University of Tartu

Faculty of Science and Technology

Institute of Computer Science

Musa Guliyev

The Effects of COVID-19 on Consumption of Animal and Plant-based Food: An Analysis of Twitter Data

Master's Thesis (20 ECTS)

ITM

Supervisors:

Rajesh Sharma

Tartu 2021

The Effects of COVID-19 on Consumption of Animal and Plant-based Food: An Analysis of Twitter Data

Abstract:

This study finds the the impact of COVID-19 on the consumption of meat and plant-based food by analyzing the opinions of the supporters of both diets on Twitter. The study uses around 10 million tweets gathered from November 2019 to June 2020 to cover 3 time periods: Pre-pandemic, when COVID-19 was spreading silently (November 2019 - January 2020), Transition, when governments around the world started mass information spread and took action (February - March 2020), and Pandemic, when almost all people lived under some sort of restriction (April - June 2020). The study also analyzes the opinions of the users on both types of food and groups them into 2 groups: meat-lovers and veggies. Tweets of each group in each time period are analyzed and compared to see the change over time. The results show that, neither of the groups were more popular than the other in all time periods. However, there is a change when tweets are grouped by topics. Tweets about diets became more popular during pandemic, while the number of very positive and very negative tweets about animal-based food increased. In addition, towards the pandemic, both groups became more in contact with each other, despite being previously isolated, and started paying more attention to risks related to meat and links with the virus. Finally, the pandemic has increased negativity in veggies' tweets on food while not affecting meat-lovers.

CERCS: P170 Computer science, calculation methods, systems, management (automatic control theory)

Keywords: Social Media, Twitter, COVID-19, Meat, Vegan, Vegetarian

COVID-19 mõju loomse ja taimse toidu tarbimisele: Twitteri andmete analüüs

See uuring otsis COVID-19 mõju liha ja taimse toidu tarbimisele, analüüsides Twitteris mõlema dieedigrupi toetajate arvamusi. Uuringus kasutatakse umbes 10 miljonit säutsu, mis on kogutud novembrist 2019 kuni juunini 2020, et hõlmata kolme ajaperioodi: pandeemiaeelne aeg, kui COVID-19 levis vaikselt (november 2019 - jaanuar 2020), üleminek, kui valitsused kogu maailmas hakkasid massiteavet levitama ja asusid tegutsema (veebruar - märts 2020), ja pandeemiaaeg, kui peaaegu kõik inimesed elasid mingisuguste piirangute all (aprill - juuni 2020). Uuringus analüüsitakse ka kasutajate arvamusi mõlema toiduliigi kohta ja rühmitatakse nad kahte rühma: lihasõbrad ja taimetoitlased. Analüüsitakse ja võrreldakse iga ajavahemiku iga grupi säutse, et näha ajas toimunud muutusi. Tulemused näitavad, et kumbki rühm ei olnud kõigil ajaperioodidel teisest populaarsem. Siiski esineb erinevus, kui säutsud on teemade kaupa rühmitatud. Dieettoitumise kohta käivad säutsud muutusid pandeemia ajal populaarsemaks, samas kui loomset toitu käsitlevate väga positiivsete ja väga negatiivsete säutsude arv kasvas. Lisaks puutusid pandeemia suunas mõlemad rühmad omavahel hoolimata varasemast isoleeritusest rohkem kokku ja hakkasid rohkem tähelepanu pöörama lihaga seotud riskidele ja selle seostele viirusega. Lõpetuseks on pandeemia suurendanud taimetoitlaste säutsude negatiivsust, mõjutamata samas lihasõpru.

CERCS: P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Contents

1 Introduction **6**
1.1 Background . . . 6
1.2 Social media, nutrition, and COVID-19 . . . 6
1.3 The purpose of this thesis . . . 7
1.4 Contributions of the thesis . . . 7
1.5 Structure of the thesis . . . 8

2 Related work **9**
2.1 Prior research on choosing between meat and plant-based diet . . . 9
2.2 Using social media for eliciting patterns related to health . . . 9
2.3 Using social media for eliciting patterns in nutrition . . . 11
2.3.1 Analysis of data from various platforms . . . 11
2.3.2 Analysis of Twitter data . . . 12

3 Data **13**
3.1 Data Collection . . . 13
3.2 Data preparation . . . 14
3.2.1 Removing irrelevant tweets . . . 14
3.2.2 Removing non-English tweets . . . 14
3.2.3 Removing tweets by possible bot accounts . . . 14
3.2.4 Categorizing the users based on dietary preference . . . 14
3.2.5 Predicting the users' gender . . . 16

4 Analysis **17**
4.1 Popularity of users . . . 17
4.2 Topics of tweets . . . 22
4.3 Mention networks . . . 28
4.4 Random walk on mention networks . . . 34
4.5 Communities within networks . . . 35
4.6 Emotions within tweets . . . 41
4.7 Word shift within groups . . . 42

5 Discussion **46**
5.1 Key findings . . . 46
5.2 Limitations . . . 47
5.3 Future directions . . . 47

6 Conclusion **49**

References 50

Licence 54

1 Introduction

1.1 Background

The fast spread and easy access to the internet have enabled everyone to connect to the global network and join social media. Social media enables its users to generate content, share it on the platform, and have discussions with other users. The highly interactive nature of social media has pushed people to stay connected and share as often as they can. By publishing their content or opinion, social media users willingly surrender their personal data and let the platforms store and use it. This, in turn, has turned social media into a strong tool for collecting a vast amount of data from all around the world.

Various industries have already realized and taken advantage of social media's potential. On one hand, there are private organizations that use social media and its data for increasing profits. Social media has helped companies to have more communication channels, as well as is used a tool for marketing, PR, and targeted advertising [1]. On the other hand, the social media data has found applications in various academic fields. Available literature ranges from simple studies such as detecting tourism activity [2] to more complex research such as developing models for detecting fake news on social media [3].

1.2 Social media, nutrition, and COVID-19

Among the social media data, posts related to food and nutrition habits have a quite significant share. The willingness of people to share information about their meals on social media has led to the accumulation of quite large data. In fact, as of May 2021, there are more than 442 million publicly available photos tagged as "#food" alone on Instagram and it will only keep increasing [4]. Each of those posts contains a bit of information about the user. When used together, such posts can help to discover eating habits, nutrition information, as well as a healthy or unhealthy lifestyle of not only the individual users but also the location, region, gender, age group, and so on they belong to. Therefore, social media data plays an important role in analyzing and discovering the food and nutrition habits of its users.

With the help of social media, researchers have been able to collect food and nutrition information from millions of people within a few months. A quick look at the related work shows that there have been a lot of studies that have tackled this issue. Most of the research has focused on collecting food and nutrition-related information from social media platforms like Instagram, Facebook, and Twitter. Among the platforms, Instagram is the most used source for research and analysis [5] [6]. Some studies have gathered Twitter data to discover the most-used food-related hashtags and identify the most popular food based on each state in the USA, along with creating a model that would predict obesity and diabetes rate based on Twitter data [7]. Another work has also included the restaurants to the equation using restaurant type and location data from Foursquare and discovered more on healthy/unhealthy eating habits [8].

Meanwhile, there is not much work done on analyzing the pandemic aspect of food consumption. As the world battles with COVID-19, more people share information regarding it on social media. Although data on COVID-19 hasn't been unnoticed by researchers, studies have only focused on using social media as a surveillance tool for tracking the pandemic [9]. Those that have analyzed the impact of COVID-19 or any other pandemic on food have only focused on production and supply issues. Their main focus was to determine how much a pandemic can impact and endanger the availability of food. In the case of those studies, social media data was not used at all [10]. In summary, the prior works show that the researchers have used social media food data to analyze the healthy/unhealthy eating habits and to create various prediction models.

1.3 The purpose of this thesis

As the previous studies have shown, the food related posts on the social media are aligned with the regional health levels and people's food consumption patterns [11] [12]. As a result, this study assumes that people's opinion on food reflects their eating habits. This thesis aims to analyze the impact of COVID-19 on changes in food consumption habits through the interaction of people on Twitter. The main goal of the study is to determine how the opinion of people on meat and plant-based food has changed after the pandemic hit. The study has gathered around 10 million tweets about meat and plant-based food from November 2019 to June 2020. The tweets are divided into 3 periods: Pre-pandemic (November 2019 - January 2020), Transition period (February - March 2020), and Pandemic (April - June 2020). Prior to analysis, the users are categorized into Meat-lovers and Veggies based on the support/opposition towards meat and plant-based food in their tweets. Then the study conducts both qualitative and quantitative analyses on each time period and compares the results to identify changes. The study mainly focuses on the changes in the popularity of the groups, the intensity of communication between the groups, topics discussed within individual tweets, as well as network communities, and the emotions expressed in the tweets.

1.4 Contributions of the thesis

The contributions of this thesis are two-fold. First, it aims to fill the gap in the literature. As mentioned in the section 1.2, the literature hardly contains any work on the impact of a pandemic on food consumption habits. Additionally, there aren't enough studies that focus on changes in people's food choices through social media. Thus, this research provides the missing information.

Second, it creates a ground for further studies. The results of this study not only show changes in people's dietary choices but also elicit patterns that are outside of the scope of this research. For instance, by changing the scope from nutrition directly towards COVID-19, it's possible to identify more behavioral changes reflected in social media due to the pandemic.

1.5 Structure of the thesis

1. Chapter 2 discusses the previous academic works on meat and plant-based diets, as well as, using social media for analyses on health and nutrition.

2. Chapter 3 discusses how the data was collected and prepared for the analysis.

3. Chapter 4 contains the analyses and their results. Each subsection first describes which analysis method is applied, then shows the results of the analysis.

4. Chapter 5 describes the key findings from the analyses, the limitations of this study, and the recommendations for further studies.

5. Chapter 6 concludes the study.

2 Related work

This chapter of the thesis discusses the academic literature on using social media platforms for eliciting patterns on health and nutrition. Section 2.1 briefly discusses the literature on meat and plant-based diet. Section 2.2 focuses on using the social media data for detecting patterns of healthy or unhealthy behaviors, as well as identifying the patterns of spread of information, misinformation, and the interactions among the people who share such information. Section 2.3 shifts this focus onto information related to food and nutrition. While section 2.3.1 provides how various social media platforms were used for nutrition-related research, section 2.3.2 discusses how Twitter data was analyzed in previous studies.

2.1 Prior research on choosing between meat and plant-based diet

Although the academic literature contains an abundance of studies on consumption of meat and plant-based food, there are hardly any studies that directly compare the both dietary preferences and show a change in patterns. For decades, the scientists and researchers have mainly focused on the benefits and harms of choosing one of the diets mainly through biological analyses. The studies that have focused on the social aspect of it have mainly done it through surveys and interviews. The main focus of such studies was to discover how the public understands the benefits and obstacles of switching to one diet [13] [14]. In addition, there also are studies that analyze the social media aspect of dietary preferences. However, such studies only focus on what and how people are carrying out the spread of information through digital media channels [15] [16]. The research does not analyze how successful social media is in affecting people's dietary choices, nor if any changes have occurred.

2.2 Using social media for eliciting patterns related to health

As Web 2.0 became more popular, the social media platforms that provided the services also became owners of the largest collections of data. This led to researchers gather a large amount of data through various means, whether it's directly from the publicly available website or the authorized use of the API. It enabled them to conduct analyses on various topics on a scale that was not previously possible. Several years into the popularity of social media, researchers began determining how viable source its data was for identifying health issues. Researchers Denecke and Nejdl [17] directed their focus on health-related question and answer (Q&A) forums in order to test the value of social media in their research. By analyzing the data from forums, they quickly discovered that social media is able to provide health-related data on a much higher scale compared to traditional surveys. This idea was further supported by Paul and Dredze [11] who stated that social media data can be used to analyze allergies, obesity, and even insomnia patterns.

The study conducted by Prier et al [18] widened the scope and tested the viability of a larger pool of data, Twitter. They came to a conclusion that social media is a potential source for public health-related analyses, but with a caveat. The researchers argued that the related data is not readily available. However, they provided that Natural Language Processing (NLP) tools can be used to detect and filter the required data prior to beginning the actual analysis. An issue with such approach is the ability of a machine to understand written text. If this method is followed, the results of the study will be as accurate as the reliability of the NLP tool. Aramaki et al [19] used Twitter data to identify communications on influenza and identify the regions, as well as communities that are involved in discussions. In addition, they have also mentioned the reliability of NLP. Their conclusion was that, at the time, NLP tools were not enough to detect sarcasm through text, which could cause inaccurate results in further research.

The previously mentioned literature played an important role in laying the groundwork for future analyses. As the social media got more popular, it generated data on people from various backgrounds and social groups on various health issues and behavior patterns. Using the opportunity, researchers were able to identify patterns of drug use or vaccination through social media. For example, Zhou et al [20] used Instagram data to not only determine the time and day of the week for peak drug use but also detect the common interest among drug users. Meanwhile, researchers Betti et al [21] used social media data to identify whether parents followed recommended vaccination schedules or not.

On the other hand, some researchers were not sufficed with only identifying the health patterns in individuals. The literature contains studies that analyze the bigger picture that shows how people are grouped or divided, and what causes the divide. For instance, the studies conducted on the measles outbreak in Italy [22] [23] discovered that there was a massive polarization between people who supported and opposed vaccination. The network of people showed that people of each group are closely in contact with one another, and the center of the debates did not only consist of scientists, but also politicians and activists.

Additionally, due to its public access and inadequate moderation, people are allowed to submit anything they wish, even if it contains false facts. As a result, the misinformation causes the polarization of people on issues like public health. However, using the same misleading information, it's possible to fight against it. For instance, Ghenai and Mejova [24] did not only use Twitter to identify information on fake (scientifically disproven) methods of curing cancer, but also applied NLP tools to determine their linguistic characteristics to create a prediction model. The prediction model was able to analyze the texts on cancer cures and determine whether there was misinformation.

Moreover, as the Web became more complex, the services started gathering more data on people in different forms. For example, the data does not only contain what people share on social media, but also which social media platforms they use, what type of content they consume and what the topics of the contents are. All of these can be combined to create both individual and mass profiles for people and analyze various aspects of their lifestyles, including health. For example, a novel approach was analyzed by researchers Araújo et al [25]. Instead of focusing on Facebook posts, their study analyzed Facebook Advertisement settings, which is based on tracking individuals' online activity and offering them products they might consume. Using such data, they were able to predict obesity rates on regional levels. Although this approach had a great potential, the closure of Facebook API to the public has hindered the progress on it.

2.3 Using social media for eliciting patterns in nutrition

Nevertheless, the Web still has plenty of social media platforms where researchers can find publicly available data. The researchers were not sufficed with analyzing general public health issues on social media and have used this opportunity to analyze a crucial part of public health – nutrition.

2.3.1 Analysis of data from various platforms

Nutrition and food consumption are some of the overlooked topics hidden in social media data. Social media users have provided information on their food consumption habits in various ways. One of the simplest forms is food recipes. There are plenty of social media platforms that host recipes provided by the users. These recipes do not only contain which ingredients go well with one another but also show which region the dishes belong to. Researchers have used this data to analyze mostly used ingredients and flavor networks and changes based on geographical proximity [26] and to identify patterns and to predict nutritional values [27]. The study done by Ahn et al [28] have discovered the differences between eastern and western cuisine in terms of food pairing. West et al [29], on the other hand, discovered that holidays are when there are major nutritional shifts. In addition, they have found out that majority of people that start a healthy diet return to using the old, less healthy recipes within weeks.

Some social media platforms provide more complex datasets. Though they are capable of providing previously mentioned results, such as finding cultural boundaries of food consumption habits using Foursquare data [30], more researchers have used them for more complex analyses of nutrition and food consumption patterns. Among them, one of the most common approaches is identifying which food people talk about and categorize the food into healthy and unhealthy [6]. With the help of additional information such as national statistics, it is possible to determine who tends to to speak about healthy or unhealthy food [8] [31]. On one hand, the literature states that people living in areas with little to no access to affordable healthy food and with high levels of obesity tend to post about unhealthy food [32], and people from low-obesity areas tend to post more about physical activities [33]. On the other hand, the data shows that people from the same regions who post more about healthier food receive more support (in forms of likes and follows) [5] [34].

2.3.2 Analysis of Twitter data

Although Twitter is not the first choice when it comes to sharing opinions on food, its vast database also stores complex information useful for food and nutrition analysis. At the simplest level, the researchers have used its data for pattern detection. For example, Kershaw et al considered Twitter a "self-reporting tool" to identify nationwide temporal alcohol consumption patterns across the UK [35], while researchers Widener and Li [36] and Gore et al [12] have used the data to discover how positively or negatively people talk about healthy and unhealthy food on social media. The latter study came to a conclusion that people from low-obesity areas post happier tweets and mention physical activities often. In addition, similar to other platforms, Twitter data is also used for discovering obesity patterns [37]. However, studies have not only identified the patterns, but also developed prediction algorithms that detect the regional obesity level of the people by analyzing their tweets [7] [38]. Moreover, using the Twitter data, it is also possible to identify how people's nutritional habits are affected by certain events. For example, every year, National Eating Disorders Association holds an awareness week on Twitter to raise awareness on this issue and educate people on healthier food consumption behavior. Though the impact of the awareness campaign was previously unknown, researchers Mejova and Lledo [39] have collected the tweets of affected people before and after the campaign and applied linguistic analysis methods to determine how much their opinion, as well as eating behavior has changed. Their conclusion was that the impact of government organizations and NGOs was significantly more than social media influencers.

3 Data

This chapter of the thesis focuses on the data that is used for the analyses. Section 3.1 shows how the data was collected and how it fits the narrative of this study. Section 3.2 describes how the data was prepared for the analysis. The preparation consists of two main tasks, cleaning the data and generating new variables by applying algorithms and various logic on the existing variables. The data preparation steps are written in the order they were done in.

3.1 Data Collection

The main data of this thesis is the tweets about food. The tweets were collected from Twitter's Application Programming Interface (API) using a Python package called Tweepy. The program written in Python ran from November 2019 to the end of June 2020 to retrieve tweets with certain keywords related to food. The keywords are divided into 2 groups based on dietary preference (Table 1).

Table 1: Keywords for collecting tweets

Veggie	vegan, govegan, veg, veggie, vegetarian, nomeat
Meat-lover	meatlover, meat-lover, beef, pork, chicken, mutton, steak

Overall, 8 months of data were collected for this study and it is divided into 3 intervals.

1. **Pre-Pandemic:** This period covers from November 2019 – January 2020. It refers to the time when the general public was going on with their life as usual. In addition, it covers two major holidays, Thanksgiving and Christmas, which are the times when people host dinners and share opinions about different food and recipes.

2. **Transition Period:** This period covers from February - March 2020 . It refers to the time when both people and governments started taking safety measures to prevent or slow down the spread of the virus. The reason why this period takes 2 months in this thesis is the speed at which people started accepting the seriousness of the situation. Not all governments acknowledged the graveness of the threat at the same time, therefore, different parts of the world went under lockdown at different times.

3. **Pandemic:** This period covers from April – June 2020. It focuses on the time when the majority of the world was under some type of restriction. This includes the time when people started paying more attention to their health and comparing this data against pre-pandemic data can show how much the eating habits and the opinion of people on food have changed.

Table 2: Number of tweets after each step

Step	Pre-Pandemic	Transition	Pandemic
Initial dataset	3924593	2101448	4016000
After removing keywords from usernames	2273230	1404596	2673771
After removing non-English tweets	2110307	1305750	2486074
Number of frequently posting users	201	154	229
After removing frequent users	2060224	1282385	2438619

3.2 Data preparation

This subsection gives a detailed description of how the data was processed. The number of tweets remained after each step is shown in Table 2.

3.2.1 Removing irrelevant tweets

The text of the tweet contains various attributes, including links, hashtags, and the usernames of mentioned people. In this case, if the mentioned users have the search keywords in their names, then the program also considers such tweets as relevant and adds them to the dataset. In this step, the tweets where the keywords are used in the mentioned usernames but not in the main text are removed from the dataset.

3.2.2 Removing non-English tweets

In order to be able to apply text-mining algorithms such as sentiment analysis, topic modeling, etc, this study only focuses on tweets in English. Twitter already provides the language in which the tweets are. Using the provided variable, the non-English tweets are removed from the dataset.

3.2.3 Removing tweets by possible bot accounts

In this study, a bot account is considered a user that tweets more than 24 tweets in one day. Tweets are grouped by users and calendar days, and tweets of users who have posted more than 24 tweets in one day are removed from the dataset.

3.2.4 Categorizing the users based on dietary preference

To conduct the required analyses, the collected data was not enough. A key variable, Dietary Preference, is needed to separate the Twitter users into 2 groups: people who eat meat and people who don't. One way of doing it is dividing the dataset into tweets about meat and plant-based food, then analyzing whether a tweet contains negative, positive, or neutral opinions on the food type. While the dataset contains tweets about food, it does not specify whether the user talks positively or negatively about it. As a result, R library called Syuzhet is used for analyzing the sentiment of the tweets. However, to perform the sentiment analysis, the text of the tweets had to be cleaned from unnecessary attributes. Before the calculation of scores, these actions were performed:

- Text was converted from UTF-8 to ASCII to remove non-standard characters
- White space was removed
- All letters were transformed to lower case
- Stopwords in English were removed
- Usernames were removed
- “RT” in front of retweets were removed
- Punctuation and digits were removed
- Links were removed

Once the sentiment score for each tweet is calculated, the tweets dataset is separated into time intervals and topics. For example, the Pre-pandemic Veggie dataset would show tweets about plant-based food in the first interval. Then for each separated dataset, sentiment scores are grouped by each user and the average is calculated. Based on the score and which dataset they belonged to, users were labeled Veggie or Meat-lover. Here are the cases:

- If User has an average positive score in Veggie dataset, and no score in Meat dataset, then that user is labeled as Veggie. If his score was negative, then that'd make him a Meat-lover.
- If User has a positive score in Veggie and a negative score in Meat dataset, then that would make him Veggie and vice versa.
- If User has made positive or negative opinions on both types of food, then whichever is the strongest would determine the preference. If the user has a 5.0 score in Veggie and 2.5 in Meat datasets, then that'd make him a Veggie.
- If User has a score of 0, but only tweeted about one type of food, then he'd be labeled based on that type of food. The logic is that a person who posts a neutral opinion on vegan food, then that person is more likely to be a vegan.
- If User has a score of 0 in both of the datasets, then he'd be identified as true neutral and be removed from the dataset. For each time interval, there are around 1500 true neutrals.

Table 3: Number of unique users based on time interval and preference

Interval	Preference	Unique users
Pre-Pandemic	Meat-lover	494159
Pre-Pandemic	Veggie	505338
Transition	Meat-lover	363463
Transition	Veggie	356933
Pandemic	Meat-lover	631554
Pandemic	Veggie	589179

3.2.5 Predicting the users' gender

The genders of the users are not disclosed by Twitter. However, it's possible to derive the gender using the first names of the users. R library called Gender, which has historical datasets for prediction is used for this process. In this research, the US Social Security data from 1912 to 2012 is used to match the first names to a gender. The pro of this dataset is that it is able to predict gender from names with non-European origins. The con of this approach is the availability of the first names. Not all users prefer to use their real names, which makes their gender undefined. In addition, for people with unisex names, the package may predict either wrong results or no result at all.

Table 4: Number of unique users based on time interval and gender

Interval	Unique users	Females	Males
Pre-Pandemic	1000929	201018	279781
Transition	721386	144824	196533
Pandemic	1222764	242110	330362

4 Analysis

This chapter is dedicated to methods used for analyzing the data and the results of the analyses. In order to maintain coherence, the thesis couples each method with its result, instead of having separate chapters for methods and results. The chapter consists of 7 subsections where each subsection is related to one type of analysis and the methods required to do it. Section 4.1 focuses on the popularity and positivity of users based on dietary preference and gender of the users. Section 4.2 discusses topics of the tweets and the popularity of each topic. Section 4.3 visualizes the communication between users in the form of network graphs. Section 4.4 quantitatively analyzes the changes in the communication with random walks and supports the results in Section 4.3. Section 4.5 further investigates the networks to determine communities within them and topics discussed. Section 4.6 describes the emotions within tweets based on dietary preference. Section 4.7 uses quantitative methods to calculate the shift in the positivity of the tweets and support the results in Section 4.6.

4.1 Popularity of users

The popularity of users is measured by the average number of likes, replies, and retweets a user has got while positivity is measured with the average sentiment score. Likes, replies, and retweets variables exist in the tweets dataset while the dietary preference, gender, and average sentiment score of each user are available at unique users dataset created during the data preparation step. To measure the popularity:

1. Tweets dataset of each time interval is merged with unique users dataset of the same interval.
2. It is grouped by username, gender, preference, and sentiment score to calculate the averages of likes, replies, and retweets count.
3. The dataset is cleaned from outliers. To do so, observations that are above 99^{th} percentile are removed.

After these steps, the data is ready for visualization. They are visualized using the boxplot function of the ggplot2 library of R. In this study, boxplot is used instead of mere average in order to show a more clear distribution with the minimum, maximum values, and the interquartile range. Additionally, since the datasets contain more than a million observations, the Y-axis is shown on a logarithmic scale.

The results of the analyses are shown in Figures 1 - 7. An initial glance at the results shows that, from November to June, there were either no or very subtle changes in the popularity and the positivity of the users. Average like counts of meat-lover and veggie users have not changed over time (Figure 1) while average replies and retweets received by meat-lovers seen a slight movement. Based on the inter-quartile range, meat-lovers have gotten slightly more replies and retweets during the transition period, though, the results returned back to the status quo after transition (Figures 2, 3). Average numbers based on genders show slight differences. The average number of likes was a little higher for females during Pre-pandemic and transition periods but reached equal levels during pandemic (Figure 4). The average number of replies increased slightly during transition for both men and women but returned to the previous state in the pandemic period (Figure 5). In terms of average retweet count, men had a slight advance in the transition period only (Figure 6). Lastly, the sentiment scores based on gender and preference show weak but noticeable trends over time. For females, the average sentiment score of veggies has gone down a little during the pandemic despite being a little higher during the pre-pandemic period, while meat-lovers have kept the same level. For males, while meat-lovers have always had slightly more scores than veggies, the inter-quartile range for veggies has dropped over time, similar to females (Figure 7).

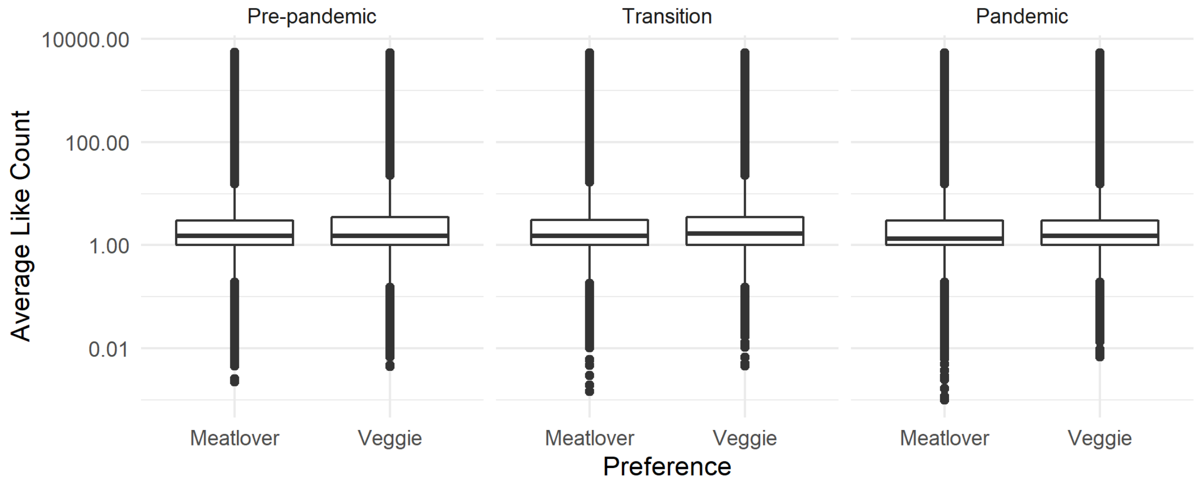

Figure 1: Average number of likes by dietary preference.

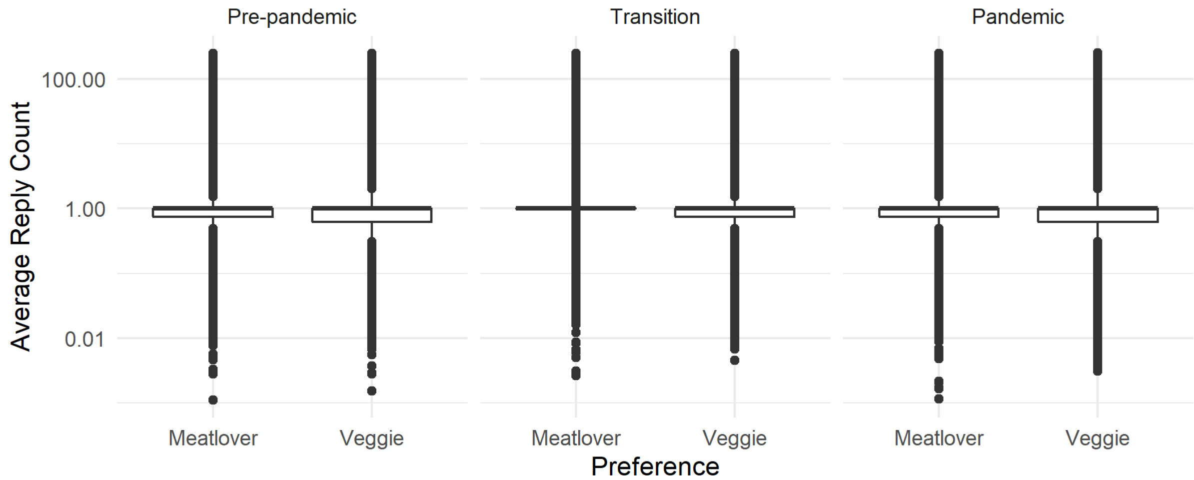

Figure 2: Average number of replies by dietary preference.

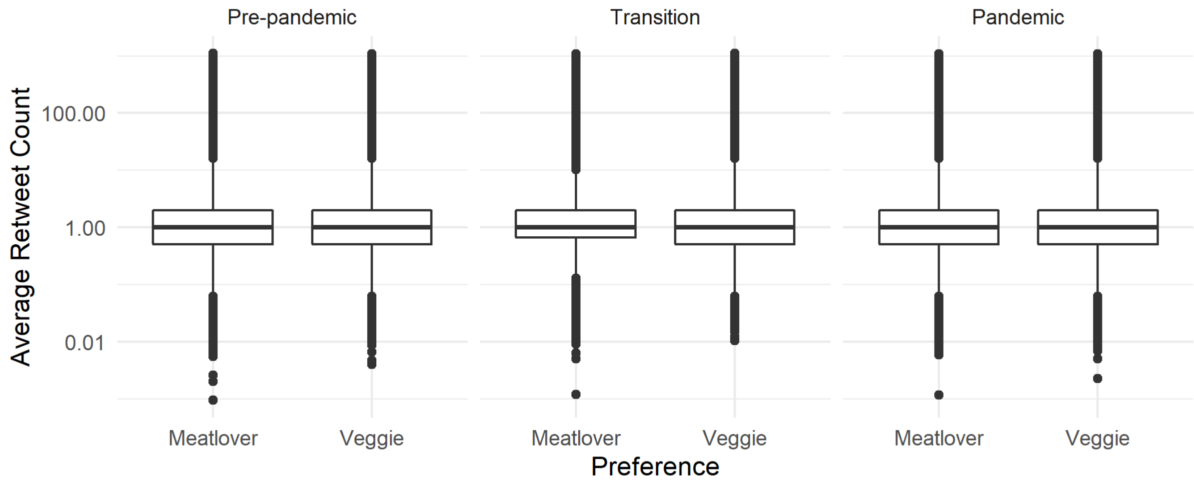

Figure 3: Average number of retweets by dietary preference.

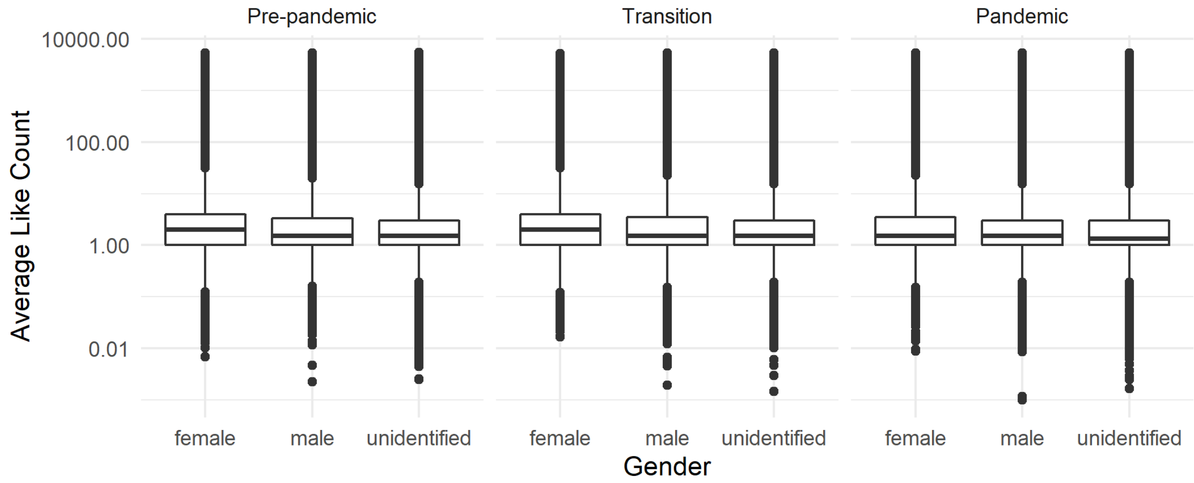

Figure 4: Average number of likes by gender.

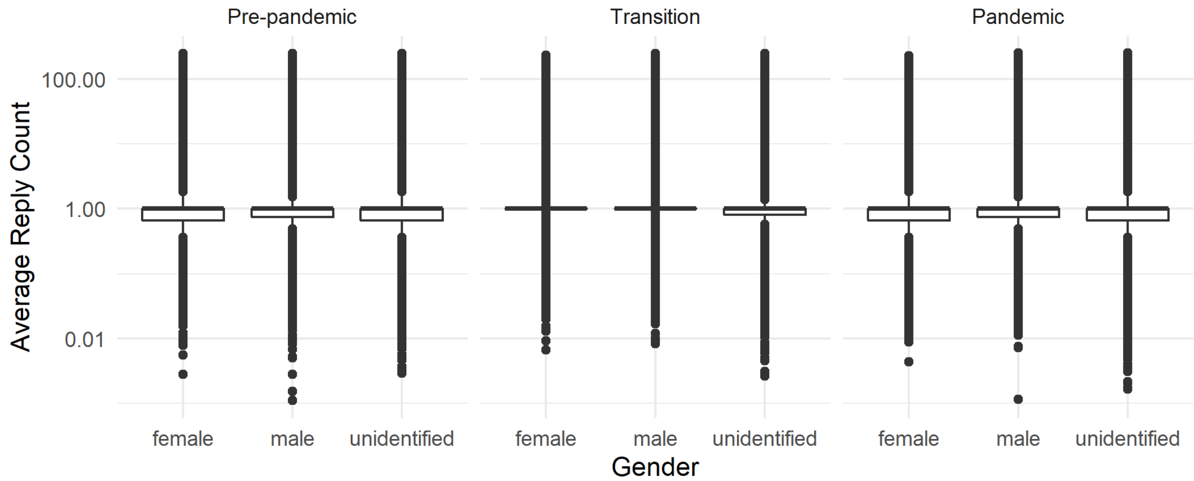

Figure 5: Average number of replies by gender.

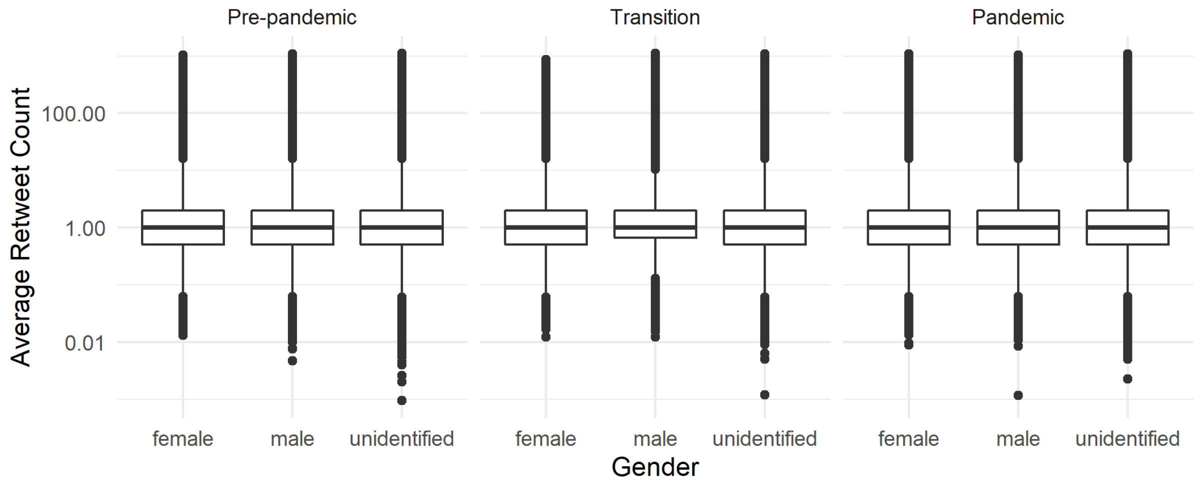

Figure 6: Average number of retweets by gender.

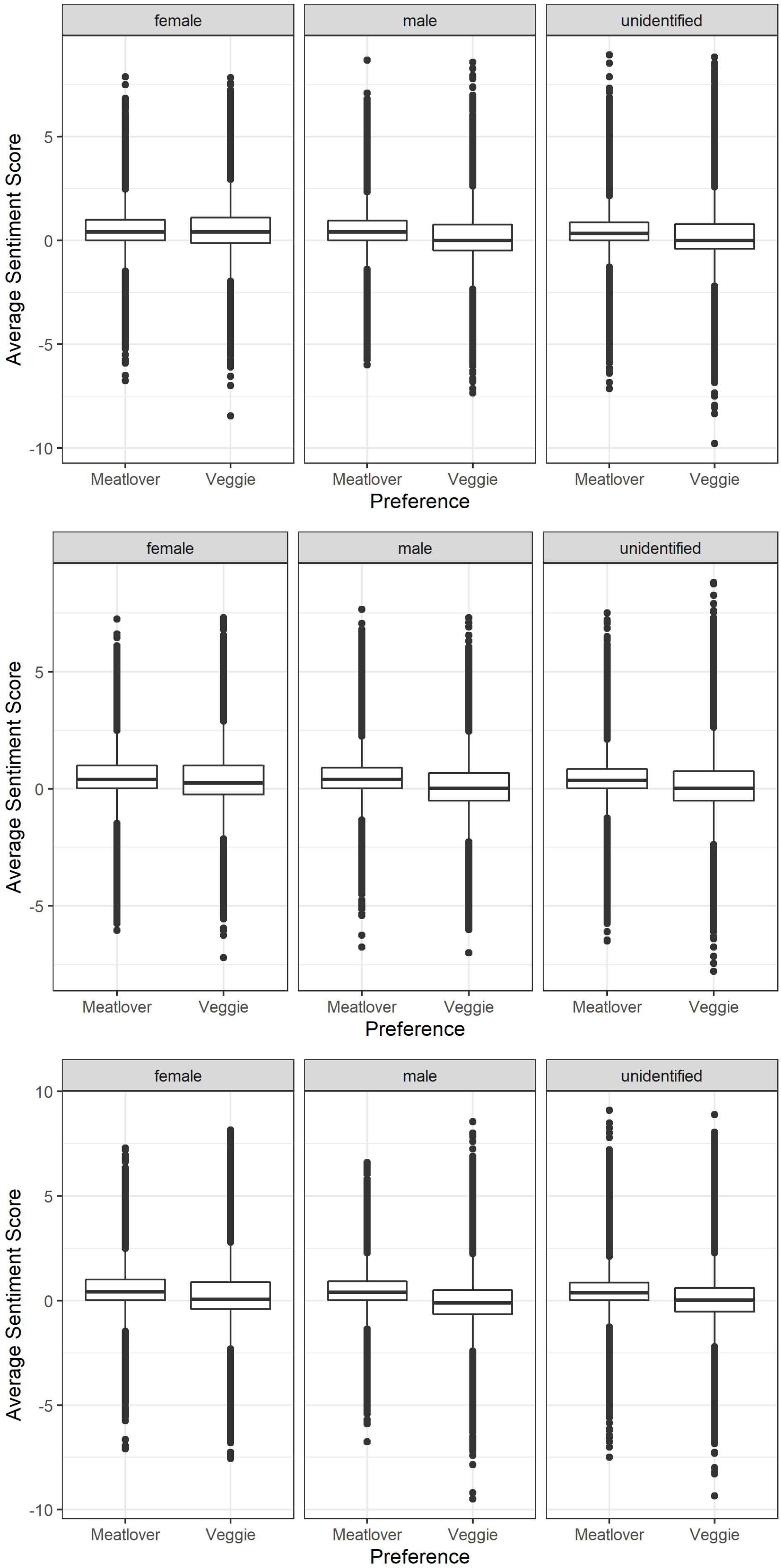

Figure 7: Average sentiment score by gender and dietary preference. Top to bottom: Pre-pandemic, Transition, Pandemic

4.2 Topics of tweets

The previous analysis focuses on the users and the changes in their characteristics. This subchapter provides an in-depth analysis of the tweets, instead of the users. Specifically, it analyzes what people were talking about during the 3 periods, and how popular were the tweets by themselves.

This analysis consists of two main steps. First, the tweets dataset is divided into time intervals and for each interval, topics are determined. To do so:

1. Using TM library, the cleaned text of tweets are turned into a corpus, a collection of machine-readable text [40].
2. The corpus is turned into a document term matrix where each row is a tweet and each column is a word. Every cell shows whether a word exists in a tweet or not.
3. To clear unnecessary information, the sparse terms are removed from the matrix. After this, the matrix is ready for topic modeling.
4. With the help of the Topicmodels library, the Latent Dirichlet Allocation (LDA) algorithm with Gibbs sampling is applied on the matrix to assign each document to a topic [41]. In this case, the algorithm was programmed to sort the tweets into 5 topics for each time interval.

Figure 8 shows the result of LDA topic modeling in the form of identified keywords for each topic of each time interval. Table 5 summarizes those keywords into topics. According to the table, 3 topics repeat across the 3 time periods. These topics are:

- **Animal-based food:** Topics #1 in Pre-pandemic, Transition, Pandemic period
- **Diet and health:** Topic #3 in Pre-pandemic, #5 in Transition, #4 in Pandemic
- **Plant-based food recipes:** Topic #2 in Pre-pandemic, #4 in Transition, #5 in Pandemic

Table 5: Topics for each time interval

Topic	Pre-Pandemic	Transition	Pandemic
Topic #1	Meat and industry	Animal-based food	Animal-based food
Topic #2	Plant food and recipes	Vegan, right thing to do	Doing the right thing
Topic #3	Diet and health	Vegan and time of the year	Food industry in pandemic
Topic #4	Vegetarian food	Plant-based recipes	Choice between meat and plants
Topic #5	Vegan food	Diet and health	Vegan recipes

The second step is to determine the popularity and positivity of the tweets and the topics they belong to. To do so:

1. Using the keywords of the 3 popular topics, tweets are filtered into Animal, Diet, and Recipe subsets.

2. Unlike the previous analysis, the scores are put on a histogram, using the ggplot2 library.

3. The Y-axis is in logarithmic scale (except for the sentiment), and the observations above 99.99th percentile are removed.

The histograms for each topic and time interval are calculated separately and merged based on the topic. Each row shows a variable and each column shows a time interval (Figures 9 - 11).

According to Figure 9, the sentiment scores of tweets have changed significantly over time. In the pre-pandemic period, there were more than 300,000 tweets about animal-based food that had 0 sentiment scores, while this number is a little over 30,000 during the pandemic. A look at the X-axis shows that the scores have increased in terms of both positivity and negativity. In short, people have started talking less neutrally about animal-based food during the pandemic. Meanwhile, the popularity of these tweets has dropped significantly as they have received fewer likes, replies, and retweets over time.

According to Figure 10, which shows the statistics of tweets on diet and health, the majority of the sentiment scores distribution falls between -4 and 4, and the number of neutral tweets has increased over time. In addition, there is a slight increase in like, reply, and retweet counts, as there are more diet-related tweets in the pandemic period with higher counts.

Lastly, according to Figure 11, the tweets about recipes have not experienced a change in their sentiment over time. Based on the graph, both the numbers of tweets and the distribution range of pandemic results are almost the same as the pre-pandemic results. In terms of popularity, there has been a slight decrease in the numbers.

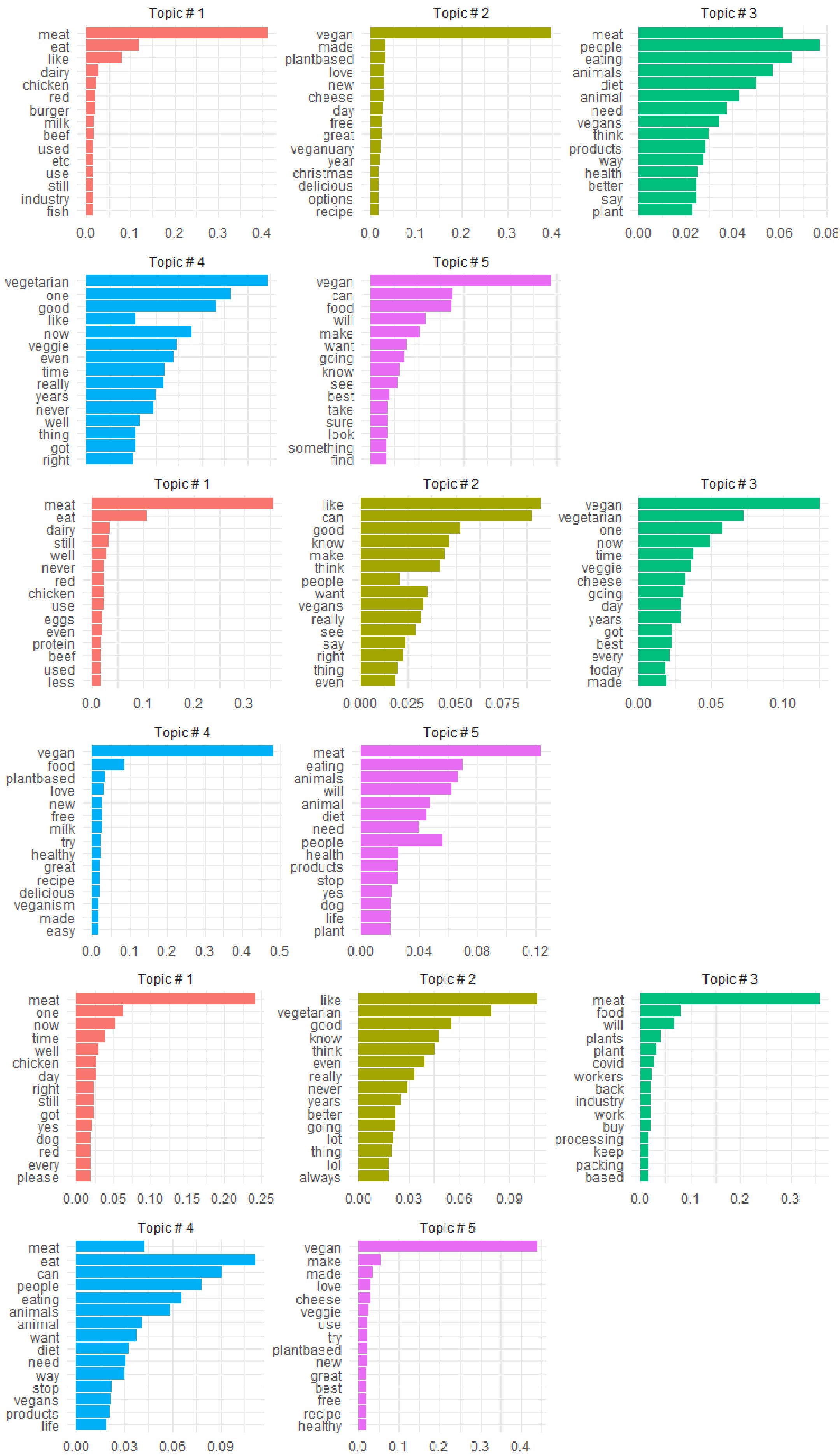

Figure 8: Keywords for each topic. Top to bottom: Pre-pandemic, Transition, Pandemic

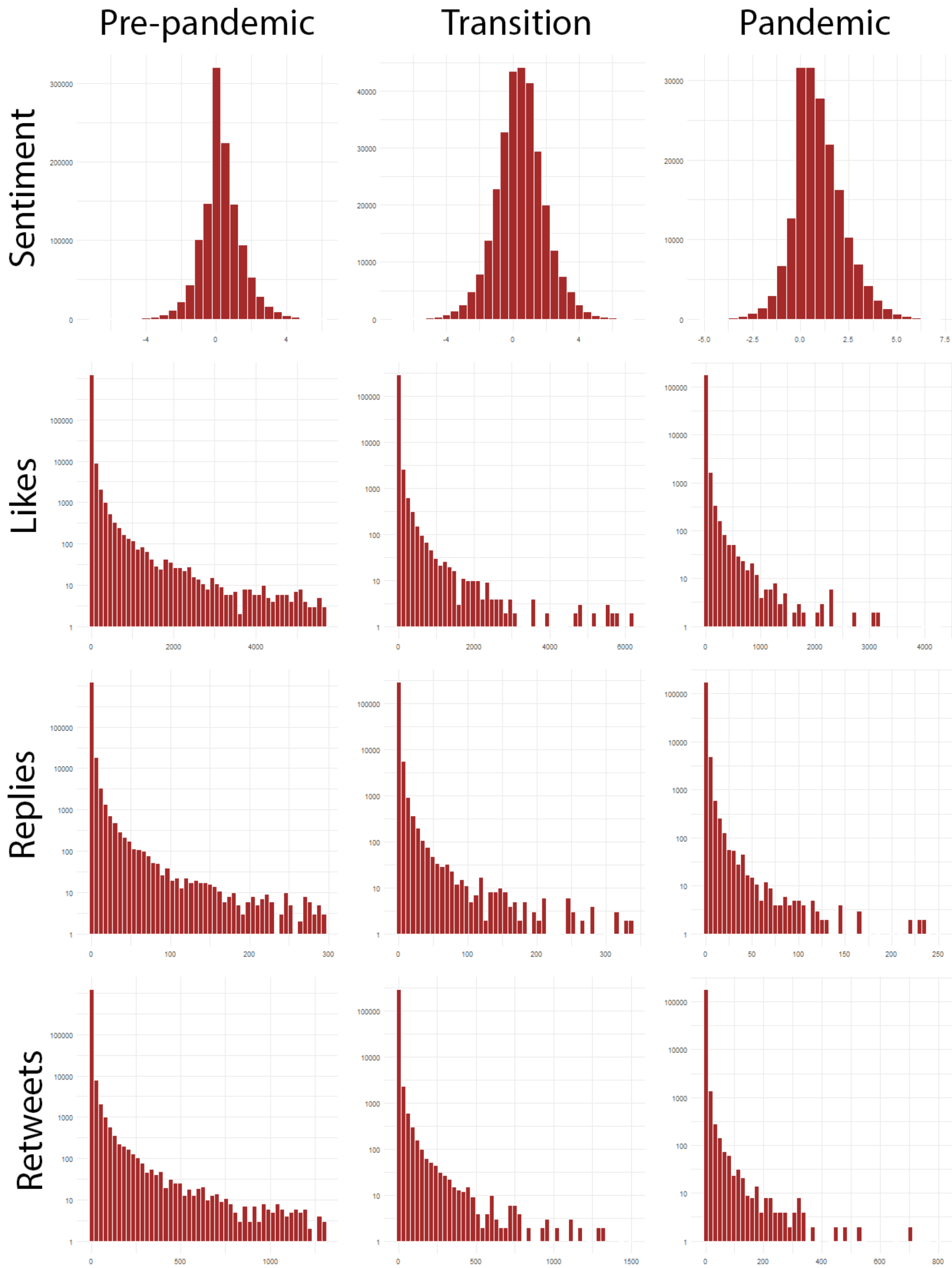

Figure 9: Results of Animals topic.

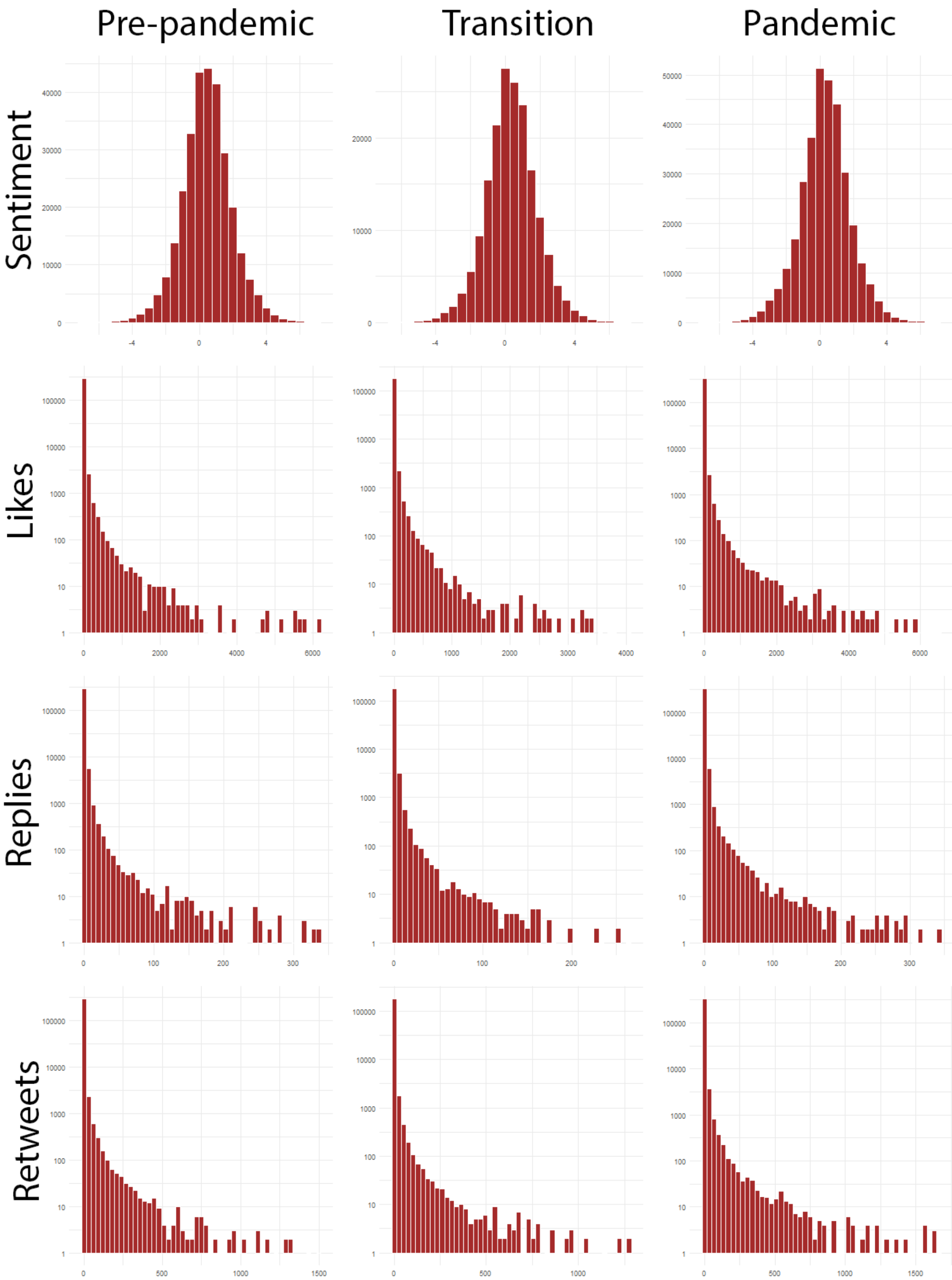

Figure 10: Results of Diet and health topic.

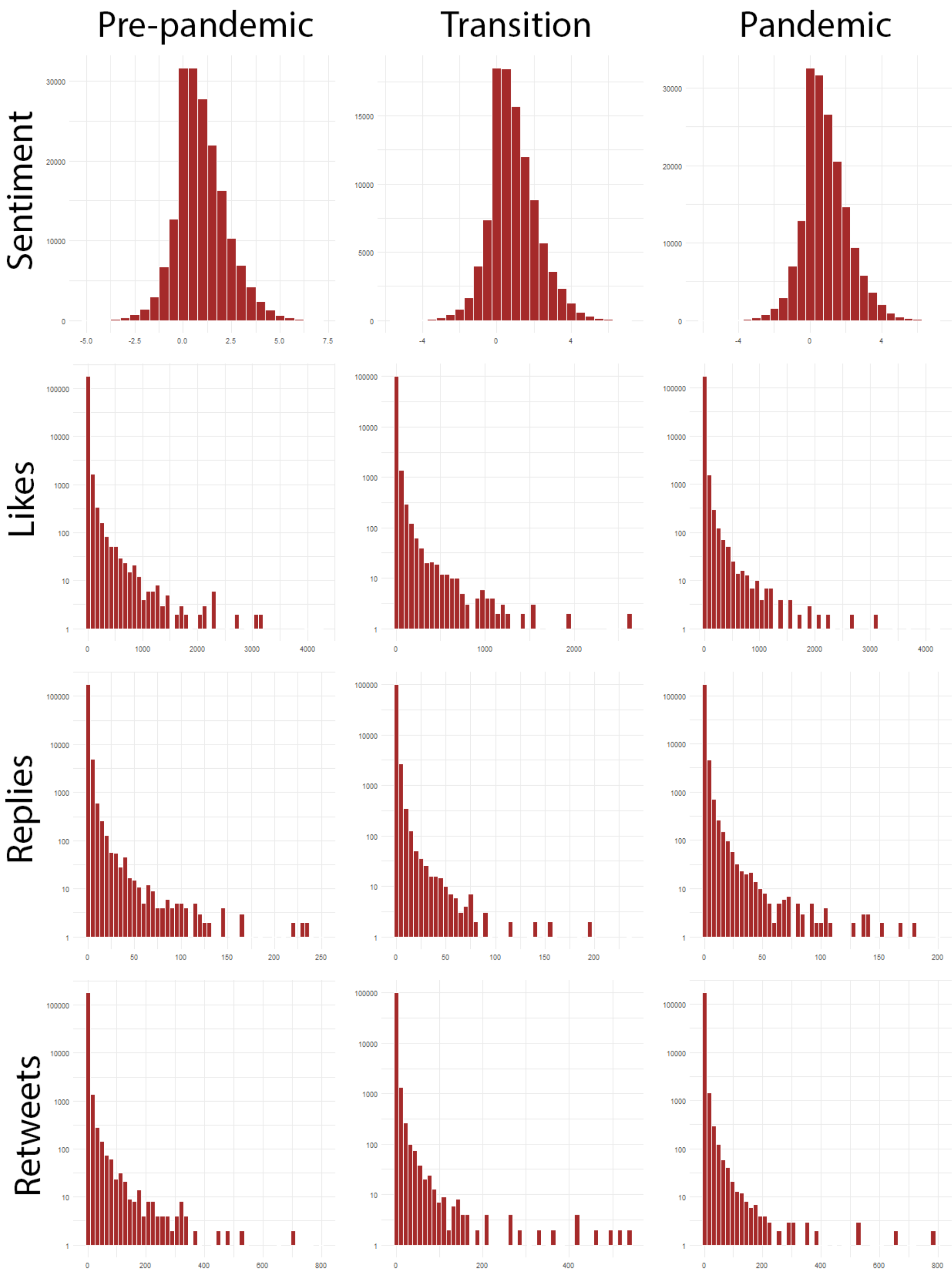

Figure 11: Results of Recipe topic.

4.3 Mention networks

Another way to visualize the changes in dietary preferences is to analyze the communication between groups of people with similar dietary preferences. In this study, we analyze the mention network which is based on users mentioning one another in their tweets. Specifically, the aim of this analysis is to determine the communication between meat-lovers and veggies over time and to find out whether one group became dominant or isolated or in direct communication with the other.

In order to do so:

1. The data was prepared to find the mentions. A new dataset is created where each row is one instance of mention. For example, if User A were to mention 3 users, there would be 3 separate rows for each mention.
2. The dataset is grouped by mentioning and mentioned users to find the weight of mentions, in other words, how many times has one user mentioned another.
3. Mentions with weight less than 10 are removed to keep stronger bonds.
4. Dietary preference of the mentioning users are directly taken from unique users dataset.
5. Some of the mentioned users already exist in the main dataset, and their dietary preference is also taken from the unique users dataset.
6. For mentioned users that has no tweets in the main dataset, a short python script is run to retrieve up to 100 tweets in each time interval.
7. For the retrieved tweets, steps in Section 3.2 are run to find their preference.

Once both the users and mentions are ready, it is possible to construct some networks. The networks are created using Gephi which is an open-source, free, and one of the most famous network visualization programs [42]. For this analysis, the colors of the nodes (users) are set based on their preferences. To make it intuitive, red represents meat-lovers and green represents veggies. In some rare cases, blue represents true neutrals.

Initially, there are 2 networks: Pre-pandemic and pandemic (top left and bottom right networks in Figure 12). However, it is possible that those networks are not made of the same people. To acquire more accurate results, the behavior of the same people needs to be found. Therefore:

1. The users that make up the pre-pandemic and pandemic networks are put into a list.
2. Their tweets in the other 2 time periods are retrieved using the same python script mentioned above.

3. Once again, steps in Section 3.2 and this section are run to determine the preference of both mentioning and mentioned users.

4. Using Gephi, network graphs for Pre-pandemic and Pandemic users in other time intervals are generated.

5. Network graphs are put together in Figure 12.

Although Figure 12 shows the change in two groups of people, it is not possible to draw a conclusion. To avoid confusion, the networks of respective time periods need to be merged:

1. For each time period, the users and mentions datasets are joined and the duplicates are removed.

2. The mentions in the mentions dataset are once again grouped and the weights are summed to remove duplicates and show final weight.

3. Using the igraph library of R, the network datasets are analyzed and the 10 central nodes are found for each network.

4. For every central node, a manual search of their Twitter profiles is done to determine their profession or role and the result is added to the label variable. The centrality of the nodes is used to determine who are the most influential in the flow of a network

5. The datasets are fed into Gephi and the results are merged in Figures 13 - 15.

According to Figure 13, meat-lovers and veggies were mainly homogeneous groups, with meat-lovers being a little more in contact with veggies. The central nodes of the network consist of mainly political activists, while a fashion stylist is a central node for the meat-lovers. The patterns of homogeneity can also be observed in Figure 14, which shows that the isolation of the groups has reached its peak during the transition period. During this time, there was a dietitian and the same fashion stylist among the central nodes. Lastly, according to Figure 15, the communities became more heterogeneous during the pandemic period. Additionally, similar to previous periods, central nodes are mainly activists, with some nutritionists among them. To put it simply, based on the Figures 13 - 15, meat-lovers and veggies were relatively isolated groups of people before the pandemic and during the transition period. A significant level of communication between them is not observed until the pandemic period where people were under lockdowns and quarantines.

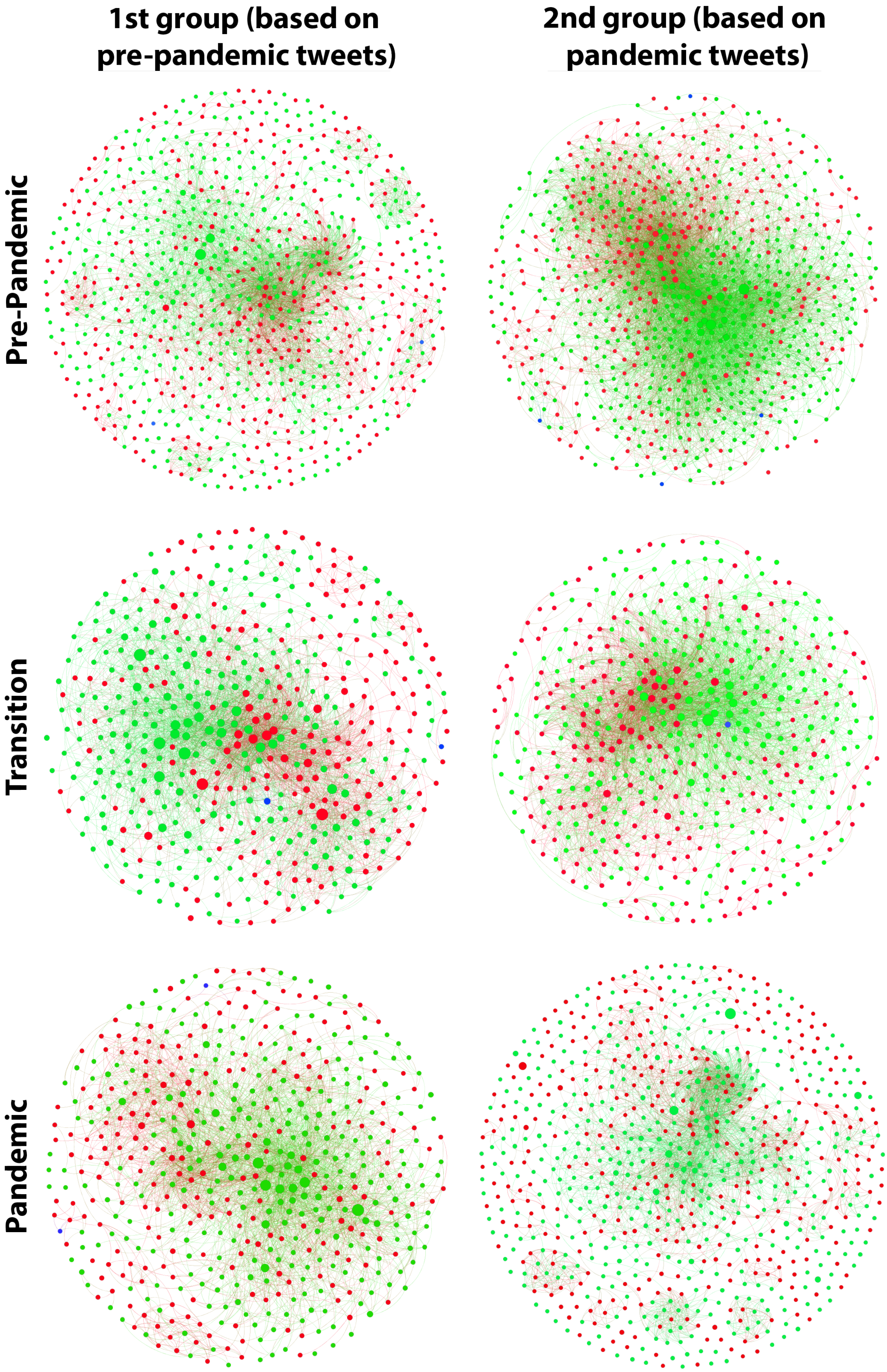

Figure 12: Mention network based on preferences.

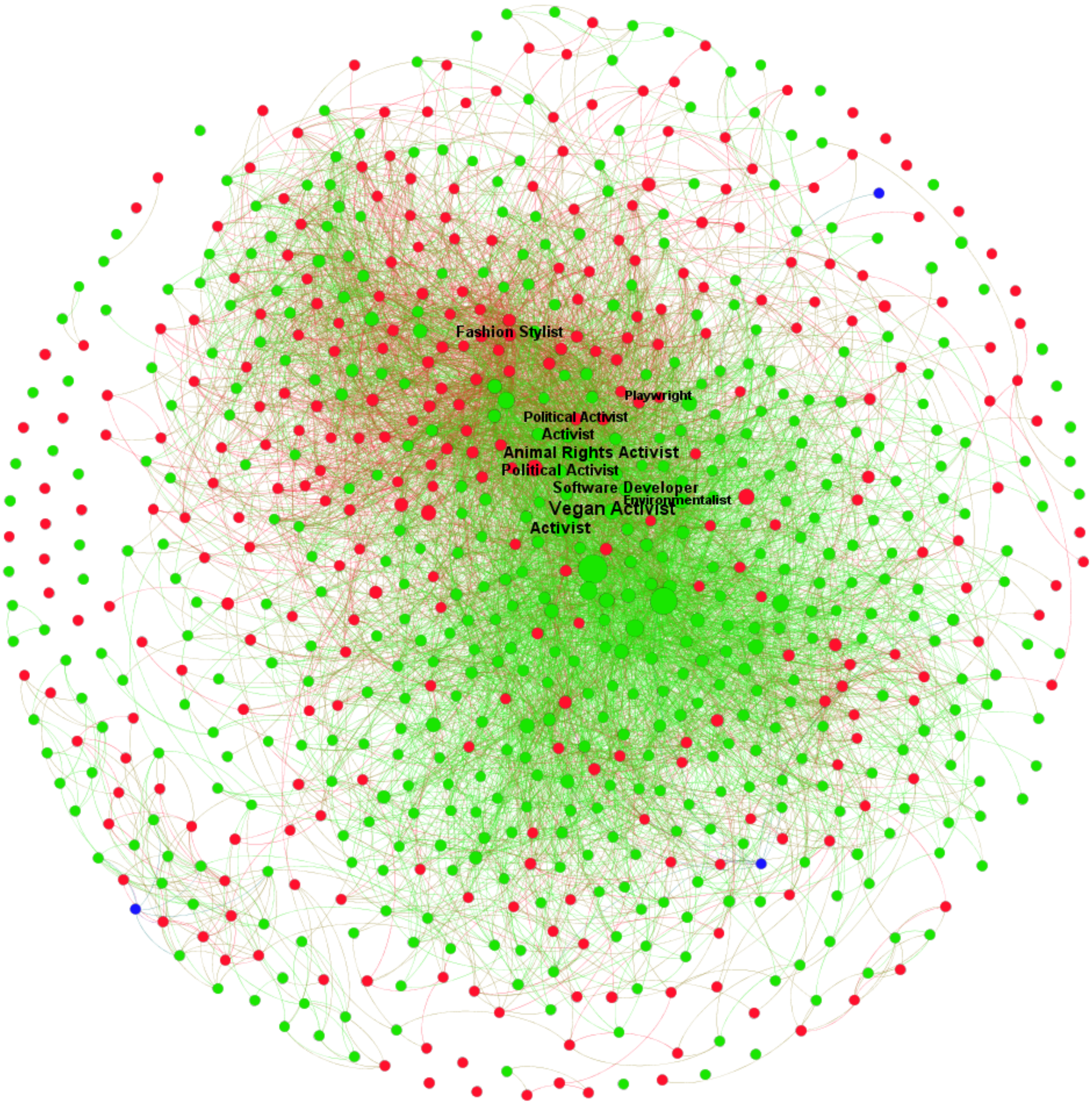

Figure 13: Joined mention network for pre-pandemic period based on preferences.

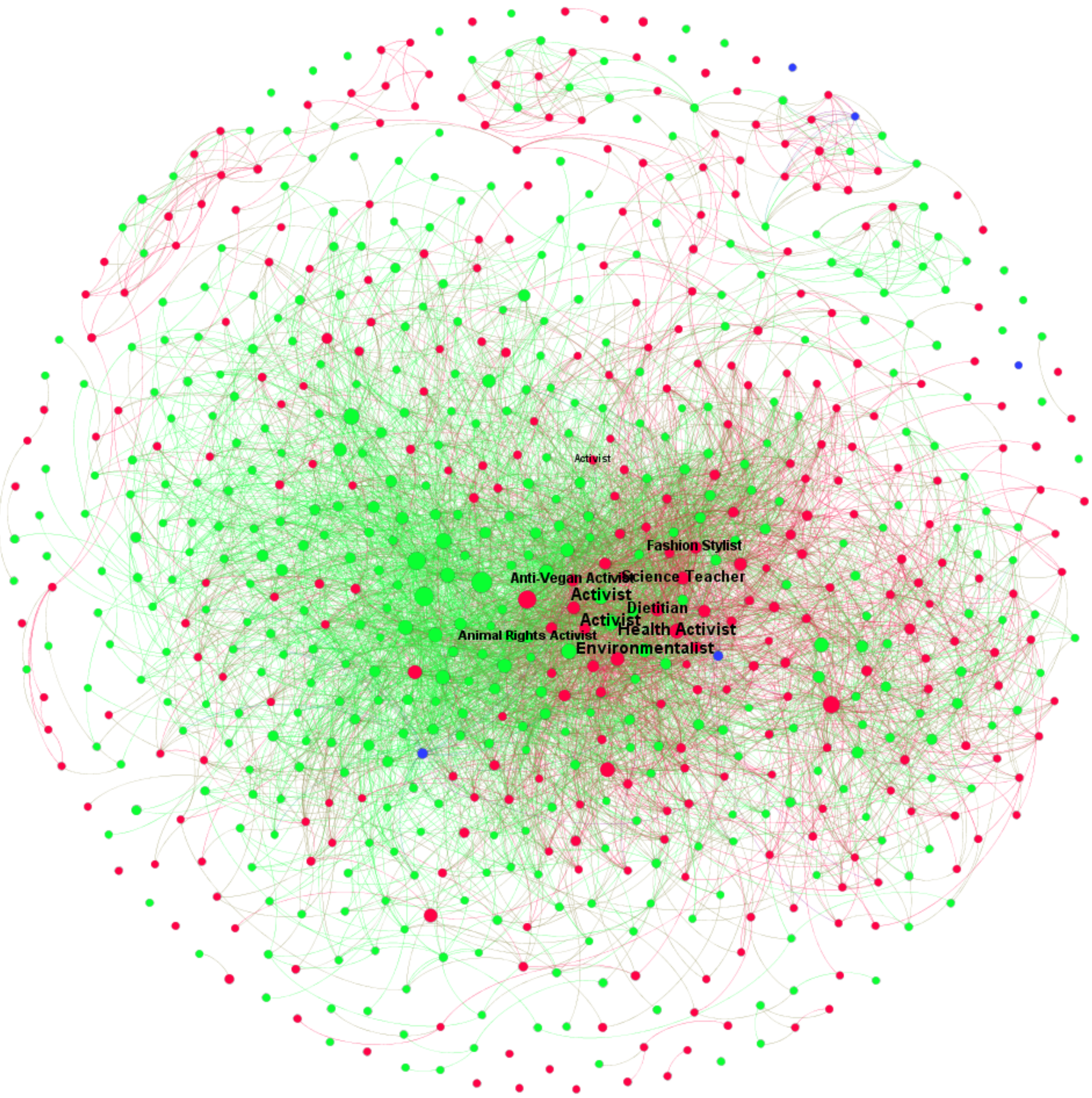

Figure 14: Joined mention network for transition period based on preferences.

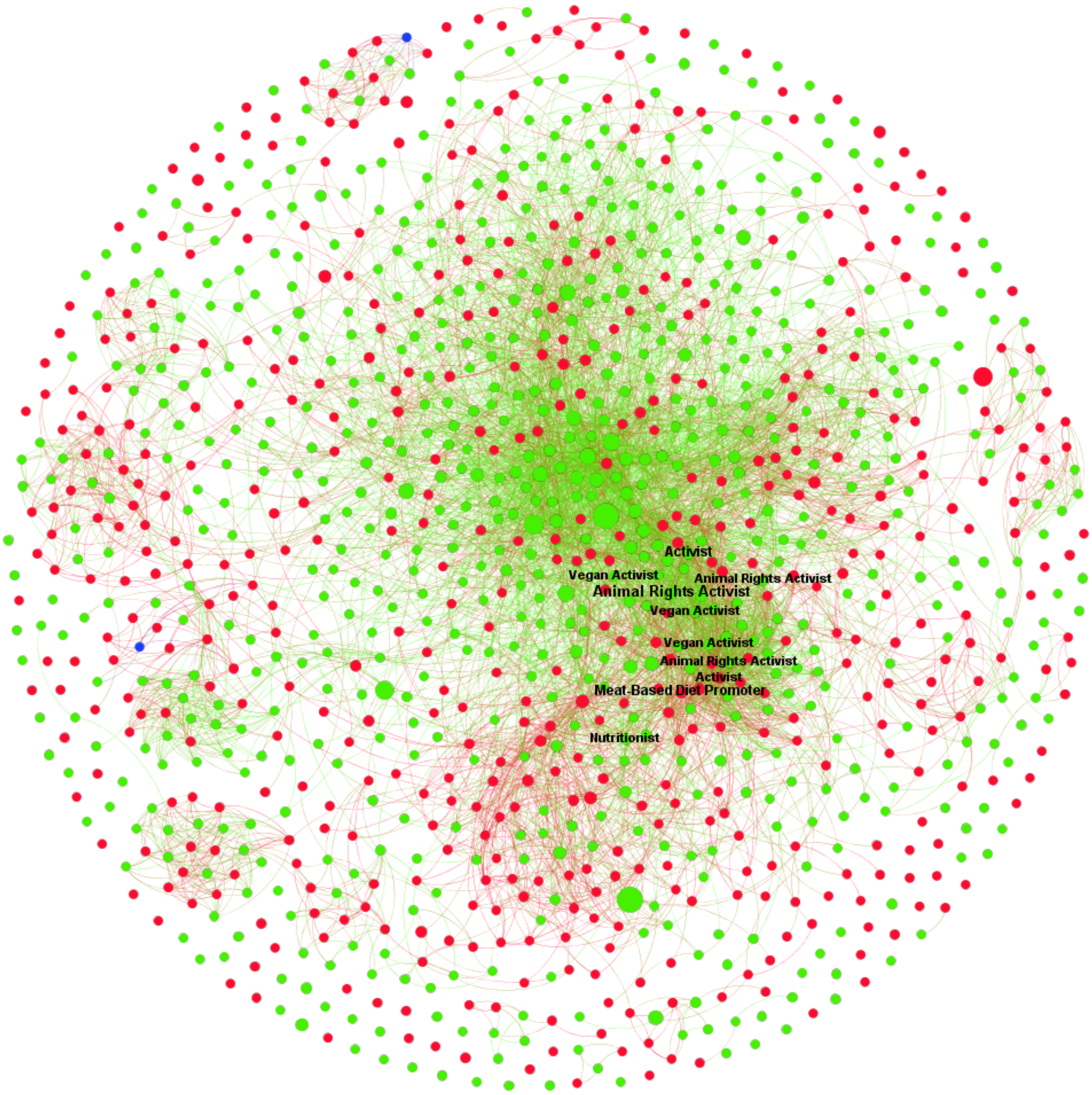

Figure 15: Joined mention network for pandemic period based on preferences.

4.4 Random walk on mention networks

The methods used in section 4.3 analyze the change in communication through qualitative means. Although the network graphs visualize a strong result, they can also be tested with quantitative methods. One way to test the change in the networks is using a random walk algorithm. According to Garimella et al [43], random walks can be used to calculate the probability of a group of users being exposed to the other group. In other words, if we select a random node from one group in the network, what is the probability that after a certain number of steps, we will end up on a node from the other group. The score is called Random Walk Controversy (RWC) and is calculated using this formula:

$$\text{RWC} = P_{XX}P_{YY} - P_{XY}P_{YX}$$

where P shows probability, and the following letters indicate the beginning and ending groups. For example:

$$P_{XY} = \text{P[Start in group X | End in group Y]}$$

If the score is close to 1, then there is strong isolation in the network, meaning there is a high probability that a random walk will start and end in the same group. If the score is closer to 0, then a random walk is likely to end in the other group, showing the connection between groups. Garimella et al [43] have used this formula to calculate whether the randomly selected nodes are exposed to authoritative figures which are the nodes with a high degree in a network. However, this study focuses on communication between any nodes. Therefore, when the random walk lands on a node, the algorithm won't check whether the node is high-degree or not.

To calculate the RWC:

1. Nodes and Edges are loaded into R and the network graph is made using igraph library.
2. The networks consist of smaller communities, where some of them may have 2-3 nodes. Having a random walk using those nodes will end up in a loop of the same nodes. Therefore, for the random walk, only the nodes of the biggest connected component in the network are filtered.

Period	Total nodes	Nodes in biggest component
Pre-Pandemic	41396	21519
Transition	12811	8801
Pandemic	49246	22017

3. Number of steps are needed for the random walk and they are determined by calculating the Diameter and the Average Path Length (mean) of the networks. The diameter is 26 for pre-pandemic, 53 for transition, and 89 for the pandemic. The mean is around 6 for all

3 networks. The number of steps should be less than the diameter and around the mean. There is no single best number that is proven to give the most consistent results, therefore, the random walk is run using 5, 10, 15, 25, 35 steps and the results are compared to see if they are consistent.

4. Random walk algorithm is run on the network. For Meat-lovers and Veggies each, the algorithm is run 1000 times in each time period. The result shows the number of each type of walk (XX, YY, XY, YX).

5. The number of walks is put into the formula to calculate RWC. The results are in Table 6

Table 6: RWC scores for each period and step count

Period	5-step	10-step	15-step	25-step	35-step
Pre-pandemic	0.0158	0.0205	0.0175	0.0297	0.0180
Transition	0.0080	0.0238	0.0200	0.0225	0.0180
Pandemic	0.0145	0.0150	0.0115	0.0200	0.0173

These results bring a couple of key findings. First, all scores are below 0.1, showing that the networks have always been more heterogeneous, and people from each group have always been in contact with one another. Second, aside from the 5-step walk, the results of the calculations somewhat align with the results in Section 4.3. The scores increase in pre-pandemic and transition periods and see a drop in pandemic. This also shows that people of the different groups were relatively more in contact with one another during the pandemic.

4.5 Communities within networks

Identifying meat-lovers and veggies and visualizing their communication network isn't enough to determine how much the food preferences have changed over time. To get a better understanding, we need to dive deep into the tweets and see what the members of each group have talked about over time. It should be taken into account that, not all members are directly in contact with one another. Instead, based on their communication patterns, it can be observed that they have formed groups of different sizes within networks. Therefore, in this sub-chapter, we analyze the communities within networks and the topics they have discussed in their tweets.

To do so:

1. The joined networks are separated based on preference. In this case, both the mentioning and the mentioned users have to belong to the same preference. For each time interval, there are meat-lover to meat-lover and veggie to veggie networks.

2. Using Gephi's modularity function, the communities within the networks are detected and the nodes are assigned to a community.

3. The node colors are adjusted based on the community, where each color represents one community. In this study, the largest 8-12 communities in each network are looked at, which covers more than 90% of the nodes.

4. The nodes tables with communities in Gephi are extracted to R. Using nodes, tweets dataset is filtered into subsets for each community of each time period.

5. Using the same methods mentioned in Section 4.2, topics are predicted for each community.

6. The topics are matched with the corresponding colors of the communities on the network graphs (Figures 16 - 18).

An initial look at the networks shows that in all time periods, veggie networks are more densely packed than meat-lovers. This shows that, regardless of the pandemic, veggies were more in contact with one another than meat-lovers. Another noticeable pattern is communication between different types of nodes which also holds itself over time. In meat-lover networks, people who belong to one community tend to mention the people of the same community. Whereas, people of veggie networks are in contact with different communities.

In terms of topic-specific patterns, there are unexpected results. To begin with, during pre-pandemic and transition periods (Figures 16 and 17), both groups have communities that discuss becoming a vegetarian or vegan. While it's expected from veggies, it's not usual to see meat-lovers discuss such issues. For example, here are some tweets by meat-lovers posted during the pre-pandemic period:

- @PaulBMcGill @timgortonz @MailOnline I did a month last year of no meat or dairy and I had a lot more energy, felt better and lost a couple of kilos. Ultimately I want to balance things out (I love meat) but reckon I'll try a year to see how it goes. I reserve the right to cancel should I miss meat too much!

In addition, it can be observed that meat-lovers not only discuss nutritional aspects of meat-based diets but also include their beliefs as well. One of such topics is halal meat. Being an Arabic word, halal means permitted and in this case, it contains certain Islamic regulations on meat. More specifically, it focuses on the source of meat and the slaughtering techniques. Example tweets include:

- @GazWatty1 @SajdaMughal Halal certified meat is the most humane way to kill an animal. The animal is given a painless death as the 2 vessels in the neck are cut which instantly kills the animals. I've done an entire case study on this and I've switched to eating halal meats because of the results

- @TheJavanDuke So will we be able to slaughter our own meat if we use halal methods as you don't need any special equipment not that I want to

As time goes on, both groups continue discussing various aspects of each type of food, instead of sticking to topics related to their groups. However, COVID-19 significantly affects the discussions and the topics. During the transition period, meat-lovers posted their opinions on the relation between meat and the virus. For instance, some meat-lovers have defended meat consumption:

- #coronavirus It impossible this pandemic to be originated from a poor, dirty unsanitized meat market in #Wuham This virus is not food poisoning, this virus is engineered: look at pictures of it. This is beyond Chernobyl! We will not know the global outcome from it for a time.

Meanwhile, during the same time, veggies emphasized the link between viruses and meat and promoted abstinence from meat consumption. The example tweet below includes the mentioned ideas, along with promoting safe meat consumption.

- @michaellistman all virus are same , they come from the animals which are resistant to them .. humans should go vegetarian or eat only known domesticated animals.

Moreover, in the heat of the transition, some people found an excuse to focus on stereotypes and promote hateful messages. For example, this tweet was posted by a veggie during the transition period:

- @ScribeUndead WTF is it with Chinese and eating live animal/raw meat. Let coronavirus be a lesson to them (that they won't learn)

Coming to the pandemic period, it can be observed that meat-lovers have started paying relatively more attention to vegetarian and vegan alternatives, as well as focusing on poultry and coronavirus.

- I decided I'd cut out red meat and poultry from my diet. I started Monday and today I weighed myself and lost 2 pounds .. pretty neat

In the meantime, veggies directed their attention to the food industry, particularly, the working conditions in the meat industries and plant farms. It appears that the pandemic made both groups pay more attention to not only the nutritional benefits of certain types of food but also the production quality and safety standards.

- Beyond Meat swings to profit as meat supply chain slammed by coronavirus https://t.co/6eLD5KavSF #MONEY https://t.co/XGXcEJL3Xg

All in all, it is clear that, due to the pandemic, meat-lovers have started discussing the alternatives for meat while veggies have tried to promote abstinence from meat due to COVID-19. Both groups, however, seem to have paid more attention to safety standards than ever before.

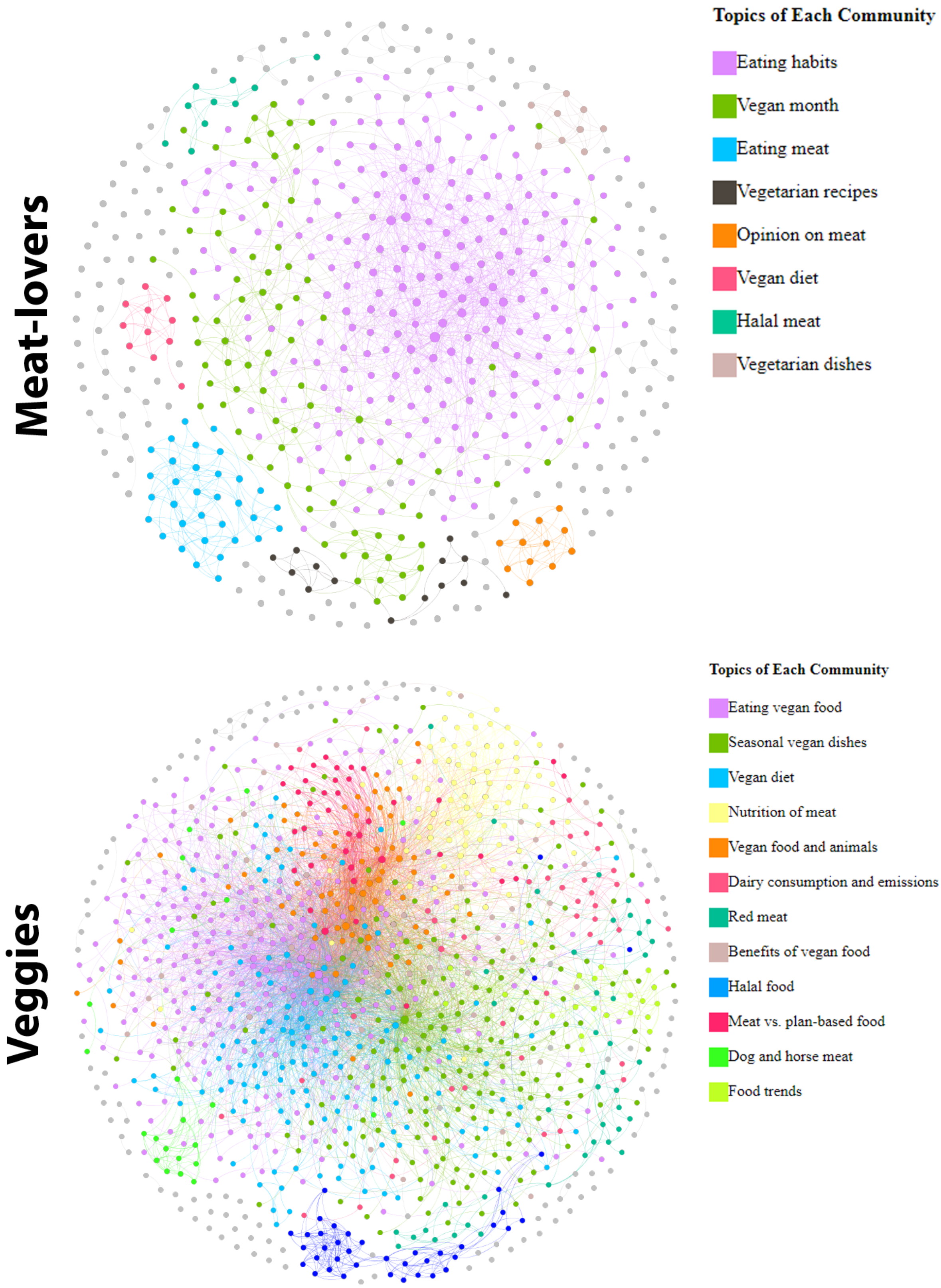

Figure 16: Louvain communities and their topics in pre-pandemic period.

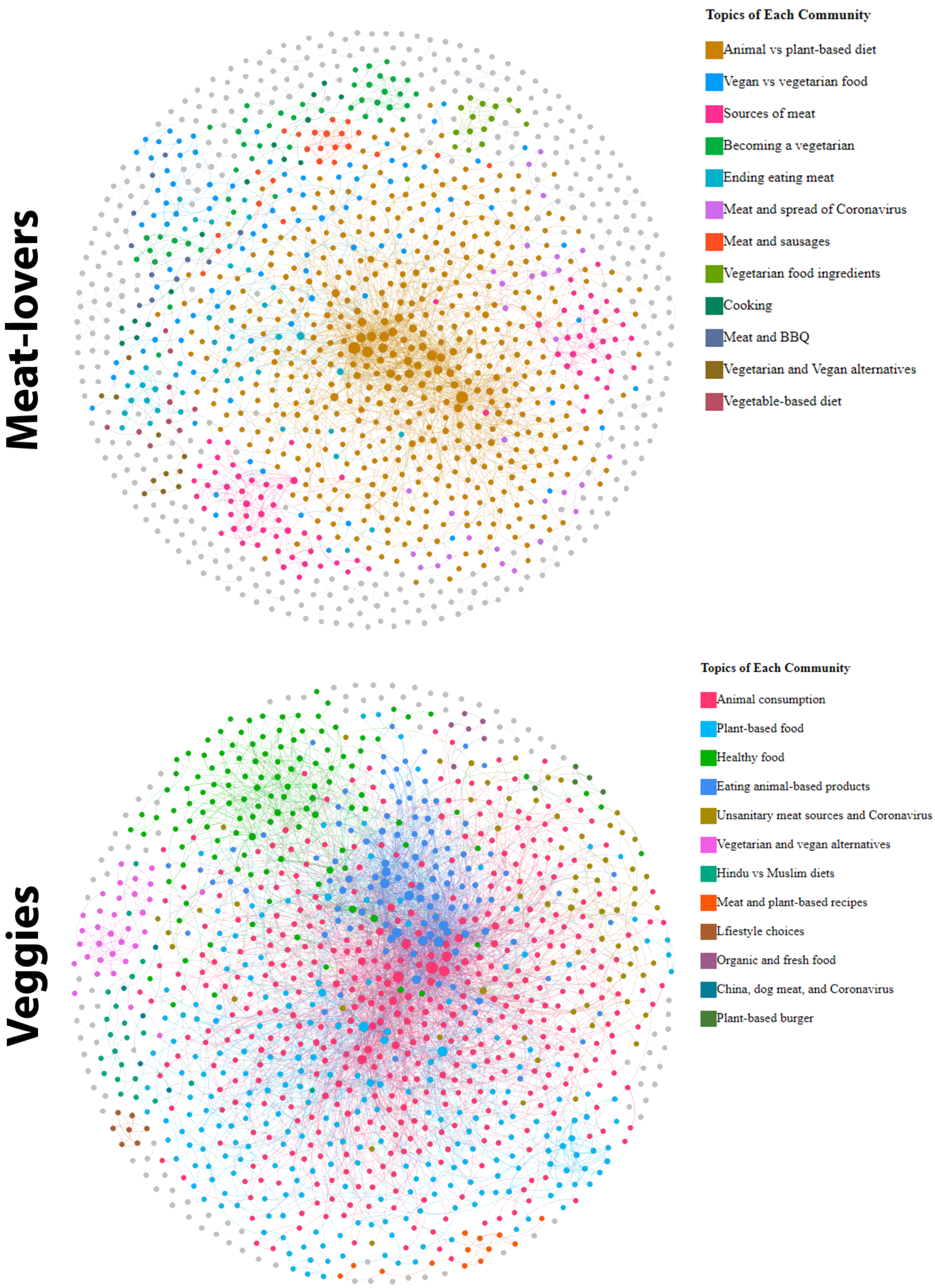

Figure 17: Louvain communities and their topics in transition period.

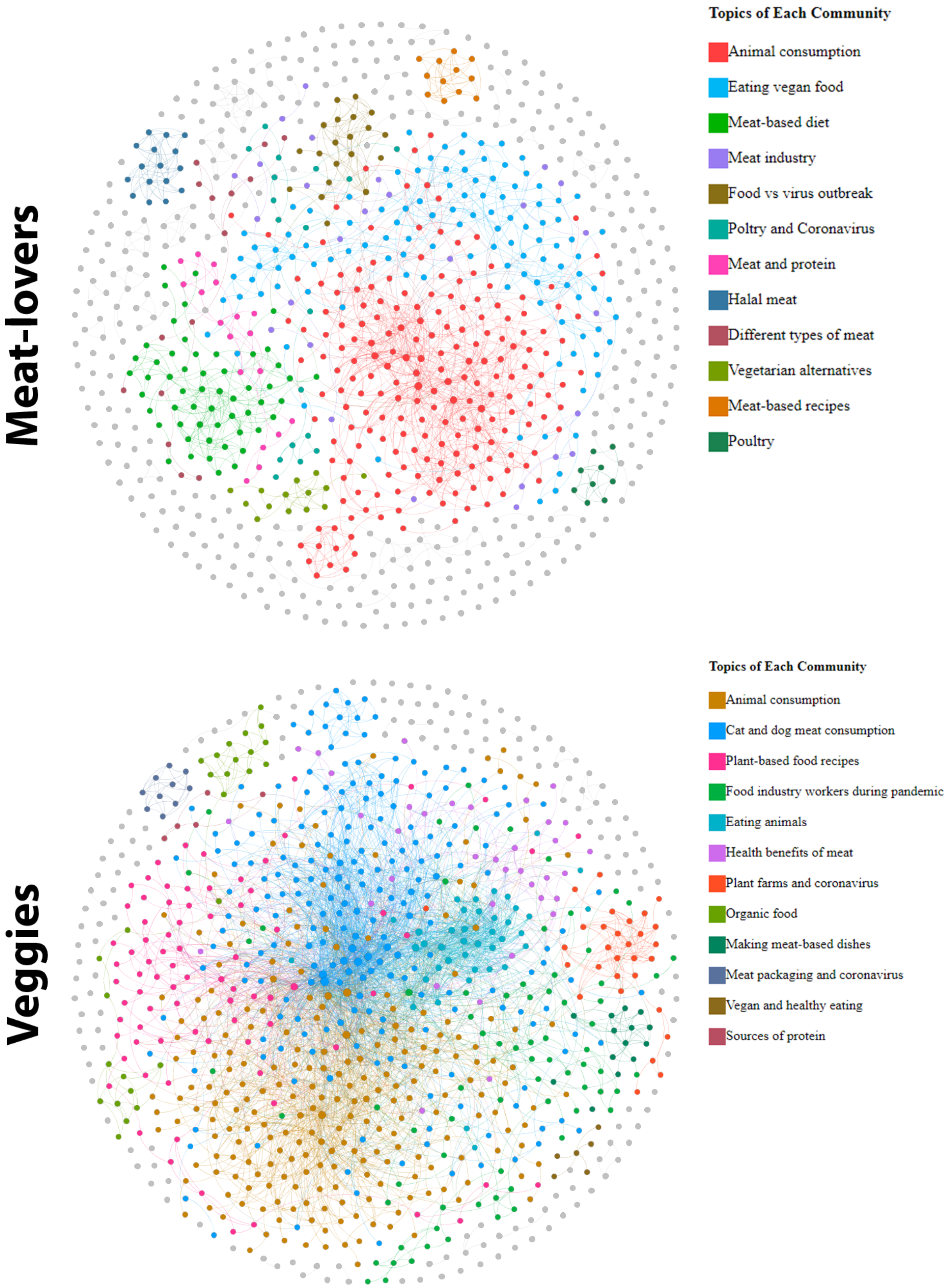

Figure 18: Louvain communities and their topics in pandemic period.

4.6 Emotions within tweets

The study analyzes the emotions expressed in the tweets to determine how the attitude of the people in food-related tweets on food have changed over time. To conduct this analysis, the joined dataset of tweets and unique users are used:

1. The tweets are divided into meat-lover and veggie tweets (based on the preference of the user).

2. Using Syuzhet library, which was also used previously to predict sentiment scores, the emotions of each tweet are predicted. The output was a dataset where each row represented one tweet and 10 columns each represented one emotion (as shown in the Y-axis of Figures 19 and 20). Each column had a value of 0 or 1 indicating whether each tweet had a certain emotion or not.

3. The scores in each column are summed to show how many tweets contained each emotion.

4. Using those numbers, percentage of tweets containing each emotion is calculated.

5. Using Ggplot2 library, Figures 19 and 20 are created.

According to the Figure 19, meat-lovers have not shown significant changes in their emotions. Out of 10, 7 of them increased or decreased slightly, but not enough to be significant. They have experienced a small increase in trust, anticipation, as well as some fear.

Veggies, on the other hand, experienced more negative changes. According to Figure 20, tweets of veggies had a significant decrease in joy and positivity and a significant increase in fear and negativity. In addition, they have shown a small decrease in trust and anger while having a small increase in sadness and anger.

In general, the pandemic has not affected the emotions of meat-lovers as much as veggies. In transition and pandemic periods, veggies have experienced more negative feelings, including higher levels of fear and lower levels of trust. Lastly, it appears that surprise and disgust were the only emotions that had the same result for both groups. The pandemic has not caused significant changes in these emotions.

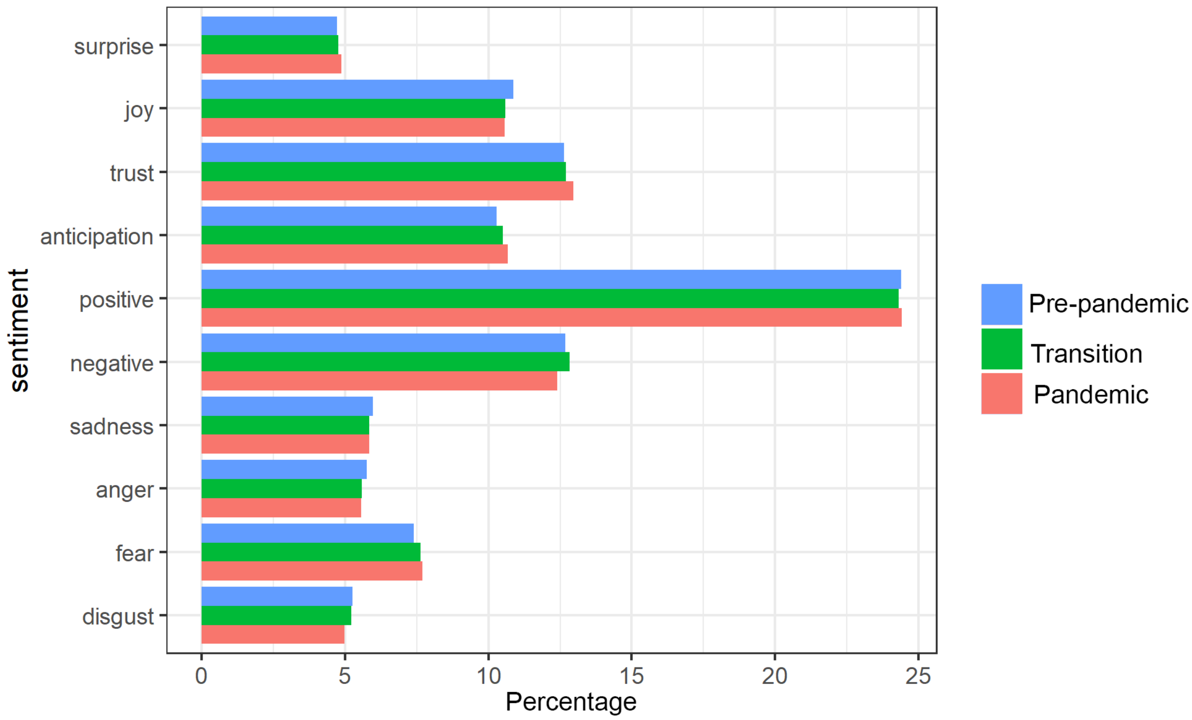

Figure 19: Emotion analysis of meat-lovers.

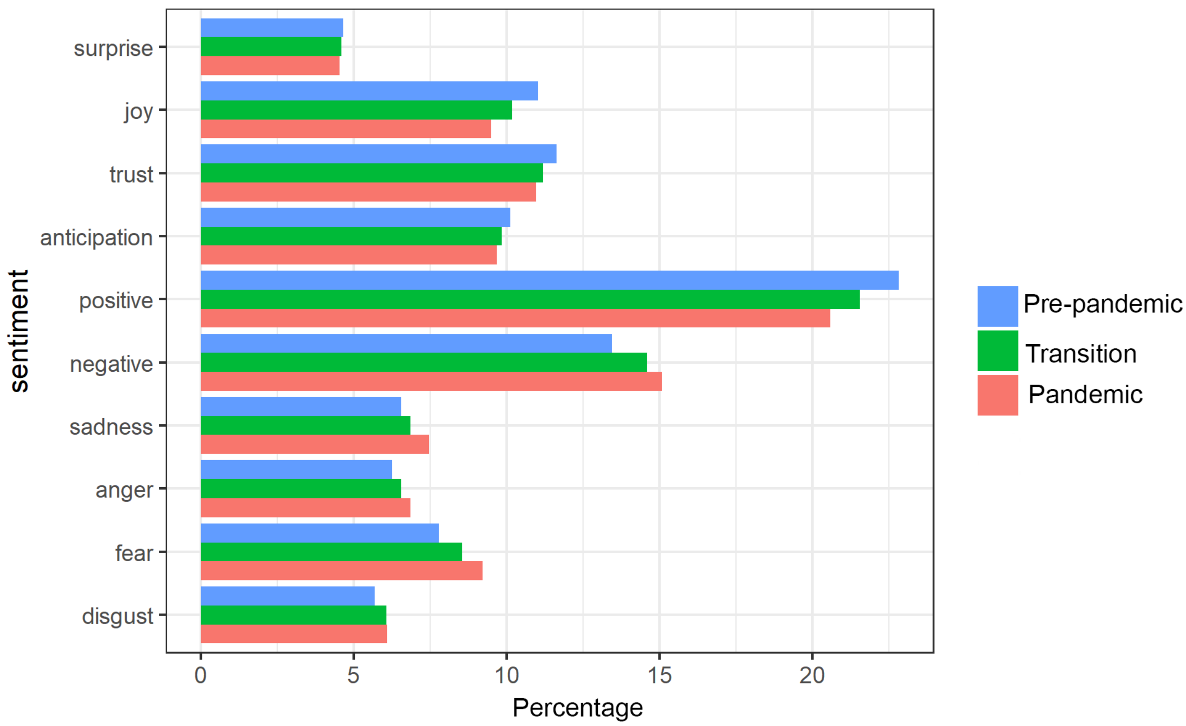

Figure 20: Emotion analysis of veggies.

4.7 Word shift within groups

While Section 4.6 qualitatively analyzes how the positivity of the tweets has changed over time, it is also possible to quantitatively show which words have caused the most change. This is done with Word Shift (WS) graphs introduced by Gallagher et al [44]. While there are several types of WS graphs, this study focuses on Weighted Average Shifts that use sentiment scores of words as weights and calculates the impact on positivity. The shift score is calculated using the following formula

$$\delta\phi = \sum_{\tau \in \mathcal{T}} \phi_\tau (p_\tau^{(2)} - p_\tau^{(1)}) = \sum_{\tau \in \mathcal{T}} \delta\phi_\tau$$

where $p_\tau^{(1)}$ is the probability of a word τ being in the 1st text, $p_\tau^{(2)}$ is the same in 2nd text and ϕ_τ is the weight of that word. In this study, ϕ_τ is the sentiment and $(p_\tau^{(2)} - p_\tau^{(1)})$ is the frequency of the word.

To generate the graphs:

1. Tweets are divided based on period and dietary preference. For this analysis, the difference between only pre-pandemic and pandemic periods is considered.

2. Cleaned texts of the tweets are loaded into python using Pandas library.

3. A new dictionary is generated for each period and preference to show the frequency of each word.

4. The dictionaries are entered into the word shift generator. For this analysis, shifts within the same groups are examined.

5. labMT sentiment library is entered as the weight input [45].

The results of the analysis are shown in Figures 21 and 22. The figures are divided into 2 vertical sections and contain 4 color codes. The X-axis shows that the words on the right side contribute towards the positive shift and the ones on the left side towards the negative shift. The color codes explain how each word affects the shift:

- **Bright yellow:** positive words that have increased in frequency.
- **Violet:** negative words that have decreased in frequency.
- **Light yellow:** positive words that have decreased in frequency.
- **Blue:** negative words that have increased in frequency.

According to Figure 21, the WS score of meat-lovers has dropped by 0.01, showing almost no change in their sentiment. A closer look at the graph shows that this was due to less frequent use of positive words such as Christmas, and more frequent use of negative words such as virus. According to Figure 22, WS scores of veggies have seen small, but relatively bigger drop in comparison to meat-lovers. The biggest contributors to this change were an increase in negative words such as virus, death, sick, and a decrease in positive words such as Christmas, new, food, delicious. The detailed investigation of the words in Figure 22 shows which words might have caused an increase in negativity and fear shown in Figure 19

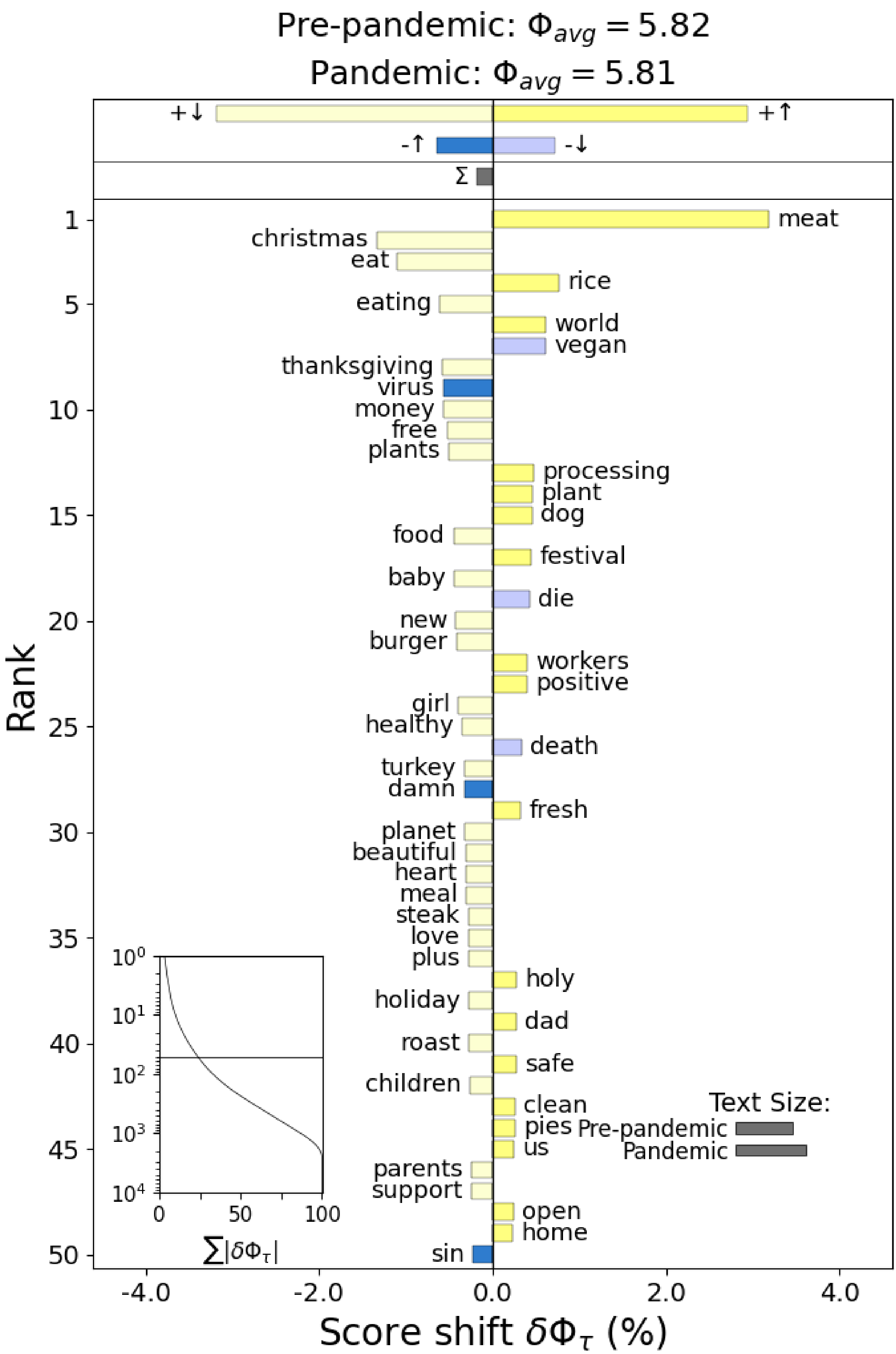

Figure 21: Word Shift graph of meat-lovers.

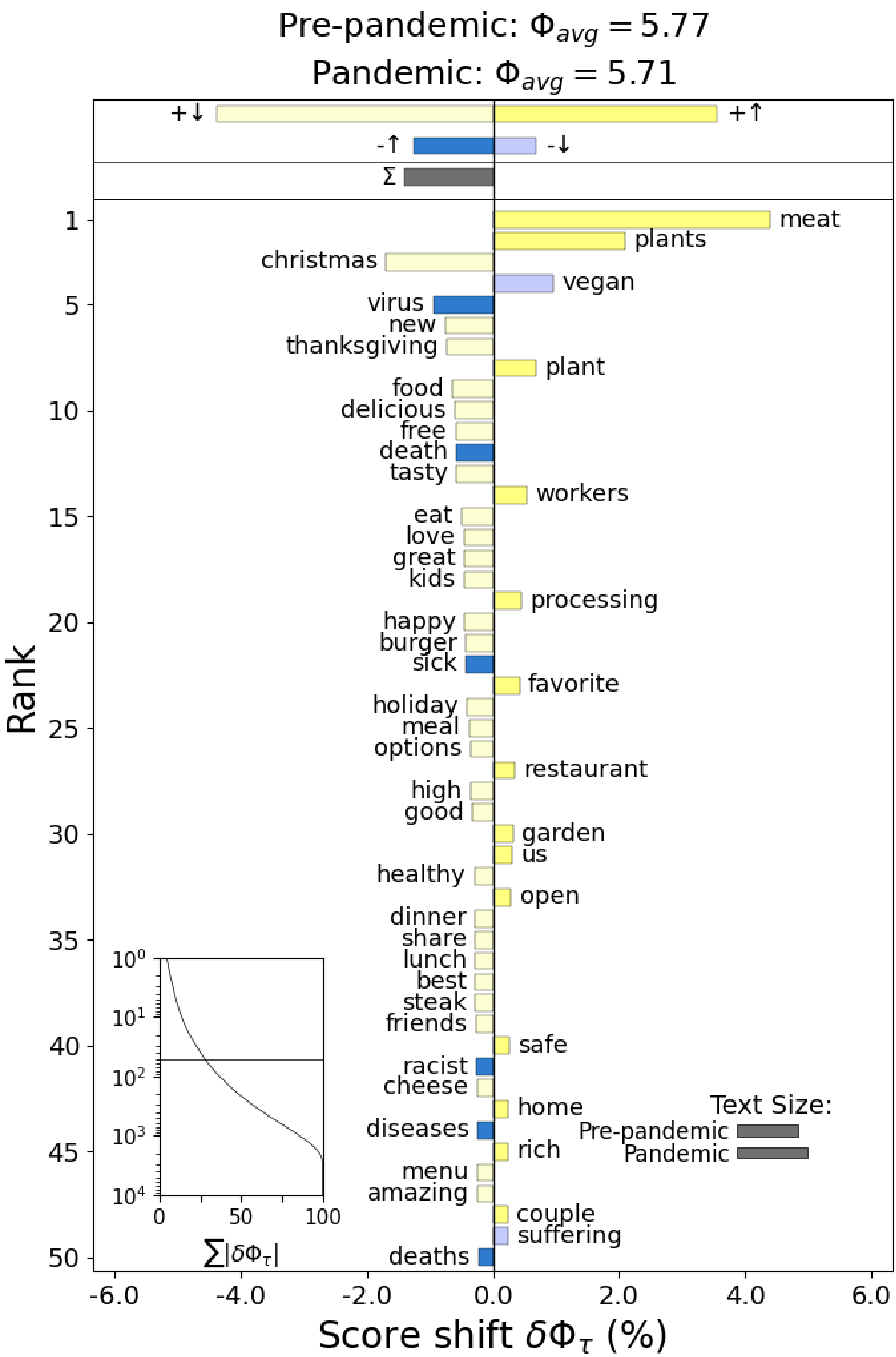

Figure 22: Word Shift graph of veggies.

5 Discussion

This chapter aims to connect the results, as well as discuss the strength and weaknesses of the study. Section 5.1 dives into the results, connects the dots, and depicts the bigger picture. Section 5.2 discusses the limitations of the study. Section 5.3 provides recommendations for future studies.

5.1 Key findings

1. **Change in popularity of dietary preference:** There was not much difference in popularity of one group of people, either based on gender or preference. Some groups may have had a smaller lead during the transition period, but, during the pandemic, meat-lovers and veggies, and males and females had the same level of popularity of tweets (Figures 1 - 7).

2. **Change in popularity of topics:** During all three periods, people have talked about animal consumption, diet, and food recipes. With the pandemic, tweets on animal consumption became either more negative or more positive, while losing popularity. Tweets on diet and health had more neutral ideas over time, and their popularity increased during the pandemic. Tweets about food recipes, on the other hand, have not experienced a significant change in sentiment or popularity (Figures 8 - 11).

3. **Change in communication between the groups:** In comparison to meat-lovers, veggies were less in contact with the other group during pre-pandemic and transition periods. This trend has reached its peak during the transition, but both groups became heterogeneous and more in contact with one another during pandemic (Figures 12 - 15). The density of networks in Figures 16 - 18 further proves it. However, the RWC calculations show that the changes are not significant.

   In addition, the general topic modeling predicted animals, diets, and recipes were the hot topics across periods. However, these figures show that over time, on a community level, the pandemic shifted the focus from discussing the health benefits of each type of food to safety and standards of food production, as well as the alternative options. These findings could indicate that the pandemic has significantly affected the opinions of meat-lovers on meat making them second-guess the benefits and pleasure of meat consumption while giving more reasons for veggies to promote switching to plant-based food.

   Furthermore, topic modeling on communities has discovered an off-topic finding. Once the lockdowns started around the world, people began connecting hypothetical dots on Twitter. It reached to a point where tweets contained odd racial stereotypes regarding consumption of bat and dog meat.

4. **Change in emotions:** Figures 19 and 20 show that, in comparison to meat-lovers, veggies have shown more negative changes in their emotions. Although the previous results show

that meat-lovers have experienced changes in opinion on meat, the pandemic has not affected how they feel about different types of food. In the meantime, it has made veggies more skeptical of meat consumption. Figures 21 and 22 further support the results and show which words may have caused the changes in the emotions.

5.2 Limitations

The limitations of this study can be broken down into 3 perspectives - data, software, and hardware, some of which can be analyzed from multiple perspectives.

1. **Data limitations:** The data used in this thesis was not enough to conduct more thorough analyses. For instance, there aren't enough quantifiable variables to conduct more quantitative analyses and statistical tests.

2. **Software limitations:** The data derived from the collected data is not absolutely accurate, since it's a prediction. For example, Twitter does not provide the gender of its users, and the preference is only calculated based on sentiment score. With improved versions of prediction algorithms, more accurate data can be derived from the tweets dataset.

3. **Hardware limitations:** The limitations on hardware affect both the data and the software. Working with large datasets requires faster machines in order to have the capability to both load and analyze them. Due to the technical limitations of the machine on which the analyses were done, pre-pandemic and pandemic periods were limited to 3 months of data, both of which occupied at least 1.5 GB of space on the drive and consumed almost 16GB of RAM during analyses.

   In addition, insufficient hardware made some analyses slow while making others nearly impossible to complete on time. For example, the emotion analysis took around 3.5 hours to complete for one preference in one time period. Meanwhile, the study had to use simple filtering methods for determining whether a tweet was about meat or plant-based food because the system kept running out of usable memory when utilizing more sophisticated methods such as bi-gram analysis.

5.3 Future directions

1. **Limitations:** For further studies, data, software, and hardware limitations should be taken into account to accommodate for analysis of larger datasets using more sophisticated methods.

2. **Social media platform:** This study has answered the research question from Twitter's perspective only. While different social media platforms have different policies and restrictions on the usage of their data, with proper permissions, it should be possible to conduct the same analyses or more from other platforms' perspectives.

3. **COVID-19 and the social media:** For further studies, the scope can be shifted from food directly onto COVID-19 and the effects of it on online communications. As discovered in Section 4.5, tweets related to food in the Pandemic period contained various topics on the origins of the virus, relation to meat, as well as, food-related stereotypes. Using similar methods, it is possible to analyze the effect of COVID-19 on social media, such as on the spread of fake facts, mockery or promotion of stereotypes, and hateful messages via online communication.

6 Conclusion

In summary, this study fills the gap in the academic literature by analyzing the impact of a pandemic on food consumption from the perspective of the consumers. In this case, it focuses on how the COVID-19 pandemic affected the consumption of meat and plant-based food by analyzing the Twitter data. The academic literature contains studies on using Twitter and other social media data for finding food consumption and nutrition patterns, though there is no extensive study on comparison of meat and plant-based diets. In addition, the literature analyzes the impact of a pandemic on food only from the production perspective and focuses on the supply chain. This study fills such gaps in the literature.

In order to conduct the analyses, Twitter data regarding meat and plant-based food was gathered from November 2019 to June 2020. The data was divided into Pre-pandemic (November 2019 - January 2020), Transition (February - March 2020), and Pandemic (April - June 2020) periods. Additionally, the users were grouped into meat-lovers and veggies, based on their dietary preference, which was acquired by analyzing their positive and negative opinions on each type of food. The analyses were done on each time period and dietary preference and the changes were observed by comparing the results.

Based on the analyses, neither of the groups had a clear lead in popularity in any time period. Tweets about diets became more popular, while tweets about animals began to contain significantly more positive and negative opinions during the pandemic. Meat-lovers and veggies were relatively isolated groups in terms of being in contact with each other. The isolation reached its peak during the transition period, but both groups became more in contact with one another during the pandemic, though the statistical analysis shows the change is small. Additionally, although the health aspect of both diets was discussed before the pandemic, users began to put more emphasis on the safety of meat and its relation to the virus during the pandemic. Lastly, the analysis showed that the pandemic has increased the levels of negativity, anger, fear, and sadness in the tweets of veggies, while its effect on meat-lovers was insignificant.

The study focuses only on Twitter and analyzes the impact of COVID-19 only on food consumption. Using this study as a reference point, future studies can direct the focus towards other social media platforms and find how COVID-19 affected food consumption on different platforms. In addition, the focus can be shifted from food to other topics that can be analyzed through text such as hate speech. In short, the methodology of this study can be used for analyzing the impact of a pandemic on a wide range of topics.

References

[1] Yung-Ming Li, Lienfa Lin, and Shih-Wen Chiu. Enhancing targeted advertising with social context endorsement. *International Journal of Electronic Commerce*, 19(1):99–128, 2014.

[2] Huy Quan Vu, Gang Li, Rob Law, and Yanchun Zhang. Tourist activity analysis by leveraging mobile social media data. *Journal of travel research*, 57(7):883–898, 2018.

[3] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

[4] food hashtag on instagram: Photos and videos.

[5] Yelena Mejova, Sofiane Abbar, and Hamed Haddadi. Fetishizing food in digital age:# foodporn around the world. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, 2016.

[6] Thanh-Trung Phan and Daniel Gatica-Perez. Healthy# fondue# dinner: analysis and inference of food and drink consumption patterns on instagram. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pages 327–338, 2017.

[7] Daniel Fried, Mihai Surdeanu, Stephen Kobourov, Melanie Hingle, and Dane Bell. Analyzing the language of food on social media. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 778–783. IEEE, 2014.

[8] Yelena Mejova, Hamed Haddadi, Anastasios Noulas, and Ingmar Weber. # foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health 2015*, pages 51–58, 2015.

[9] Jiawei Li, Qing Xu, Raphael Cuomo, Vidya Purushothaman, and Tim Mackey. Data mining and content analysis of the chinese social media platform weibo during the early covid-19 outbreak: retrospective observational infoveillance study. *JMIR public health and surveillance*, 6(2):e18700, 2020.

[10] David Laborde, Will Martin, Johan Swinnen, and Rob Vos. Covid-19 risks to global food security. *Science*, 369(6503):500–502, 2020.

[11] Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, 2011.

[12] Ross Joseph Gore, Saikou Diallo, and Jose Padilla. You are what you tweet: connecting the geographic variation in america's obesity rate to twitter content. *PloS one*, 10(9):e0133505, 2015.

[13] Emma J Lea, David Crawford, and Anthony Worsley. Public views of the benefits and barriers to the consumption of a plant-based diet. *European journal of clinical nutrition*, 60(7):828–837, 2006.

[14] Malou FS Reipurth, Lasse Hørby, Charlotte G Gregersen, Astrid Bonke, and Federico JA Perez Cueto. Barriers and facilitators towards adopting a more plant-based diet in a sample of danish consumers. *Food quality and preference*, 73:288–292, 2019.

[15] Dennis Lawo, Katharina Litz, Christina Gromov, Hannah Schwärzer, and Gunnar Stevens. Going vegan: The use of digital media in vegan diet transition. In *Proceedings of Mensch und Computer 2019*, pages 503–507. 2019.

[16] Şeyma Esin Erben and Jale Balaban-Sali. Social media use of vegan activists in turkey. *Global Media Journal*, 6(12):75–88, 2016.

[17] Kerstin Denecke and Wolfgang Nejdl. How valuable is medical social media data? content analysis of the medical web. *Information Sciences*, 179(12):1870–1880, 2009.

[18] Kyle W Prier, Matthew S Smith, Christophe Giraud-Carrier, and Carl L Hanson. Identifying health-related topics on twitter. In *International conference on social computing, behavioral-cultural modeling, and prediction*, pages 18–25. Springer, 2011.

[19] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter catches the flu: detecting influenza epidemics using twitter. In *Proceedings of the 2011 Conference on empirical methods in natural language processing*, pages 1568–1576, 2011.

[20] Yiheng Zhou, Numair Sani, Chia-Kuei Lee, and Jiebo Luo. Understanding illicit drug use behaviors by mining social media. *arXiv preprint arXiv:1604.07096*, 2016.

[21] Lorenzo Betti, Gianmarco De Francisci Morales, Laetitia Gauvin, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. Detecting adherence to the recommended childhood vaccination schedule from user-generated content in a us parenting forum. *PLOS Computational Biology*, 2021.

[22] Alessandro Cossard, Gianmarco De Francisci Morales, Yelena Mejova, and Daniela Paolotti. *5th International Conference on Computational Social Science*. 2019.

[23] Alessandro Cossard, Gianmarco De Francisci Morales, Kyriaki Kalimeri, Yelena Mejova, Daniela Paolotti, and Michele Starnini. Falling into the echo chamber: the italian vaccination debate on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 130–140, 2020.

[24] Amira Ghenai and Yelena Mejova. Fake cures: user-centric modeling of health misinformation in social media. *Proceedings of the ACM on human-computer interaction*, 2(CSCW):1–20, 2018.

[25] Matheus Araujo, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 2017 ACM on Web science conference*, pages 253–257, 2017.

[26] Claudia Wagner, Philipp Singer, and Markus Strohmaier. Spatial and temporal patterns of online food preferences. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 553–554, 2014.

[27] Tomasz Kusmierczyk and Kjetil Nørvåg. Online food recipe title semantics: Combining nutrient facts and topics. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 2013–2016, 2016.

[28] Yong-Yeol Ahn, Sebastian E Ahnert, James P Bagrow, and Albert-László Barabási. Flavor network and the principles of food pairing. *Scientific reports*, 1(1):1–7, 2011.

[29] Robert West, Ryen W White, and Eric Horvitz. From cookies to cooks: Insights on dietary patterns via analysis of web usage logs. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1399–1410, 2013.

[30] Thiago Silva, Pedro Vaz De Melo, Jussara Almeida, Mirco Musolesi, and Antonio Loureiro. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.

[31] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. When personal tracking becomes social: Examining the use of instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on human factors in computing systems*, pages 1674–1687, 2017.

[32] Munmun De Choudhury, Sanket Sharma, and Emre Kiciman. Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th acm conference on computer-supported cooperative work & social computing*, pages 1157–1170, 2016.

[33] Rumi Chunara, Lindsay Bouton, John W Ayers, and John S Brownstein. Assessing the online social environment for surveillance of obesity prevalence. *PloS one*, 8(4):e61373, 2013.

[34] Sanket S Sharma and Munmun De Choudhury. Measuring and characterizing nutritional information of food and ingestion content in instagram. In *Proceedings of the 24th International Conference on World Wide Web*, pages 115–116, 2015.

[35] Daniel Kershaw, Matthew Rowe, and Patrick Stacey. Towards tracking and analysing regional alcohol consumption patterns in the uk through the use of social media. In *Proceedings of the 2014 ACM conference on Web science*, pages 220–228, 2014.

[36] Michael J Widener and Wenwen Li. Using geolocated twitter data to monitor the prevalence of healthy and unhealthy food references across the us. *Applied Geography*, 54:189–197, 2014.

[37] Aron Culotta. Estimating county health statistics with twitter. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1335–1344, 2014.

[38] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3197–3206, 2015.

[39] Yelena Mejova and Víctor Suarez-Lledó. Impact of online health awareness campaign: Case of national eating disorders association. In *International Conference on Social Informatics*, pages 192–205. Springer, 2020.

[40] Ingo Feinerer. Introduction to the tm package, text mining in r, 2018.

[41] Topic models, 2021.

[42] Gephi, 2018.

[43] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.

[44] Ryan J Gallagher, Morgan R Frank, Lewis Mitchell, Aaron J Schwartz, Andrew J Reagan, Christopher M Danforth, and Peter Sheridan Dodds. Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Science*, 10(1):4, 2021.

[45] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and Christopher M Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, 6(12):e26752, 2011.

Licence

Non-exclusive licence to reproduce thesis and make thesis public

I, **Musa Guliyev**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   The Effects of COVID-19 on Consumption of Animal and Plant-based Food: An Analysis of Twitter Data,

   supervised by Rajesh Sharma.

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Musa Guliyev
14/05/2021