Testing: This strategy measures the strength and quality of Bluetooth signals between a mobile device and a Bluetooth receiver. The Bluetooth signal testing aims to estimate the range of Bluetooth signals and how they are reachable, ensure that the Bluetooth signals are stable and reliable, and support the required data transfer rate for given use cases. RSSI values captured from the signals of two Bluetooth-enabled devices are used to indicate the strength of a Bluetooth signal at a specific location. Generally, the common range of RSSI values is between -100 dBm and -20 dBm. An RSSI value of -30 dBm indicates a strong Bluetooth signal, while an RSSI value of -90 dBm indicates a weak signal. In addition, a reasonably acceptable value is -70 dBm to -80 dBm.

2) Network Latency Testing: This strategy involves measuring the time it takes for a data packet to travel from its source to its destination across a network. This measurement is expressed in milliseconds (ms). The network latency can be measured by determining the Round Trip Time (RTT), which is defined as the amount of time it takes for requested data to be transferred from a client to a server and for the server to respond to the client after the request has been processed. Ideally, the response time of network latency should be as close to zero as possible, which impacts the user experience of real-time applications, including online gaming, video conferencing, and financial trading.

3) Application Service Testing: This strategy is referred to as concurrent user testing, which assesses the performance of an application service under heavy loads, with multiple users accessing the service simultaneously for a specified amount of time. Ideally, the optimal average load time for accessing an application service is a few seconds, improving the user experience and making it easier for users to access the required information. Concurrent user testing is an important part of software testing because it helps organizations ensure that application services can perform scalability and reliability efficiently under multiple concurrent users.

These strategies could help assess the quality of application services and local network communication between a mobile device and a sandbox server and identify any issues affecting performance.

B. Experimental Setup

To show that the proposed framework is proofed and achieved, all experiments were conducted on a personal computer (PC) with an Intel (R) Core (TM) i5-4570 CPU at 3.20 GHz and 8 GB of RAM as a sandbox server in an isolated environment placed in an indoor office room with a length of 10 meters and a width of 6 meters. The Xiaomi Redmi Note 11 Pro, based on the Android 11 operating system, was used as a mobile device in all of the experimental testing. The Bluetooth receiver comes with Bluetooth version 5.0. Nginx version 1.23.3 was set up as a web server to serve dynamic web pages and web applications written in Python and Java. The backend data was stored in MySQL Server version 8.0.31, which was running on an Ubuntu 22.04 LTS Linux server.

The proposed framework was evaluated based on the three strategies mentioned above. In the Bluetooth signal testing, the mobile device and the Bluetooth receiver were used within the indoor office room to record raw RSSI values at varying distances from 1 to 10 meters. The average was calculated within a certain time frame from the raw RSSI values obtained from multiple samplings. For the network latency testing, the mobile device was set up as a client. The process starts when the client sends a data packet to the sandbox server and measures the time it takes for the sandbox server to respond. Then, the results, which contain the amount of time it takes for every packet to reach its destination and return, are used to calculate the average network latency value. The number of sending packets in each process is limited to 100. For parameter setting, the time interval of each request is set from 0.001 to 1000 milliseconds. The value of the packet size is set at 1000 bytes. For the application service testing, the testing procedure begins by simulating a large number of concurrent users accessing an application service at the same time, where the number of concurrent users ranges from 1 to

1000. The result then displays the average response time for each procedure.

VI. Results and Discussions

To verify the performance of the proposed framework, the experimental results tested on three main strategies are provided.

A. The Result of Bluetooth Signal Testing

The RSSI measurement is considered to determine the strength of Bluetooth signals. It is a metric that represents the relative quality level of a Bluetooth signal received on two Bluetooth-enabled devices. RSSI values also influence the distance range of a reliable Bluetooth connection.

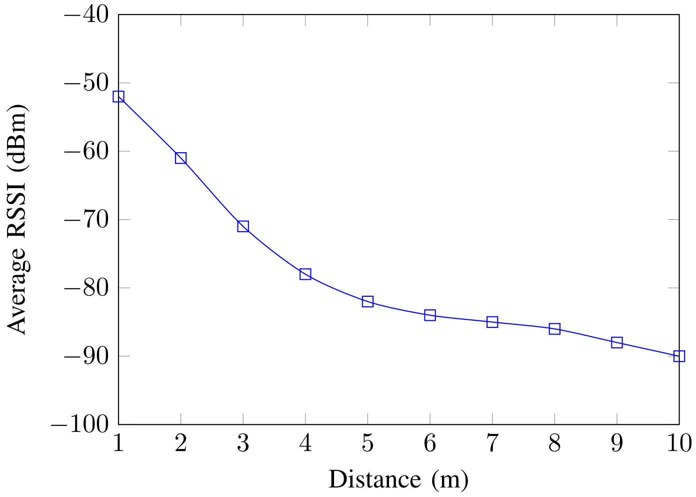


Fig. 6. The relationship between average RSSI and distance

In Figure 6, the graph demonstrates the average of the RSSI values corresponding to the different distances ranging from 1 to 10 meters. From the graph, the average RSSI values are around -50 dBm to -90 dBm. The graph is observed that the average RSSI values significantly increase when the value of the distance range is changed. Therefore, from this experimental result, the proposed framework for Bluetooth signals can support Bluetooth connections within 10 meters stably.

B. The Result of Network Latency Testing

To evaluate the network performance, network latency measurement is an important factor for network communication between a mobile device and a sandbox server.

Figure 7 shows the graph of network latency measurement using the round trip time metric by ranging the time interval from 0.001 to 1000 milliseconds. The graph indicates that the average network latency is relatively constant between 0.05 and 0.06 milliseconds for time intervals of less than one millisecond. The average network latency then gradually increases. Since the average network latency should be close to zero, it is possible to conclude from this experimental result that the proposed framework in the network connection section has the ability to quickly transfer data packets inside the pipe as they travel from client to server and back again.

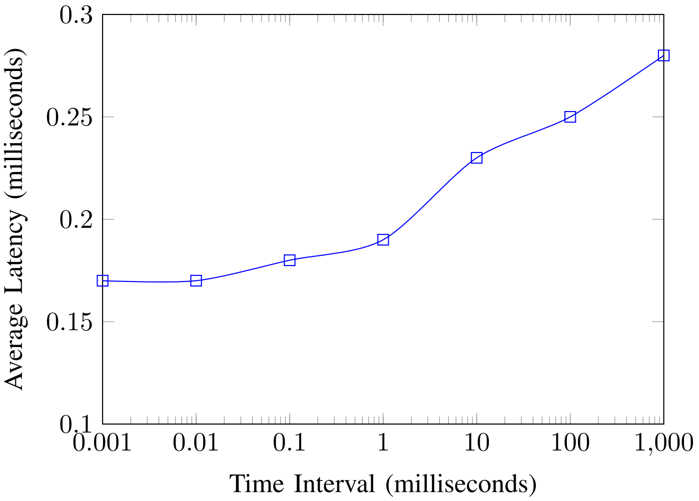


Fig. 7. The average response time with varying the number of concurrent users

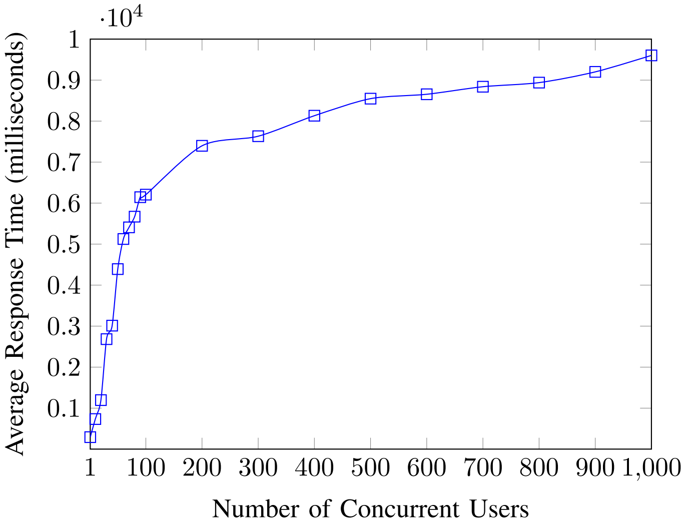


Fig. 8. The average network latency

C. The Result of Application Service Testing

To evaluate the performance of an application service running on a sandbox server under different levels of user concurrency, the metric of concurrent user testing is used.

Figure 8 shows the average response time varying with the number of concurrent users. The graph indicates that the average response time gradually increases when the number of concurrent users grows. Therefore, the number of concurrent users affects the growth of average response times because when multiple users access the application service simultaneously, they compete for the same resources, such as memory, CPU, disk I/O, network bandwidth, and database connections, which can cause system performance to degrade and increase response times. However, since the average load time should be within a few seconds, this experimental result can be interpreted as meaning that the sandbox server has enough capacity to run the application services and handles several users accessing the sandbox server efficiently.

The experimental results and proofs conclude that the proposed framework could be deployed in a production environment because it can efficiently maintain several factors at acceptable levels, including network connectivity, reliability, scalability, and security. This means that the sandbox server is built to do different things and meet the needs of its local users. It can support many users with large amounts of data while keeping sensitive information safe and ensuring the sandbox server is not open to attack.

VII. Use Cases

In this section, a local conference room and shop located in its geographic area are examples of scenarios used to describe the storytelling and planning processes of the proposed framework. For instance, participants who enter a conference room and open a mobile application connected to a local Wi-Fi network will receive real-time information, such as a list of participants with profiles, scheduled programs, presentations, documents, videos, and any other relevant material, which will be reflected immediately in the mobile application for users who are granted access to the network. Moreover, the mobile application can be customized to allow participants to collaborate with each other, meet new friends, make comments, and share information while displaying relevant information, providing a personalized experience for each participant. Therefore, the ability to access information and interact with others through the mobile application can increase user engagement and participation at the conference. In another case, customers who enter a shop and open an application connected to the free Internet access in the venue will receive notifications on their smartphones about promotions and menu items related to the shop. This is beneficial for customers because it provides them with convenient access to valuable information. Moreover, it can also enhance the positive user experience in terms of accessibility and user satisfaction.

VIII. Conclusion and Future Directions

This article proposes a step towards a conceptual framework of the future Web that will help local users access and contribute web content and services on the local and global webs directly and safely through their mobile devices over a local network. The proposed framework is designed and implemented to run locally on a sandbox server located in a specific area to provide security, scalability, and reliability to local users. This would make the Web more secure and private for people while helping users maintain control over their privacy and avoid the risk of hacking attacks or theft. In addition, the proposed framework aims to enhance the Web evolution into an intelligent Web by leveraging the power of Web 3.0 technologies, including AI, NLP, and ML. These innovative technologies make the Web more intelligent because it can understand meaning through data analysis to automatically provide users with highly personalized and appropriate suggestions of items according to environmental contexts such as user profiles, weather, and locations. This way, the owner of their data can have great user experiences. In addition, it is crucial to consider potential legal and regulatory issues that may arise as a result of Web usage and to propose solutions to address these problems.

The future directions of the proposed framework have great potential for use in various smart environments within geographical areas (e.g., smart homes, smart universities, smart cities, and smart industries) [50]–[52]. From a home automation scenario perspective, when a user visits a home as a guest, a smart home enables the user to access and control various smart home devices via a local Wi-Fi network. When entering the home, the user must first authenticate with a proximity-based mechanism to be recognized and permitted. The user is then given access to the network and can control these smart home devices easily and safely, such as turning on the lights, air conditioners, and TVs, using an app on their phone or voice commands. When the user leaves home, he or she can no longer access and control the smart home devices. In the context of smart cities, the proposed framework could offer opportunities to fully integrate into smart cities, towns, or villages through a smartphone application to improve the overall quality of life for citizens and make their lives more efficient and convenient [53]. The mobile application serves targeted information to citizens in real time, enabling them to engage with their city. On the other hand, the proposed framework may enable city commerce to send notifications, promotions, activities, and events, which each business publishes in the mobile application, leading to increased visits. Also, all businesses get direct sales related to the offers sent.

One innovative way of using smart city apps that impact daily life is the use of navigation city tours with augmented reality technology [54], [55]. Mobile applications can help tourists visualize and navigate around cities through the lens of smartphones. The tourists receive information of interest about the history of the city and important places while walking nearby. Therefore, it is particularly useful for those cities that provide tourists with location-based augmented reality experiences using GPS coordinates.

References

[1] S. Thaiprayoon and H. Unger, "Towards personalized context-aware recommendation agent in mobile social networks," in *AFIN 2022, The Fourteenth International Conference on Advances in Future Internet*. IARIA, 2022, pp. 1–8. [Online]. Available: https://www.thinkmind.org/index.php ?view=article&articleid=afin_2022_1_10_40004

[2] K. Nath, S. Dhar, and S. Basishtha, "Web 1.0 to web 3.0-evolution of the web and its various challenges," in *2014 International Conference on Reliability Optimization and Information Technology (ICROIT)*. IEEE, 2014, pp. 86–89.

[3] P. J. Weiser, "The future of internet regulation," *UC Davis L. Rev.*, vol. 43, p. 529, 2009.

[4] T. Seymour, D. Frantsvog, S. Kumar *et al.*, "History of search engines," *International Journal of Management & Information Systems (IJMIS)*, vol. 15, no. 4, pp. 47–58, 2011.

[5] N. Choudhury, "World wide web and its journey from web 1.0 to web 4.0," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 6, pp. 8096–8100, 2014.

[6] D. Kaufmann, A. Kraay, and M. Mastruzzi, "The worldwide governance indicators: Methodology and analytical issues1," *Hague journal on the rule of law*, vol. 3, no. 2, pp. 220–246, 2011.

[7] C. Jaksch, "Digital personal assistants with ai and data protection gdpr & e-privacy-reg," in *Law and Technology in a Global Digital Society: Autonomous Systems, Big Data, IT Security and Legal Tech.* Springer, 2022, pp. 135–161.

[8] M. Hodgson, "A decentralized web would give power back to the people online," Oct 2016. [Online]. Available: https://techcrunch.com/2016/10/09/a-decentralized-web-would-give-power-back-to-the-people-online/

[9] S. Vojir and J. Kucera, "Towards re-decentralized future of the web: Privacy, security and technology development," *Acta Informatica Pragensia*, vol. 10, no. 3, pp. 349–369, 2021. [Online]. Available: https://aip.vse.cz/artkey/aip-202103-0009.php

[10] H. K. M. Al-Chalabi, "Evaluation of a multi-parameter e-learning system using web 3.0 technologies," in *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI).* IEEE, 2021, pp. 1–4.

[11] H. C. Salar, U. Başarmak, and M. E. Sezgin, "Educational integration of the metaverse environment in the context of web 3.0 technologies: A critical overview of planning, implementation, and evaluation," *Shaping the Future of Online Learning: Education in the Metaverse*, pp. 154–173, 2023.

[12] M. Breeding, "Web 2.0? let's get to web 1.0 first." *Computers in Libraries*, vol. 26, no. 5, pp. 30–33, 2006.

[13] K. Zdravkova, M. Ivanović, and Z. Putnik, "Experience of integrating web 2.0 technologies," *Educational Technology Research and Development*, vol. 60, pp. 361–381, 2012.

[14] R. Rudman and R. Bruwer, "Defining web 3.0: opportunities and challenges," *The electronic library*, 2016.

[15] C. Chen, L. Zhang, Y. Li, T. Liao, S. Zhao, Z. Zheng, H. Huang, and J. Wu, "When digital economy meets web 3.0: Applications and challenges," *IEEE Open Journal of the Computer Society*, 2022.

[16] F. A. Alabdulwahhab, "Web 3.0: the decentralized web blockchain networks and protocol innovation," in *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*. IEEE, 2018, pp. 1–4.

[17] K. Pahlavan and P. Krishnamurthy, "Evolution and impact of wi-fi technology and applications: A historical perspective," *International Journal of Wireless Information Networks*, vol. 28, pp. 3–19, 2021.

[18] Y. Guo, S. Zhang, and D. Xiao, "Overview of wi-fi technology," in *2012 International Conference on Computer Application and System Modeling*. Atlantis Press, 2012, pp. 1293–1296.

[19] S. Ahmed, A. N. Sakib, and S. Rahman, "Wpa 2 (wi-fi protected access 2) security enhancement: Analysis," *Global Journal of Computer Science and Technology*, vol. 12, no. 6, pp. 83–89, 2012.

[20] M. Islam and S. Jin, "An overview research on wireless communication network," *Networks*, vol. 5, no. 1, pp. 19–28, 2019.

[21] A. I. Al-Alawi, "Wifi technology: Future market challenges and opportunities," *Journal of computer science*, vol. 2, no. 1, pp. 13–18, 2006.

[22] P. Suresh, J. V. Daniel, V. Parthasarathy, and R. Aswathy, "A state of the art review on the internet of things (iot) history, technology and fields of deployment," in *2014 International conference on science engineering and management research (ICSEMR)*. IEEE, 2014, pp. 1–8.

[23] S. Zeadally, F. Siddiqui, and Z. Baig, "25 years of bluetooth technology," *Future Internet*, vol. 11, no. 9, p. 194, 2019.

[24] S. S. Chadha, M. Singh, and S. K. Pardeshi, "Bluetooth technology: Principle, applications and current status," *International Journal of Computer Science & Communication*, vol. 4, no. 2, pp. 16–30, 2013.

[25] R. Shi, "The world of the bluetooth," in *Third International Conference on Electronics and Communication; Network and Computer Technology (ECNCT 2021)*, vol. 12167. SPIE, 2022, pp. 161–167.

[26] J. Haartsen, M. Naghshineh, J. Inouye, O. J. Joeressen, and W. Allen, "Bluetooth: Vision, goals, and architecture," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 2, no. 4, pp. 38–45, 1998.

[27] S. Chai, R. An, and Z. Du, "An indoor positioning algorithm using bluetooth low energy rssi," in *2016 International Conference on Advanced Materials Science and Environmental Engineering*. Atlantis Press, 2016, pp. 274–276.

[28] M. del Carmen Rodríguez-Hernández and S. Ilarri, "Ai-based mobile context-aware recommender systems from an information management perspective: Progress and directions," *Knowledge-Based Systems*, vol. 215, p. 106740, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705121000034

[29] S. Raza and C. Ding, "Progress in context-aware recommender systems — an overview," *Computer Science Review*, vol. 31, pp. 84–97, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013718302120

[30] Y. Zheng, "Context-aware collaborative filtering using context similarity: An empirical comparison," *Information*, vol. 13, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2078-2489/13/1/42

[31] A. A. S. AlQahtani, H. Alamleh, and B. Al Smadi, "Iot devices proximity authentication in ad hoc network environment," in *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2022, pp. 1–5.

[32] L. Li, X. Zhao, and G. Xue, "A proximity authentication system for smartphones," *IEEE Transactions on Dependable and Secure Computing*, vol. 13, no. 6, pp. 605–616, 2015.

[33] A. Scannell, A. Varshavsky, A. LaMarca, and E. De Lara, "Proximity-based authentication of mobile devices," *International Journal of Security and Networks*, vol. 4, no. 1-2, pp. 4–16, 2009.

[34] J. Zhang, Z. Wang, Z. Yang, and Q. Zhang, "Proximity based iot device authentication," in *IEEE INFOCOM 2017-IEEE conference on computer communications*. IEEE, 2017, pp. 1–9.

[35] A. Kalamandeen, A. Scannell, E. de Lara, A. Sheth, and A. LaMarca, "Ensemble: cooperative proximity-based authentication," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, 2010, pp. 331–344.

[36] L. Bian and H. Holtzman, "Online friend recommendation through personality matching and collaborative filtering," *Proc. of UBICOMM*, pp. 230–235, 2011.

[37] H. Ning, S. Dhelim, and N. Aung, "Personet: Friend recommendation system based on big-five personality traits and hybrid filtering," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 394–402, 2019.

[38] T. Van Le, T. Nghia Truong, and T. Vu Pham, "A content-based approach for user profile modeling and matching on social networks," in *Multi-disciplinary Trends in Artificial Intelligence: 8th International Workshop, MIWAI 2014, Bangalore, India, December 8-10, 2014. Proceedings 8*. Springer, 2014, pp. 232–243.

[39] Z. Deng, B. He, C. Yu, and Y. Chen, "Personalized friend recommendation in social network based on clustering method," in *Computational Intelligence and Intelligent Systems: 6th International Symposium, ISICA 2012, Wuhan, China, October 27-28, 2012. Proceedings*. Springer, 2012, pp. 84–91.

[40] J. Salunke and M. A. Chaudhari, "Implementation of friendbook: a recommendation system for social networks," *Journal of Web Development and Web Designing*, vol. 29, no. 3, pp. 1–7, 2017.

[41] A. B. Barragáns-Martínez, E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. A. Mikic-Fonte, and A. Peleteiro, "A hybrid content-based and item-based collaborative filtering approach to recommend tv programs enhanced with singular value decomposition," *Information Sciences*, vol. 180, no. 22, pp. 4290–4311, 2010.

[42] S. Y. Ho and D. Bodoff, "The effects of web personalization on user attitude and behavior," *MIS quarterly*, vol. 38, no. 2, pp. 497–A10, 2014.

[43] E. Toch, Y. Wang, and L. F. Cranor, "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems," *User Modeling and User-Adapted Interaction*, vol. 22, pp. 203–220, 2012.

[44] A. Liu, Y. Zhang, and J. Li, "Personalized movie recommendation," in *Proceedings of the 17th ACM International Conference on Multimedia*, ser. MM '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 845–848. [Online]. Available: https://doi.org/10.1145/1631272.1631429

[45] A. Yan, C. Dong, Y. Gao, J. Fu, T. Zhao, Y. Sun, and J. Mcauley, "Personalized complementary product recommendation," in *Companion Proceedings of the Web Conference 2022*, ser. WWW '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 146–151. [Online]. Available: https://doi.org/10.1145/3487553.3524222

[46] M. Xin and C. Wan, "Poi recommendation algorithm for mobile social network based on user perference tracking," in *The 2nd International Conference on Computing and Data Science*, 2021, pp. 1–7.

[47] C.-M. Chen, "Intelligent web-based learning system with personalized learning path guidance," *Computers & Education*, vol. 51, no. 2, pp. 787–814, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131507000978

[48] A. Sharma, "Semantic web mining for intelligent web personalization," *Journal of Global Research in Computer Science*, vol. 2, no. 6, pp. 77–81, 2011.

[49] E. Mansour, A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, and T. Berners-Lee, "A demonstration of the solid platform for social web applications," in *Proceedings of the 25th International Conference Companion on World Wide Web*, ser. WWW '16 Companion. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, p. 223–226. [Online]. Available: https://doi.org/10.1145/2872518.2890529

[50] C. Yin, Z. Xiong, H. Chen, J. Wang, D. Cooper, and B. David, "A literature survey on smart cities." *Sci. China Inf. Sci.*, vol. 58, no. 10, pp. 1–18, 2015.

[51] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustainable cities and society*, vol. 38, pp. 697–713, 2018.

[52] M. Batty, K. W. Axhausen, F. Giannotti, A. Pozdnoukhov, A. Bazzani, M. Wachowicz, G. Ouzounis, and Y. Portugali, "Smart cities of the future," *The European Physical Journal Special Topics*, vol. 214, pp. 481–518, 2012.

[53] A. Hoadjli and K. Rezeg, "A scalable mobile context-aware recommender system for a smart city administration," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 36, no. 2, pp. 97–116, 2021. [Online]. Available: https://doi.org/10.1080/17445760.2019.1626855

[54] P. Yagol, F. Ramos, S. Trilles, J. Torres-Sospedra, and F. J. Perales, "New trends in using augmented reality apps for smart city contexts," *ISPRS International Journal of Geo-Information*, vol. 7, no. 12, p. 478, 2018.

[55] S. Kaji, H. Kolivand, R. Madani, M. Salehinia, and M. Shafaie, "Augmented reality in smart cities: applications and limitations," *Journal of Engineering Technology*, vol. 6, no. 1, pp. 28–45, 2018.

TCP Congestion Control Algorithm Estimation by Deep Recurrent Neural Network and Its Application to Web Servers on Internet

Takuya Sawada, Ryo Yamamoto, Satoshi Ohzahata, and Toshihiko Kato
Graduate School of Informatics and Engineering
University of Electro-Communications
Tokyo, Japan
e-mail: sawada@net.lab.uec.ac.jp, ryo-yamamoto@uec.ac.jp, ohzahata@uec.ac.jp, kato@net.lab.uec.ac.jp

Abstract — **Recently, as various types of networks are introduced, a number of Transmission Control Protocol (TCP) congestion control algorithms have been adopted. Since the TCP congestion control algorithms affect traffic characteristics in the Internet, it is important for network operators to analyze which algorithms are used widely in their backbone networks. In such an analysis, a lot of TCP flows need to be handled and so the automatic processing is indispensable. This paper proposes a machine learning based method for estimating TCP congestion control algorithms. The proposed method uses a passively collected packet traces including both data and ACK segments, and calculates a time sequence of congestion window size for individual TCP flows contained in the traces. We use a classifier based on deep recurrent neural network in the congestion control algorithm estimation. As the results of applying the proposed classifier to ten congestion control algorithms, we obtained high accuracy of classification compared with our previous work using recurrent neural network with one hidden layer. This paper also describes the results of applying the classifier to popular web servers for checking the distribution of congestion control algorithms in the real world.**

Keywords — TCP; Congestion Control; Deep Recurrent Neural Network.

I. INTRODUCTION

This paper is an extension of our previous paper [1], which was presented at the IARIA conference EMERGING 2022.

Along with the introduction of various types of networks, such as a long-haul high speed network and a wireless mobile network, a number of TCP congestion control algorithms have been designed, implemented, and widely spread [2]. Since the congestion control was introduced [3], a few algorithms, such as TCP Tahoe [4], TCP Reno [4], and NewReno [5], have been used commonly for some decades. Recently, new algorithms have been introduced and deployed. For example, HighSpeed TCP [6], Scalable TCP [7], BIC TCP [8], CUBIC TCP [9], and Hamilton TCP [10] are designed for high speed and long delay networks. TCP Westwood+ [11] is designed for lossy wireless links. While those algorithms are based on packet losses, TCP Vegas [12] triggers congestion control against an increase of Round-Trip Time (RTT). TCP Veno [13] combines loss based and delay based approaches in such a way that congestion control is triggered by packet losses but the delay determines how to grow the congestion window (cwnd). In 2016, Google proposed a new algorithm called TCP BBR (Bottleneck Bandwidth and Round-trip propagation time) [14] to solve problems mentioned by conventional algorithms.

Since TCP traffic is a majority in the Internet traffic and the TCP congestion control algorithms characterize the behaviors of individual flows, the estimation of congestion control algorithms for TCP traffic is important for network operators. It can be used in various purposes such as the traffic trend estimation, the planning of Internet backbone networks, and the detection of malicious flows violating congestion control algorithms.

The approaches for congestion control algorithm estimation are categorized into the passive approach and the active approach. The former estimates algorithms from packet traces passively collected in the middle of networks. In the latter approach, a test system communicates with a target system with a specially designed test sequence. Although the active approach is capable to identify various congestion control algorithms proposed so far, it does not fit the algorithm estimation of real TCP flows. On the other hand, generally speaking, the detecting capability of passive approaches is relatively low comparing with the active approach.

Previously, we proposed a passive method that can estimate a number of congestion control algorithms [15][16]. In this proposal, we focused on the relationship between the estimated congestion window size and its increment. Their relationship is indicated as a graph and the congestion control algorithm is estimated based on the shape of the graph. Our proposal succeeded to identify eight congestion control algorithms implemented in the Linux operating system, including recently introduced ones.

However, the identification is performed manually by human inspectors, and so it is difficult to deal with a large number of TCP flows. So, we proposed a machine learning based classifier estimating the TCP congestion control algorithms using TCP packet traces in two steps [17][1]. In our first trial [17], we used a conventional Recurrent Neural Network (RNN) with one hidden layer. From a packet trace, we estimate the relationship of cwnd values and their increment with congestion control algorithm labels, and apply the results to a RNN classifier for training. Using the RNN classifier, we estimate the algorithms for other packet traces. We obtained a relatively good estimation result from the RNN classifier, but we could not classify similar algorithms, such as TCP Reno and Vegas. In our second trial [1], we proposed a revised version of machine learning classifier for automatic estimation of congestion control algorithms. Here, we adopted a Deep Recurrent Neural Network (DRNN) with multiple hidden layers. We also applied a hyper parameter tuning for the classifier. We picked up ten congestion control algorithms mentioned above and showed that our new approach can estimate those algorithms better than our first trial.

This evaluation of our approach was performed in a simple in-lab test network environment. Applying our approach to servers in the Internet will be effective to show its effectiveness and to investigate the trends of TCP congestion control algorithms in the real world. We estimated the congestion control algorithms of 20,000 web servers listed in Alexa Top Sites 1,000,000 offered by Alexa Traffic Rank [18]. This paper also shows the results of this estimation.

The rest of this paper is organized as follows. Section II gives some background information including the conventional studies on the congestion control estimation and the machine learning applied for the network areas. Section III describes the proposed method and Section IV gives the performance evaluation results. Section V discusses the results of the estimation of Internet web servers by the proposed method. In the end, Section VI concludes this paper.

II. Backgrounds

A. Studies on TCP Congestion Control Algorithm Estimation

The proposals on the passive approach in the early stage [19-21] estimate the internal state and variables, such as cwnd and ssthresh (slow start threshold), in a TCP sender from bidirectional packet traces. They emulate the TCP sender's behavior from the estimated state/variables according to the predefined TCP state machine. But, they considered only TCP Tahoe, Reno and New Reno and did not handle any of recently introduced algorithms. Oshio et al. [22] proposed a method to discriminate one out of two different TCP congestion control algorithms randomly selected from fourteen algorithms implemented in the Linux operating system. This method keeps track of changes of cwnd from a packet trace and to extract several characteristics, such as the ratio of cwnd being incremented by one packet. Although this method targets all of the modern congestion control algorithms, they assumed that the discriminator knows two algorithms contained in the packet trace.

The active approaches, on the other hand, are able to identify more TCP congestion control algorithms including those introduced recently. Yang et al. [23] proposed a tool called CAAI (TCP Congestion Avoidance Algorithm Identification), which could identify recent TCP congestion control algorithms at first. It makes a web server send 512 data segments under the controlled network environment, and observes the number of data segments contiguously transmitted. From those results, it estimates the window growth function and the decrease parameter to determine the congestion control algorithm. Mishra et al. [24] proposed another active approach based tool called Gordon. It makes a web server send several data segments and causes a packet loss intentionally. By analyzing the retransmission and the following congestion avoidance sequence, it obtains cwnd values. It then estimates congestion control algorithms based on the shape of cwnd time sequence graph, the increase of cwnd, and the back-off at packet loss. It could estimate the recent congestion control algorithms and was applied to the 20,000 web servers listed in the Alexa Top Sites.

Our previous proposals [15][16] estimated cwnd in RTT intervals from bidirectional packet traces, in the similar way with the other methods. Different from other methods, we focused on the incrementing situation of estimated cwnd values. From the definition of individual congestion control algorithms, the graph of cwnd increments vs. cwnd has their characteristic forms. For example, in the case of TCP Reno, the cwnd increment is always one segment. In the case of CUBIC TCP, the graph of cwnd increment follows a $\sqrt[3]{cwnd^2}$ curve. In this way, we proposed a way to discriminate eight congestion control algorithms in the Linux operating system.

B. Studies on Application of Machine Learning to TCP

Recently, several papers focus on applying the machine learning to TCP analysis. Edalat et al. [25] proposed a method to estimate RTT using the fixed-share approach from measured RTT samples. Mirza et al. [26] estimated the future throughput of TCP flow using the support vector regression from measured available bandwidth, queueing delay, and packet loss rate. Chung et al. [27] proposed a machine learning based multipath TCP scheduler based on the radio strength in wireless LAN level, wireless LAN data rate, TCP throughput, and RTT with access point, by the random decision forests.

III. Proposed Method

A. Estimation of cwnd values at RTT interval

In the passive approach, packet traces are collected at some monitoring point inside a network. So, the time associated with a packet is not the exact time when the node focused sends/receives the packet. Our scheme adopts the following approach to estimate cwnd values at RTT intervals using the TCP time stamp option.

- Pick up an ACK segment in a packet trace. Denote this ACK segment by *ACK1*.
- Search for the data segment whose TSecr (time stamp echo reply) is equal to TSval (time stamp value) of *ACK1*. Denote this data segment by *Data1*.
- Search for the ACK segment that acknowledges *Data1* for the first time. Denote this ACK segment by *ACK2*. Denote the ACK segment prior to *ACK2* by *ACK1'*.
- Search for the data segment whose TSecr is equal to TSval of *ACK2*. Denote this data segment by *Data2*.

From this result, we estimate a cwnd value at the timing of receiving *ACK1* as in (1).

$$cwnd = \left\lfloor \frac{seq\ in\ Data2 - ack\ in\ ACK1'}{MSS} \right\rfloor \text{ (segments)} \qquad (1)$$

Here, *seq* means the sequence number, *ack* means the acknowledgment number of TCP header, and *MSS* is the maximum segment size (MSS). $\lfloor a \rfloor$ is the truncation of a.

Figure 1 shows an example of cwnd estimation. In this figure, MSS is 1024 byte. Data segments are indicated by

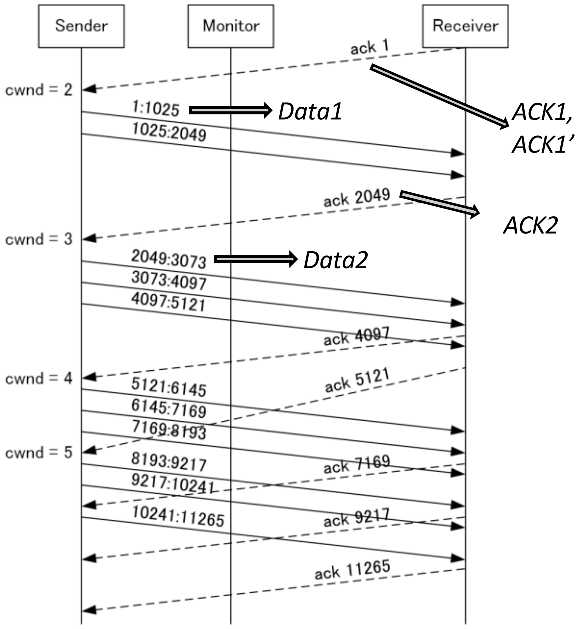


Figure 1. Example of cwnd estimation.

```
Algorithm 1
function Normalize (E, InputLength)
    S <- DivideAtLoss(E)
    Delete(S[0])
    S <- SortBySequenceLength(S)
    for t = 0 to Len(S) − 1 do
        S <- MinMaxNormalization(S)
    end for
    I <- Array(Len(S))
    for t = 0 to Len(S) − 1 do
        I[t] <- Array(InputLength)
        for u = 0 to InputLength − 1 do
            SurjectiveMap <- InputLength/Len(S[t])
            Index <- Trunc(u / SurjectiveMap)
            I[t][u] <- S[Index]
        end for
    end for
    return I
end function
```

Figure 2. Selection/normalization algorithm.

solid lines with "sequence number : sequence number + MSS." ACK segments are indicated by dash lines with acknowledgment number. When "ack 1" is picked up, data segment "1:1024" is focused on as *Data1* above. ACK segment "ack 2049" responding the data segment corresponds to *ACK2*. The ACK segment before this ACK segment (*ACK1'* above) is "ack 1" again. *Data2* in this case is "2049:3073." So, the estimated cwnd is (2049 – 1)/1024 = 2. Similarly, for the following two RTT intervals, the estimated RTT values are (5121 – 2049)/1024 = 3 and (10241 –5121) /1024= 5.

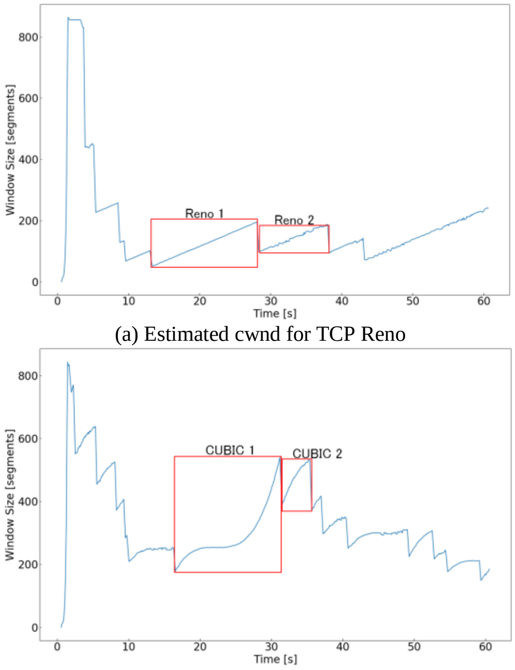
B. *Selection and Normalization of Input Data to Classifier*

When a packet is lost and retransmitted, cwnd is decreased. In order to focus on the cwnd handling in the congestion avoidance phase, we select a time sequence of cwnd between packet losses. We look for a part of packet trace where the sequence number in the TCP header keeps increasing. We call this duration without any packet losses *non-loss duration*. We use the time variation of estimated cwnd values during one non-loss duration as an input to the classifier. However, the length of non-loss duration differs for each duration, and the range of cwnd values in a non-loss duration also differs from one to another. So, we select and normalize the time scale and the cwnd value scale for one non-loss duration.

The algorithm for selecting and normalizing input to classifier is given in Figure 2. In this algorithm, the input E is as time sequence of cwnd values estimated from one packet trace. The input *InputLength* is a number of samples in one input to the classifier. In this paper, we used 128 as *InputLength*. This is because we think that the cwnd vs time curve can be drawn by 128 points. In the beginning, the time sequence of cwnd is divided at packet losses, and the divided sequences are stored in a two dimensional array *S*. Next, the first sequence *S*[0] is removed, because we focus only on the congestion avoidance phase. Then *S* is reordered according to the length of cwnd sequence. Then the cwnd values for one sequence *S*[t] are normalized between 0 and 1. The normalization is performed in the following way.

Let $w_{max}[t] = \max(S[t][u])$
for $u = 0 \cdots \mathrm{Len}(S[t]) - 1$, and
$w_{min}[t] = \min(S[t][u])$
for $u = 0 \cdots \mathrm{Len}(S[t]) - 1$.
Each cwnd value in S[t] is normalized by
$S[t][u] \leftarrow \frac{S[t][u] - w_{min}[t]}{w_{max}[t] - w_{min}[t]}$.

After that, the cwnd values are resampled into the number of *InputLength* (128 in this paper). This is done by the loop



(a) Estimated cwnd for TCP Reno

(b) Estimated cwnd for CUBIC TCP

Figure 3. Examples of cwnd estimation.

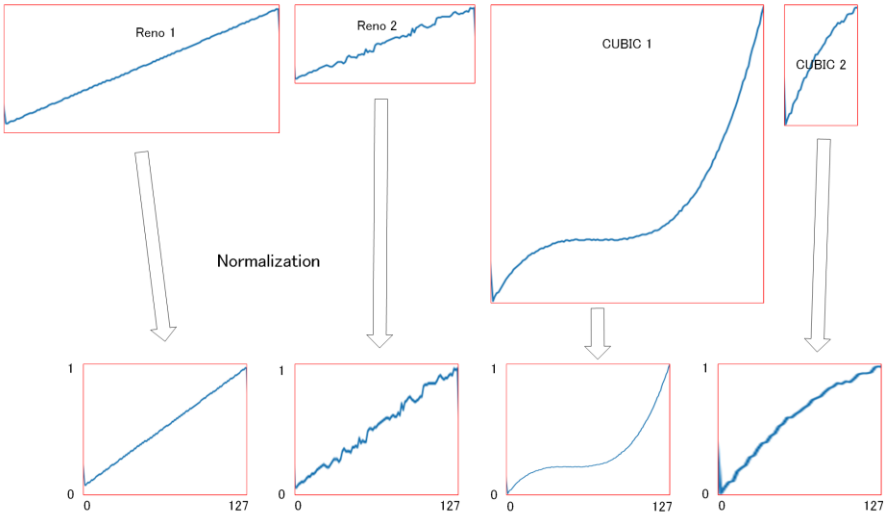


Figure 4. Examples of normalization.

between step 11 and step 15. As a result, a cwnd sequence in *S*[t] is converted to an array *I*[t] with 128 elements. By this algorithm, all of the time sequences of cwnd values are the arrays with 128 elements whose value is between 0 and 1.

Figure 3 shows some examples of cwnd estimation. Figure 3 (a) and (b) show the estimated cwnd time sequences for TCP Reno and CUBIC TCP, respectively. We focus on the non-loss durations as described above. Reno 1, Reno 2, CUBIC 1, and CUBIC 2 in the figure are examples. The size of these sequences differ from each other, both for the time scale and the scale of cwnd. Therefore, it is necessary to normalize these sequences.

Figure 4 shows the results of the normalization for the examples shown in Figure 3. Different scale of cwnd time sequences are transformed into a canonical form with 128 samples in the range of 0 through 1.

C. *DRNN Based Classifier for Congestion Control Algotithm Estimation*

We used DRNN for constructing the classifier, which has three hidden layers and whose output layer defines the TCP congestion control algorithms. Among the RNN technologies, we pick up the long short-term memory mechanism [28], which was proposed to handle a relatively long time sequence of data. The input is a normalized time sequence of cwnd as described above, with using labels of congestion control algorithms represented by one-hot vector.

In our previous work, we selected the hyper parameters given in Table I. In the work presented in this paper, we select the hyper parameter ranges shown in Table II. The input length is the same as that of the previous work. We use three hidden layers and the number of neurons are as specified in the table. As for the optimizer, the learning rate and the weight decay, we propose the alternatives shown in the table. We perform the hyper parameter tuning based on the target area in this table.

In the training of the classifier, we use the mini-batch method, which selects a specified number of inputs randomly from the prepared training data. The mini-batch size will be determined for individual training. The training will be continued until the result of the loss function becomes smaller than the learning rate.

TABLE I. HYPER PARAMETERS OF CLASSIFIER IN OUR PREVIOS WORK.

Parameter	Value
Input Length	128
Hidden Layers	1
Hidden Neurons	512
Optimizer	Adam
Learning Rate	2×10^{-4}

TABLE II. HYPER PARAMETER RANGES IN THIS WORK.

Parameter	Value
Input Length	128
Hidden Layers	3
Hidden Neurons	1st./2nd: 512, 3rd: 256
Optimizer	Adam or MomentumSGD
Learning Rate	$[10^{-5}, 10^{-1}]$
Weight Decay	$[10^{-10}, 10^{-3}]$

IV. EXPERIMENTAL RESULTS

A. *Experimental Setup*

Figure 5 shows the experimental configuration for collecting time sequence of cwnd values. A data sender, a data receiver, and a bridge are connected via 100 Mbps Ethernet links. In the bridge, 50 msec delay for each direction is inserted. As a result, the RTT value between the sender and the receiver is 100 msec. In order to generate packet losses that will invoke the congestion control algorithm, packet losses are inserted randomly at the bridge. The average packet loss ratio is 0.01%. The data transfer is performed by use of iperf3 [29], executed in both the sender

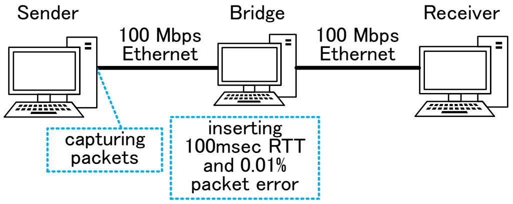


Figure 5. Experiment configuration.

and the receiver. The packet traces are collected by use of tcpdump at the sender's Ethernet interface. We use the Python 3 dpkt module [30] for the packet trace analysis. We changed the congestion control algorithm at the sender by use of the sysctl command provided by the Linux operating system.

The targeted congestion control algorithms are TCP Reno, HighSpeed TCP, BIC TCP, CUBIC TCP, Scalable TCP, Hamilton TCP, TCP Westwood+, TCP Vegas, TCP Veno, and BBR. We collected more than 1,500 samples for individual algorithms, and prepared 1,000 samples as training data, 250 samples as verifying data, and 250 samples as test data.

B. Results of Congestion Control Algorithm Estimation

First, we re-evaluated the performance of our previous approach. The result is shown in Figure 6. The total accuracy for ten congestion control algorithms was 42.8%, which is rather worse than the result described in our previous paper [17]. This means that our previous classifier will depend largely on the prepared training data.

So, we applied the same training data and verifying data for a DRNN based classifier with three hidden layers and selected optimal values for the hyper parameters mentioned in Table II. We tried to look for optimal values 100 times by the mini-batch method using 256 as the mini-batch size. Table III shows the values of hyper parameters obtained by this tuning.

TABLE III. TUNED UP HYPER PARAMETER VALUES.

Parameter	Value
Optimizer Learning Rate Weight Decay	Adam 0.0015967736 2.967486×10^{-8}

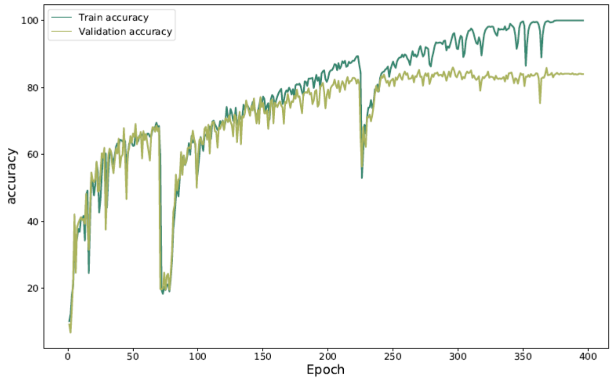


Figure 7. Learning curve for ten congestion control algorithms.

Figure 7 shows the learning curve for ten congestion control algorithms using the DRNN based classifier with the selected hyper parameter values. The horizontal axis of this figure indicates the epoch, which is the number of training and verifying trials. The vertical axis indicates the accuracy for the training process and the verifying process. The blue line is the accuracy for the training process and the green line is for the verification process. This result shows that the classifier learns the model for estimating congestion control algorithms. Figure 8 shows the confusion matrix for this experiment. By comparing Figures 6 and 8, we can conclude that the DRNN based classifier estimates the congestion control algorithms much better than our previous classifier.

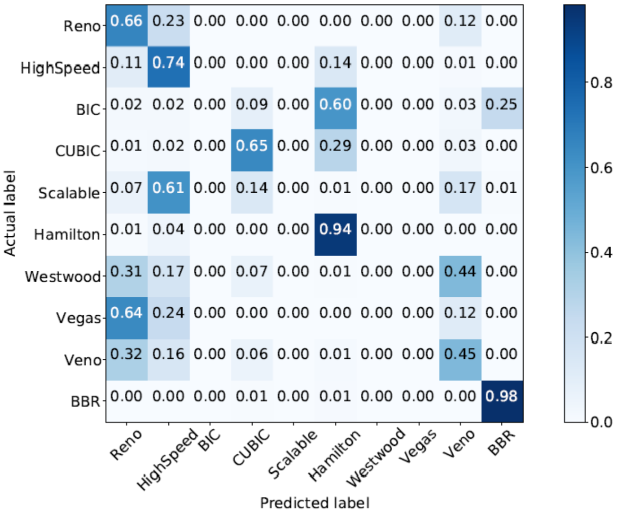


Figure 6. Confusion matrix for previous approach.

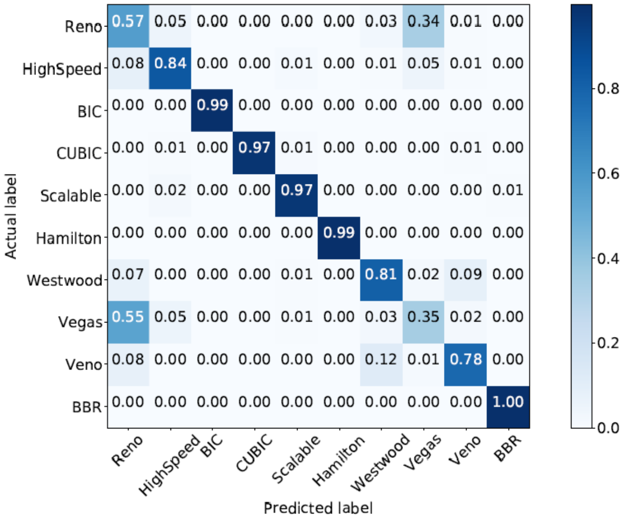


Figure 8. Confusion matrix for this approach.

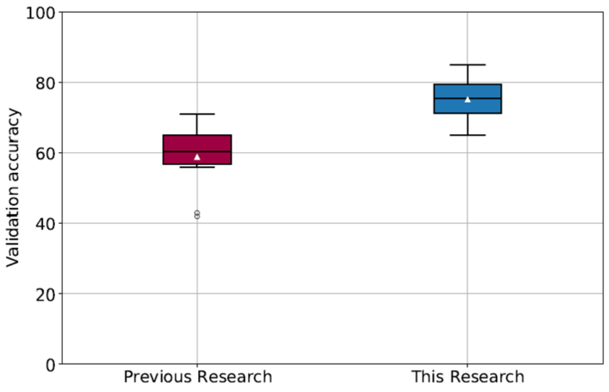


Figure 9. Generalization accuracy of previous work and this work.

The total accuracy was 82.9%, which is higher than that of our previous work. The only problem is that it still confuses TCP Reno and TCP Vegas. Further studies are required.

As the last analysis, we evaluated the generalization accuracy for our previous work and this work using the 10-fold cross-validation. We divided the training data into ten folds, and selected one fold for the validation and used the rest folds for the training. Figure 9 shows the result. The vertical axis is the validation accuracy. In our previous classifier, the validation accuracy sometimes drops to 40%, although it goes up 70%. On the other hand, our new classifier provides 70% through 80% accuracy stably.

V. Congestion Control Algorithm Estimation of Web Servers on Internet

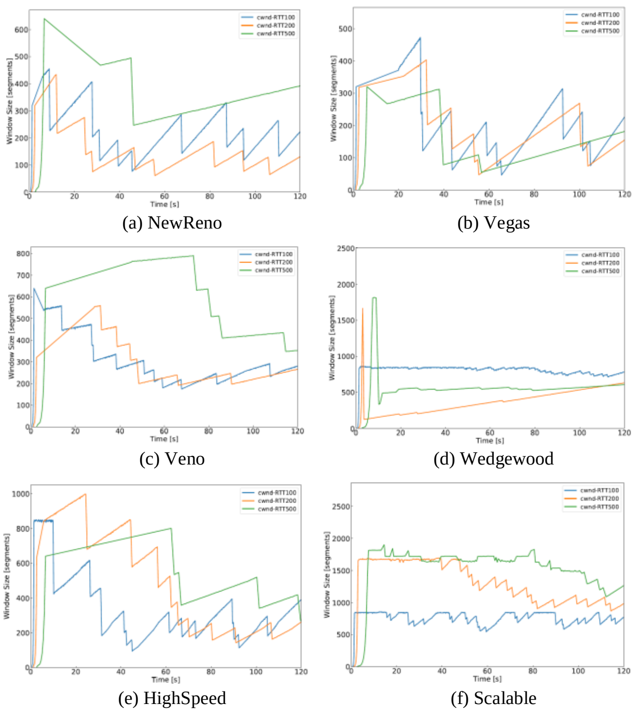
A. Steps

The next work is to apply our DRNN based classifier for estimating congestion control algorithms used by web servers on the Internet. Basically, we will use the classifier described in the previous sections. That is, the classifier is trained in the in-lab test network and applied to the estimation using data obtained from outside web servers.

For this purpose, we decided to take the following steps. The first step is to redesign our classifier to fit the estimation for real servers. Again, it should be mentioned that the training and testing of the redesigned classifier are performed using the in-lab network. This is because we do not know the congestion control algorithms of web servers on the Internet. The second step is to estimate the algorithms used by real web servers using our redesigned classifier.

B. Redesign of Classifier

As mentioned in Section II, Mishra et al. tried to estimate congestion control algorithms used by popular web servers given in the Alexa Top Site list based on an active approach. The measurements were made between July and October in 2019. This measurement found that thirteen algorithms are adopted by the target web servers. They include TCP Hybra [31], YeAH TCP [32], and TCP Illinois [33] besides ten algorithms mentioned in the previous section. It should be mentioned that CTCP (Compound TCP) [34], which was said to be adopted in Windows OS, is handled as TCP Illinois. So, we needed to extend our classifier so as to include the added three congestion control mechanisms.



(a) NewReno (b) Vegas

(c) Veno (d) Wedgewood

(e) HighSpeed (f) Scalable

Figure 10. Estimated cwnd vs. time for different RTT values.

Another redesign point is RTT. The experiment shown in the previous section used 100 msec RTT inserted at the bridge. However, the real web server accesses take various values of RTT. So, for training our classifier in various RTT situations, we adopted 100 msec, 200 msec, and 500 msec as RTT values. Specifically, the delay of the half of individual RTT values are inserted for each direction at the bridge. The average packet loss rate inserted at the bridge is 0.01% for all cases. It should be mentioned that these values are constant ones throughout one experiment run.

Figure 10 shows some examples of estimated cwnd values. The blue, orange, and green lines indicate cwnd values for 100 msec, 200 msec, and 500 msec, respectively. By introducing different RTT values, we could obtain different behaviors of cwnd. Especially, this helped to discriminate NewReno, Vegas, Veno, Westwood, and HighSpeed, for which the classifier in the previous section was suffered from erroneous estimation.

We collected more than 2,400 samples for individual congestion control algorithms. We prepared 2,000 samples as training data and 400 samples as test data. Table IV shows the selected hyper parameter values for the redesigned classifier. LAMB (Layer-wise Adaptive Moments optimizer for Bach training) [35], adopted as the optimizer, is a training method for deep neural networks with large batch size.

Figure 11 shows the learning curve of the redesigned DRNN classifier for thirteen congestion control algorithms. The horizontal axis is the epoch and the vertical axis is the accuracy, similarly to Figure 7. This result shows that the redesigned classifier also learns the model for estimating the

TABLE IV. TUNED UP HYPER PARAMETER VALUES OF REDSIGNED CLASSIFIER.

Parameter	Value
Optimizer	LAMB
Learning Rate	0.01866181607680214
Weight Decay	$3.094144144895362 \times 10^{-7}$

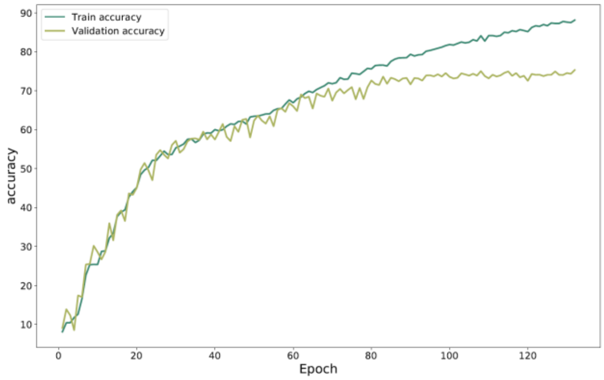


Figure 11. Learning curve for thirteen congestion control algorithms.

targeted algorithms. More specifically, the accuracies for training and validation become different from 80 epochs, and so the trained model will keep the status around that area.

Figure 12 shows the confusion matrix of the redesigned classifier. The total accuracy is 73.2 %, which is a little lower than the result of Figure 8, but it will be high enough to estimate a server's congestion control algorithm from a passively collected packet trace. The misestimation between Reno and Vegas decreased compared with the result of Figure 8. This is a result of introducing different RTT values. The confusion among Scalable, Illinois, and YeAH, the latter two are of which introduced newly, is one reason of decreasing the total accuracy. We used this classifier to categorize the congestion control algorithms used by actual web servers over the Internet.

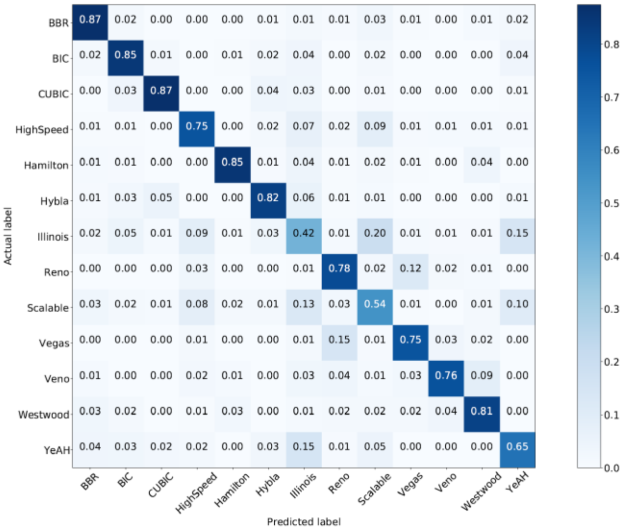


Figure 12. Confusion matrix for thirteen congestion control algorithms.

C. Results of Estimation for Real Web Servers

We applied our redesigned classifier to the web servers listed in the Alexa Top Site list. We examined the list in December 2020, and it should be noted that this service was terminated at May 2022. We selected 20,000 popular web servers in the list, and communicated with each server for 30 seconds through three minutes to obtain cwnd time sequences. Then, we applied the results to our classifier to estimate the congestion control algorithm of the individual server.

Figure 13 shows the distribution of estimated congestion control algorithms for 20,000 servers. TCP BBR and CUBIC TCP are popular algorithms, which occupy 29.9 % and 40.7 %, respectively. TCP Illinois and YeAH TCP follow them. As described above, it is possible that Illinois includes CTCP. On the other hand, NewReno, which used to be dominant, is not used any more.

Figure 14 shows the distribution of estimated algorithms for top 100 servers. In this case, TCP BBR occupies 41.0 % in the first place. CUBIC TCP is in the second place and occupies 26.0 %. The third is TCP Illinois. The reason that BBR and CUBIC change the place in top 100 servers will be that a lot of video distribution sites and Google related sites are included in top 100 servers. This means that the large portion of Internet TCP traffic will include BRR and CUBIC.

Several studies conducted the TCP congestion control algorithm classification in the past. Padhye et al. [36] proposed an active method called TBIT (TCP Behavior Inference Tool), in which a client test tool communicates with a web server with dropping data segments intentionally. It was applied to several web servers during 2000 and 2001. Medina et al. [37] measured the TCP variants by use of TBIT in 2004. Yang et al. applied their tool, CAAI, to several web servers to determine their TCP variants in 2011 [23]. Most recently, Mishra et al. applied their tool, Gordon, to the Alexa Top 20,000 servers as mentioned above [24].

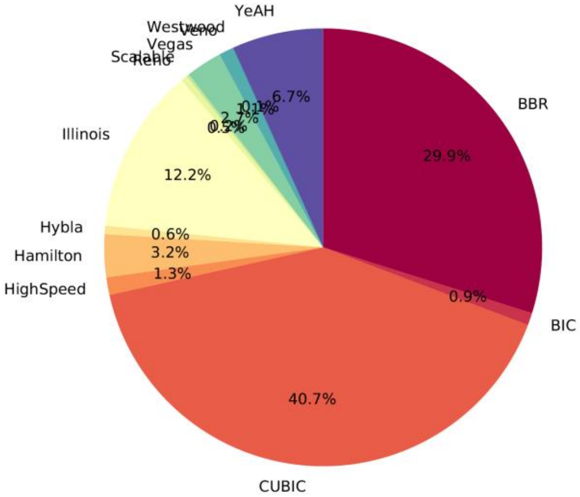


Figure 13. Distribution of estimated congestion control algorithms for AlexaTop-20000.

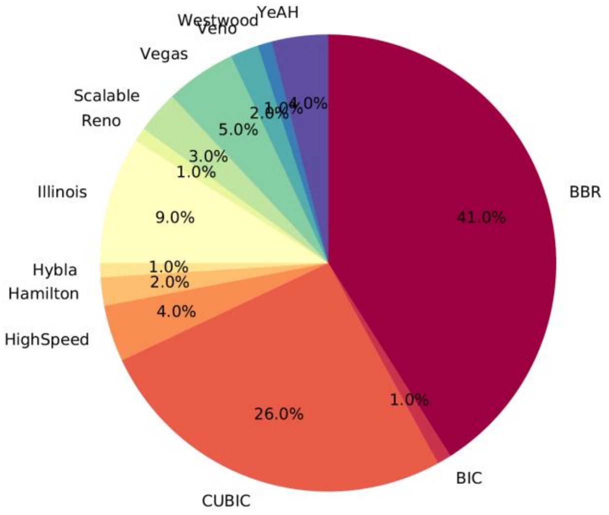


Figure 14. Distribution of estimated congestion control algorithms for AlexaTop-100.

Table V shows the results of classifying TCP variants by the previous studies and by us. This table is a modified version of the table given in [24] and our results are added to the modified one. Each entry contains a TCP variant and its rate/number. Variants are categorized into loss-based, delay-based, and rate-based ones. "Unknown etc." indicates the case that the method could not estimate the algorithm.

In the time frame of 2000s, only Reno-like TCP variations are identified. In the beginning of 2000s, NewReno was more popular than Reno and Tahoe [36]. In the middle of 2000s, the variants other than Reno types were increasing, that is, the unknown variants occupied 67% [37].

At the beginning of 2010s, a variety of TCP congestion control algorithms were introduced [23]. The Reno-type (AIMD) variants decreased, and instead, CUBIC was the most popular variant, and BIC, HighSpeed, and CTCP were adopted by some servers.

At the end of 2010s, CUBIC and BBR were top two variants [24]. Besides them, a few web servers were using Illinois including CTCP, YeAH, Hamilton, and Vegas/Veno. Our experiment used the measurement results performed in December 2020, and so the used web servers may be different from those used in the Gordon measurement, and the same server may change its algorithm. So, the results of ours and the Gordon's are a little different from each other. However, the trends of both results are similar. The most popular algorithms are CUBIC and BBR, and Illinois and YeAH follow. A certain amount of servers keep using the delay-based algorithms such as Vegas and Veno.

VI. Conclusions

This paper is an extended version of our conference paper [1] presented in IARIA EMERGING 2022. This paper provides two contributions.

The first, which is the contribution provided by the conference paper, is that we showed a result of TCP congestion control algorithm estimation using a Deep Recurrent Neural Network (DRNN) based classifier. From packet traces including both data segments and ACK segments, we derived a time sequence of cwnd values at RTT intervals without any packet retransmissions. By ordering the time sequences and normalizing in the time dimension and the cwnd value dimension, we obtained the input for the

TABLE V. CLASSIFICATIONS OF TCP VARIANTS IN SEVERAL STUDIES

	2001 [36]		2004 [37]		2011 [23]		2019 [24]		2020 (our results)	
Loss-based	NewReno	35% (1,571)	NewReno	25% (21,266)	AIMD	12.46% (623)	NewReno	0.80% (160)	NewReno	0.49% (98)
	Reno	21% (945)	Reno	5% (4,115)			Reno	-		
	Tahoe	26% (1,211)	Tahoe	3% (2,164)			Tahoe	-		
					CUBIC	22.30% (1,115)	CUBIC	30.70% (6,139)	CUBIC	40.69% (8,139)
					BIC	10.62% (531)	BIC	0.90% (181)	BIC	0.90% (181)
					HighSpeed	7.38% (369)	HighSpeed	in NewReno	HighSpeed	1.30% (260)
					Scalable	1.38% (69)	Scalable	0.20% (39)	Scalable	0.20% (39)
									Hybra	0.62% (124)
Delay-based	-	-	-	-	Vegas	1.16% (58)	Vegas	2.82% (564)	Vegas	2.65% (531)
					Westwood	2.08% (104)	Westwood	0% (0)	Westwood	0.05% (10)
					Illinois	0.56% (28)	Illinois	5.74% (1,148)	Illinois	12.23% (2,445)
					Veno	0.90% (45)	Veno	in Vegas	Veno	1.07% (214)
					YeAH	1.44% (72)	YeAH	5.81% (1,162)	YeAH	6.71% (1,342)
					Hamilton	0.36% (18)	Hamilton	2.28% (560)	Hamilton	3.15% (630)
					CTCP	6.68% (334)	CTCP	in Illinois		
Rate-based	-	-	-	-	-	-	BBR	17.75% (3,550)	BBR	29.84% (5,967)
							BBR G1.1	0.84% (167)		
							Akamai CC	5.51% (1,103)		
Unknown etc.		18% (822)		67% (56,479)		32.84% (1,642)		26.14% (5,227)		-
Total		100% (4,550)		100% (84,394)		100% (5,000)		100% (20,000)		100% (20,000)

note: BBR G1.1 indicates the Google dialect of BBR, and Akamai CC is a rate-based congestion control used by the Akamai content delivery network.

DRNN classifier. As the results of applying the proposed classifier for ten congestion control algorithms implemented in the Linux operating system, we showed that the DRNN based classifier can estimate ten algorithms effectively, with a problem that TCP Reno and TCP Vegas are difficult to discriminate. This result is much better than our previous classifier that used a simple recurrent neural network.

The second is an original contribution newly provided in this paper. We applied our DRNN based classifier to the estimation of congestion control algorithms used by 20,000 frequently accessed web servers identified by the Alexa Top Sites list. For this purpose, we redesigned our classifier so as to handle TCP Hybra, YeAH TCP, and TCP Illinois variants, and to include the situations with different RTT values. We confirmed that the redesigned classifier estimates thirteen variants with the accuracy of 73.2 %. Then we applied our classifier to 20,000 web servers listed in Alexa Top Sites. The results were that the top two variants were CUBIC and BBR, and that Illinois (including CTCP) and YeAH followed them. The results have the similar trends with the study conducted by Mishra et al. in 2019 [24], and this indicates that our estimation will be reasonable.

References

[1] T. Sawada, R. Yamamoto, S. Ohzahata, and T. Kato, "Estimation of TCP Congestion Control Algorithms by Deep Recurrent Neural Network," Proc. IARIA EMERGING 2022, pp. 19-24, 2022.

[2] A. Afanasyev, N. Tilley, P. Reiher, and L. Kleinrock, "Host-to-Host Congestion Control for TCP," IEEE Commun. Surveys & Tutorials, vol. 12, no. 3, pp. 304-342, 2010.

[3] V. Jacobson, "Congestion Avoidance and Control," ACM SIGCOMM Comp. Commun. Review, vol. 18, no. 4, pp. 314-329, 1988.

[4] W. R. Stevens, "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algotithms," IETF RFC 2001, 1997.

[5] S. Floyd, T. Henderson, and A. Gurtov, "The NewReno Modification to TCP's Fast Recovery Algorithm," IETF RFC 3728, 2004.

[6] S. Floyd, "HighSpeed TCP for Large Congestion Windows," IETF RFC 3649, 2003

[7] T. Kelly, "Scalable TCP: Improving Performance in High-speed Wide Area Networks," ACM SIGCOMM Comp. Commun. Review, vol. 33, no. 2, pp. 83-91, 2003.

[8] L. Xu, K. Harfoush, and I. Rhee, "Binary increase congestion control (BIC) for fast long-distance networks," Proc. IEEE INFOCOM 2004, vol. 4, pp. 2514-2524, 2004.

[9] S. Ha, I. Rhee, and L. Xu, "CUBIC: A New TCP-Friendly High-Speed TCP Variant," ACM SIGOPS Operating Systems Review, vol. 42, no. 5, pp. 64-74, 2008.

[10] D. Leith and R. Shorten, "H-TCP: TCP for high-speed and long distance networks," Proc. Int. Workshop on PFLDnet, pp. 1-16, 2004.

[11] L. Grieco and S. Mascolo, "Performance evaluation and comparison of Westwood+, New Reno, and Vegas TCP congestion control," ACM Computer Communication Review, vol. 34, no. 2, pp. 25-38, 2004.

[12] L. Brakmo and L. Perterson, "TCP Vegas: End to End Congestion Avoidance on a Global Internet," IEEE J. Selected Areas in Commun., vol. 13, no. 8, pp. 1465-1480, 1995.

[13] C. Fu and S. Liew, "TCP Veno: TCP Enhancement for Transmission Over Wireless Access Networks," IEEE J. Sel. Areas in Commun., vol. 21, no. 2, pp. 216-228, 2003.

[14] N. Cardwell, Y. Cheng, C. S. Gumm, S. H. Yeganeh, and V. Jacobson, "BBR: Congestion-Based Congestion Control," ACM Queue vol. 14 no. 5, pp. 20-53, 2016.

[15] T. Kato, A. Oda, S. Ayukawa, C. Wu, and S. Ohzahata, "Inferring TCP Congestion Control Algorithms by Correlating Congestion Window Sizes and their Differences," Proc. IARIA ICSNC 2014, pp.42-47, 2014.

[16] T. Kato, A. Oda, C. Wu, and S. Ohzahata, "Comparing TCP Congestion Control Algorithms Based on Passively Collected Packet Traces," Proc. IARIA ICSNC 2015, pp. 145-151, 2015.

[17] N. Ohzeki, R. Yamamoto, S. Ohzahata, and T. Kato, "Estimating TCP Congestion Control Algorithms from Passively Collected Packet Traces using Recurrent Neural Network," Proc. ICETE DCNET 2019, pp. 33-42, 2019.

[18] "Alexa Top Sites 1M," http://s3.amazonaws.com/alexa-static/top-1m.csv.zip. (Accessed on 12/03/2020).

[19] V. Paxson, "Automated Packet Trace Analysis of TCP Implementations," ACM Comp. Commun. Review, vol. 27, no. 4, pp.167-179, 1997.

[20] T. Kato, T. Ogishi, A. Idoue, and K. Suzuki, "Design of Protocol Monitor Emulating Behaviors of TCP/IP Protocols," Proc. IWTCS '97, pp. 416-431, 1997.

[21] S. Jaiswel, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley, "Inferring TCP Connection Characteristics Through Passive Measurements," Proc. INFOCOM 2004, pp. 1582-1592, 2004.

[22] J. Oshio, S. Ata, and I. Oka, "Identification of Different TCP Versions Based on Cluster Analysis," Proc. ICCCN 2009, pp. 1-6, 2009.

[23] P. Yang, W. Luo, L. Xu, J. Deogun, and Y. Lu, "TCP Congestion Avoidance Algorithm Identification," In Proc. ICDCS '11, pp. 310-321, 2011.

[24] A. Mishra, et al., "The Great Internet TCP Congestion Control Census," Proc. ACM Meas. Anal. Comput. Syst., vol. 3, no. 3, article 45, pp. 1-24, 2019.

[25] Y. Edalat, J. Ahn, and K. Obraczka, "Smart Experts for Network State Estimation," IEEE Trans. Network and Service Management, vol. 13, no. 3, pp. 622-635, 2016.

[26] M. Mirza, J. Sommers, P. Barford, and X. Zhu, "A Machine Learning Approach to TCP Throughput Prediction," IEEE/ATM Trans. Networking, vol. 18, no. 4, pp. 1026-1039, 2010.

[27] J. Chung, D. Han, J. Kim, and C. Kim, "Machine Learning based Path Management for Mobile Devices over MPTCP," Proc. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp 2017), pp. 206-209, 2017.

[28] S. Hochreiter and J. Schimidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[29] iPerf3, "iPerf - The ultimate speed test tool for TCP, UDP and SCTP," https://iperf.fr/.

[30] dpkt, "dpkt," https://pkt.readthedocs.io/en/latest/.

[31] C. Caini and R. Firrincieli, "TCP Hybla: a TCP enhancement for heterogeneous networks," Int. J. Satell. Commun. Network., vol. 22, no. 5, pp. 547‑566, 2004.

[32] A. Baiocchi, A.P. Castellani, and F. Vacirca, "YeAH-TCP: Yet Another Highspeed TCP," Proc. PFLDnet, vol.7, pp. 37‑42, 2007.

[33] S. Liu, T. Basar, and R. Srikant, "TCP-Illinois: A loss-and delay-based congestion control algorithm for high-speed networks," Performance Evaluation, vol. 65, no. 6-7, pp. 417-440, 2008.

[34] K. Tan, J. Song, Q. Zhang, and M. Sridharen, "A Compound TCP Approach for High-speed and Long Distance Networks," Proc. IEEE INFOCOM 2006, pp. 1-12, 2006.

[35] Y. You, et al., "Large Batch Optimization for Deep Learning: Training BERT in 76 minutes," Proc. ICLR 2020, pp. 1-37, 2020.

[36] J. Padhye and S. Floyd, "On Inferring TCP Behavior," SIGCOMM Comput. Commun. Rev., vol. 31, no. 4, pp. 287-298, 2001.

[37] A. Medina, M. Allman, and S. Floyd, "Measuring the Evolution of Transport Protocols in the Internet," SIGCOMM Comput. Commun. Rev., vol. 35, no. 2, pp. 37-52, 2005.

Review Ranking to Support Selection of Recommended Items

Short Paper

Taketoshi Ushiama
Faculty of Design
Kyushu University
Fukuoka, Japan
email: ushiama@design.kyushu-u.ac.jp

Daichi Minami
Graduate School of Design
Kyushu University
Fukuoka, Japan
email: minami@kyudai.jp

***Abstract*—This paper proposes a novel approach to aid product selection in e-commerce through the effective ranking of online reviews. Often, users find it challenging to identify the most valuable information amidst a sea of reviews. Our approach addresses this by ranking reviews based on the user's empathy towards reviewers. By taking user feedback on reviews of known products, we estimate the level of empathy towards the reviewer, subsequently ranking reviews of unknown items accordingly. This enables users to easily pinpoint the most relevant reviews amidst the multitude of information. Our evaluation experiments have revealed this new approach to be superior to traditional comparative methods.**

Keywords - online reviews; recommendations; rankings; natural language processing; machine learning.

I. Introduction

In recent years, with the spread of the Internet, various web services such as E-commerce sites and social media have appeared. It has become common to purchase products and research product information online. In addition, the number of items present on the Internet is increasing, and such "information overload" has become an important issue in recent years [1], [2]. Under these circumstances, it becomes difficult for users to discover items on the Internet that match their preferences. To solve this problem, many web services often provide information recommendation functions.

In recent years, social media-style Web sites (online review sites) that collect reviews about items in specific fields have attracted many users. Many e-commerce sites also offer the ability for users to post reviews on items, and many reviews on each item have been posted. Reviews play an important role in users' selection of items. However, when many reviews about the same item are included, it is impractical to browse through all reviews [3]. It has been reported that 80% of users read only a maximum of 10 reviews when purchasing an item on online review sites such as Amazon. Therefore, functions that rank reviews are essential to assist users in merchandising recommended items [1].

Some reviews are helpful to the user, while others are not. Therefore, e-commerce sites provide a mechanism for rating reviews and ranking highly rated reviews at the top. However, there exist cases where reviews that are valuable to one user are not valuable to another. Since the existing review ranking method is not personalized, reviews that are not valuable to a user may appear at the top of the list. Therefore, the review ranking mechanism is expected to reflect the values and preferences of users.

This paper proposes a method for recommending online reviews about a target item based on the user's empathy. We assume that reviews that are useful for item evaluation are reviews by reviewers who have a high degree of empathy with the target user. The proposed method predicts the target user's confidence level about a reviewer based on the reviews the user has rated in the past. The proposed method then considers the opinions of reviewers with whom the target user can empathize based on the user's level of trust in the reviewer to be of high value, estimates the value of the review to the target user, and recommends reviews based on that value.

Section II describes related works. Section III explains the trust-based collaborative filtering for reviewers that forms the background for this study. Section IV introduces a method for ranking item reviews based on empathy. Section V shows the experimental results for evaluating the effectiveness of the proposed method. Section VI describes the summary and future work.

II. Related Work

A. Information Recommendation using Reviews

In recent years, many studies on information recommendation using reviews have been conducted [4]–[15]. There are three main directions of recommendation using reviews [16].

- Tries to solve the data sparseness problem by extracting user preferences information.
- Tries to solve the cold-start problem of the inability to make high-performance recommendations when user evaluations are not sufficiently collected.
- Tries to derive useful information for a recommendation other than evaluation values, such as user context and latent preference factors, from the reviews.

Several methods have been proposed for extracting user preferences and recommending information using reviews. Hayashi et al. [17] extract interest words representing user preferences and their polarity from user-written reviews and recommend movies with reviews that contain interest words and have matching polarity. In the above information recommendation method using reviews, recommendations are made based on the degree to which the recommended items match the preferences of the target user. They do not support the selection of items recommended to the user by the recommendation system. The purpose of this research is to support the act of sorting recommended items by the user.

The approach of acquiring user preferences by having users directly input reviews for items they have selected in the past is not realistic because it places a heavy burden on the user. Therefore, the proposed method adopts an approach to estimate user preferences indirectly by using feedback from user reviews.

B. Review Ranking Methodology

Amazon has a voting button for whether a review is helpful. Users vote for reviews, and the quality of the reviews is determined based on the results. However, only about 10% of all reviews on Amazon are rated [18]. In addition, older reviews have many votes and appear at the top of the list, while newer reviews that do not yet have votes are considered useless [19].

Reviews with many positive votes are more likely to attract more positive votes. To solve these problems, studies have been conducted to estimate the quality score of reviews [20].

Some studies have been conducted on the reliability of textual information such as reviews [21]–[23]. The main purpose of these studies is to address the problem of malicious contributors who post unfair reviews, such as spam, and to automatically determine whether a post is a spam or not or whether a post is fake or not. Thus, most of the conventional studies on ranking and filtering of reviews are concerned with the objective value of reviews, and not much discussion has been given to personal preferences for reviews, such as "whether they are useful to the target user" or "whether they are favorable to the target user". This paper differs from the above studies in that it focuses on the value of reviews to each individual.

III. Collaborative Filtering Based on the Trust in reviewers

In general, collaborative filtering methods can be classified into user-based collaborative filtering and item-based collaborative filtering. User-based collaborative filtering calculates user similarity from a user-item evaluation matrix based on the assumption that "users who select the same item have similar preferences in item selection". For example, "Dokusho Meter" [24], which is one of the leading online review sites in Japan, allows users to register the books they have read, so the similarity between users can be calculated based on the books they have read. It then recommends books that are likely to be of interest to the target user based on their similarity.

However, when looking at or thinking about various things, including books, people evaluate them from their own standpoints. How people perceive a subject often differs depending on their individual sensitivities. In online review sites, how users perceive an item is expressed in their reviews. Even if the reviews are about the same item, there are often differences in the topic of each review. For example, in the case of a review of a comic book, some reviewers evaluate the drawings, while others evaluate the story. This suggests that not all reviewers are sympathetic to the user, but sometimes some reviewers are not sympathetic to the user. Therefore, even among users who have selected the same item, there may be cases where their tastes are not similar.

Therefore, when recommending book information using collaborative filtering on online review sites for books, it is possible to improve the recommendation accuracy by changing the weights of reviewers and calculating the recommended books based on the target user's preferences for reviewers [25].

Figure 1 shows a diagram of the recommendation system using trust information. First, on a page related to a book selected by the target user, reviews posted for the book are displayed. The target user reads the reviews and enters a rating for each reviewer as to whether or not he or she supports the reviewer. By calculating the similarity between the features of the rated reviewers and those of the reviewers who have not yet been rated, we estimate the target user's confidence in the unknown reviewers. Reviewer features are vectorized from submitted documents using the pre-trained Doc2Vec [26]. Then, the recommended books are determined by using the confidence level of each reviewer as the reviewer's weight.

The predicted evaluation value of book B_2 for user u when book B_1 is selected is calculated by the following equation.

$$\mathrm{Pred}(u, b_1, b_2) = \sum_{r \in \mathbf{R}(b_1)} \big(\mathrm{trust}(u, r) \times \mathrm{read}(r, b_2)\big) \quad (1)$$

where $\mathbf{R}(b)$ represents the set of reviewers who posted reviews on book b and $\mathrm{trust}(u, r)$ represents the trust level of reviewer r for target user u. Also, $\mathrm{read}(r, b)$ is determined by whether or not reviewer r has read book b.

$$\mathrm{read}(r, b) = \begin{cases} 1 & (r \text{ have read } b.) \\ 0 & (r \text{ haven't read } b.) \end{cases} \quad (2)$$

The sum of the trust of the reviewers who read book b_2 among the set of reviewers who read book b_1 becomes the predicted rating of book b_2 for user u, and thus the recommendation by collaborative filtering considering the user's trust level.

IV. Proposed Method

A. Evaluation of Reviews for Known Items

The proposed method assumes that users rate reviews. Figure 2 shows an example of an online review site's input interface for rating reviews. The user gives a rating for a review of a known item on this interface. In this figure, the two icons at the bottom right of the review are buttons for entering ratings. The "Sympathetic" icon indicates a positive rating, and the "Not sympathetic" icon indicates a negative rating. If the user's evaluation of the review is neither "Like" nor "Dislike" the user does not click either icon. The system registers the review ratings entered by the user as user profile information.

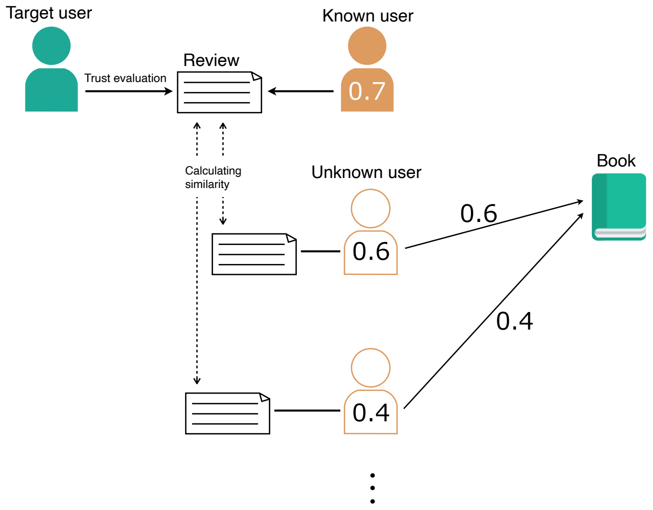
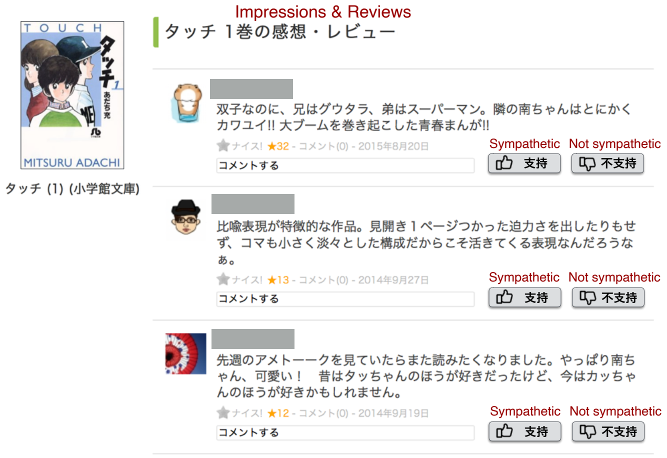


Fig. 1. Recommendation system using trust



タッチ 1巻の感想・レビュー

タッチ (1) (小学館文庫)

[redacted]

双子なのに、兄はグウタラ、弟はスーパーマン。隣の南ちゃんはとにかくカワユイ!! 大ブームを巻き起こした青春まんが!!

ナイス! ★32 - コメント(0) - 2015年8月20日

[redacted]

比喩表現が特徴的な作品。見開き１ページつかった迫力さを出したりもせず、コマも小さく淡々とした構成だからこそ活きてくる表現なんだろうなあ。

ナイス! ★13 - コメント(0) - 2014年9月27日

[redacted]

先週のアメトーークを見ていたらまた読みたくなりました。やっぱり南ちゃん、可愛い！　昔はタッちゃんのほうが好きだったけど、今はカッちゃんのほうが好きかもしれません。

ナイス! ★12 - コメント(0) - 2014年9月19日

Fig. 2. Example of a review rating on the interface

B. Value Estimation for Evaluated Reviewers

This section describes a method for estimating the value of a user's rating to a reviewer based on the rating information described in Section IV-A.

In the proposed method, users evaluate the reviews posted by reviewers, and based on the evaluations, the reviewers' trust in the users is estimated and used for recommending items. Typically, reviewers post reviews for a single item or multiple items. Each review posted by the same reviewer is expected to reflect the reviewer's characteristics. However, as reviewers post reviews for various items, they may post reviews with content they would not usually post. For example, if a review that a user gave a negative rating was of a type that the author of that review does not often post, it does not mean that other reviews written by that reviewer are less valuable to the user.

We propose a method for estimating the trust of a reviewer that considers the certainty of how reviewer-like a given rated text is in its utterances when calculating the value of a reviewer. The confidence of a reviewer is a value that is higher when a given review is a reviewer-like text and lower when it is not reviewer-like. It is used to determine how much the reviewer's confidence value should reflect the reviewer's evaluation of the review itself when computing the confidence value for the reviewer. Figure 3 is a conceptual diagram for estimating the value of a reviewer whose review is directly rated. The trust level $\mathrm{trust}(u, r)$ for a reviewer r who posted a review rated by a user u is defined by the following equation based on the rating.

$$\mathrm{trust}(u, r) = \sum_{d in D_r(pos)} \mathrm{conf}(d, r) - \sum_{d in D_r(neg)} \mathrm{conf}(d, r) \tag{3}$$

This formula estimates the user's trust of a reviewer by subtracting the sum of the confidence levels of the negatively rated document sets from the sum of the confidence levels of the document sets that received positive ratings from the user. In this equation, $D_r(pos)$ represents the set of documents posted by reviewer r that received positive ratings from users, $D_r(neg)$ represents the set of documents posted by reviewer r that received negative ratings from users, and d represents a single document. This study defines $\mathrm{conf}(d, r)$ as the confidence of document d in reviewer r. The details of the confidence level are described in Section IV-D.

C. Value Estimation for Unevaluated Reviewers

This section describes a method for estimating the trust of reviewers who have not yet been evaluated by the user among the reviewers of the item selected by the user in Section IV-A. The method calculates the cosine similarity between the feature vectors of the reviewers that have been rated by the user and the feature vectors of the reviewers that have not yet been rated. This value affects the trust of reviewers who have not yet been evaluated. The method is based on the idea that the value of a reviewer who is similar to a reviewer with high trust is high, while the value of a reviewer who is similar to a reviewer with low trust is low. Figure 4 is a conceptual diagram for estimating the value of an unevaluated reviewer. The trust $\mathrm{trust}(u, r)$ of any reviewer r for a user u is calculated using the following formula.

$$\begin{aligned}\mathrm{trust}(u, r) =& \Bigg(\sum_{d \in D(pos)} \mathrm{sim}(d, d_r) - \sum_{d \in D(neg)} \mathrm{sim}(d, d_r) \Bigg) \\ & \times \mathrm{conf}(d_r, r)\end{aligned} \tag{4}$$

In this formula, we first obtain the value obtained by subtracting the sum of the similarity between the reviews that user u has rated negatively and the review d_r of unrated reviewer r from the sum of the similarity between the reviews that user u has rated positively and review d_r. Then, by multiplying the obtained value by the confidence of the review, we estimate user u's confidence in reviewer r. Where $\mathbf{D}(pos)$ is the set of reviews that received positive ratings from users, and $\mathbf{D(neg)}$ is the set of reviews that received negative evaluations from the users. $\mathrm{sim}(d, d_r)$ represents the similarity between a review d and a review d_r of an unrated reviewer r, and is obtained by computing the Cosine similarity between $\mathbf{d}$ and $\mathbf{d_r}$, which are vectorized by Doc2Vec. The similarity is obtained by calculating the cosine similarity between $\mathbf{d}$ and $\mathbf{d_r}$ vectorized by Doc2Vec.

D. Computing the Confidence of a Review Using an Author-Estimation Model

This section describes a method for computing the confidence level of a review using an author estimation model for reviews based on Deep Learning. Specifically, a neural network with two hidden layers is used to train the model using a document vector that represents the semantic and lexical information of a single review using Doc2Vec and character-based unigrams as input data, and ID information associated with the user as the correct answer label. The output obtained by inputting arbitrary reviews to the trained model is the probability for each reviewer. This probability is higher when the review is a document that is typical of the target reviewer and lower when the review is not typical of the reviewer.

E. Ranking of Reviews for a Recommended Item

It is thought that the reviews that are useful in the item selection process are the reviews of reviewers with whom the user can empathize. This section proposes a method to support efficient item selection by estimating the reviews that users can relate to and displaying them at the top of the list. Figure 5 shows a conceptual diagram of the proposed method. Based on the feedback of ratings on reviews of known items A, we estimate the value of reviews of recommended items B using the confidence values for reviewers obtained in Section IV-B. Based on the assumption that reviewers who are similar to reviewers with high trust values can also be trusted, the recommendation score of review d for user u is calculated by the following formula.

$$\mathrm{score}(u, d) = \frac{\sum_{r \in RVR(PR,NR)} \mathrm{trust}(u, r) \times \mathrm{sim}(r, r_d)}{\sum_{r \in RVR(PR,NR)} \mathrm{sim}(r, r_d)} \tag{5}$$

where PR represents the set of reviews posted by user u that received positive ratings from users, and NR represents the set of reviews posted by user u that received negative ratings from users. The R_D represents the reviewer who posted review D. By multiplying the similarity between reviewer r and reviewer r_d by the confidence value of reviewer r and calculating the weighted average, we can compute the recommendation score of review d for user u. The recommendation score of d is calculated by multiplying the similarity between reviewers r and r_d and calculating the weighted average. The similarity $\mathrm{sim}(r, r_d)$ between reviewers is calculated using Doc2Vec [26], which can acquire a distributed representation of sentences. Specifically, all reviews submitted by reviewers are concatenated into a single document and converted into a vector by Doc2Vec, and the similarity between vectors is calculated by cosine similarity. The reviewer set RVR is determined by the following formula:

$$\mathrm{RVR}(PR, NR) = \bigcup_{d \in PR \cup NR} \mathrm{reviewer}(\mathrm{item}(d)) \tag{6}$$

Where $\mathrm{item}(d)$ denotes the item set to which d reviews are posted and $\mathrm{reviewer}(\mathrm{item}(d))$ denotes the set of reviewers who post reviews on the item set to which d reviews are posted.

V. Evaluation

A. Experimental Setup

Experiments were conducted to evaluate the effectiveness of the proposed method. The dataset used for the experiments

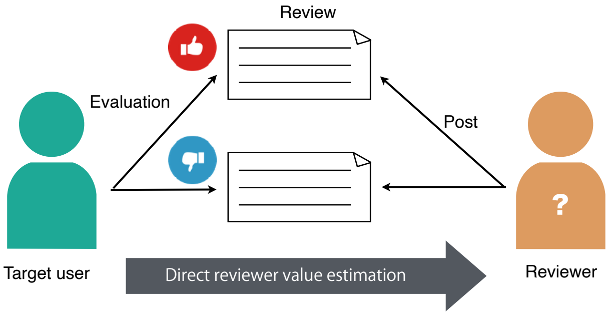


Fig. 3. Value estimation for Evaluated Reviewers

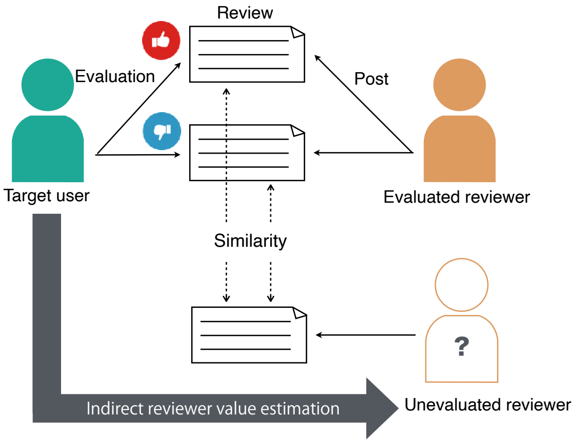


Fig. 4. Value estimation for unevaluated reviewers

was obtained by crawling from one of the famous Japanese online book review sites “Dokusho meter”. MeCab[1] was used for Japanese analysis, and mecab-ipadic-neologd[2] was used for the dictionary. The number of training epochs for the author estimation model was set to 30.

The participants were asked to select two books from among those they had recently read and to input their evaluation feedback for the reviews of the two books. The participants were asked to rate the reviews of two books on a three-point scale of “agree,” “don’t know,” and “don’t agree,” based on the question “do you agree with this review?” The evaluation data for one book review was used as training data to predict the recommendation score for the other book review.

The experiment participants were six men and six women in their twenties. The average of the recommendation results for 12 samples crossed between the training data and the validation data was calculated. In addition, a validation test was conducted on the accuracy of the proposed author estimation method on the reviewers of the books selected by the participants.

To evaluate the effectiveness of the proposed method, we compared the results with those of three different methods. They are random sampling, vote ranking, and Support Vector Regression (SVR). In the vote ranking, we compared the top 10 reviews with the highest number of votes with the top 10 reviews using the proposed method. In SVR, the explanatory variables for the regression analysis were the Term Frequency–Inverse Document Frequency (TF-IDF) vector of words, the percentage of each part of speech in the reviews, the total number of words, and the number of word types.

[1] https://taku910.github.io/mecab/
[2] https://github.com/neologd/mecab-ipadic-neologd

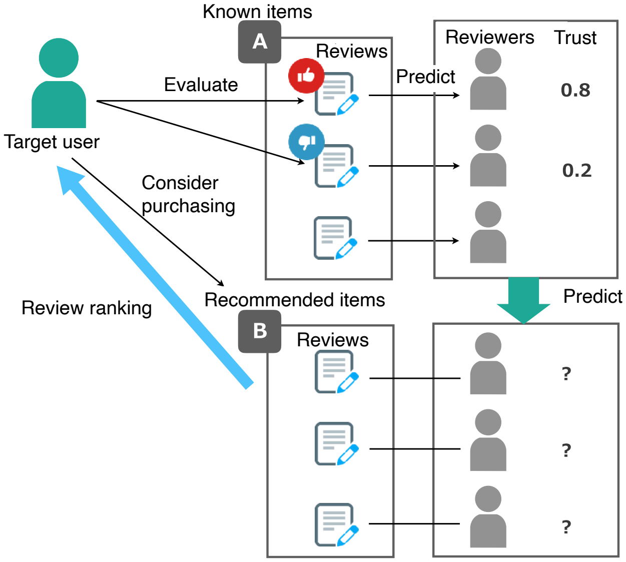


Fig. 5. Conceptual diagram of the proposed method

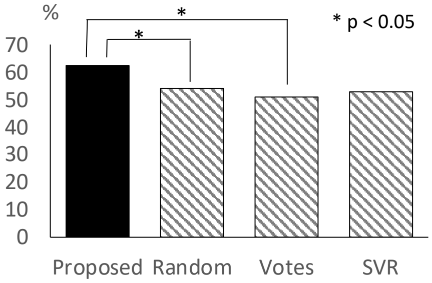


Fig. 6. Percentage of ”sympathetic” reviews in top 10 ranked.

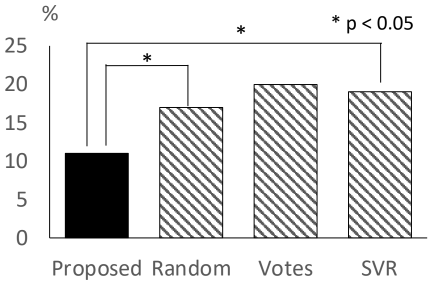


Fig. 7. Percentage of ”not sympathetic” reviews in top 10 ranked.

B. Experimental Results

We calculated the percentage of reviews that participants rated as “sympathetic” and the percentage of reviews that they rated as “not sympathetic” out of the top 10 ranked reviews in the proposed and comparative methods.

The results for the reviews evaluated as “sympathetic” are shown in Figure 6. When significant differences were confirmed by T-test, significant differences were observed between the proposed method and the random sampling method, and between the proposed method and the order of the number of votes, at a significance level of 5 percent.

The results for the reviews evaluated as “not sympathetic” are shown in Figure 7. When the T-test was used to confirm the significant differences, significant differences were observed between the proposed method and the random sampling method and between the proposed method and the support vector regression at a significance level of 5 percent.

C. Discussion

The top 10 reviews sorted based on the proposed method had the largest percentage of reviews rated as ”sympathetic” among all the methods shown in Figure 6. Significant differences were observed between the proposed method and the order of votes. Thus, it was found that the review ranking considering subjective preferences by the proposed method presented reviews that were more useful in terms of product selection at the top than the review ranking based on objective indices. The proposed method is more effective than the regression-based review recommendation since there is no significant difference between the regression method and the order of the number of votes. In addition, among all the methods shown in Figure 7, the proposed method had the lowest percentage of reviews that were input with a rating of “not sympathetic.” There was a significant difference between the proposed method and the support vector regression, indicating that the proposed method effectively filters reviews that are not useful for the target user in determining the product. The proposed method was found to be effective in filtering out reviews that are not useful for the target user’s evaluation. The percentage of reviews for which the rating of ”do not agree” was entered was significantly different as a percentage between the proposed method and the order of the number of votes but was not significantly different. The reason for this may be the small number of participants, which could be improved by improving the number of participants.

VI. Conclusion

This paper proposed a method to assist users in efficiently selecting items for recommendation by collaborative filtering on online review sites, focusing on the act of selection expected to occur after recommended items are presented to the user. To predict the value of each reviewer in the target user, the method uses the feedback of ratings on reviews of known items as input to predict the trust of the reviews as to whether the reviews are reviewer-like statements. The ranking of reviews is then based on the user's trust. To verify whether the review ranking sorted by the proposed method is helpful for users' product evaluation, we conducted a subject experiment using reviews on a reading meter. The experimental results showed that the proposed method is effective for users in ranking items because it gave higher priority to the reviews that the users could identify with and lower priority to the reviews that the users could not identify with.

As a future issue, we can conduct validation experiments using data from online review sites besides books. The proposed method can be adapted for books, movies, music, and other items because users' preferences for reviewers will likely differ. Therefore, we are considering developing a method that considers objective and subjective values. There is also a possibility that the proposed method can be adapted to social networking services and the sharing economy. Specifically, the proposed method could be applied to timeline filtering in Social Networking Services (SNSs), review recommendations in the sharing economy, and so on. Additionally, we are considering using advanced resource language models such as BERT [27] to more accurately predict the degree of empathy of unknown users.

References

[1] T. Ushiama, D. Minami, "Personalized Item Review Ranking Method Based on Empathy," In Proc. of The Sixteenth International Conference on Digital Society, 2022, pp. 42–43.

[2] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[3] M. L. Anderson and J. R. Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *Economic Journal*, vol. 122, issues 563, pp. 957–989, 2012.

[4] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "Effective product recommendation using the real-time web," In Proc. of the 30th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, 2010, pp. 5–18.

[5] S. G. Esparza, M. P. O'Mahony, and B. Smyth, "A multi-criteria evaluation of a user-generated content based recommender system," In Proc. of the 3rd Workshop on Recommender Systems and the Social Web in RecSys'11, 2011, pp. 49–56.

[6] C.W.K. Leung, S.C.F. Chan, and F. Chung, "Integrating collaborative filtering and sentiment analysis: A rating inference approach," In Proc. of the ECAI 2006 Workshop on Recommender Systems, 2006, pp. 62–66.

[7] W. Zhang, G. Ding, L. Chen, C. Li, and C. Zhang, "Generating virtual ratings from chinese reviews to augment online recommendations," *ACM Trans. Intell. Syst. Technol*, vol. 4, no. 1, 2013.

[8] D. Poirier, F. Fessant, and I. Tellier, "Reducing the cold-start problem in content recommendation through opinion classification," In Proc. of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 204–207.

[9] C.C. Musat, Y. Liang, and B. Faltings, "Recommendation using textual opinions," In Proc. of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), 2013, pp. 2684–2690.

[10] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," In Proc. of the 7th ACM International Conference on Recommender Systems (RecSys'13), 2013, pp. 165–172.

[11] Y. Seroussi, F. Bohnert, and I. Zukerman, "Personalised rating prediction for new users using latent factor models," In Proc. of the 22nd ACM Conference on Hypertext and Hypermedia (HT'11), 2011, pp 47-56.

[12] Y. Wang, Y. Liu, X. Yu, "Collaborative filtering with aspect-based opinion mining: A tensor factorization approach," In Proc. of the IEEE International Conference on Data Mining (ICDM'12), 2012, pp. 1152–1157.

[13] H. Liu, J. He, T. Wang, W. Song, and X. Du, "Combining user preferences and user opinions for accurate recommendation," *Electron. Commer. Res. Appl.*, vol. 12, no. 1, 2013, pp.14–23.

[14] L. Chen and F. Wang, "Preference-based clustering reviews for augmenting e-commerce recommendation," *Knowl. Based Syst.*, vol. 50, pp. 44–59, 2013.

[15] A. Levi, O. Mokryn, C. Diot, and N. Taft, "Finding a needle in a haystack of reviews: Cold start context-based hotel recommender system," In Proc. of the 6th ACM International Conference on Recommender Systems (RecSys'12), 2012, pp. 115-122.

[16] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: the state of the art," *User Modeling and User-Adapted Interaction*, vol. 25, pp 99-154, 2015.

[17] T. Hayashi and R. Onai, "Movie Recommendation Using Reviews on the Web," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 30, no. 1, pp. 102–111, 2015.

[18] A Statistical Analysis of 1.2 Million Amazon Reviews: http://minimaxir.com/2014/06/reviewing-reviews, [Accessed June 1, 2023].

[19] S. Moghaddam, M. Jamali, and M. Ester. ETF, "Extended Tensor Factorization model for personalizing prediction of review helpfulness," In Proc. of the fifth ACM international conference on Web search and data mining (WSDM'12), 2012, pp. 163–172.

[20] S. Raghavan, S. Gunasekar, J. Ghosh, "Review quality aware collaborative filtering," In Proc. of the 6th ACM Conference on Recommender systems, 2012, pp. 123–130.

[21] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," In Proc. 21st International Conference on World Wide Web, 2012, pp. 191–200.

[22] S. Xie, G. Wang, S. Lin, and P.S. Yu, "Review SpamDetection via Temporal Pattern Discovery," Proc. 18th ACM International Conference on Knowledge Discovery and Data Mining, 2012, pp. 823–831.

[23] G. Wang, S. Xie, B. Liu, and S. Yu, "Review Graph based Online Store Review Spammer Detection," In Proc. 11th IEEE International Conference on Data Mining, 2011, pp. 1242–1247.

[24] Dokusho Meter, http://bookmeter.com/, [Accessed June 1, 2023].

[25] D. Minami and T. Ushiama, "Can you trust the user?: Collaborative trust estimation model for recommendations," In Proc. 2017 Twelfth International Conference on Digital Information Management, 2017, pp. 252–256.

[26] Q.L. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," In Proc. of The 31st International Conference on Machine Learning, 2014, pp. 1188–1196.

[27] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," In Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 4171–4186, 2019.