NLP midterm

December 30, 2023

1 Podcast Topic Modeling at the Episode Level from Transcripts

1.0.1 UoL NLP Midterm

by *Richard Neuboeck*

1.1 Introduction

A word before I begin the requested content: This report sports many more pages than anticipated. I'm sticking to the given word limits for each section but as my analysis is based on interpretation of the visual output I left many - not all - of the visuals as appendix in. See the note below for details. Thank you!

1.1.1 Problem Area

Since the year 2000 podcasts have grown from a curiosity into a mainstream media. With currently about 4.3 million podcasts available and 64% of people in the US who listend at least once, on demand media overtook radio recently. Naturally a platform as huge attrackts advertisers. As enourmous as the number of podcasts and listeners is the amount of ad revenue which exceeded $2 billion in 2023.

Advertising agencies match advertisers and podcasts by trying to align their interests and brand identities. There are different approaches depending on the agency ranging from self service (ie AudioGo) to full campaign planning and implementation by an agency (ie. Content Allies). What they all have in common is the need to classify podcasts in some way to be able to facilitate the matching process.

I'm proposing a topic modeling approach based on automatically produced podcast transcripts. Generally the podcast format is audio with some offering additional video content. Each episode also has metadata containing keywords and description. In *Topic Modeling on Podcast Short-Text Metadata* Valero et al. has attempted to compare different classification methods on podcast (not episode level) metadata. Their approach was in large successful but they also noted that metadata provided by the content creators is unreliable and noisy making it difficult to extract usable information without intervention.

Unfortunately automatically generated podcast transcripts come with problems too. A transcript generated or at least verified by a human is extremely costly and therefore unfeasable for large numbers of podcasts. As podcast hosts are not always professional speakers and also different people might speak at the same time automatically generated transcripts tend to have errors recognizing distinct words. However with recent advances in voice recognition the neural net based Whisper system offered by OpenAI approaches human level detection. There is still the problem that

even human level detection (at least with the model I was able to run on my machine) makes a considerable amount of mistakes and without human intervention those mistakes are not corrected. However the results are good enough for an episode level based analysis.

1.1.2 Objective

Advertisement agencies have podcasters under contract to provide their service but they rely on metadata provided by the podcast hosts describing the shows. Even though there are attempts by hosting services to constrain the variability in metadata it usually is noisy and iadequate with the content description of questionable qaulity and the categorization generally unreliable as described in *Current Challenges and Future Directions in Podcast Information Access"* by Jones et al.. Therefore agencies need to manually curate their own databases containing the information they need to match potential advertisers with shows. This is a time consuming and costly process.

A topic analysis of automatically generated podcast transcripts at the episode level will reveal themes within a show and also commonalities between shows. Depending on the necessities the algorithm can be tuned to discover a large number of topics or a few broader genres. But in each case the discovery step is solely based on the actual podcast content side stepping the metadata issues Valero et al. had in their approach.

With the below described pipeline at hand an advertisement agency could run a topic analysis in a very short amount of time with minimum human oversight that will produce an overview of the general direction of a shows content. Since having humans doing this work is usually expensive this automatism will save money and in the long run will provide better advertiser/podcast matches.

The pipeline below is obvioulsy only a prototype and would need (much) refinement for an actual deployment.

1.1.3 Dataset

The data set is 38MB in size, consists of 663 UTF-8 encoded text files (transcripts) from 5 podcasts within a period of three years. Each file holds the transcript of one podcast episode. There is one directory per show as shown below:

```
podcasts/
   AWeekOfMornings
   JJHO
   Shmanners
   WgT
   YoureDeadToMe
```

File names within the show directories have the structure of *release-date episode title.mp3.txt*. In example *2022.05.25 My Legal Pony (RERUN).mp3.txt*.

The podcasts are popular shows with regular releases and even though I'm listening to most of them and they are mostly commedies for this analysis it can be said that they have been randomly chosen.

The five podcasts in no particular order are:

- A week of Mornings : A compilation of the WRSI radio host Monte Belmonte's morning shows without proprietary content (music, commercials).

- Judge John Hodgman : A comedic court show adjudicating real-life disputes in a fake court room.
- Shmanners : Wife and husband discussing extraordinary etiquette for ordinary occasions (paraphrased from their podcast).
- We Got This with Mark and Hal : Small debates definitively settled by two actors.
- You're dead to me : Comedy podcast about history.

Data Origin Sources for the *JJHO*, *WgT* and *Shmanners* podcasts are Maximumfun.org (using simplecast as provider), the BBC (using their own resources to provide the podcasts) for *You are dead to me* and *A Week of Mornings* is self published via Soundcloud.

The podcasts are licensed to their respective networks or person but are freely available to listen to. Further processing of the podcast data is not mentioned on the Maximumfun network pages and the self published podcast page. The BBC is very clear that it is not allowed but they offer special access for educational institutions. In any case I inquiered to use specific podcast episodes on each network. The BBCs response was swift and can be seen below. Maximum Fun and Monte Belmonte did not respond so far but I'm still going ahead with the analysis counting on a delayed reply since this paper is for educational purposes only and will remain unpublished anyway.

> Hi Richard,
>
> Thank you for contacting BBC Enquiries.
>
> I understand you wish to transcribe You're Dead To Me podcast part of your language analysis research. This will feature part of your unpublished project for the University of London. With regard to all rights owned by the BBC in this programme, we would have no objection to it being transcribed, purely for these educational purposes.
>
> This does not apply to publishing material outside of the education setting.
>
> Thank you for raising this request with the BBC, we wish you all the best with your research project.
>
> Kind regards, Ciaran Black BBC Enquiries Team

The download of the episodes using getpodcast happend only once so no out of the ordinary strain was put on the hosting provider.

1.1.4 Evaluation Methodology

The topic modeling algorithms used are statistics based and do not necessarily lead to topics consisting of words that seem to belong together with humans in mind. From a statistics point of view these "random" topics may still represent valid categories. Yet the first evaluation step is human perception based as the top words for each topic are compared. Furthermore a major part is a visual analysis of the output of pyLDAvis. This library uses the topic model and creates a lower dimensional representation that can be interactively explored. One of the benefits is that it shows the coverage of a topic as size of a circle and also topic overlaps (when circles overlap). Going forward the visual inspection of the model output will be the basis for the evaluation. The ideal would be to find the highest number of topic with little to no topic overlaps.

The two models used, namely Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) do not allow a direct comparison on the model level. However another way to

evaluate the generated topics themselves would be to calculate a coherence score. This would be a measure of how closely related terms within a topic are. But it seems *sklearn* doesn't provide a model for that. *Gensim* can calculate these values but only from a *gensim* LDA/NMF model since *gensim* and *sklearn* are not directly compatible.

NB: the documentation of the tmtookit library states that it should be able to handle sklearn model inputs to calculate coherence scores but the output was not significant as it only ranged from 0.6 to 0.7 no matter how the model parameters have been tuned. This library was therefor not used.

Like previous research suggested the main points for evaluating the model output is visually inspecting the intertropic distance map, topic frequency pie charts and topic terms.

1.2 Implementation

1.2.1 Preprocessing

Steps outside of the Jupyter notebook have been:

- find the RSS feed of each podcast (websearch)
- configure getpodcast and download episodes (as MP3 files) and build the file and directory structure

```
(getpodcast.venv) [hawk@den getpodcast]$ python getpodcasts.py --run
```

- use a shell script and Whisper to transcribe audio to text

```
(whisper.venv) [hawk@den podcasts]$ for i in *; do cd $i; echo $i; find . -type f -name '*.mp3
```

All the prepcrocessing steps in this Jupyter notebook are described close to the processing step and within the code cell. The pipeline follows these steps:

- Import podcasts (including some meta data) into a Pandas data frame
- Clean the transcript text
  - Conversion of all characters to lower case
  - Substiution of newlines with one space character
  - Removal of certain dynamic ad inserts using a regular expression. This step has been integrated in the cleaning step after an inital analysis attempt that resulted in skewed output since all ads were holiday based (due to the time of the download).
  - Fixing all contractions
  - Word tokenize the text
  - Remove stop words (list build from NLTK stop words, the list from the lectures and a manually currated list after several analysis runs. This step also ensures that only words containing letters from a-z are included (no digits or symbols)
  - POS tag all words and only keep nouns, also for simplification of the classification
  - Stem the word list to simplify the classification

1.2.2 Baseline performance

There is no heuristic to predict a meanigful baseline for the ideal number of topics. LDA itself is used with a "good guess" as number of topics to start with and as basis to improve upon.

Using the meta data keywords of the podcasts three out of the five podcasts are categorized as "comedy", one is "news" and one is "history". One could therefore argue that three should be the

minimum number of topics. However the analysis is not on the show level but episode level and each episode no matter if it's comedy or history can explore different topics so it's likely that the number of topics is higher.

1.2.3 Classification approach

Two models have been compared:

- Latent Dirichlet Allocation (LDA)
- Non-Negative Matrix Factorization (NMF)

The most popular model for unsupervised topic modeling is LDA. It is a generative probabilistic model meaning that it assumes a document is built out of latent topics and each topic consists of a certain set of words. The distribution of words and topics is what the algorithm analyses but to do that it needs a number of topics to be specififed beforehand. Since the number of topics is usually unknown sever analysis iterations with different topic numbers need to be done to find an ideal topic count.

NMF in another unsupervied learning algorithm. While LDA is mostly used for text analysis NMF also has applications in image and audio analysis. The idea is to find two lower raking matrices that combined approximate the higher dimensional source matrix. One of the lower dimensional matrices represents the extracted features the other the coefficients. If the dataset is very large NMF gets computationaly very expensive in comparison to LDA.

The two models take as input a vectorized form of the text documents. Two vectorization methods have been used:

- Term frequency (TF) : TF counts each unique word in a (set of) document(s) and divides by the total number of terms.
- Term frequency - inverse document frequency (TF-IDF) : combines the term frequency (TF) with the inverse document frequency (IDF). IDF represents the importance of a term by taking the logarithm of the number documents divided by the number of documents containing a specific term. The combination TF-IDF emphasises terms that occure often within a document but are rare in the document set.

A general note on the combinations of vectorizers and models: LDA performs very well with the count vectorizer but does not do so well using the TF-IDF vectorizer whereas NMF seems to perform better in combination with the tfidf vectorizer. The problem with LDA and TF-IDF seems to be that the algorithm found the same word combinations for multiple topics leading to many equal word combinations for topics no matter the given topic count. LDA/TF-IDF was therefore only used up to topic count 4.

Allowing a larger n-gram range for the TF-IDF vectorizer had also some visible benefits but needs considerable more computational power. Due to limited computational resources available single words, bi-grams and tri-grams have been explored.

1.2.4 Coding Style

Unless otherwise marked in the cell the code is my own built from prior experience, reading documentation and tutorials.

1.3 Conclusion

1.3.1 Evaluation

The model output is evaluated based on the visual inspection of the intertropic distance map (created with pyLDAvis) and the topic frequency over all and across episodes. The meaningfulness of the word combination of the top words within a topic is in large classes not always a significant feature for inspection.

The baseline of three large and independent topics as output from both models has been visually confirmed. With the exception of the problematic combination of LDA and TF-IDF vectorized data all other model, vectorizer and n-gram combinations yielded the same overall result. A closer look at the topic distribution across shows and episodes show that the LDA model and count vectorizer match one aspect of the keyword description of the podcasts. *A week of mornings* as news recap podcast has its own category. *JJHO* and *WgT* have a lot in common as they are true comedy podcasts. *You're dead to me* is another separate group that identifies it as history podcast. *Shmanners* is a mixture of history and comedy as it truly is. The NMF model using TF-IDF as vectorizer did reveal a better categorization. With the exception of *A week of mornings* all podcasts even *You're dead to me* which is mainly concerned with history but with a comedic twist are identified as mostly comedy. Changing the n-gram range for the vectorizer shows a different topic frequency across the episodes.

Scaling up to four topics shows that those are also completely separated in the distance map. *A week of mornings* with its news content still remains (mostly) a separate group. While the topic distribution in the other four podcasts changes with the LDA model drastically. *JJHO* and also *WgT* content establishes itself as distinct topics that also show up in *Shmanners* which as a listener to the podcasts I can confirm. The NMF model still sees all but *A week of mornings* mostly as having the same topic with a couple of episodes with other content.

At a topic count of five the first topic overlap in the distance map occurs. Less so using the NMF model compared to LDA. Also a first at this level is a clearer distinction using LDA in content between the comedy (and comedy-history) podcasts *JJHO*, *WgT* and *You're dead to me*. *Shmanners* which is concerned about etiquette and presenting it in a comedic way is a combination of all the other topics. NMF shows a more fine grained separation of the comedy podcast content and starts to separate the history podcast.

At level six there is much more overlap visually but the separation into distinct podcast groups continues. The trend that NFM shows a more detailed picture is seen here too. Now *JJHO* splits off from the other comedy podcasts.

The overlap at topic count level seven starts to be larger and even though there is now a clearer separation of the shows the compositional picture starts to get confusing with LDA and low n-gram range TF-IDF NMF. This trend continues in the following level with at first no clear sign of usable information. With increasing topic numbers and at the higher n-gram range for TF-IDF for NMF the pie charts begin to show a new trend. Each show has a distinct topic but still features many epidodes with varying topical content.

1.3.2 Summary and conclusion

Acquiring the data source, transcription and preprocessing worked as planned with multiple minor modifications to the text clean up due to knowledge gained from the following analysis. The idea

of comparing different models and vectorizers turned out to be much more difficult than the initial research suggested. Everything is possible but within the given timeframe there was no attainable way to directly compare the models themselves. A mathematical comparison of the topic coherence would have been a possibility where it not for the lack of knowledge that a student without much insight into this field of research brings. To attain this goal with a different library would have meant to rewrite the whole analysis as the libraries (*sklearn* and *gensim*) are not compatible. However other research (ie. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts) in this area showed that it is not uncommon and even favourable due to the particular insight a human mind brings over a strictly mathematical solution to analyse the topic composition and lower dimensional topic layout in a visual manner.

In that regard it was very insightful that the initial analysis results matched the baseline to the point. Increasing the topic count did at first reveal a clearer distinction between shows which would allow for a very fine grained match of advertiser and podcast even on an episode level. With higher topic counts a recognisable topical direction appeared but the looking at the many smaller topics did not provide more insight. At much higher topic levels the intertropic distance map showed a very interesting feature. Topics tended to group up in certain regions of the map. This could be an indication of a larger trend that the baseline already predicted. Ie. instead of 25 small topics that grouped up into three or four regions, three or four larger topics. In this case I would suggest a follow up analysis using the *gensim* library and a coherence analysis of the topic terms. The coherence is presumably high within each topic but it would be interesting to see the coherence when these close topics are joined into one larger topic compared to the baseline.

As noted in the previous sections there has been research that compared different topic model algorithms obviously with different data sets but none that focused on vectorizers and their parameter tuning too. This showed up some interesting aspects especially using the TF-IDF vectorizer and higher n-gram ranges. Using a data set that has a known number of topics one could further explorer the impact of the n-gram range (and other parameter) tuning on the precision of the topic detection.

Given the raw text files the analysis inside the Jupyter notebook can be re run depending on the hardware within a couple of hours. The transcription process using *Whisper* from MP3 to TXT took about a week on "regular" workstation. Using a *requirements.txt* files to replicate the virtual environment it would be easy to replicate the whole analysis on most modern OSs.

2 Workflow

The following shows the workflow of the anlysis ending in a (very) long section of topic predictions and visualisations.

[1]:
```
# preparing the python environment
import os
from pathlib import Path
import re
import nltk
import contractions
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
import statistics
from nltk.stem.snowball import SnowballStemmer
import csv
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.decomposition import LatentDirichletAllocation, NMF
import warnings
import pyLDAvis
import pyLDAvis.lda_model
import math
import numpy as np
```

[2]:
```
# disable warnings ONLY after testing JUST for nicer visuals
# pyLDAvis throws a lot of warnings otherwise
warnings.filterwarnings('ignore')

# setup of pyLDAvis
pyLDAvis.enable_notebook()
```

[3]:
```
nltk.download('punkt')
nltk.download('stopwords')
# nltk.download('wordnet')
```

```
[nltk_data] Downloading package punkt to
[nltk_data]     /home/mescalin/hawk/nltk_data…
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     /home/mescalin/hawk/nltk_data…
[nltk_data]   Package stopwords is already up-to-date!
```

[3]:
```
True
```

2.1 Preprocessing

[4]:
```
# raw data path relative to the jupyter notebook

# limited dataset for testing
#raw_data = './data_raw/podcasts'
# full dataset
raw_data = './data_raw_full/podcasts'
```

[5]:
```
# prepare an empty dataframe for the podcasts
df = pd.DataFrame(columns=['podcast', 'release', 'title', 'text_raw'])
```

[6]:
```
# iter through all podcast files and fill the data frame
for podcast in Path(raw_data).iterdir():
    # check if the entry is a directory (which it should be)
    if podcast.is_dir():
        # go through all items in the podcast directory
```

```
        for file in Path(raw_data + '/' + podcast.name).iterdir():
            # get the year, month, day and title from the file name using a␣
↪regex
            match = re.match(r'(\d{4})\.(\d{2})\.(\d{2})\s(.+)\.mp3\.txt', file.
↪name);
            # check if a match exists
            if match:
                # if so read the data
                with open(file, 'r') as f:
                    data = f.read()
                # prepare the new dataframe row
                row = {
                    'podcast'  : podcast.name,
                    'release'  : pd.to_datetime(match.group(1) + '-' + match.
↪group(2) + '-' + match.group(3), format='%Y-%m-%d'),
                    'title'    : match.group(4),
                    'text_raw' : data
                }
                # append row to dataframe
                df.loc[len(df)] = row
```

[7]:
```
# test output
df
```

[7]:
```
             podcast    release                                  title  \
0    AWeekOfMornings 2020-07-17     A WEEK OF MORNINGS July 17th 2020
1    AWeekOfMornings 2020-01-03   A WEEK OF MORNINGS January 3rd 2020
2    AWeekOfMornings 2020-07-24     A WEEK OF MORNINGS July 24th 2020
3    AWeekOfMornings 2020-01-23  A WEEK OF MORNINGS January 17th 2020
4    AWeekOfMornings 2020-08-07    A WEEK OF MORNINGS August 7th 2020
..               …          …                                     …
658    YoureDeadToMe 2022-10-07         The Tang Dynasty (Radio Edit)
659    YoureDeadToMe 2022-10-14             Mary Shelley (Radio Edit)
660    YoureDeadToMe 2022-10-14     The Ancient Olympics (Radio Edit)
661    YoureDeadToMe 2022-10-22                  Saladin (Radio Edit)
662    YoureDeadToMe 2022-10-22   The Haitian Revolution (Radio Edit)

                                              text_raw
0    Monty is the coolest, funniest, most handsomes…
1    Monty is the coolest, funniest, most handsomes…
2    Monty is the coolest, funniest, most handsomes…
3    Monty is the coolest, funniest, most handsomes…
4    Monty is the coolest, funniest, most handsomes…
..                                                 …
658  This is the BBC.\nThis podcast is supported by…
659  This is the BBC.\nThis podcast is supported by…
660  This is the BBC.\nThis podcast is supported by…
```

```
661  This is the BBC.\nThis podcast is supported by…
662  This is the BBC.\nThis podcast is supported by…

[663 rows x 4 columns]
```

Methods to read stop word files, generate stop word lists and clean the raw text.

[8]:
```
# create the stemmer
stemmer = SnowballStemmer('english')
```

[9]:
```
# read a CSV file containing stopwords
# return a list of words
# code from the NLP lecture with modifications
def read_in_csv(csv_file):
    with open(csv_file, 'r', encoding='utf-8') as f:
        reader = csv.reader(f, delimiter=',', quotechar='"')
        data_read = [' '.join(row) for row in reader]
    return data_read
```

[10]:
```
# default stop word file and word from the podcasts that should be excluded
stopwords_file_paths = ['stopwords.csv', 'podwords.csv']

# assemble and return a list of stopwords
# code from the NLP lecture with modifications
def get_stopwords(path_list=stopwords_file_paths):
    stopwords = []
    # read a custom file
    for list in path_list:
        stopwords += read_in_csv(list)
    # append the NLTK default stopwords
    stopwords = stopwords + nltk.corpus.stopwords.words('english')
    # stem all the stopwords and add those to the list too
    stemmed_stopwords = [stemmer.stem(word) for word in stopwords]
    stopwords = stopwords + stemmed_stopwords
    return stopwords
```

[11]:
```
# total number of stop words
len(get_stopwords())
```

[11]:
```
1352
```

[12]:
```
# assemble the stop word list
stop_words = get_stopwords()

# method to clean the input text
# takes the string to clean
# returns the cleaned string
def clean_text(data):
```

```
    # check if it's a string
    if isinstance(data, str):
        # convert to lower case and remove leading and trailing whitespace
        data = data.lower().strip()
        # subsitute newlines with space
        data = re.sub(r'\n', ' ', data)
        # remove known dynamic ad inserts that caused problems
        data = re.sub(r'(\. [\w\s\']*?(quick )??break.*?(welcome|are|re)␣
↪back[^\.]*[\.!])', ' ', data, flags=re.I)
        # fix contractions
        data = contractions.fix(data)
        # tokenize the data by word
        data = nltk.word_tokenize(data)
        # remove stop words and make sure it's a word
        data = [ word for word in data if not word in stop_words and word.
↪isalpha() and re.search('[a-z]', word) ]
        # pos tag and only keep NN
        data = [ word for word, tag in nltk.pos_tag(data) if tag == 'NN' ]
        # stem the result
        data = [ stemmer.stem(word) for word in data ]

        return ' '.join(data)
    # if nothing else return an empty string
    return ''
```

[13]:
```
# sanity check the the custom stop word file is interpreted correct
if not 'monti' in stop_words: print('ok')
```

The following cleans the *text_raw* column entries and stores them in a new column named *text_clean*.

[14]:
```
# add a new dataframe column with cleaned up podcast text
df['text_clean'] = df['text_raw'].map(clean_text)
```

[15]:
```
df
```

[15]:

	podcast	release	title	\
0	AWeekOfMornings	2020-07-17	A WEEK OF MORNINGS July 17th 2020	
1	AWeekOfMornings	2020-01-03	A WEEK OF MORNINGS January 3rd 2020	
2	AWeekOfMornings	2020-07-24	A WEEK OF MORNINGS July 24th 2020	
3	AWeekOfMornings	2020-01-23	A WEEK OF MORNINGS January 17th 2020	
4	AWeekOfMornings	2020-08-07	A WEEK OF MORNINGS August 7th 2020	
..	…	…	…	
658	YoureDeadToMe	2022-10-07	The Tang Dynasty (Radio Edit)	
659	YoureDeadToMe	2022-10-14	Mary Shelley (Radio Edit)	
660	YoureDeadToMe	2022-10-14	The Ancient Olympics (Radio Edit)	
661	YoureDeadToMe	2022-10-22	Saladin (Radio Edit)	
662	YoureDeadToMe	2022-10-22	The Haitian Revolution (Radio Edit)	

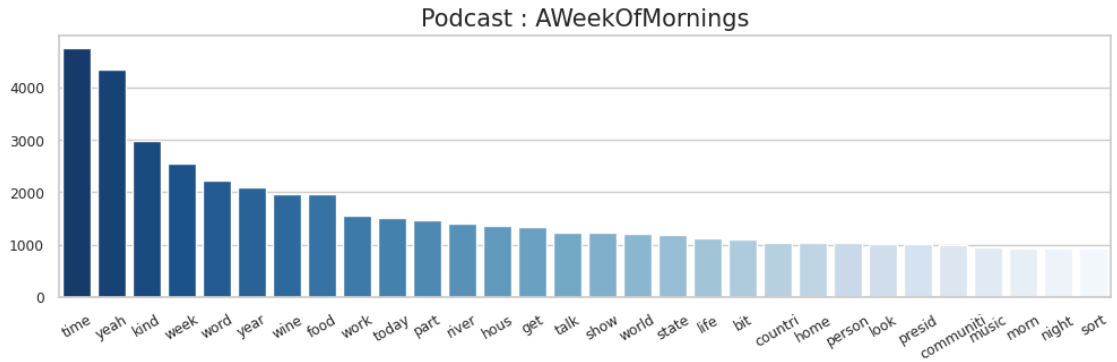
```
                                                text_raw  \
0    Monty is the coolest, funniest, most handsomes…
1    Monty is the coolest, funniest, most handsomes…
2    Monty is the coolest, funniest, most handsomes…
3    Monty is the coolest, funniest, most handsomes…
4    Monty is the coolest, funniest, most handsomes…
..                                                 …
658  This is the BBC.\nThis podcast is supported by…
659  This is the BBC.\nThis podcast is supported by…
660  This is the BBC.\nThis podcast is supported by…
661  This is the BBC.\nThis podcast is supported by…
662  This is the BBC.\nThis podcast is supported by…

                                              text_clean
0    coolest person week compil podcast radio show …
1    coolest person hello valmonti week compil podc…
2    coolest person hello belmonti week compil podc…
3    coolest person week compil podcast radio show …
4    coolest person hello bell week compil podcast …
..                                                 …
658  podcast advertis hello greg radio edit episod …
659  podcast advertis hello greg radio edit episod …
660  podcast advertis hello greg radio edit episod …
661  podcast advertis hello greg radio edit episod …
662  podcast advertis hello greg radio edit episod …

[663 rows x 5 columns]
```

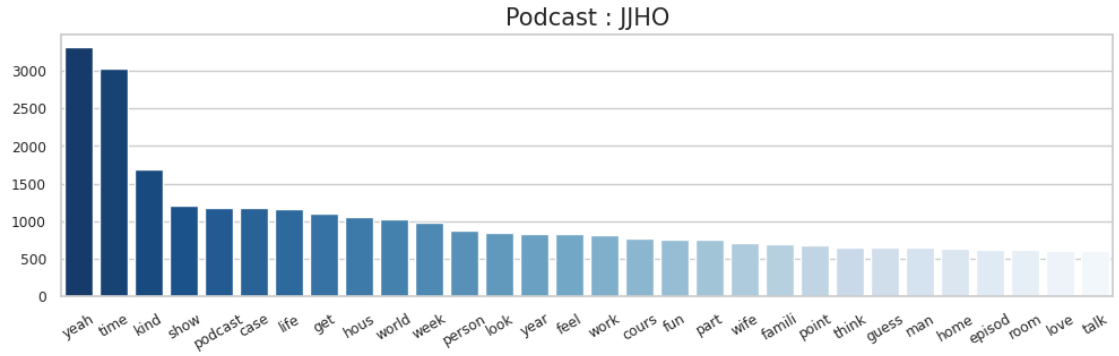
[16]:
```python
# method to do the frequency analysation and plotting
def analyse_and_plot(words, title, pout=False):
    # calc frequency distribution from the most common 50 words
    fdist = nltk.FreqDist(words).most_common(30)

    # convert to pandas series for easier handling
    fdist = pd.Series(dict(fdist))

    if pout:
        print(fdist)

    # seaborn defaults
    sns.set(font_scale=0.8)
    sns.set_style("whitegrid")

    # setting up figure
    fig, ax = plt.subplots(figsize=(12,3))
```

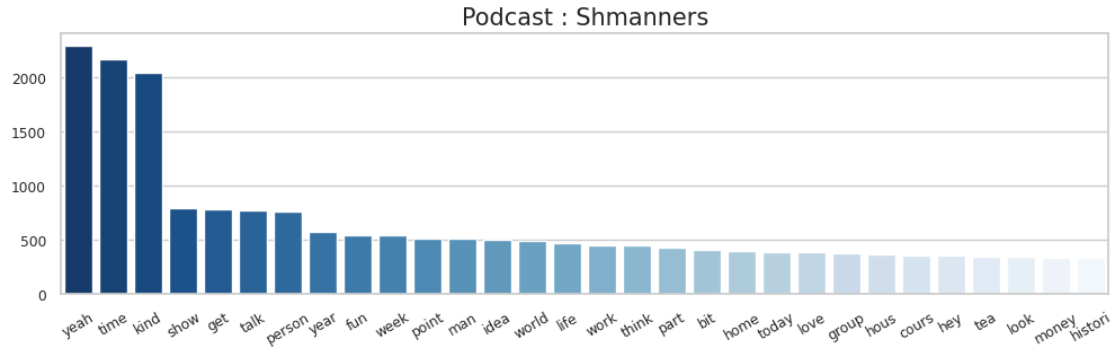
```
    # plotting using seaborn and pandas
    plot = sns.barplot(x=fdist.index, y=fdist.values, ax=ax, palette='Blues_r',␣
 ↪hue=fdist.index, legend=False)
    plot.axes.set_title(title, fontsize=15)
    plot.set(xlabel=None)
    plt.xticks(rotation=30)
    plt.show()
```

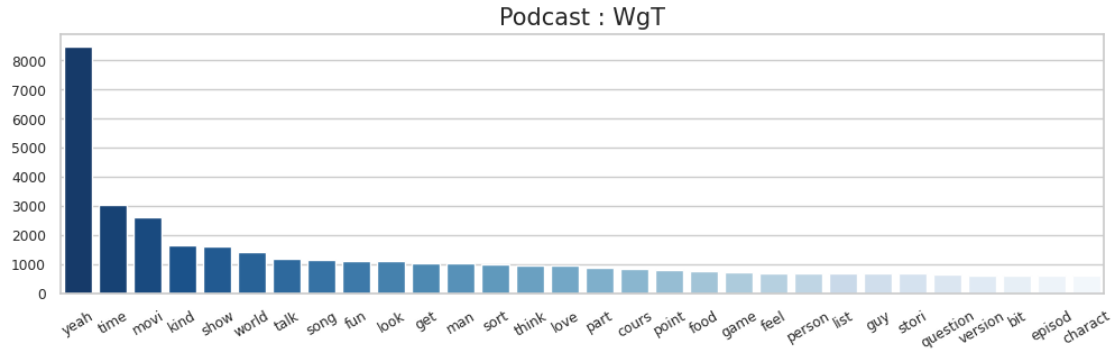
Preliminary analysis for the word frequencies in all podcasts.

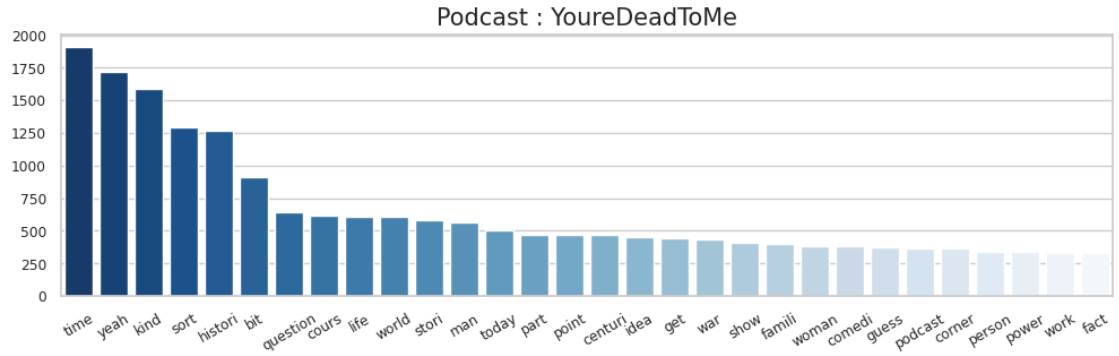
[17]:
```
# get a list of podcast names
podcasts = df['podcast'].unique()

# all words from all podcasts
all_words = []

# iterate over all podcasts
for cast in podcasts:
    # words from this podcast
    words = []

    for index, row in df[df['podcast'] == cast].iterrows():
        words.extend(nltk.word_tokenize(row['text_clean']))

    # add the words to the global list
    all_words.extend(words)

    # analyse and plot word list
    analyse_and_plot(words, 'Podcast : ' + cast)

analyse_and_plot(all_words, 'Everything')
```

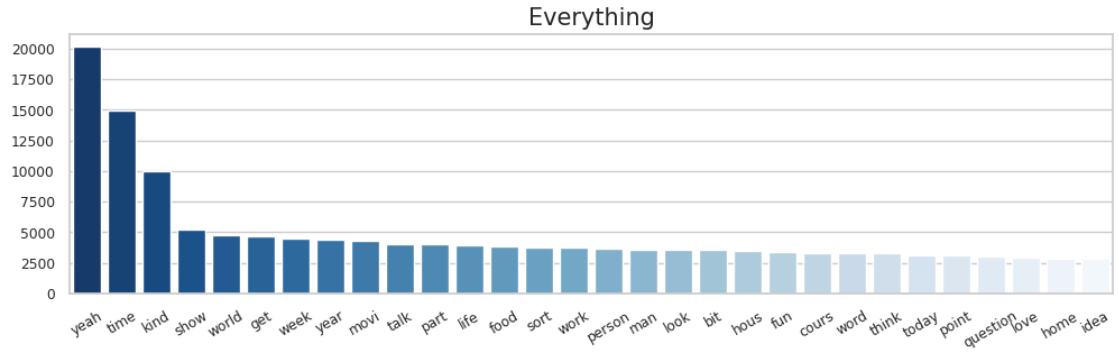



Podcast : JJHO
3000
2500
2000
1500
1000
500
0
yeah time kind show podcast case life get hous world week person look year feel work cours fun part wife famili point think guess man home episod room love talk


Podcast : Shmanners
2000
1500
1000
500
0
yeah time kind show get talk person year fun week point man idea world life work think part bit home today love group hous cours hey tea look money histori


Podcast : WgT
8000
7000
6000
5000
4000
3000
2000
1000
0
yeah time movi kind show world talk song fun look get man sort think love part cours point food game feel person list guy stori question version bit episod charact





It is interesting to see that the terms *yeah* and *time* are the most common words in five different podcasts especially considering that one of those is from the UK and the others from the US. I did not expect such an overlap.

2.2 Analysis

The following cells define methods to initialize vectorizers and models in an easy way for each analysis step. The code (LDA model and count vectorizer) is from the lectures with modifications. The NMF model and TF-IDF vectorizer code is from myself as is the topic analyis method that wraps everything up.

[18]:
```
# turn the documents into vectors
def create_count_vectorizer(documents):
    # count_vectorizer = CountVectorizer(stop_words=stopwords,␣
 ↪tokenizer=clean_text, max_features=1500)
    count_vectorizer = CountVectorizer(#token_pattern=r'\b[a-z][a-z]+\b',
                                       # binary=True,
                                       # stop_words=stop_words,
                                       # max_features=1500,
                                       max_df=0.7
                                      )
```

```
    data = count_vectorizer.fit_transform(documents)
    return (count_vectorizer, data)
```

[19]:
```
# turn the documents into a TF-IDF vector using the given n-gram range
def create_tfidf_vectorizer(documents, ng_range=(1,1)):
    tfidf_vectorizer = TfidfVectorizer(ngram_range=ng_range, max_df=0.7,␣
 ↪smooth_idf=True)
    data = tfidf_vectorizer.fit_transform(documents)
    return (tfidf_vectorizer, data)
```

[20]:
```
# create the LDA model (note that usually num_topics is unknown)
def create_and_fit_lda(data, num_topics):
    lda = LatentDirichletAllocation(n_components=num_topics, n_jobs=-1,␣
 ↪max_iter=100, random_state=1202)
    lda.fit(data)
    return lda
```

[21]:
```
# create and fit the NMF model
def create_and_fit_nmf(data, num_topics):
    nmf = NMF(n_components=num_topics, random_state=1202)
    nmf.fit(data)
    return nmf
```

[22]:
```
# wrapper method to easily run an analysis (vectorize and model) in one line
def topic_analyse(a_documents, a_num_topics=3, a_model='lda',␣
 ↪a_vectorizer='count', a_ngram_range=(1,1)):
    # create the vectorizer
    if a_vectorizer == 'count':
        (vectorizer, data) = create_count_vectorizer(a_documents)
    elif a_vectorizer == 'tfidf':
        (vectorizer, data) = create_tfidf_vectorizer(a_documents,␣
 ↪ng_range=a_ngram_range)

    # create the model
    if a_model == 'lda':
        model = create_and_fit_lda(data, a_num_topics)
    elif a_model == 'nmf':
        model = create_and_fit_nmf(data, a_num_topics)

    print('-'*80)
    print('Vectorizer:', a_vectorizer, ', Model:', a_model, ', Number of Topics:
 ↪', a_num_topics, ', tidf ngram range:', a_ngram_range)
    print('-'*80)

    # print the top words for each topic
    for topic_idx, topic in enumerate(model.components_):
```

```
        print(f"Topic #{topic_idx}: {', '.join([vectorizer.
↪get_feature_names_out()[i] for i in topic.argsort()[:-15 - 1:-1]])}")

    # return the processed features
    return (vectorizer, data, model)
```

The following methods generate the bar and pie charts. The code is my own.

[23]:
```
def prep_topic_df(model, data, vectorizer):
    model_doc_matrix = model.transform(data)

    topic_columns = []

    for topic_idx, topic in enumerate(model.components_):
        print(f"Topic #{topic_idx}: {', '.join([vectorizer.
↪get_feature_names_out()[i] for i in topic.argsort()[:-15 - 1:-1]])}")
        topic_columns.append(f'Topic #{topic_idx}')

    topic_df = pd.DataFrame(model_doc_matrix, columns=topic_columns)
    topic_df['label'] = topic_df.idxmax(axis=1)
    topic_df['podcast'] = df['podcast']
    topic_df['release'] = df['release']
    topic_df['title'] = df['title']
    topic_df['text_raw'] = df['text_raw']

    return topic_df, topic_columns
```

[24]:
```
# plot a bar chart displaying the episode count per topic
def plot_topic_frequencies(topic_df):
    # count and topic label extraction
    temp_df = pd.DataFrame(list(topic_df['label'].value_counts(sort=False)),␣
↪columns=['count'])
    temp_df['topic'] = topic_df['label'].unique()
    temp_df.sort_values('topic', ascending=True, inplace=True)

    # seaborn defaults
    sns.set(font_scale=0.8)
    sns.set_style('whitegrid')
    f, ax = plt.subplots(figsize=(10, 0.4*len(temp_df['topic'])))
    sns.barplot(x=temp_df['count'], y=temp_df['topic'], color='b')
    ax.set(ylabel='', xlabel='Podcast Count')
    plt.title('Topic Frequency', fontsize=11)
    sns.despine(left=True, bottom=True)
    plt.show()
```

[25]:
```
# method to extract the topic frequency per show and topic
def get_freq_list(podname, topic_df):
```

```
    label_freq = []

    # loop over all topics
    for t in topic_columns:
        # count occurences of a particular topic for a particular podcast
        freq = list(topic_df[(topic_df['podcast'] == podname) &␣
 ↪(topic_df['label'] == t)]['label'].value_counts())
        if len(freq) == 0:
            label_freq.extend([0])
        else:
            label_freq.extend(freq)

    return label_freq
```

[26]:
```
# plot the given information at the given position in the plot
def plot_pod_topic_freq(axs, podname, row, col, topic_df, topic_columns):
    # get the label counts
    label_freq = get_freq_list(podname, topic_df)
    # sum up all the counts
    podcast_sum = sum(label_freq)

    # conditionally create labels
    labels = [f'{category}: {value}' if value != 0 else '' for category, value␣
 ↪in zip(topic_columns, label_freq)]

    # format and plot a pie chart
    axs[row, col].pie(label_freq, labels=labels)
    axs[row, col].set_title(f'{podname}', fontsize=11)
    axs[row, col].set(ylabel='', xlabel=f'{podcast_sum} podcasts')
```

[27]:
```
# method to plot the topic counts per podcast in pie chart subplots
def plot_all_pod_topic_freq(topic_df, topic_columns):
    # get a list of podcast names
    podcasts = topic_df['podcast'].unique()

    cols = 3
    rows = math.ceil(len(podcasts) / cols)


    # prepare subplot layout
    fig, axs = plt.subplots(rows, cols, figsize=(12, 10))

    # row and column count of subplots
    row = 0
    col = 0

    # iterate over all podcasts
```

```
    for cast in podcasts:
        # plot the podcast statistics
        plot_pod_topic_freq(axs, cast, row, col, topic_df, topic_columns)
        # increment row and column count
        if col == 2: row +=1
        col = (col + 1) % 3

    # hide the last subplots (ugly hack probably)
    for i in range(1, (rows * cols) - len(podcasts) + 1):
        axs.flat[-i].set_visible(False)

    plt.show()
```

The code below doesn't work because *pyLDAvis* can't be called multiple times from one cell due to *display* limitations in Jupyter cells.

[28]:
```
# # extract the cleaned text
# documents = df['text_clean']

# # analysation loop
# # model choice
# for m in ['lda']:#, 'nmf']:
#     # vectorizer choice
#     for v in ['count']:#, 'tfidf']:
#         # number of topic choice
#         for t in [3, 4]:#, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 30]:
#             # if v == 'tfidf':
#             #     # ngram range choice for tfidf
#             #     for n in [(1,1), (1,2)]:#, (1,3), (1,4)]:
#             #         topic_analyse(documents, t, m, v, n)
#             # else:

#             vectorizer, data, model = topic_analyse(documents, t, m, v)
#             pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

Manually running each analysis step in a separate cell since the above cell didn't work.

The iterations follow this scheme:

- topic count 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30
  - model LDA using count vectorizer
  - model NMF using count vectorizer
  - model LDA using TF-IDF vectorizer with n-gram range (1,1) -> ONLY used for topic count 3 and 4 (as described in the documentation)
  - model NMF using TF-IDF vectorizer with n-gram range (1,1), (1,2), (1,3)

2.3 Analysis Execution and Output Visualisation

Note: The following is only a subset of the whole analysis described above. While the written report clings to the given word count there was no viable way of compressing

the visuals. I'm well aware that it's pushing the limits and I would not put that in a final report but as we are supposed to show our work I'll leave it in for the midterm for a quick perusal after skrutinizing the above report. It turns out that exporting the Jupyter notebook as PDF doesn't allow to supress the warning messages of the visualization which blew the output up tremendously. However it's nice to see how the output changes with increased topic count. I hope you'll forgive me without downgrading my paper. The figures are just inteded for skimming. Thank you.

[124]:
```
# extract the information to work on from the dataframe
documents = df['text_clean']
```

As described above the following cells repeat themselves over and over with different settings for the vectorizer, n-gram range and model. Only the first cell is documented since they all follow the same scheme!

[30]:
```
# topic count for the upcoming analysis
topic_count = 3
```

[31]:
```
# running the analysis with the given model, vectorizer and later n-gram range
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
# visualising the output
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

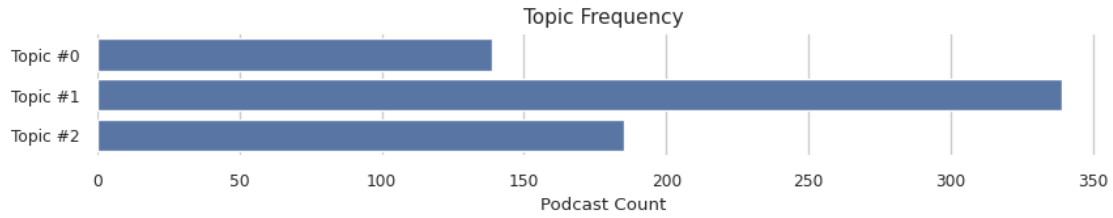
```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 3 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, communiti, presid, countri, morn, song, radio,
book, connect, bill, yesterday, street, support
Topic #1: song, game, book, car, water, christma, watch, cream, eat, ice,
friend, list, dog, slash, charact
Topic #2: war, centuri, woman, power, comedi, book, corner, citi, period, hand,
art, busi, age, radio, franc
```

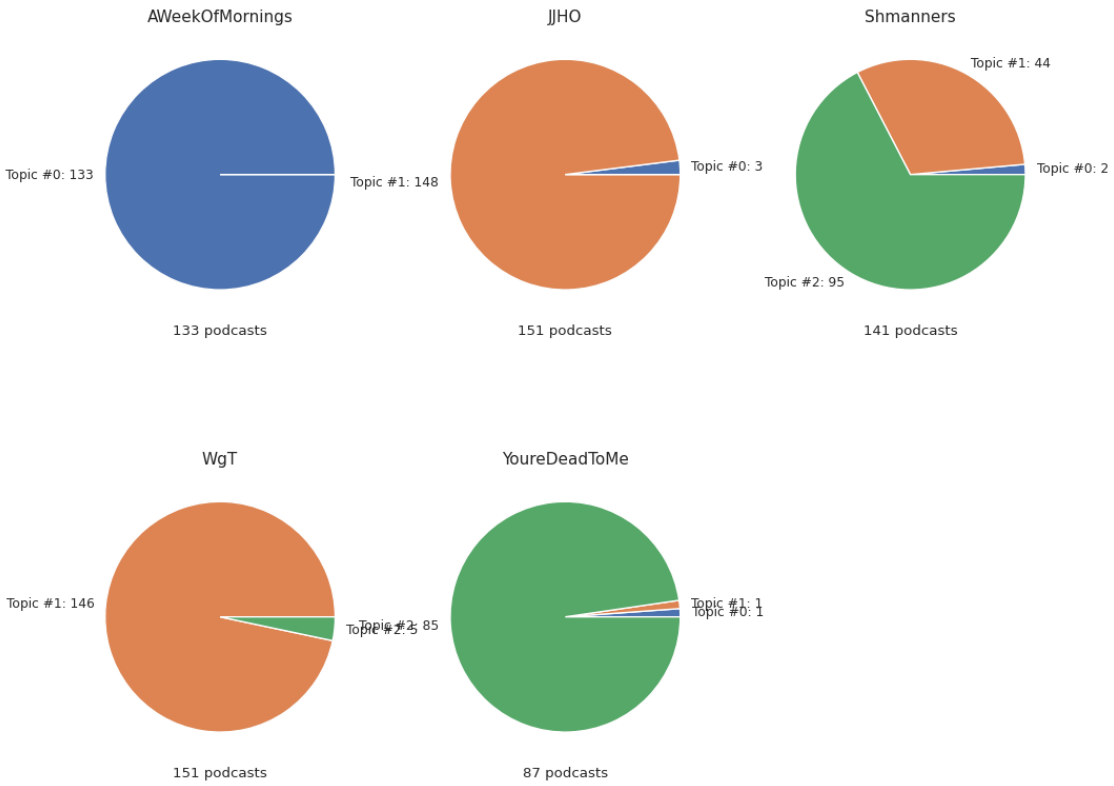
[31]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.058803  0.077356       1        1  46.612225
0     -0.056870 -0.078223       2        1  34.958125
2      0.115673  0.000867       3        1  18.429650, topic_info=          Term
Freq        Total Category  logprob  loglift
28601     wine  2030.000000  2030.000000  Default  30.0000  30.0000
22067    river  1553.000000  1553.000000  Default  29.0000  29.0000
4443   centuri   865.000000   865.000000  Default  28.0000  28.0000
28149      war  1207.000000  1207.000000  Default  27.0000  27.0000
20435   presid  1176.000000  1176.000000  Default  26.0000  26.0000
…         …           …            …       …       …        …
4654     check   250.902348   880.093295   Topic3  -6.1576   0.4362
12707  husband   239.865757   791.400578   Topic3  -6.2026   0.4975
19075    parti   252.080277  1150.537625   Topic3  -6.1529   0.1730
```

```
28222    water   259.136111  1756.673026   Topic3  -6.1253  -0.2226
10295     game   233.627915  2156.856219   Topic3  -6.2289  -0.5315

[274 rows x 6 columns], token_table=       Topic      Freq       Term
term
256           1  0.262612      advic
256           2  0.183546      advic
256           3  0.553462      advic
326           1  0.400617        age
326           2  0.227706        age
…       …        …         …
28937         3  0.008615  yesterday
28990         1  0.586054       york
28990         2  0.333125       york
28990         3  0.080968       york
29124         3  0.984782      zhang

[505 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 1, 3])
```

[32]:
```
# further visualisation of the analysis using methods previously defined
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, communiti, presid, countri, morn, song, radio,
book, connect, bill, yesterday, street, support
Topic #1: song, game, book, car, water, christma, watch, cream, eat, ice,
friend, list, dog, slash, charact
Topic #2: war, centuri, woman, power, comedi, book, corner, citi, period, hand,
art, busi, age, radio, franc
```





[33]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

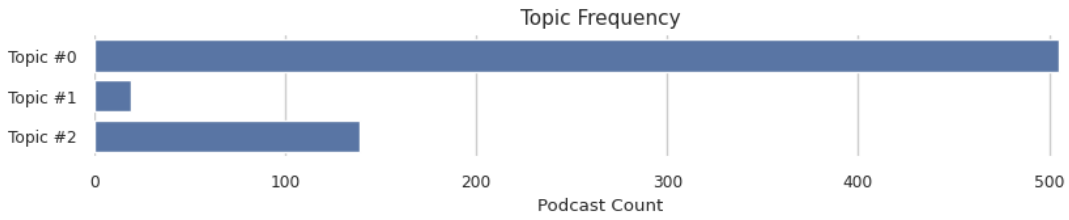
```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 3 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: game, song, book, christma, citi, watch, list, car, york, charact,
version, audienc, water, hand, film
Topic #1: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, tea
Topic #2: wine, river, state, presid, communiti, connect, countri, cancer, morn,
radio, yesterday, support, bank, street, congressman
```

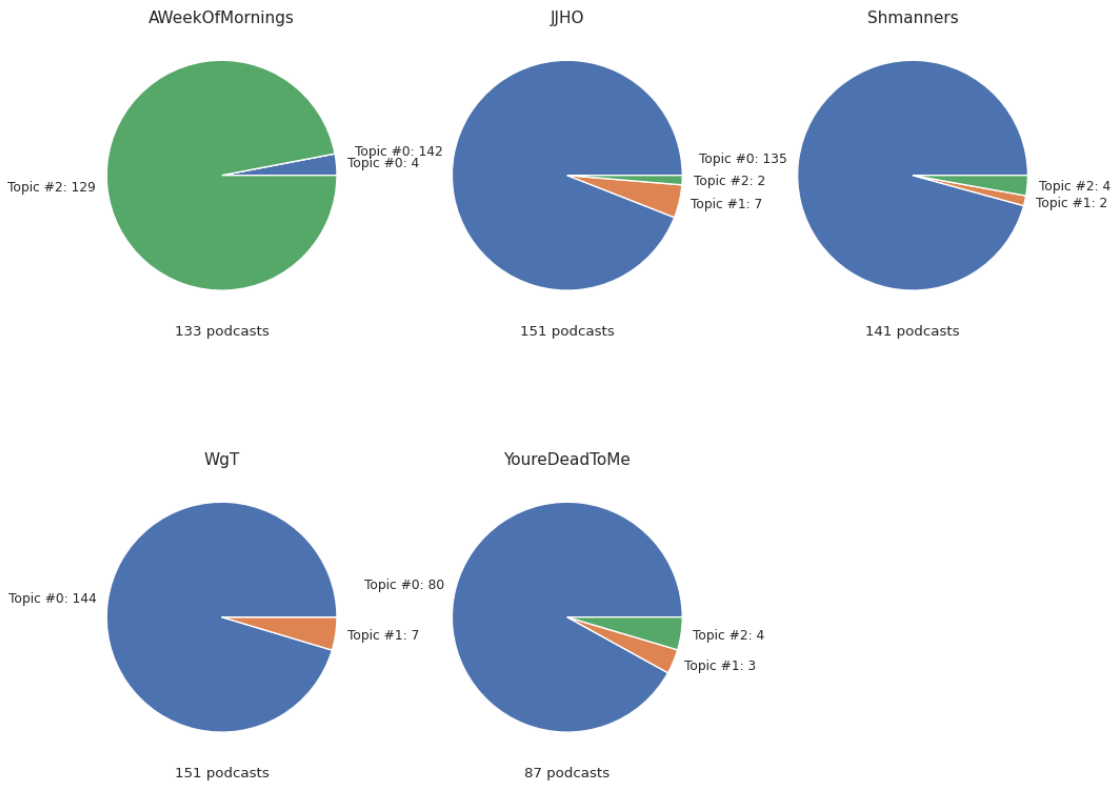
[33]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.061136 -0.120779       1        1  59.673157
2     -0.179680  0.086729       2        1  33.235610
1      0.240816  0.034049       3        1   7.091233, topic_info=         Term
Freq        Total Category  logprob  loglift
6255    cream  3292.000000  3292.000000  Default  30.0000  30.0000
12801     ice  3275.000000  3275.000000  Default  29.0000  29.0000
```

```
28222   water  2127.000000  2127.000000  Default  28.0000  28.0000
4877   chocol  1339.000000  1339.000000  Default  27.0000  27.0000
3833     cake   917.000000   917.000000  Default  26.0000  26.0000
…         …          …            …        …        …        …
11601    half   135.206297  1063.431406   Topic3  -5.8208   0.5839
19075   parti   120.935316  1146.004781   Topic3  -5.9323   0.3975
10742   glass   109.225244   551.926591   Topic3  -6.0341   1.0263
1789      bar   111.840104   850.613694   Topic3  -6.0105   0.6174
15784    make   108.849528   945.003191   Topic3  -6.0376   0.4851

[309 rows x 6 columns], token_table=       Topic      Freq       Term
term
4          2  0.727650   aassenha
9          1  1.384932  abakhumid
10         2  0.755695   abalmant
26         2  1.063001        abd
28         2  1.042861     abduct
…        …         …          …
28601      2  0.986557       wine
28601      3  0.013622       wine
28937      2  0.999498  yesterday
28990      1  0.783240       york
28990      2  0.216918       york

[440 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 3, 2])
```

[34]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, song, book, christma, citi, watch, list, car, york, charact,
version, audienc, water, hand, film
Topic #1: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, tea
Topic #2: wine, river, state, presid, communiti, connect, countri, cancer, morn,
radio, yesterday, support, bank, street, congressman
```





[35]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: lda , Number of Topics: 3 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: burlesqu, rammel, slipper, eccc, abinn, doo, ju, pumpernicl, hearth,
heng, iago, backcountri, ottoman, baguett, inlin
Topic #1: wine, song, book, game, river, christma, water, communiti, state, car,
citi, presid, watch, countri, war
Topic #2: burlesqu, rammel, slipper, eccc, abinn, doo, ju, pumpernicl, hearth,
heng, iago, backcountri, ottoman, baguett, inlin
```

[35]:
```
PreparedData(topic_coordinates=              x    y  topics  cluster       Freq
topic
1      0.043211 -0.0       1        1  95.151488
2     -0.021605 -0.0       2        1   2.424256
0     -0.021605 -0.0       3        1   2.424256, topic_info=            Term
Freq     Total Category  logprob  loglift
20795     pumpkin  0.000000  0.000000  Default   30.000  30.0000
1852       barnum  0.000000  0.000000  Default   29.000  29.0000
```

```
1703      balloon  0.000000  0.000000  Default   28.000  28.0000
6443        crust  0.000000  0.000000  Default   27.000  27.0000
28782        worm  0.000000  0.000000  Default   26.000  26.0000
…             …       …       …       …       …
4693   cheltenham  0.007229  0.235518   Topic3  -10.281   0.2360
7305         dhah  0.007229  0.218656   Topic3  -10.281   0.3103
23333      sentar  0.007229  0.244676   Topic3  -10.281   0.1978
21411      reclin  0.007229  0.334875   Topic3  -10.281  -0.1160
16385    mccallet  0.007229  0.249110   Topic3  -10.281   0.1799

[290 rows x 6 columns], token_table=       Topic      Freq     Term
term
73         1  1.430296  absolut
1182       1  1.444974     arti
1644       1  1.422381     bail
1698       1  1.414751   ballet
1703       1  1.413772  balloon
…      …        …       …
28222      1  0.906797    water
28269      1  1.411883   weapon
28601      1  1.022266     wine
28782      1  1.408896     worm
28990      1  0.943091     york

[65 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1', 'ylab':
'PC2'}, topic_order=[2, 3, 1])
```

[36]:
```
# vectorizer, data, model = topic_analyse(documents, topic_count, 'lda',␣
↪'tfidf', (1, 2))
# pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

[37]:
```
# vectorizer, data, model = topic_analyse(documents, topic_count, 'lda',␣
↪'tfidf', (1, 3))
# pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

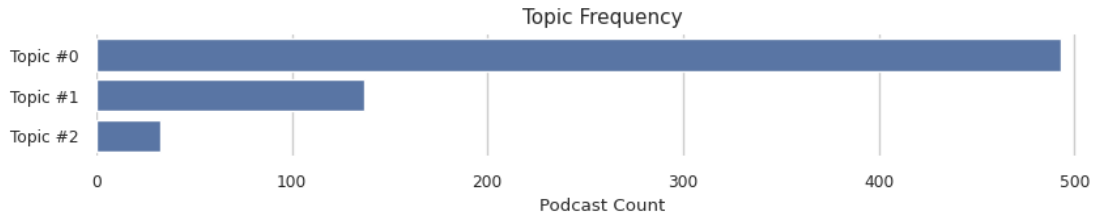
[38]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

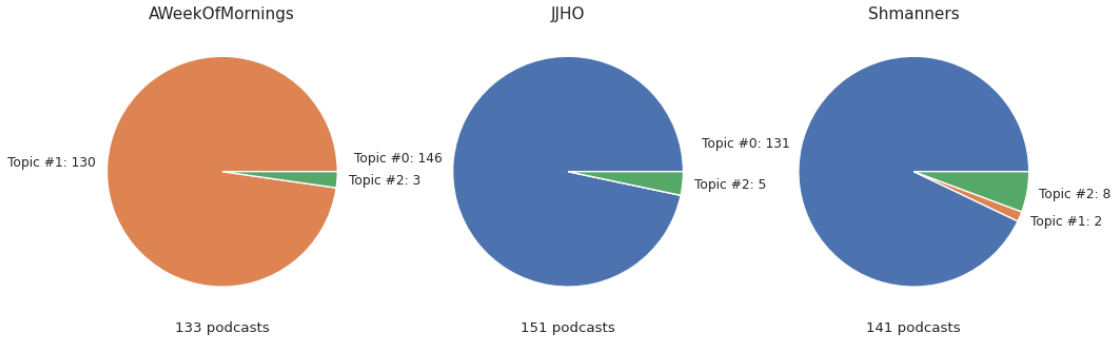
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 3 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: game, book, water, car, dog, cream, slash, eat, song, citi, parti,
ice, friend, film, comedi
Topic #1: wine, river, presid, state, yesterday, communiti, countri,
congressman, trump, morn, hampton, bank, radio, bill, vote
Topic #2: christma, holiday, santa, claus, song, chocol, scroog, tradit,
hanukkah, gift, winter, ghost, villain, tree, krampus
```

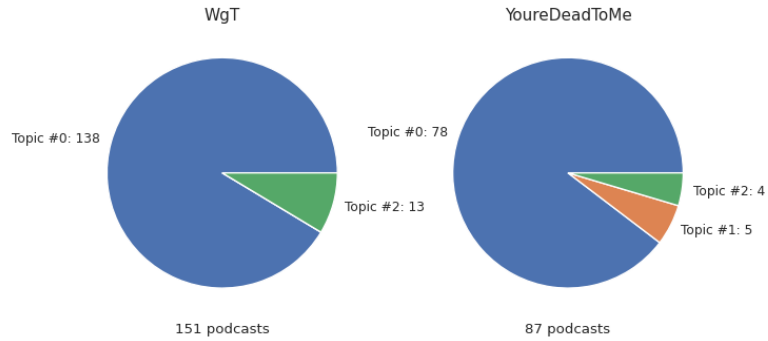
[38]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.060683 -0.137835       1        1  60.390160
1     -0.171604  0.099981       2        1  31.521587
2      0.232287  0.037854       3        1   8.088253, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  30.000000  30.000000  Default  30.0000  30.0000
12359   holiday  10.000000  10.000000  Default  29.0000  29.0000
28601      wine  20.000000  20.000000  Default  28.0000  28.0000
24514      song  16.000000  16.000000  Default  27.0000  27.0000
22765     santa   7.000000   7.000000  Default  26.0000  26.0000
…           …         …         …       …        …        …
27747   version   1.910651   7.868544   Topic3  -5.9088   1.0993
15260      list   1.887830   8.476602   Topic3  -5.9208   1.0129
23194    season   1.744476   7.619135   Topic3  -5.9998   1.0405
3970       candi   1.589177   5.791060   Topic3  -6.0930   1.2217
26131  telescop   1.536165   5.687568   Topic3  -6.1270   1.2058

[305 rows x 6 columns], token_table=       Topic      Freq        Term
term
249         3  0.767993      advent
575         1  0.388751      almond
575         3  0.777501      almond
962         1  1.087673         app
1384        3  1.877969      aubrey
…       …        …           …
28702       2  0.292210       woman
28708       3  0.634411  wonderland
28937       2  1.038416   yesterday
28990       1  0.695328        york
28990       2  0.347664        york

[253 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 3])
```

[39]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, water, car, dog, cream, slash, eat, song, citi, parti,
ice, friend, film, comedi
Topic #1: wine, river, presid, state, yesterday, communiti, countri,
congressman, trump, morn, hampton, bank, radio, bill, vote
Topic #2: christma, holiday, santa, claus, song, chocol, scroog, tradit,
hanukkah, gift, winter, ghost, villain, tree, krampus
```







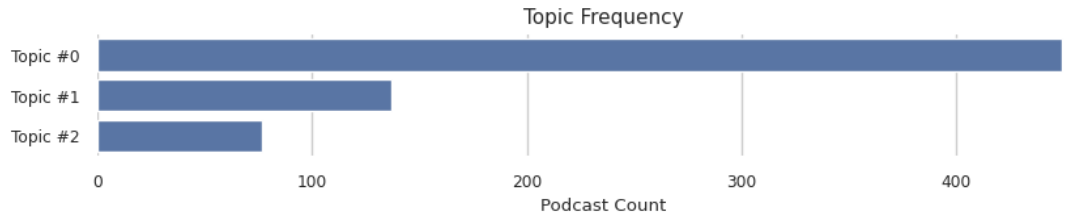
[40]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

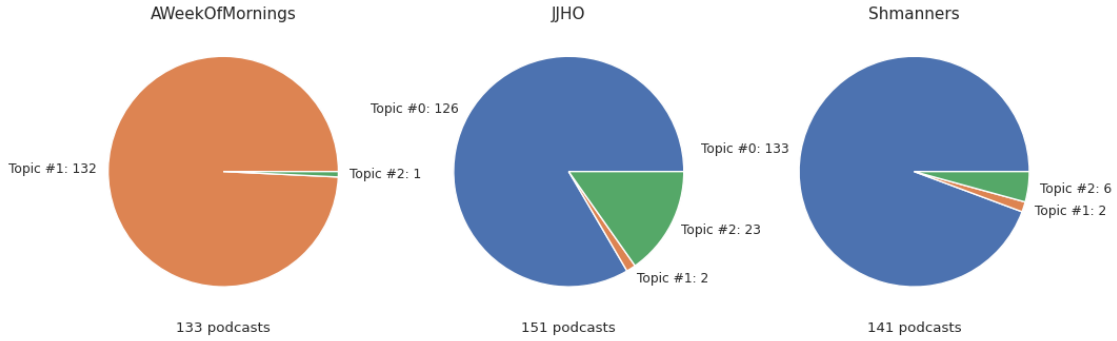
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 3 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: christma, game, book, song, car, charact, slash, film, citi, dog,
comedi, list, watch, water, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, bill
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, flavor, eat,
bread, candi, chocol chocol, dip, water
```

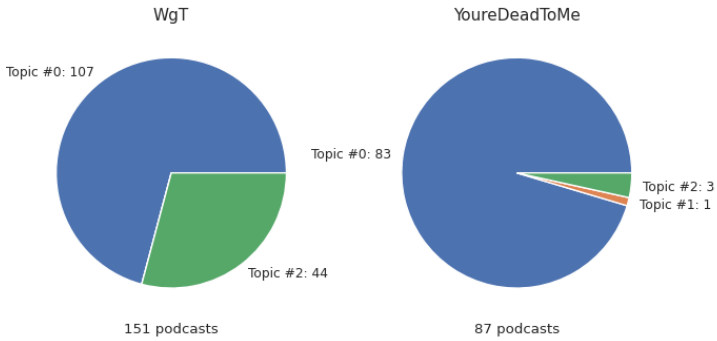
[40]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.010055 -0.252317       1        1  58.097092
1     -0.258261  0.133386       2        1  29.078287
2      0.268316  0.118931       3        1  12.824621, topic_info=
Term       Freq      Total Category  logprob  loglift
83347      chocol  11.000000  11.000000  Default  30.0000  30.0000
116310      cream   9.000000   9.000000  Default  29.0000  29.0000
240891  ice cream   9.000000   9.000000  Default  28.0000  28.0000
240842        ice   8.000000   8.000000  Default  27.0000  27.0000
535853       wine  11.000000  11.000000  Default  26.0000  26.0000
…             …         …         …       …        …        …
470537      sugar   1.380733   2.148802   Topic3  -7.5593   1.6115
418655       sauc   1.608225   3.070072   Topic3  -7.4067   1.4072
27982         bar   1.792804   4.230545   Topic3  -7.2981   1.1953
527770      water   2.093220   7.819373   Topic3  -7.1432   0.7359
359440        pie   1.553721   3.199346   Topic3  -7.4412   1.3315

[319 rows x 6 columns], token_table=        Topic      Freq         Term
term
9786        3  0.676560       almond
14135       1  0.921919          app
26729       1  0.536752       banana
26729       3  0.536752       banana
27510       2  1.038982         bank
…       …       …           …
546537      3  1.712141  yeah chocol
548599      1  0.731596    yeah movi
552419      2  0.884931    yesterday
553250      1  0.721986         york
553250      2  0.360993         york

[247 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 3])
```

[41]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: christma, game, book, song, car, charact, slash, film, citi, dog,
comedi, list, watch, water, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, bill
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, flavor, eat,
bread, candi, chocol chocol, dip, water
```







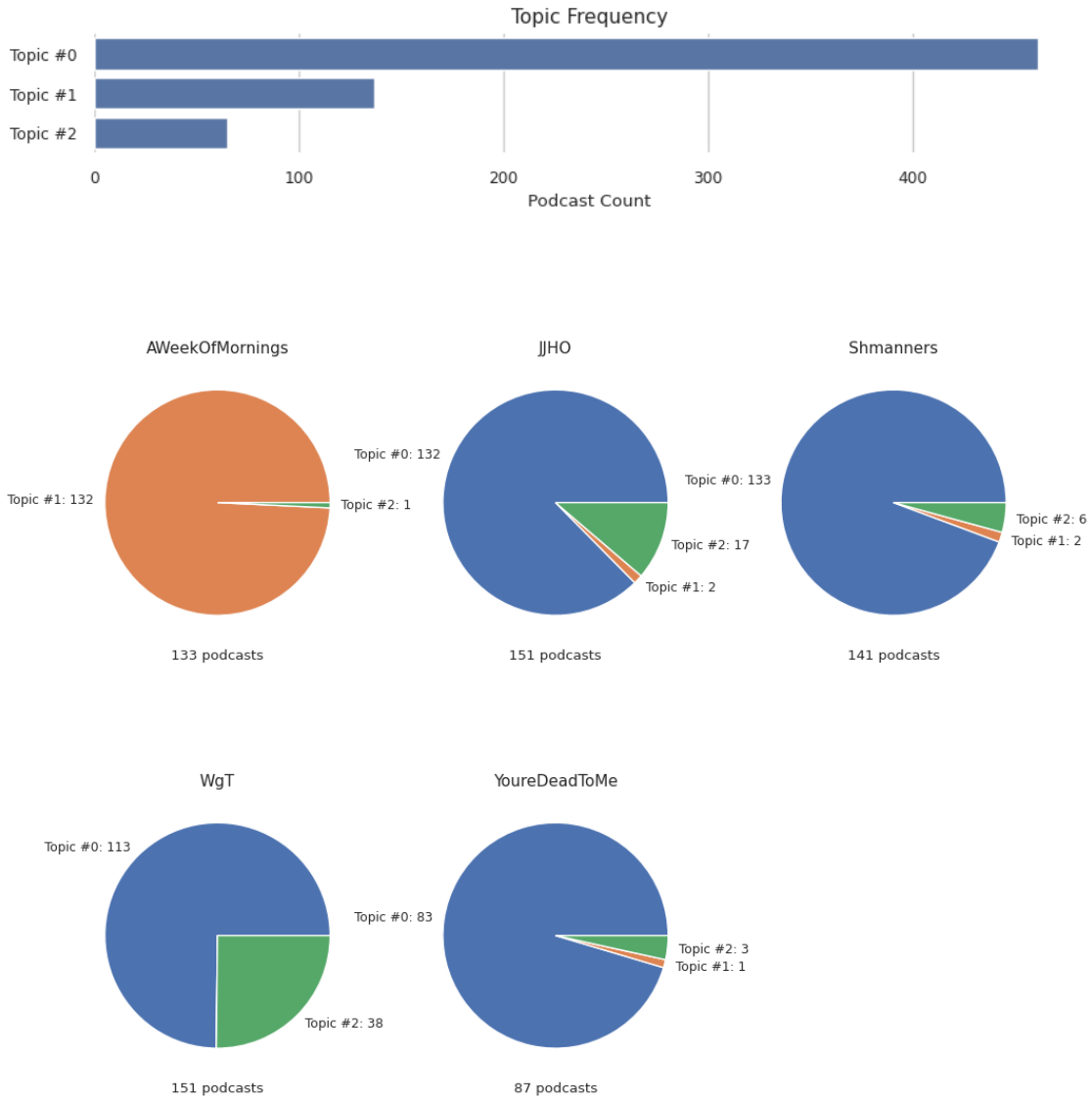
[42]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 3 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: christma, game, book, song, car, slash, charact, dog, film, citi,
water, comedi, list, watch, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, bank, morn, trump, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, water
```

[42]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0      0.069494 -0.278516       1        1  59.676749
1     -0.312741  0.087040       2        1  29.595694
2      0.243247  0.191476       3        1  10.727556, topic_info=
Term      Freq     Total Category  logprob  loglift
192266      chocol  9.000000  9.000000  Default  30.0000  30.0000
270440       cream  7.000000  7.000000  Default  29.0000  29.0000
566886   ice cream  7.000000  7.000000  Default  28.0000  28.0000
566771         ice  6.000000  6.000000  Default  27.0000  27.0000
1281272       wine  9.000000  9.000000  Default  26.0000  26.0000
…              …       …       …       …       …       …
887438         pot  0.985648  1.836662   Topic3  -8.0849   1.6099
719216        meat  1.095630  2.777351   Topic3  -7.9791   1.3022
118070   breakfast  1.064558  2.564405   Topic3  -8.0079   1.3532
1260664      water  1.324065  6.264142   Topic3  -7.7898   0.6782
853026         pie  0.992221  2.584706   Topic3  -8.0783   1.2749

[320 rows x 6 columns], token_table=         Topic      Freq       Term
term
31181        1  1.198971        app
59273        3  0.693381     banana
61131        2  0.942916       bank
62375        1  0.295798        bar
62375        2  0.295798        bar
…          …       …         …
1287155      2  0.287169      woman
1294001      2  1.143046  word week
1337670      2  0.834200  yesterday
1339625      1  0.666730       york
1339625      2  0.222243       york

[237 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 3])
```

[43]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: christma, game, book, song, car, slash, charact, dog, film, citi,
water, comedi, list, watch, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, bank, morn, trump, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, water
```



[44]:
```
topic_count = 4
```

[45]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 4 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
Topic #1: song, game, christma, list, version, charact, watch, chocol, film,
book, audienc, york, theme, citi, star
Topic #2: war, centuri, woman, comedi, power, book, citi, corner, period, radio,
```

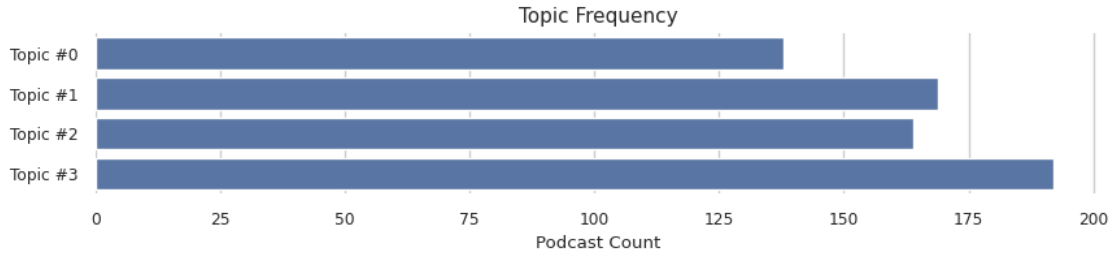
```
franc, age, art, death, god
Topic #3: water, cream, ice, car, slash, book, friend, dog, game, store,
husband, season, eat, rule, hand
```

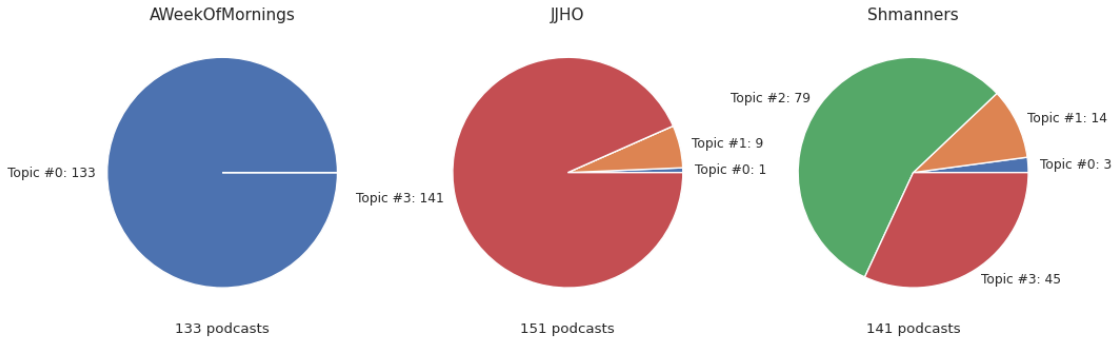
[45]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.005027 -0.105219       1        1  34.691227
3     -0.062107  0.003240       2        1  24.982775
1     -0.069351  0.069212       3        1  23.463114
2      0.136484  0.032767       4        1  16.862884, topic_info=
Term         Freq        Total Category  logprob  loglift
28601      wine  2049.000000  2049.000000  Default  30.0000  30.0000
24514      song  2391.000000  2391.000000  Default  29.0000  29.0000
22067     river  1566.000000  1566.000000  Default  28.0000  28.0000
4443    centuri   873.000000   873.000000  Default  27.0000  27.0000
4933   christma  1225.000000  1225.000000  Default  26.0000  26.0000
…         …           …           …       …      …       …
12707   husband   228.357430   790.208103   Topic4  -6.1629   0.5387
19075     parti   242.708758  1149.784071   Topic4  -6.1020   0.2246
11678      hand   234.386030  1235.328882   Topic4  -6.1368   0.1179
25778    system   219.844723   882.664064   Topic4  -6.2009   0.3900
4654      check   219.504694   879.870299   Topic4  -6.2024   0.3917

[354 rows x 6 columns], token_table=       Topic      Freq   Term
term
326        1  0.228384    age
326        2  0.242962    age
326        3  0.176147    age
326        4  0.352295    age
340        3  0.981767  agley
…      …        …    …
28990      1  0.336699   york
28990      2  0.196601   york
28990      3  0.384688   york
28990      4  0.082046   york
29124      4  0.972232  zhang

[817 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 4, 2, 3])
```

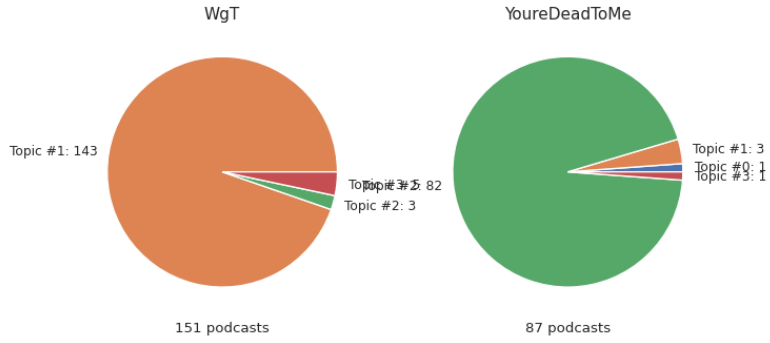
[46]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
```

```
Topic #1: song, game, christma, list, version, charact, watch, chocol, film,
book, audienc, york, theme, citi, star
Topic #2: war, centuri, woman, comedi, power, book, citi, corner, period, radio,
franc, age, art, death, god
Topic #3: water, cream, ice, car, slash, book, friend, dog, game, store,
husband, season, eat, rule, hand
```







[47]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

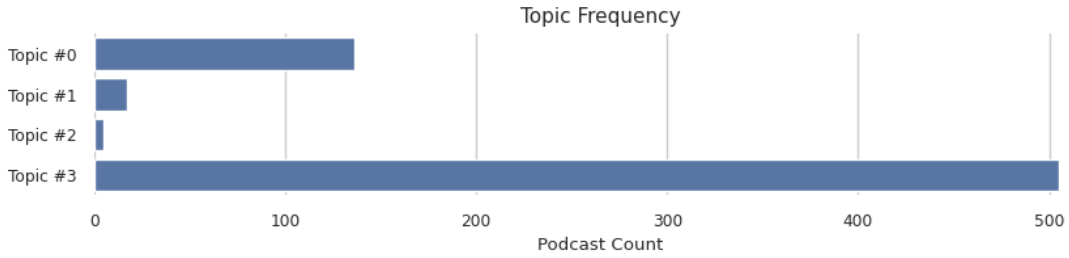
```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 4 , tidf ngram range: (1, 1)
```

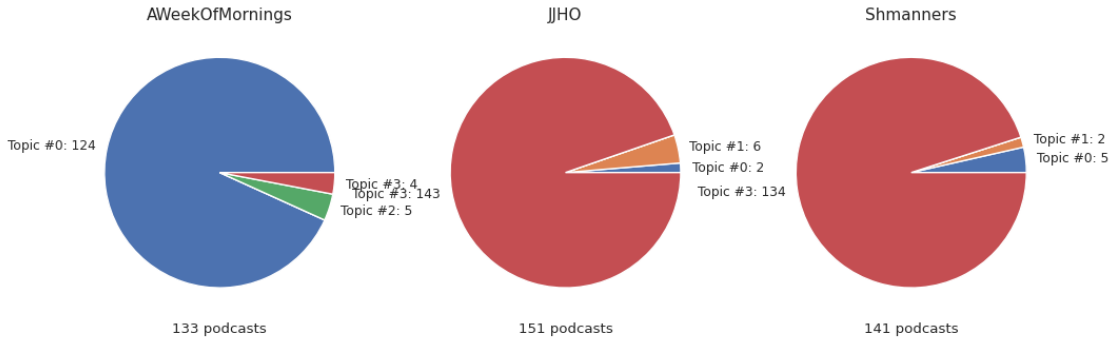
```
---------------------------------------------------------------------------
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, bill, street, song, congressman, space
Topic #1: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, tea
Topic #2: cancer, connect, support, wine, bed, car, month, communiti, organ,
diagnosi, river, hampton, pino, presid, oak
Topic #3: game, song, book, christma, car, watch, citi, list, charact, version,
audienc, york, water, hand, film
```

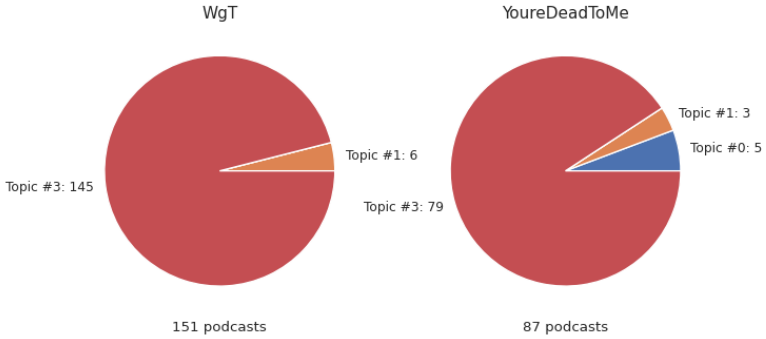
[47]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
3     -0.030582  0.122121       1        1  57.244176
0      0.098125  0.090538       2        1  32.778887
1     -0.276355 -0.080913       3        1   6.225783
2      0.208811 -0.131746       4        1   3.751154, topic_info=          Term
Freq        Total Category  logprob  loglift
6255     cream  2890.000000  2890.000000  Default  30.0000  30.0000
12801      ice  2880.000000  2880.000000  Default  29.0000  29.0000
3964    cancer  1543.000000  1543.000000  Default  28.0000  28.0000
5757   connect  1443.000000  1443.000000  Default  27.0000  27.0000
28601     wine  2494.000000  2494.000000  Default  26.0000  26.0000
…         …           …           …      …        …        …
17439     morn    80.862032  1513.084431   Topic4  -5.6980   0.3539
2887      book    87.595142  2645.621335   Topic4  -5.6180  -0.1248
25263   street    75.733682  1285.099789   Topic4  -5.7635   0.4517
15203     line    75.245116  1271.258464   Topic4  -5.7700   0.4561
12479     hope    71.111678   784.245028   Topic4  -5.8265   0.8826

[428 rows x 6 columns], token_table=       Topic      Freq      Term
term
3          1  1.090312  aaronson
4          2  0.712178  aassenha
6          1  0.788579   ababino
10         2  0.714282  abalmant
16         1  0.788579   abatino
…      …        …       …
28990      2  0.255530      york
29077      4  1.013355  zarabodi
29081      4  0.975823  zaribodi
29193      1  0.073293    zucker
29193      4  0.952810    zucker

[716 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[4, 1, 2, 3])
```

[48]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, bill, street, song, congressman, space
Topic #1: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, tea
Topic #2: cancer, connect, support, wine, bed, car, month, communiti, organ,
diagnosi, river, hampton, pino, presid, oak
Topic #3: game, song, book, christma, car, watch, citi, list, charact, version,
audienc, york, water, hand, film
```







[49]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: lda , Number of Topics: 4 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: slipper, westbrook, comedia, ju, farer, abinn, eccc, daren, dhah,
chabata, goali, panto, iago, earthwork, pumpernicl
Topic #1: wine, song, book, game, river, christma, water, communiti, state, car,
citi, presid, watch, war, countri
Topic #2: slipper, westbrook, comedia, ju, farer, eccc, abinn, daren, dhah,
chabata, goali, panto, iago, earthwork, pumpernicl
Topic #3: slipper, westbrook, comedia, ju, farer, abinn, eccc, daren, dhah,
chabata, goali, panto, iago, earthwork, pumpernicl
```

[49]:
```
PreparedData(topic_coordinates=              x    y  topics  cluster       Freq
topic
1      0.061315  0.0       1        1  94.671326
2     -0.020438  0.0       2        1   1.776225
0     -0.020438  0.0       3        1   1.776225
3     -0.020438 -0.0       4        1   1.776225, topic_info=            Term
Freq     Total Category  logprob  loglift
26178      tenni  1.000000  1.000000  Default  30.0000  30.0000
1095        argu  1.000000  1.000000  Default  29.0000  29.0000
6801        dash  1.000000  1.000000  Default  28.0000  28.0000
3855    calendar  1.000000  1.000000  Default  27.0000  27.0000
1914   basketbal  1.000000  1.000000  Default  26.0000  26.0000
…           …       …         …        …        …        …
16384     mccall  0.005292  0.209650   Topic4 -10.2819   0.3514
22084      rizzo  0.005292  0.172814   Topic4 -10.2819   0.5447
6729        dane  0.005292  0.169322   Topic4 -10.2819   0.5651
2996   botchybal  0.005292  0.185558   Topic4 -10.2819   0.4735
4689        chel  0.005292  0.185558   Topic4 -10.2819   0.4735

[330 rows x 6 columns], token_table=       Topic      Freq       Term
term
2          1  0.983505      aaron
1095       1  0.984621       argu
1914       1  0.983426  basketbal
2229       1  0.987824    benefit
2573       1  0.978493    blanket
2887       1  1.052937       book
3855       1  0.988286   calendar
4016       1  0.986115        cap
4067       1  0.925283        car
```

4158	1	0.978254	carol
4564	1	0.937795	charact
4802	1	0.985662	chili
4877	1	1.104204	chocol
4933	1	0.933051	christma
5056	1	0.930357	citi
5579	1	1.024129	communiti
5923	1	0.987505	copi
6119	1	0.977680	countri
6255	1	0.911706	cream
6648	1	0.991909	da
6801	1	0.977018	dash
7070	1	0.976682	demand
7712	1	1.017904	dog
7953	1	0.991480	drama
8291	1	1.022883	eat
9120	1	0.975760	fair
9478	1	1.005764	film
10032	1	0.934012	friend
10295	1	0.970228	game
11157	1	0.997915	greec
13491	1	0.986333	jam
15236	1	0.976318	lion
16867	1	0.993441	middl
17439	1	0.993936	morn
19075	1	1.050825	parti
20435	1	0.950109	presid
20797	1	0.983383	punch
21083	1	1.025959	radio
21675	1	0.993669	renaiss
22067	1	0.927087	river
22309	1	0.973822	ross
23985	1	0.993528	sink
24125	1	0.929541	slash
24514	1	1.025588	song
24610	1	0.923623	space
24964	1	1.097646	star
25000	1	1.035156	state
25197	1	0.990093	storag
25263	1	0.932708	street
26178	1	0.971305	tenni
26331	1	0.976593	therapist
28149	1	0.977675	war
28218	1	0.975836	watch
28222	1	0.956311	water
28586	1	0.988197	wind
28601	1	1.008655	wine

```
28802      1  0.988211       wrap
28821      1  0.989117      write
28990      1  1.038178       york
29171      1  0.985650       zone, R=30, lambda_step=0.01, plot_opts={'xlab':
'PC1', 'ylab': 'PC2'}, topic_order=[2, 3, 1, 4])
```

[50]:
```
# vectorizer, data, model = topic_analyse(documents, topic_count, 'lda',␣
 ↪'tfidf', (1, 2))
# pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

[51]:
```
# vectorizer, data, model = topic_analyse(documents, topic_count, 'lda',␣
 ↪'tfidf', (1, 3))
# pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

[52]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
 ↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

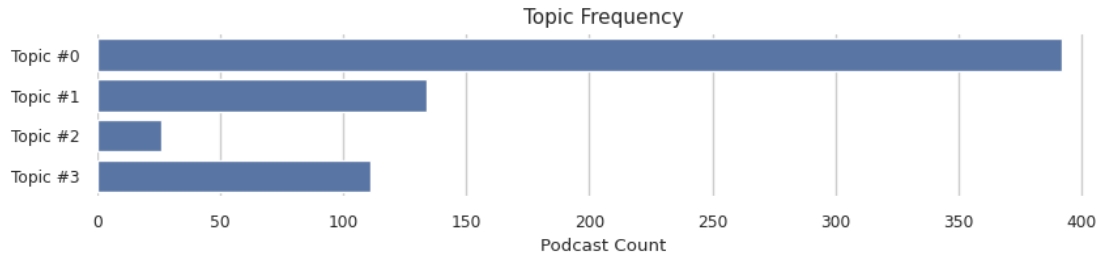
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 4 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: game, book, car, film, citi, charact, song, slash, war, comedi, host,
woman, parti, list, watch
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
Topic #2: christma, holiday, santa, claus, song, scroog, tradit, gift, hanukkah,
winter, ghost, villain, krampus, carol, tree
Topic #3: chocol, cream, ice, butter, eat, cake, bread, breakfast, water, milk,
candi, soup, pizza, flavor, dip
```

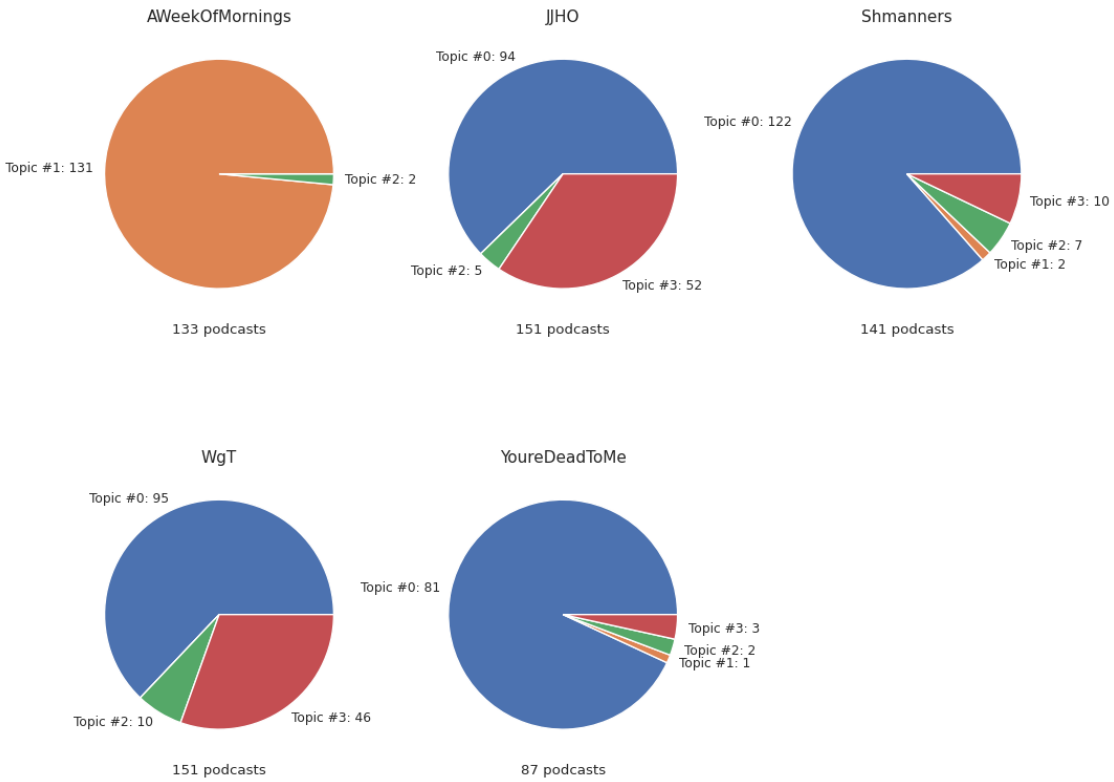
[52]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.044941  0.043200       1        1  47.155811
1     -0.115116  0.169131       2        1  27.297443
3     -0.099782 -0.212955       3        1  18.849992
2      0.259839  0.000624       4        1   6.696754, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  26.000000  26.000000  Default  30.0000  30.0000
4877     chocol  16.000000  16.000000  Default  29.0000  29.0000
6255      cream  14.000000  14.000000  Default  28.0000  28.0000
28601      wine  18.000000  18.000000  Default  27.0000  27.0000
12359   holiday   9.000000   9.000000  Default  26.0000  26.0000
…          …         …         …       …        …        …
575      almond   1.274878   3.442450   Topic4  -6.1246   1.7102
26131  telescop   1.270754   5.012284   Topic4  -6.1279   1.3313
4092       card   1.233056   4.739837   Topic4  -6.1580   1.3570
```

```
4564    charact   1.222777   7.552172   Topic4  -6.1663   0.8828
24742    spirit   1.197456   3.741512   Topic4  -6.1873   1.5643

[411 rows x 6 columns], token_table=       Topic      Freq        Term
term
249        4  0.879850      advent
256        1  1.003604       advic
326        1  0.732517         age
326        2  0.183129         age
575        3  0.580982      almond
…      …       …          …
28708      4  0.694570  wonderland
28937      2  0.984010   yesterday
28990      1  0.603971        york
28990      2  0.241588        york
28990      3  0.120794        york

[362 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 4, 3])
```

[53]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, film, citi, charact, song, slash, war, comedi, host,
woman, parti, list, watch
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
Topic #2: christma, holiday, santa, claus, song, scroog, tradit, gift, hanukkah,
winter, ghost, villain, krampus, carol, tree
Topic #3: chocol, cream, ice, butter, eat, cake, bread, breakfast, water, milk,
candi, soup, pizza, flavor, dip
```





[54]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

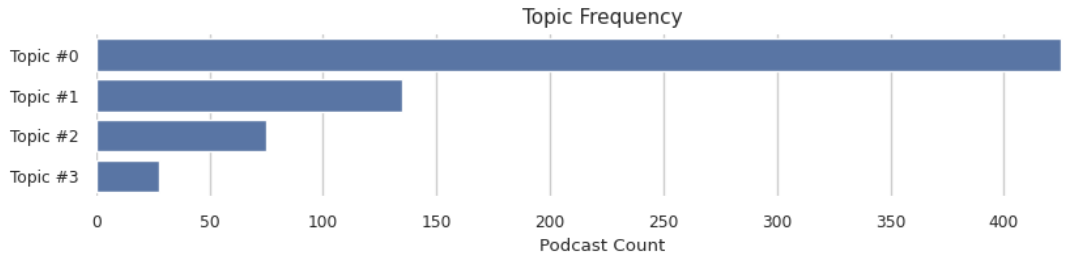
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 4 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: game, book, car, citi, song, slash, dog, water, charact, comedi, film,
war, parti, watch, york
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, connect
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, flavor, eat,
bread, candi, chocol chocol, dip, water
Topic #3: christma, holiday, santa, christma movi, claus, song, christma
christma, scroog, villain, ghost christma, santa claus, movi christma, tradit,
ghost, hanukkah
```

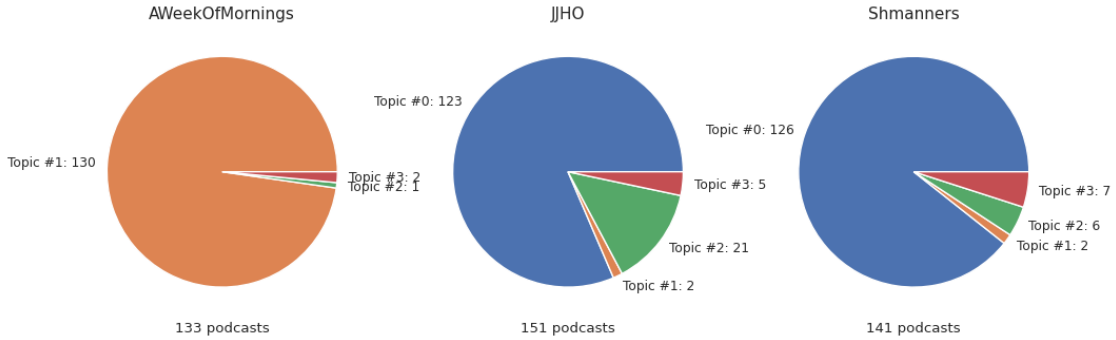
[54]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.017738 -0.007589       1        1  53.499810
1     -0.104182 -0.283826       2        1  27.532351
```

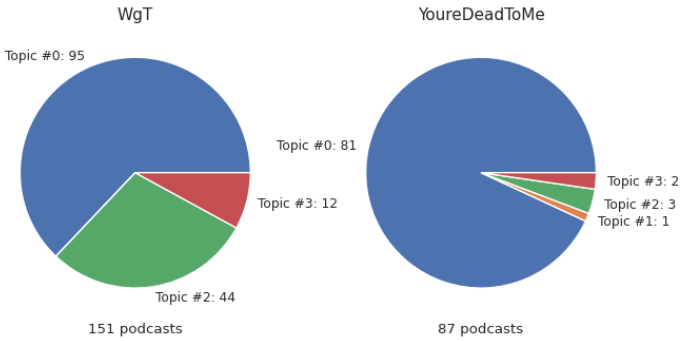
```
2     -0.196897  0.237753       3        1  12.012096
3      0.318817  0.053662       4        1   6.955744, topic_info=
Term       Freq      Total Category  logprob  loglift
84847    christma  13.000000  13.000000  Default  30.0000  30.0000
83347      chocol  11.000000  11.000000  Default  29.0000  29.0000
116310      cream   9.000000   9.000000  Default  28.0000  28.0000
240891  ice cream   8.000000   8.000000  Default  27.0000  27.0000
240842        ice   7.000000   7.000000  Default  26.0000  26.0000
…           …         …        …       …        …        …
323077  movi movi   0.842876   2.722676   Topic4  -7.4410   1.4931
518151    version   0.913622   4.757038   Topic4  -7.3604   1.0156
149984     easter   0.751877   1.723088   Topic4  -7.5553   1.8363
9786       almond   0.745936   1.773987   Topic4  -7.5632   1.7993
77996     charact   0.801504   5.017404   Topic4  -7.4913   0.8314

[410 rows x 6 columns], token_table=        Topic      Freq           Term
term
5008        4  1.790123         advent
6286        1  0.844844            age
6286        2  0.281615            age
9786        3  0.563702         almond
9786        4  0.563702         almond
…       …       …              …
546537      3  1.777548    yeah chocol
546555      4  1.448950  yeah christma
552419      2  0.915171      yesterday
553250      1  0.563639           york
553250      2  0.375760           york

[297 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 3, 4])
```

[55]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, citi, song, slash, dog, water, charact, comedi, film,
war, parti, watch, york
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, connect
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, flavor, eat,
bread, candi, chocol chocol, dip, water
Topic #3: christma, holiday, santa, christma movi, claus, song, christma
christma, scroog, villain, ghost christma, santa claus, movi christma, tradit,
ghost, hanukkah
```







[56]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 4 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: game, book, car, water, dog, slash, song, citi, comedi, charact, film,
watch, york, parti, friend
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, trump, bank, morn, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
```

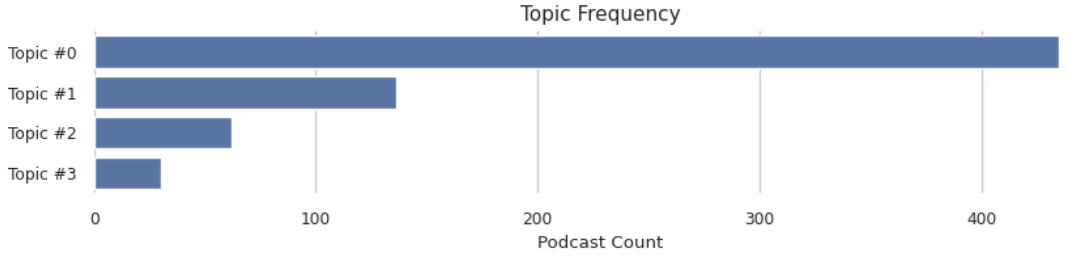
```
chocol, flavor, candi, eat, bread, vanilla, dip
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, villain, ghost christma, scroog, movi christma, santa claus, ghost,
tradit, list
```

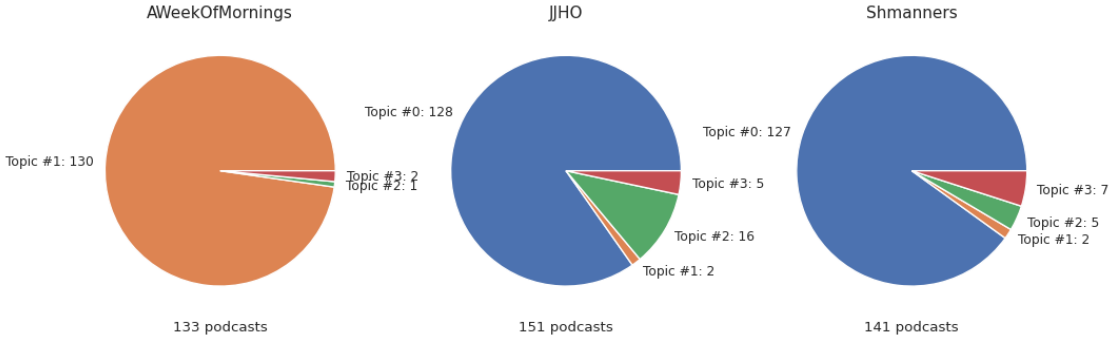
[56]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0      0.020415 -0.070281       1        1  54.651490
1     -0.183452  0.296931       2        1  28.232794
2     -0.167489 -0.262685       3        1   9.973716
3      0.330526  0.036035       4        1   7.142000, topic_info=
Term       Freq      Total Category  logprob  loglift
195860    christma  11.000000  11.000000  Default  30.0000  30.0000
192266      chocol   8.000000   8.000000  Default  29.0000  29.0000
270440       cream   7.000000   7.000000  Default  28.0000  28.0000
566886   ice cream   6.000000   6.000000  Default  27.0000  27.0000
566771         ice   6.000000   6.000000  Default  26.0000  26.0000
…              …         …         …        …        …        …
1206901       tree   0.648435   1.543023   Topic4  -8.0968   1.7722
469017        gift   0.692405   2.601842   Topic4  -8.0312   1.3154
1237517    version   0.726745   3.826348   Topic4  -7.9828   0.9781
179883     charact   0.685999   4.038822   Topic4  -8.0405   0.8663
345753      easter   0.587484   1.328537   Topic4  -8.1956   1.8232

[418 rows x 6 columns], token_table=         Topic      Freq           Term
term
21794        3  0.750089         almond
21794        4  0.750089         almond
31181        1  1.238348            app
39067        1  0.815152            art
59273        3  0.727825         banana
…          …         …              …
1294001      2  1.180578      word week
1312179      4  1.783067  yeah christma
1337670      2  0.860780     yesterday
1339625      1  0.695241          york
1339625      2  0.231747          york

[284 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 3, 4])
```

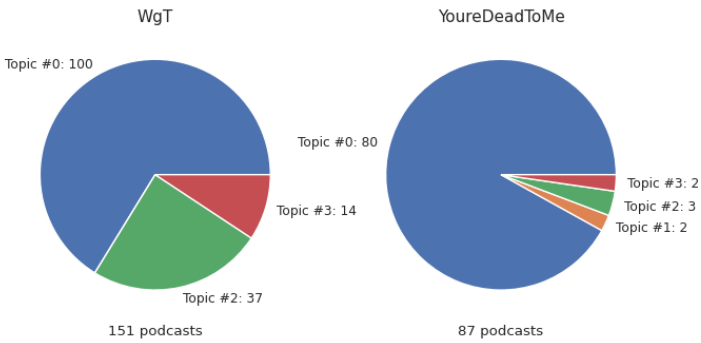
[57]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, water, dog, slash, song, citi, comedi, charact, film,
```

```
watch, york, parti, friend
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, trump, bank, morn, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, dip
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, villain, ghost christma, scroog, movi christma, santa claus, ghost,
tradit, list
```







[58]:
```
topic_count = 5
```

[59]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

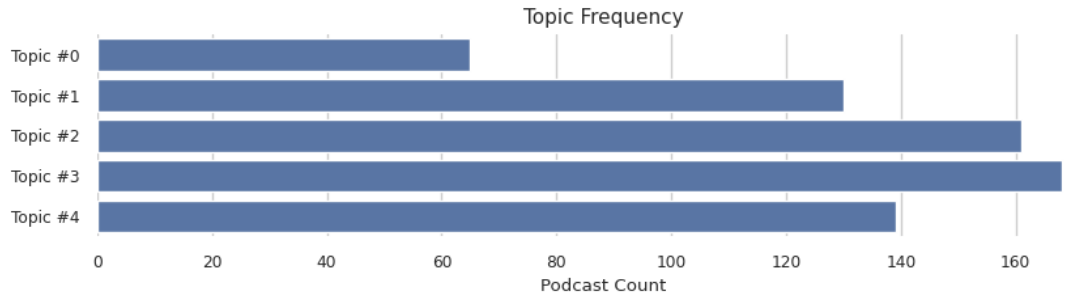
```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 5 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: cream, chocol, ice, water, eat, butter, tea, bread, meat, flavor,
cake, milk, sauc, tast, appl
Topic #1: song, game, christma, charact, list, film, watch, version, audienc,
york, star, theme, disney, car, citi
Topic #2: war, centuri, woman, comedi, citi, power, corner, book, period, radio,
art, age, franc, death, host
Topic #3: book, car, dog, slash, friend, season, store, husband, game, rule,
support, cat, style, box, drive
Topic #4: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
```

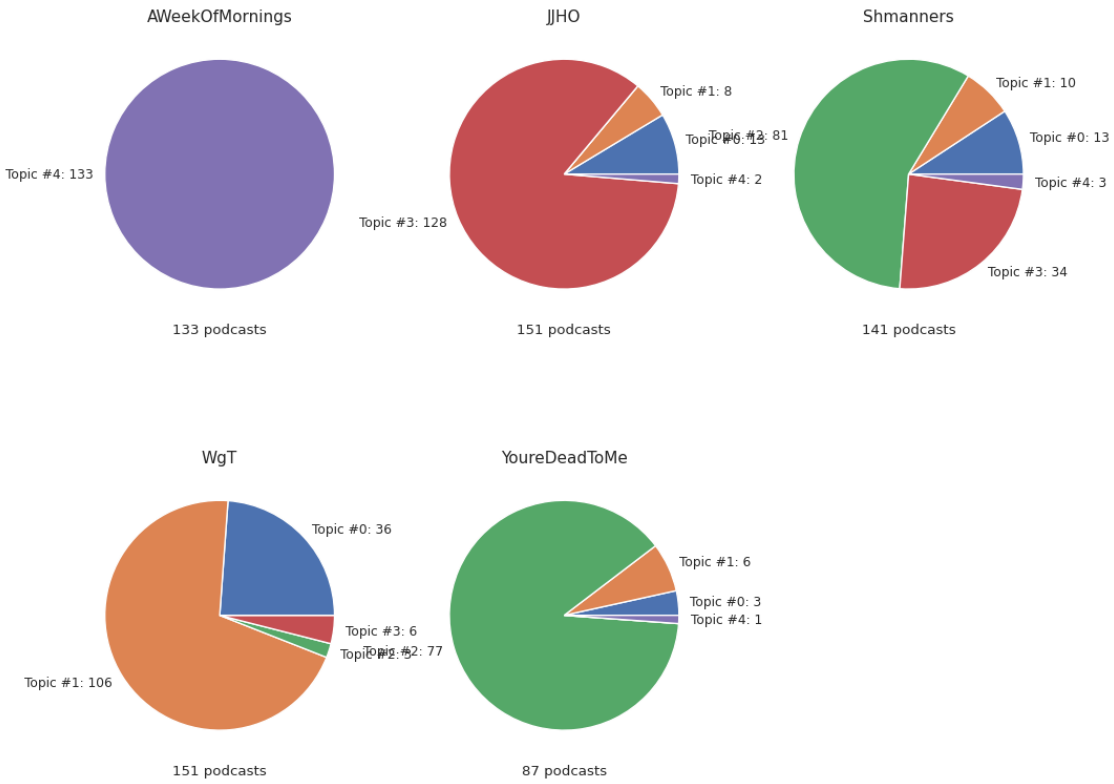
[59]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4     -0.042361  0.020976       1        1  34.468214
3     -0.027934  0.074261       2        1  21.280342
1     -0.025345  0.083625       3        1  17.816612
2     -0.115707 -0.131123       4        1  16.143430
0      0.211347 -0.047739       5        1  10.291401, topic_info=
Term         Freq        Total Category  logprob  loglift
6255      cream   972.000000   972.000000  Default  30.0000  30.0000
28601      wine  2059.000000  2059.000000  Default  29.0000  29.0000
4877     chocol   852.000000   852.000000  Default  28.0000  28.0000
12801       ice   957.000000   957.000000  Default  27.0000  27.0000
24514      song  2391.000000  2391.000000  Default  26.0000  26.0000
…           …           …           …      …       …        …
7431     dinner   225.802876   592.885390   Topic5  -5.6803   1.3085
2108       beef   185.731651   351.471930   Topic5  -5.8757   1.6360
8005      drink   216.708454   640.335053   Topic5  -5.7215   1.1904
4933   christma   248.057129  1216.247454   Topic5  -5.5863   0.6840
6592        cut   204.845200   518.042362   Topic5  -5.7778   1.3461

[433 rows x 6 columns], token_table=       Topic      Freq   Term
term
165        1  0.160382  actor
165        2  0.127762  actor
165        3  0.595316  actor
165        4  0.108734  actor
165        5  0.005437  actor
…      …        …    …
28990      2  0.159384   york
28990      3  0.343527   york
```

```
28990      4  0.106772   york
28990      5  0.051065   york
29124      4  0.986054  zhang

[1130 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 4, 2, 3, 1])
```

[60]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: cream, chocol, ice, water, eat, butter, tea, bread, meat, flavor,
cake, milk, sauc, tast, appl
Topic #1: song, game, christma, charact, list, film, watch, version, audienc,
york, star, theme, disney, car, citi
Topic #2: war, centuri, woman, comedi, citi, power, corner, book, period, radio,
art, age, franc, death, host
Topic #3: book, car, dog, slash, friend, season, store, husband, game, rule,
support, cat, style, box, drive
Topic #4: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
```





[61]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

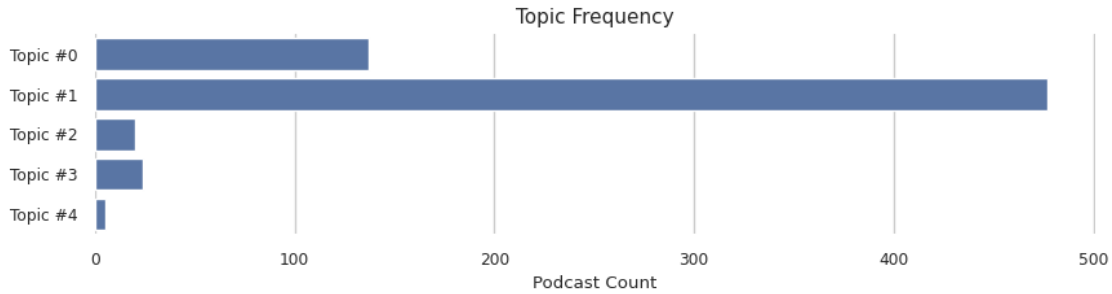
```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 5 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, song, bill, street, congressman, space
Topic #1: game, book, song, car, citi, water, york, watch, charact, audienc,
hand, list, version, dog, friend
Topic #2: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, bowl
Topic #3: christma, song, holiday, ghost, chocol, list, santa, version, claus,
tradit, villain, watch, charact, weapon, winter
Topic #4: cancer, connect, support, wine, bed, month, car, communiti, organ,
diagnosi, river, hampton, pino, presid, oak
```

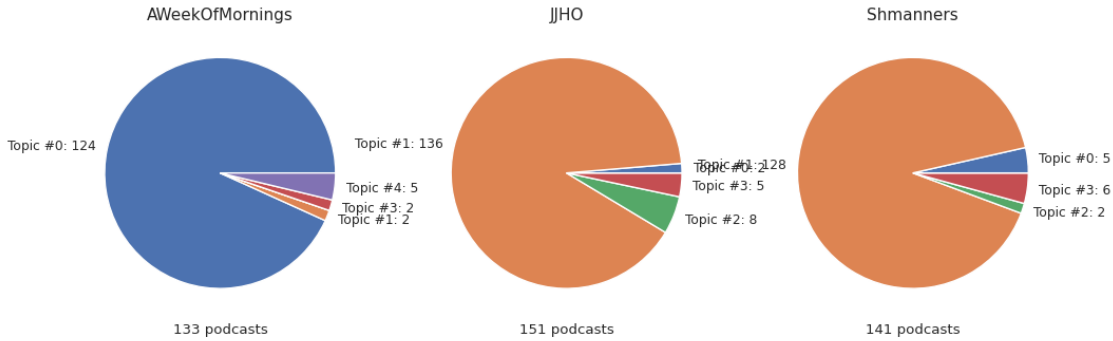
[61]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1      0.018383 -0.002682       1        1  51.013608
0     -0.123957 -0.003498       2        1  31.959266
```

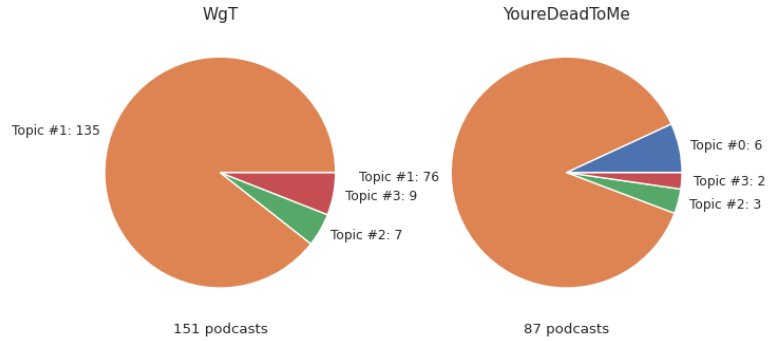
```
2      0.196531  0.219265       3        1   6.417980
3      0.159187 -0.236431       4        1   6.337241
4     -0.250144  0.023347       5        1   4.271905, topic_info=
Term         Freq        Total Category  logprob  loglift
4933   christma  3871.000000  3871.000000  Default  30.0000  30.0000
6255      cream  3018.000000  3018.000000  Default  29.0000  29.0000
12801       ice  2998.000000  2998.000000  Default  28.0000  28.0000
3964     cancer  1781.000000  1781.000000  Default  27.0000  27.0000
5757    connect  1617.000000  1617.000000  Default  26.0000  26.0000
…          …           …            …       …       …       …
17439      morn    93.571437  1565.770261   Topic5  -5.6820   0.3357
2887       book    97.428330  2487.747407   Topic5  -5.6416  -0.0869
25263    street    86.630899  1285.883368   Topic5  -5.7591   0.4556
15203      line    85.677235  1252.505664   Topic5  -5.7702   0.4708
12479      hope    81.579255   778.452591   Topic5  -5.8192   0.8974

[523 rows x 6 columns], token_table=       Topic      Freq       Term
term
26          2  0.994056        abd
57          2  1.003161      abort
93          2  1.127000  acapelago
96          2  0.694910        acc
106         2  0.940036      accol
…       …        …         …
29016       4  1.037325        yul
29077       5  0.976264   zarabodi
29081       5  0.976264   zaribodi
29193       1  0.063412     zucker
29193       5  0.951178     zucker

[970 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 1, 3, 4, 5])
```

[62]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, song, bill, street, congressman, space
Topic #1: game, book, song, car, citi, water, york, watch, charact, audienc,
hand, list, version, dog, friend
Topic #2: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, tast, cone, salt, bowl
Topic #3: christma, song, holiday, ghost, chocol, list, santa, version, claus,
tradit, villain, watch, charact, weapon, winter
Topic #4: cancer, connect, support, wine, bed, month, car, communiti, organ,
diagnosi, river, hampton, pino, presid, oak
```







[63]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 5 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: game, book, car, song, slash, film, dog, charact, list, watch, parti,
friend, york, host, max
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
```

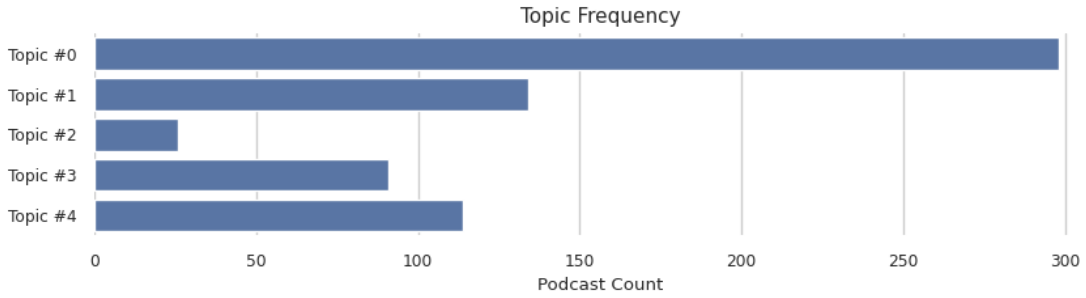
```
Topic #2: christma, holiday, santa, claus, song, scroog, tradit, gift, hanukkah,
winter, ghost, villain, krampus, carol, tree
Topic #3: chocol, cream, ice, butter, cake, eat, bread, milk, candi, breakfast,
flavor, water, dip, soup, meat
Topic #4: war, empir, franc, emperor, centuri, woman, corner, power, henri,
armi, greg, battl, comedi, citi, revolut
```

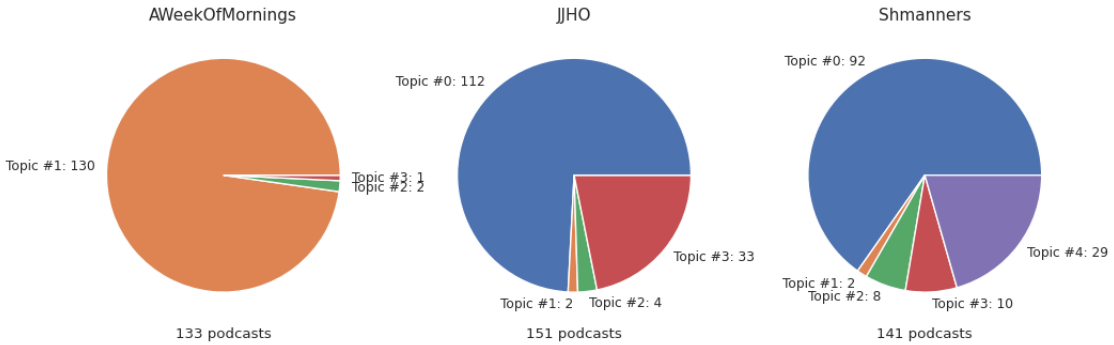
[63]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.015579 -0.029594       1        1  36.654533
1     -0.080204 -0.046453       2        1  25.381129
4     -0.236431  0.026106       3        1  17.416939
3      0.180672 -0.191636       4        1  14.529051
2      0.151542  0.241577       5        1   6.018348, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  24.000000  24.000000  Default  30.0000  30.0000
4877     chocol  17.000000  17.000000  Default  29.0000  29.0000
6255      cream  14.000000  14.000000  Default  28.0000  28.0000
12801       ice  11.000000  11.000000  Default  27.0000  27.0000
12359   holiday   8.000000   8.000000  Default  26.0000  26.0000
…          …         …         …       …       …        …
23194    season   1.401072   6.667271   Topic5  -5.9234   1.2504
575      almond   1.199816   3.386315   Topic5  -6.0785   1.7728
26131  telescop   1.199900   4.817896   Topic5  -6.0784   1.4203
24742    spirit   1.095013   3.395764   Topic5  -6.1699   1.6786
4092       card   1.100003   4.347103   Topic5  -6.1653   1.4362

[525 rows x 6 columns], token_table=       Topic      Freq        Term
term
249        5  0.828818      advent
326        1  0.506431         age
326        2  0.168810         age
326        3  0.337621         age
428        3  1.211071       akbar
…      …       …          …
28708      5  0.720614  wonderland
28937      2  0.889651   yesterday
28990      1  0.656039        york
28990      2  0.262415        york
29124      3  0.894453       zhang

[480 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 5, 4, 3])
```

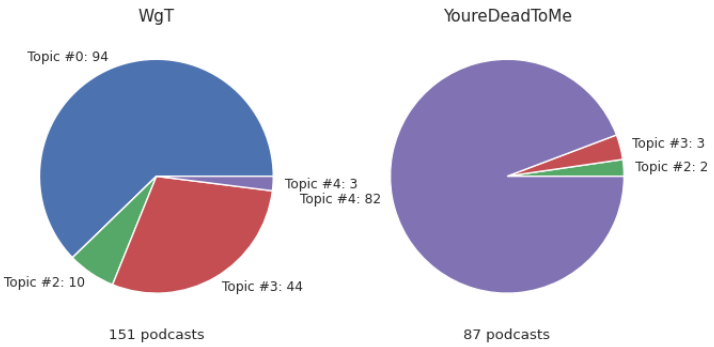
[64]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
```

```
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, song, slash, film, dog, charact, list, watch, parti,
friend, york, host, max
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
Topic #2: christma, holiday, santa, claus, song, scroog, tradit, gift, hanukkah,
winter, ghost, villain, krampus, carol, tree
Topic #3: chocol, cream, ice, butter, cake, eat, bread, milk, candi, breakfast,
flavor, water, dip, soup, meat
Topic #4: war, empir, franc, emperor, centuri, woman, corner, power, henri,
armi, greg, battl, comedi, citi, revolut
```







[65]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

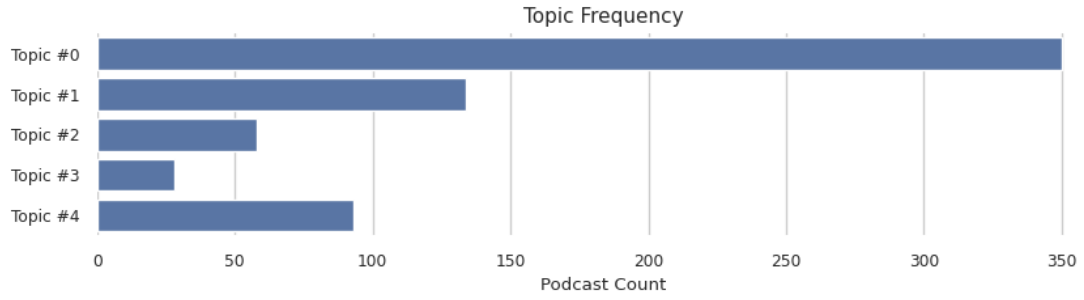
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 5 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: game, book, car, song, slash, dog, water, charact, watch, friend,
film, list, york, max, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, bill
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, dip
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
song, scroog, villain, ghost christma, santa claus, movi christma, tradit,
ghost, hanukkah
Topic #4: empir, war, emperor, franc, centuri, power, corner, woman, armi,
henri, battl, greg, comedi, byron, shelley
```

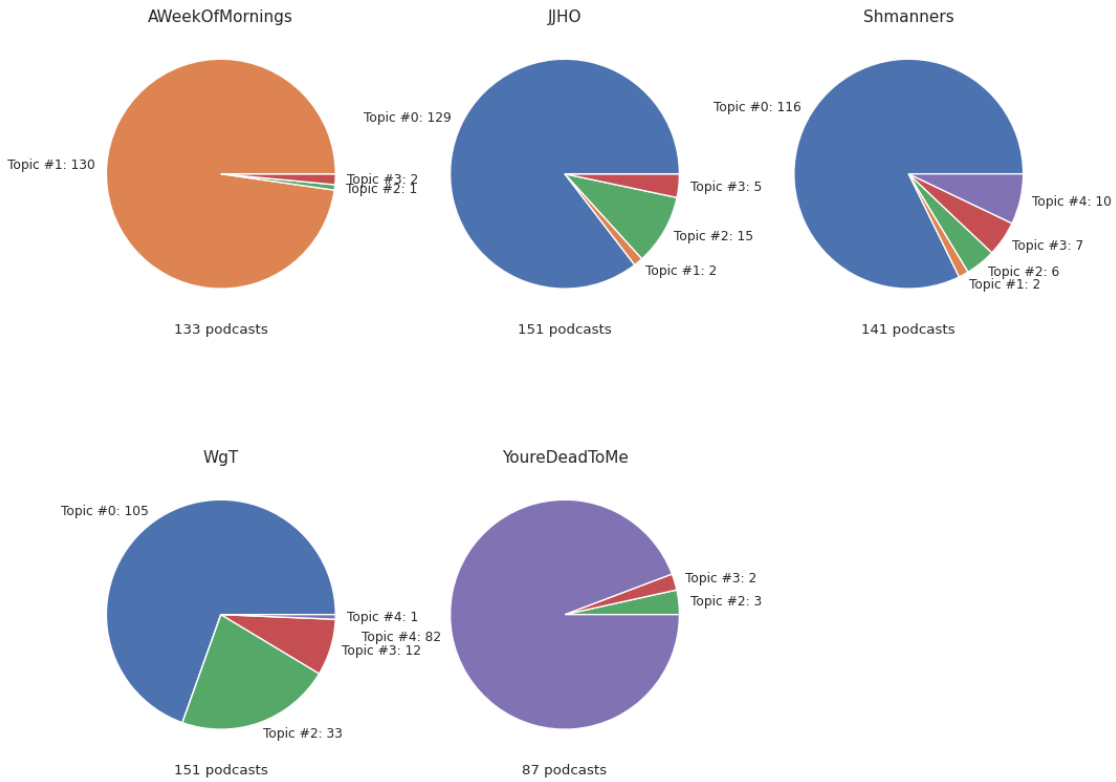
[65]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0      0.006990 -0.026616       1        1  42.045656
1     -0.051472 -0.071994       2        1  26.389491
4     -0.321531 -0.019015       3        1  15.730092
2      0.241048 -0.209614       4        1   9.397313
3      0.124965  0.327239       5        1   6.437447, topic_info=
Term       Freq      Total Category  logprob  loglift
84847    christma  13.000000  13.000000  Default  30.0000  30.0000
83347      chocol  10.000000  10.000000  Default  29.0000  29.0000
116310      cream   8.000000   8.000000  Default  28.0000  28.0000
240891  ice cream   7.000000   7.000000  Default  27.0000  27.0000
240842        ice   7.000000   7.000000  Default  26.0000  26.0000
…           …        …        …      …      …      …
9786       almond   0.717060   1.626241   Topic5  -7.5252   1.9242
323077  movi movi   0.772985   2.531557   Topic5  -7.4501   1.5567
518151    version   0.839849   4.472372   Topic5  -7.3672   1.0706
149984     easter   0.715710   1.622151   Topic5  -7.5271   1.9248
77996     charact   0.724878   4.717385   Topic5  -7.5144   0.8700

[526 rows x 6 columns], token_table=        Topic      Freq           Term
term
5008        5  1.642464         advent
9786        4  0.614915         almond
9786        5  0.614915         almond
```

```
11610       3  0.631706        ancient
13760       1  1.264686          apart
…         …        …              …
546555      5  1.517283  yeah christma
552419      2  0.946437      yesterday
553250      1  0.602540           york
553250      2  0.401693           york
554602      3  1.274985         zhang

[387 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 5, 3, 4])
```

[66]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, song, slash, dog, water, charact, watch, friend,
film, list, york, max, parti
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, bill
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, dip
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
song, scroog, villain, ghost christma, santa claus, movi christma, tradit,
ghost, hanukkah
Topic #4: empir, war, emperor, franc, centuri, power, corner, woman, armi,
henri, battl, greg, comedi, byron, shelley
```





[67]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 5 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: game, book, car, song, slash, dog, water, charact, watch, friend,
york, cat, season, max, drive
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, trump, morn, bank, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
chocol, flavor, candi, vanilla, eat, cream cake, bread
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, villain, ghost christma, scroog, movi christma, santa claus, ghost,
tradit, list
Topic #4: empir, emperor, war, franc, centuri, power, corner, woman, armi,
battl, henri, greg, byron, comedi, shelley
```

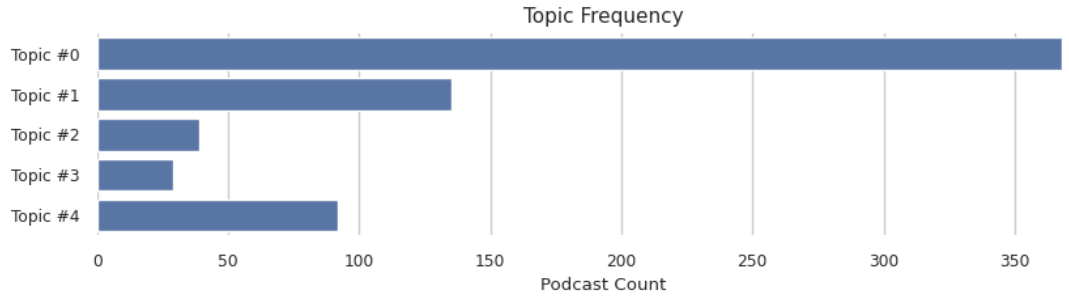
[67]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
```

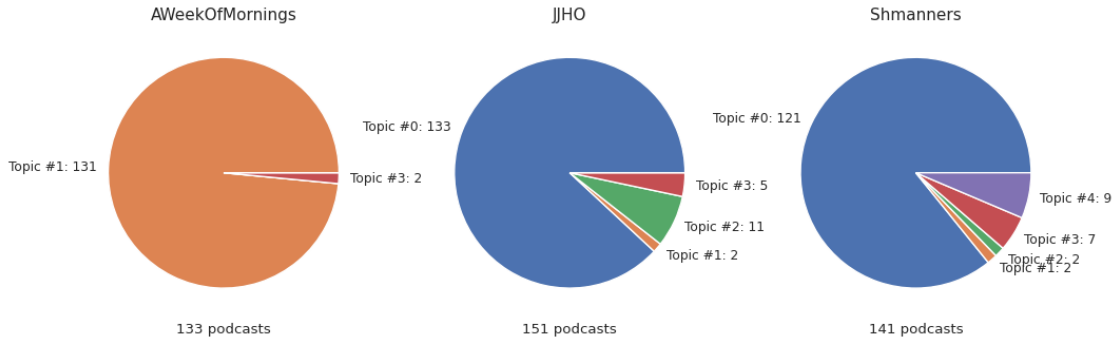
```
0     -0.028144  0.012530       1        1  42.637442
1      0.029589 -0.142901       2        1  27.269715
4      0.352181 -0.021788       3        1  15.339165
2     -0.250426 -0.189894       4        1   8.144872
3     -0.103200  0.342053       5        1   6.608806, topic_info=
Term       Freq      Total Category  logprob  loglift
195860    christma  10.000000  10.000000  Default  30.0000  30.0000
192266      chocol   8.000000   8.000000  Default  29.0000  29.0000
270440       cream   6.000000   6.000000  Default  28.0000  28.0000
566886   ice cream   6.000000   6.000000  Default  27.0000  27.0000
566771         ice   5.000000   5.000000  Default  26.0000  26.0000
…              …        …         …       …        …        …
1206901       tree   0.613722   1.514196   Topic5  -8.0743   1.8137
469017        gift   0.646681   2.464912   Topic5  -8.0220   1.3787
1237517    version   0.670892   3.587057   Topic5  -7.9852   1.0403
179883     charact   0.634155   3.779441   Topic5  -8.0415   0.9317
345753      easter   0.554554   1.241117   Topic5  -8.1757   1.9112

[510 rows x 6 columns], token_table=         Topic      Freq           Term
term
18624        3  1.949398          akbar
21794        4  0.794901         almond
21794        5  0.794901         almond
25717        3  0.782369        ancient
30376        1  0.769480          apart
…          …        …              …
1312179      5  1.886820  yeah christma
1337670      2  0.887311      yesterday
1339625      1  0.496925           york
1339625      2  0.248462           york
1342965      3  1.487451          zhang

[354 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 5, 3, 4])
```

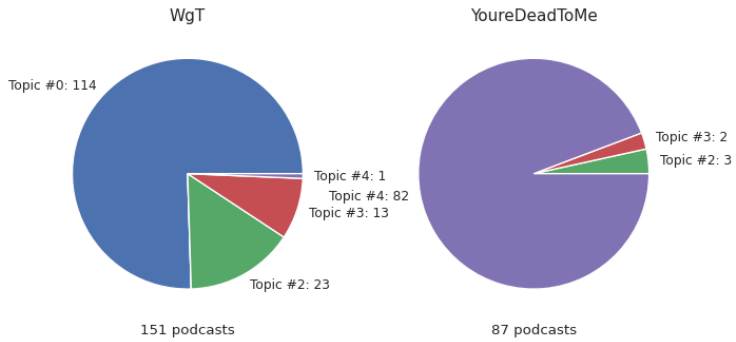
[68]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, book, car, song, slash, dog, water, charact, watch, friend,
york, cat, season, max, drive
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, trump, morn, bank, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cake, cocoa, butter, milk, chocol
chocol, flavor, candi, vanilla, eat, cream cake, bread
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, villain, ghost christma, scroog, movi christma, santa claus, ghost,
```

```
tradit, list
Topic #4: empir, emperor, war, franc, centuri, power, corner, woman, armi,
battl, henri, greg, byron, comedi, shelley
```







[69]:
```
topic_count = 6
```

[70]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 6 , tidf ngram range: (1, 1)
```

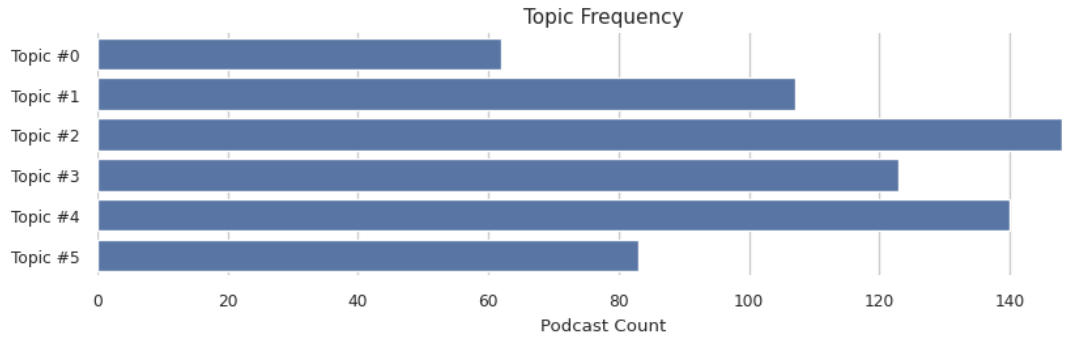
```
--------------------------------------------------------------------------------
Topic #0: cream, chocol, ice, eat, water, butter, tea, bread, flavor, meat,
christma, milk, sauc, cake, chicken
Topic #1: song, game, christma, charact, list, film, watch, version, car, star,
theme, disney, audienc, book, topic
Topic #2: book, york, parti, host, art, busi, slash, check, hand, citi, chair,
etiquett, danc, audienc, pictur
Topic #3: dog, book, rule, friend, season, slash, car, husband, justic, cat,
evid, water, letter, style, court
Topic #4: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
Topic #5: war, centuri, comedi, woman, power, corner, radio, franc, citi,
period, empir, age, god, death, battl
```

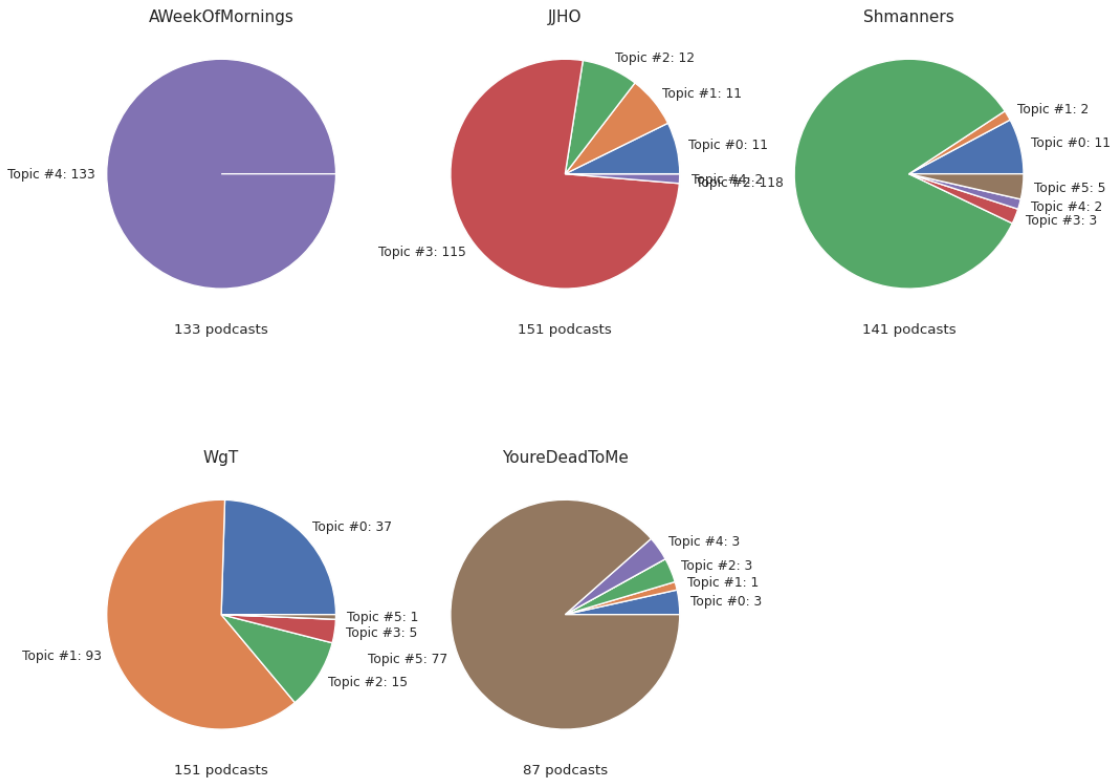
[70]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4     -0.043386  0.007929       1        1  34.549541
3     -0.023804  0.061817       2        1  16.047758
1     -0.012160  0.098229       3        1  15.393132
2     -0.036841  0.049176       4        1  12.680534
5     -0.119614 -0.157637       5        1  11.809026
0      0.235805 -0.059514       6        1   9.520009, topic_info=
Term         Freq        Total Category  logprob  loglift
6255      cream   979.000000   979.000000  Default  30.0000  30.0000
4877     chocol   858.000000   858.000000  Default  29.0000  29.0000
28601      wine  2068.000000  2068.000000  Default  28.0000  28.0000
12801       ice   965.000000   965.000000  Default  27.0000  27.0000
24514      song  2396.000000  2396.000000  Default  26.0000  26.0000
…           …           …           …       …       …       …
5365      coffe   243.061105   592.896058   Topic6  -5.5288   1.4601
4933   christma   328.911467  1216.071715   Topic6  -5.2263   1.0442
7431     dinner   212.936809   593.761472   Topic6  -5.6611   1.3263
8005      drink   215.150463   643.161512   Topic6  -5.6508   1.2567
6592        cut   199.177240   519.592151   Topic6  -5.7279   1.3929

[527 rows x 6 columns], token_table=       Topic      Freq   Term
term
165        1  0.165928  actor
165        2  0.097925  actor
165        3  0.546747  actor
165        4  0.027201  actor
165        5  0.157768  actor
…      …        …    …
28990      3  0.181276   york
28990      4  0.320718   york
28990      5  0.004648   york
```

```
28990      6  0.051904   york
29124      5  0.992756  zhang

[1520 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 4, 2, 3, 6, 1])
```

[71]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: cream, chocol, ice, eat, water, butter, tea, bread, flavor, meat,
christma, milk, sauc, cake, chicken
Topic #1: song, game, christma, charact, list, film, watch, version, car, star,
theme, disney, audienc, book, topic
Topic #2: book, york, parti, host, art, busi, slash, check, hand, citi, chair,
etiquett, danc, audienc, pictur
Topic #3: dog, book, rule, friend, season, slash, car, husband, justic, cat,
evid, water, letter, style, court
Topic #4: wine, river, state, communiti, presid, countri, morn, song, radio,
connect, book, yesterday, bill, street, support
Topic #5: war, centuri, comedi, woman, power, corner, radio, franc, citi,
period, empir, age, god, death, battl
```





[72]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 6 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, bill, street, congressman, space, trump
Topic #1: game, book, water, car, dog, hand, watch, slash, friend, charact,
comedi, eat, season, box, busi
Topic #2: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, cone, tast, salt, bowl
Topic #3: christma, holiday, ghost, chocol, santa, tradit, claus, list, villain,
watch, version, song, charact, weapon, mom
Topic #4: cancer, connect, support, wine, bed, month, car, communiti, organ,
diagnosi, river, hampton, song, pino, presid
Topic #5: song, audienc, citi, york, list, danc, particip, album, rock, band,
version, da, perform, film, theme
```

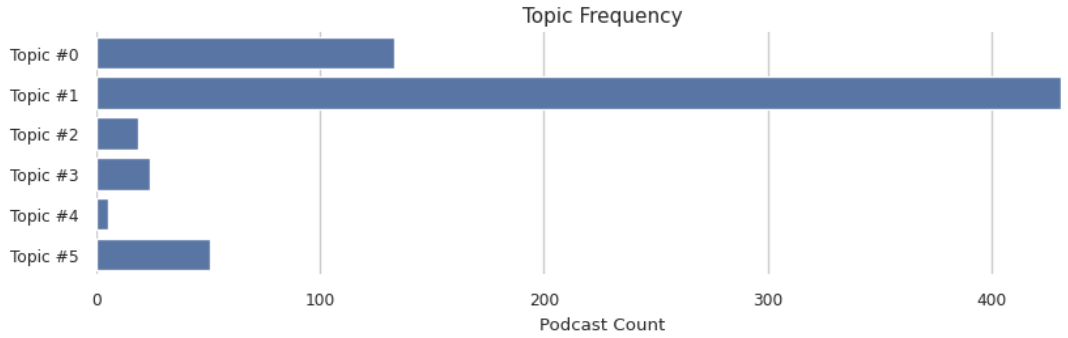
[72]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
```

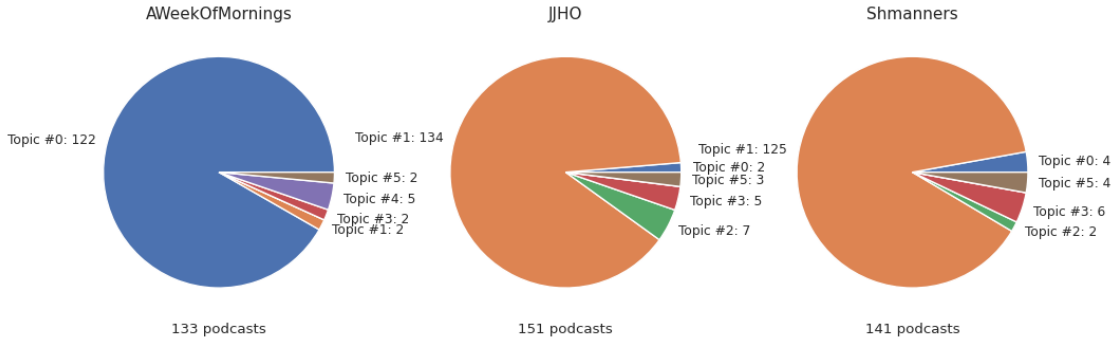
```
1      0.021285 -0.021568       1        1  42.487785
0     -0.116609  0.045499       2        1  28.934631
5     -0.047249 -0.237048       3        1  13.873080
2      0.239682  0.167826       4        1   5.468365
3      0.137844 -0.093937       5        1   5.323640
4     -0.234952  0.139228       6        1   3.912499, topic_info=
Term         Freq        Total Category  logprob  loglift
4933   christma  3330.000000  3330.000000  Default  30.0000  30.0000
6255      cream  2701.000000  2701.000000  Default  29.0000  29.0000
12801       ice  2677.000000  2677.000000  Default  28.0000  28.0000
24514      song  5234.000000  5234.000000  Default  27.0000  27.0000
3964     cancer  1655.000000  1655.000000  Default  26.0000  26.0000
…          …           …           …      …       …       …
17439      morn    86.022423  1461.782305   Topic6  -5.6783   0.4082
2887       book    85.730079  2309.355419   Topic6  -5.6817  -0.0525
25263    street    80.310771  1305.518249   Topic6  -5.7470   0.4525
15203      line    78.825438  1249.017163   Topic6  -5.7656   0.4781
12479      hope    75.452763   756.732161   Topic6  -5.8094   0.9355

[628 rows x 6 columns], token_table=       Topic      Freq       Term
term
3          1  1.195487   aaronson
4          2  0.761445   aassenha
9          1  1.455091  abakhumid
36         3  0.927726   aberprob
83         1  0.895168        abu
…      …       …         …
28990      3  0.589162       york
28990      4  0.011990       york
29016      5  0.988953        yul
29077      6  1.049743   zarabodi
29081      6  0.979760   zaribodi

[1237 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 1, 6, 3, 4, 5])
```

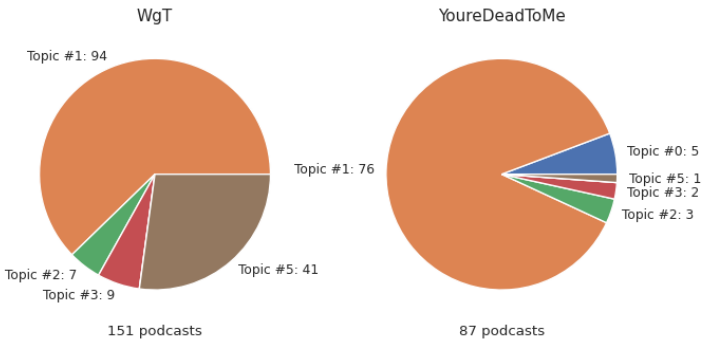
[73]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, presid, countri, communiti, morn, radio,
yesterday, bank, bill, street, congressman, space, trump
Topic #1: game, book, water, car, dog, hand, watch, slash, friend, charact,
comedi, eat, season, box, busi
Topic #2: cream, ice, cake, chocol, water, milk, eat, flavor, centuri, butter,
pie, cone, tast, salt, bowl
Topic #3: christma, holiday, ghost, chocol, santa, tradit, claus, list, villain,
```

```
watch, version, song, charact, weapon, mom
Topic #4: cancer, connect, support, wine, bed, month, car, communiti, organ,
diagnosi, river, hampton, song, pino, presid
Topic #5: song, audienc, citi, york, list, danc, particip, album, rock, band,
version, da, perform, film, theme
```







[74]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

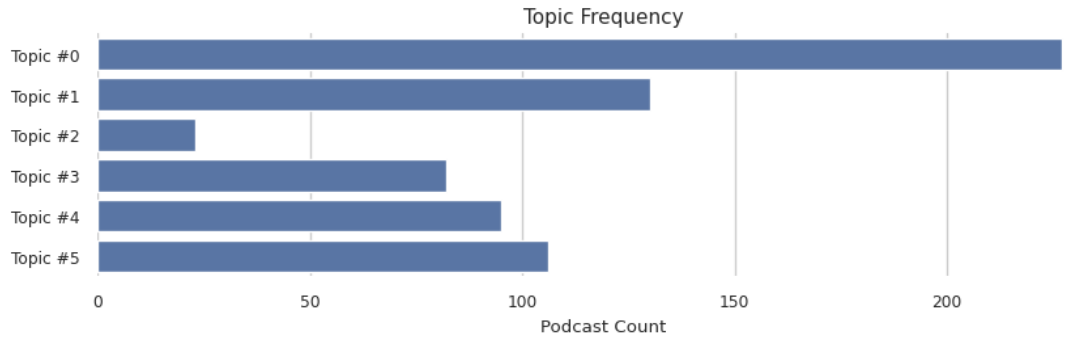
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 6 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: book, game, slash, dog, car, fund, cat, max, water, friend, parti,
husband, letter, etiquett, host
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
Topic #2: christma, holiday, santa, claus, scroog, tradit, song, gift, hanukkah,
winter, ghost, krampus, tree, carol, villain
Topic #3: chocol, cream, ice, butter, cake, milk, candi, eat, bread, flavor,
breakfast, cocoa, dip, cooki, meat
Topic #4: war, empir, emperor, franc, centuri, woman, corner, power, henri,
armi, greg, battl, comedi, citi, revolut
Topic #5: song, film, list, charact, disney, star, watch, game, version,
theater, rock, theme, tv, danc, car
```

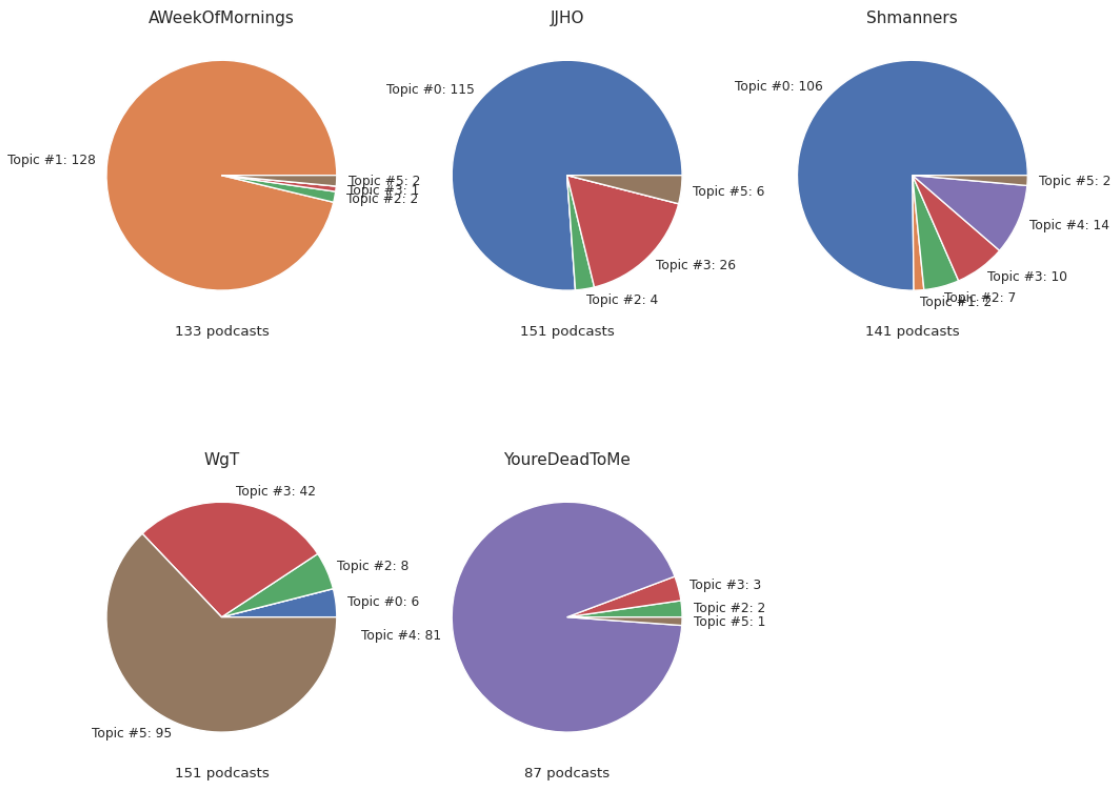
[74]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.026471  0.047181       1        1  26.490907
1     -0.080596  0.029308       2        1  23.851350
5     -0.046909  0.035240       3        1  18.048575
4     -0.221210 -0.063867       4        1  14.874388
3      0.187851  0.198210       5        1  11.763165
2      0.187335 -0.246072       6        1   4.971615, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  22.000000  22.000000  Default  30.0000  30.0000
4877     chocol  16.000000  16.000000  Default  29.0000  29.0000
6255      cream  12.000000  12.000000  Default  28.0000  28.0000
12801       ice  10.000000  10.000000  Default  27.0000  27.0000
28601      wine  16.000000  16.000000  Default  26.0000  26.0000
…           …         …         …       …        …        …
26131  telescop   1.097157   4.615250   Topic6  -5.9769   1.5648
27747   version   1.108084   7.939645   Topic6  -5.9670   1.0322
24742    spirit   0.968391   3.297035   Topic6  -6.1017   1.7763
4092       card   0.964542   4.351080   Topic6  -6.1057   1.4949
4393     celebr   0.960004   4.892376   Topic6  -6.1104   1.3729

[638 rows x 6 columns], token_table=       Topic      Freq       Term
term
161        2  0.243976     action
161        3  0.731927     action
165        3  0.783608      actor
165        4  0.261203      actor
249        6  0.906607     advent
…      …        …         …
28937      2  0.920448  yesterday
```

```
28990       1  0.243636       york
28990       2  0.243636       york
28990       3  0.365453       york
29124       4  0.999732      zhang

[594 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 6, 5, 4, 3])
```

[75]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: book, game, slash, dog, car, fund, cat, max, water, friend, parti,
husband, letter, etiquett, host
Topic #1: wine, river, presid, state, yesterday, communiti, congressman,
countri, trump, morn, hampton, bank, bill, radio, vote
Topic #2: christma, holiday, santa, claus, scroog, tradit, song, gift, hanukkah,
winter, ghost, krampus, tree, carol, villain
Topic #3: chocol, cream, ice, butter, cake, milk, candi, eat, bread, flavor,
breakfast, cocoa, dip, cooki, meat
Topic #4: war, empir, emperor, franc, centuri, woman, corner, power, henri,
armi, greg, battl, comedi, citi, revolut
Topic #5: song, film, list, charact, disney, star, watch, game, version,
theater, rock, theme, tv, danc, car
```





[76]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

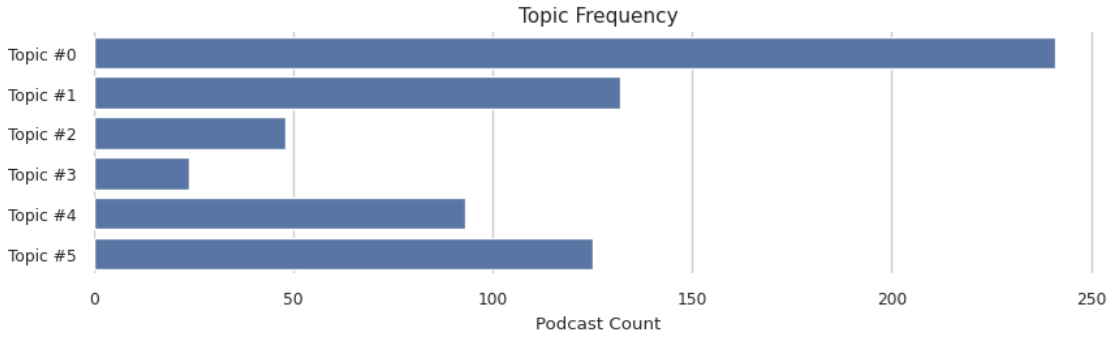
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 6 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: book, game, slash, dog, water, car, fund, cat, pizza, max, husband,
friend, birthday, letter, docket
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, connect
Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, chocol
chocol, flavor, candi, eat, bread, vanilla, peanut butter
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, song, ghost christma, tradit, santa claus, movi christma, villain,
ghost, hanukkah
Topic #4: empir, war, emperor, franc, centuri, corner, power, woman, armi,
henri, battl, greg, comedi, shelley, byron
Topic #5: song, film, list, charact, disney, game, star, version, watch, movi
movi, theme song, york, theme, topic, car
```

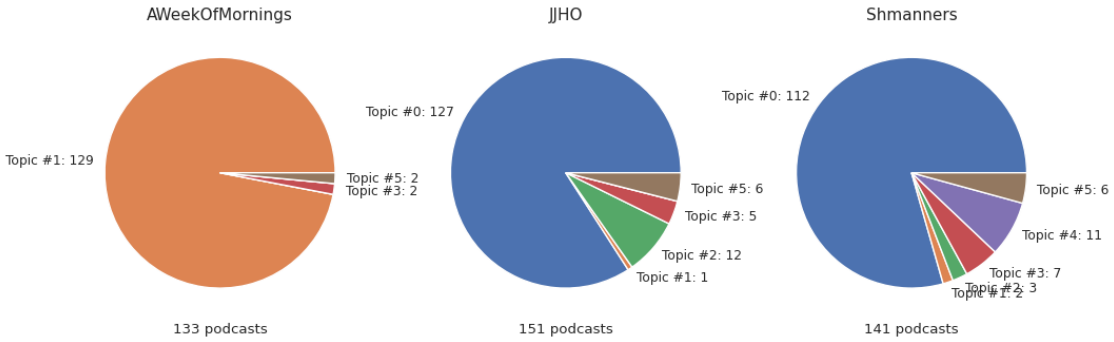
[76]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0      0.028610  0.060674       1        1  27.672821
1      0.047760  0.008975       2        1  24.763989
5      0.017156  0.104632       3        1  20.489112
4      0.311612 -0.074269       4        1  14.463132
2     -0.204589  0.216245       5        1   7.930590
3     -0.200549 -0.316256       6        1   4.680357, topic_info=
Term       Freq      Total Category  logprob  loglift
84847    christma  10.000000  10.000000  Default  30.0000  30.0000
83347      chocol   9.000000   9.000000  Default  29.0000  29.0000
116310      cream   7.000000   7.000000  Default  28.0000  28.0000
240891  ice cream   6.000000   6.000000  Default  27.0000  27.0000
240842        ice   6.000000   6.000000  Default  26.0000  26.0000
…             …         …         …       …        …        …
9786       almond   0.563318   1.386579   Topic6  -7.4478   2.1610
446912       song   0.944528   9.940100   Topic6  -6.9310   0.7081
149984     easter   0.548337   1.459442   Topic6  -7.4747   2.0829
256078      jesus   0.526067   1.790500   Topic6  -7.5162   1.8370
518151    version   0.527222   4.695380   Topic6  -7.5140   0.8751

[636 rows x 6 columns], token_table=        Topic      Freq       Term
term
2599        2  0.404369     action
2599        3  0.808739     action
3021        3  0.839833      actor
8302        4  1.851515      akbar
8667        2  0.379550      album
…       …        …         …
553250      1  0.192672       york
553250      2  0.192672       york
553250      3  0.385344       york
553571      3  1.681891  york movi
554602      4  1.380389      zhang

[449 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 6, 5, 3, 4])
```

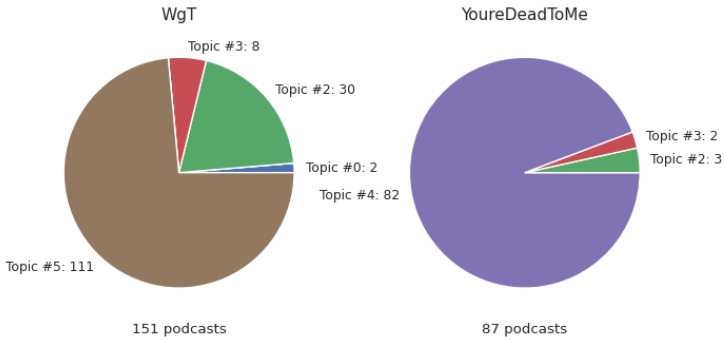
[77]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: book, game, slash, dog, water, car, fund, cat, pizza, max, husband,
friend, birthday, letter, docket
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, food bank, trump, morn, bank, hampton, cancer, connect
```

Topic #2: chocol, cream, ice cream, ice, cake, butter, cocoa, milk, chocol chocol, flavor, candi, eat, bread, vanilla, peanut butter
Topic #3: christma, holiday, christma movi, santa, claus, christma christma, scroog, song, ghost christma, tradit, santa claus, movi christma, villain, ghost, hanukkah
Topic #4: empir, war, emperor, franc, centuri, corner, power, woman, armi, henri, battl, greg, comedi, shelley, byron
Topic #5: song, film, list, charact, disney, game, star, version, watch, movi movi, theme song, york, theme, topic, car







[78]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

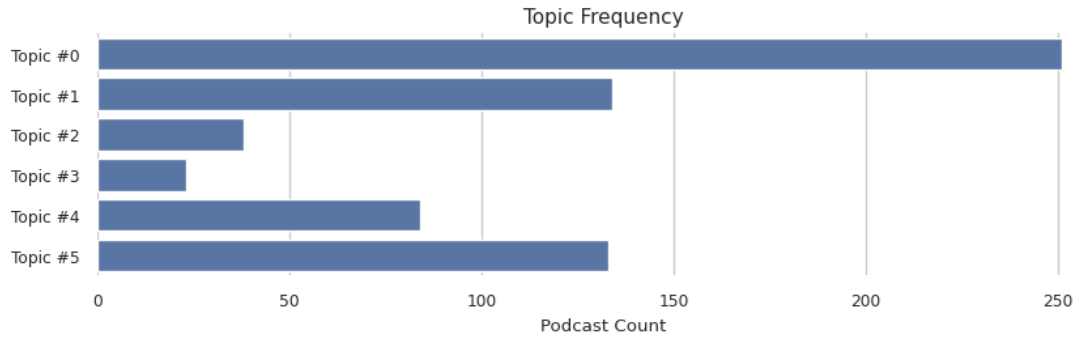
```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 6 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: game, song, list, charact, book, film, host, disney, topic, parti,
york, car, version, star, watch
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, morn, bank, trump, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cocoa, cake, butter, milk, chocol
chocol, flavor, candi, vanilla, eat, bread, cream cake
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, ghost christma, villain, movi christma, scroog, santa claus, ghost,
tradit, place christma
Topic #4: empir, emperor, war, franc, corner, power, centuri, armi, woman,
battl, greg, henri, comedi, byron, shelley
Topic #5: dog, turkey, evid, cat, rule, justic, season, pizza, car, birthday,
rudi, slash, courtroom, disput, mom
```

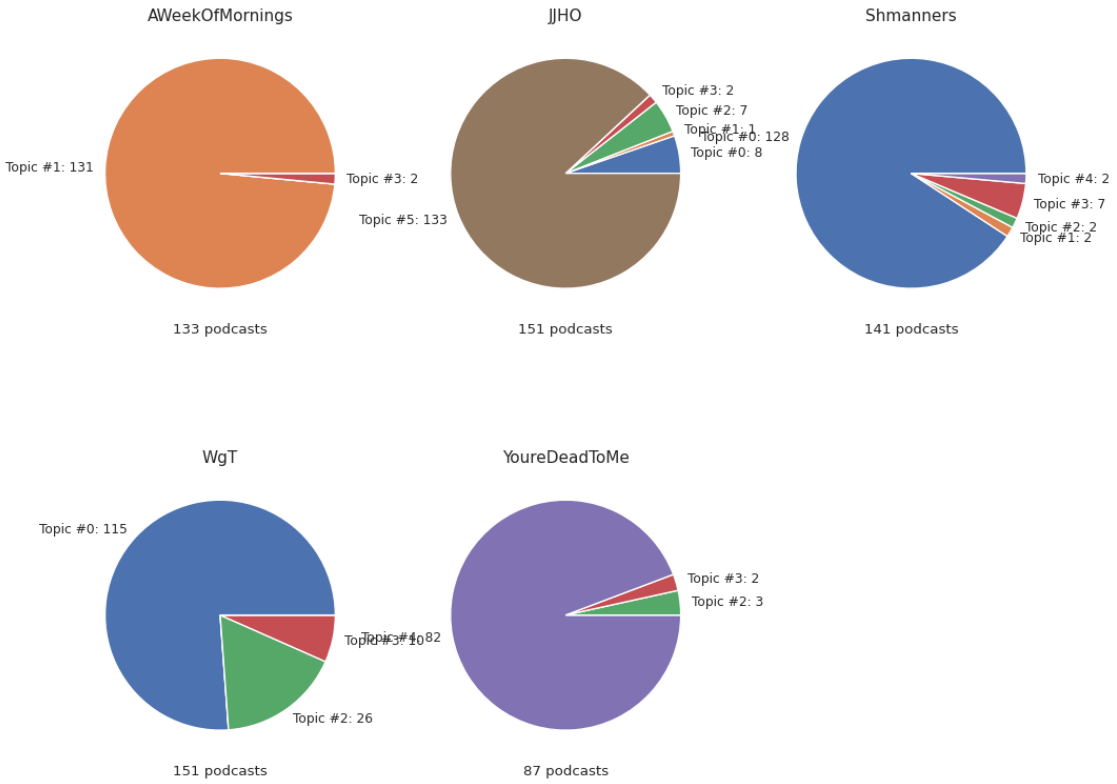
[78]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0      0.001353 -0.069593       1        1  30.729602
1     -0.052334  0.020278       2        1  26.348785
5      0.131991 -0.123235       3        1  18.243417
4     -0.369810 -0.056969       4        1  12.647337
2      0.215731 -0.143121       5        1   7.221365
3      0.073071  0.372640       6        1   4.809494, topic_info=
Term      Freq     Total Category  logprob  loglift
195860    christma  8.000000  8.000000  Default  30.0000  30.0000
192266      chocol  7.000000  7.000000  Default  29.0000  29.0000
270440       cream  6.000000  6.000000  Default  28.0000  28.0000
566886   ice cream  5.000000  5.000000  Default  27.0000  27.0000
566771         ice  5.000000  5.000000  Default  26.0000  26.0000
…              …       …       …       …       …       …
469017        gift  0.509280  2.340963   Topic6  -7.9430   1.5093
764996   movi movi  0.456838  2.002062   Topic6  -8.0517   1.5570
669287        list  0.468017  3.651936   Topic6  -8.0275   0.9801
1237517    version  0.451502  3.465833   Topic6  -8.0634   0.9965
345753      easter  0.408527  1.191030   Topic6  -8.1634   1.9646

[623 rows x 6 columns], token_table=         Topic      Freq       Term
term
31181        1  0.657901        app
38195        4  1.120868       armi
```

```
47741         1  0.830138    audienc
61131         2  1.027802       bank
62375         1  0.341116        bar
…           …        …          …
1337670       2  0.912498  yesterday
1339625       1  0.493993       york
1339625       2  0.246997       york
1339625       3  0.246997       york
1342965       4  1.652132      zhang

[385 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 2, 6, 5, 3, 4])
```

[79]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, song, list, charact, book, film, host, disney, topic, parti,
york, car, version, star, watch
Topic #1: wine, river, presid, state, communiti, yesterday, congressman,
countri, cancer, food bank, morn, bank, trump, hampton, connect
Topic #2: chocol, cream, ice cream, ice, cocoa, cake, butter, milk, chocol
chocol, flavor, candi, vanilla, eat, bread, cream cake
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
song, ghost christma, villain, movi christma, scroog, santa claus, ghost,
tradit, place christma
Topic #4: empir, emperor, war, franc, corner, power, centuri, armi, woman,
battl, greg, henri, comedi, byron, shelley
Topic #5: dog, turkey, evid, cat, rule, justic, season, pizza, car, birthday,
rudi, slash, courtroom, disput, mom
```





[80]:
```
topic_count = 15
```

[81]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 15 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: cream, ice, eat, butter, chocol, bread, meat, water, flavor, cake,
pizza, chicken, pie, soup, sauc
Topic #1: game, film, footbal, bar, mission, car, byron, shelley, vampir,
potato, team, paint, watch, basebal, sport
Topic #2: song, car, york, citi, disney, list, version, chair, danc, rock, roll,
band, kid, theme, wait
Topic #3: charact, audienc, film, song, star, list, watch, version, perform,
scene, televis, comedi, theme, actor, theater
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
connect, book, yesterday, bill, street, congressman
Topic #5: citi, empir, emperor, battl, peter, china, power, system, democraci,
armi, burton, water, comedi, worm, game
Topic #6: birthday, edit, caesar, poni, blame, celebr, watch, kyle, film, greg,
dragon, julius, nanci, court, answer
Topic #7: chocol, stone, beer, cocoa, clock, space, charact, power, water,
```

```
easter, hair, seri, age, milk, war
Topic #8: christma, book, dog, friend, season, slash, store, game, mom, holiday,
cat, box, husband, rule, dad
Topic #9: book, tank, tie, carniv, fish, bow, belt, sandler, burlesqu, jenga,
aquarium, gallon, jinger, star, neck
Topic #10: host, etiquett, check, slash, art, advic, hand, busi, exampl,
mcelroy, hair, pictur, appl, cover, parti
Topic #11: war, centuri, woman, comedi, franc, power, corner, radio, god, hors,
period, book, death, england, son
Topic #12: style, main, jazz, england, wood, town, tenni, radio, communiti,
news, sauc, game, fruit, ben, heat
Topic #13: rudi, smell, glitter, roller, comfort, josh, fan, gretchen, count,
denis, cameron, ink, hotel, jeremi, jerom
Topic #14: tea, water, game, sale, market, system, drink, tast, bathroom, video,
milk, experi, island, afternoon, cup
```

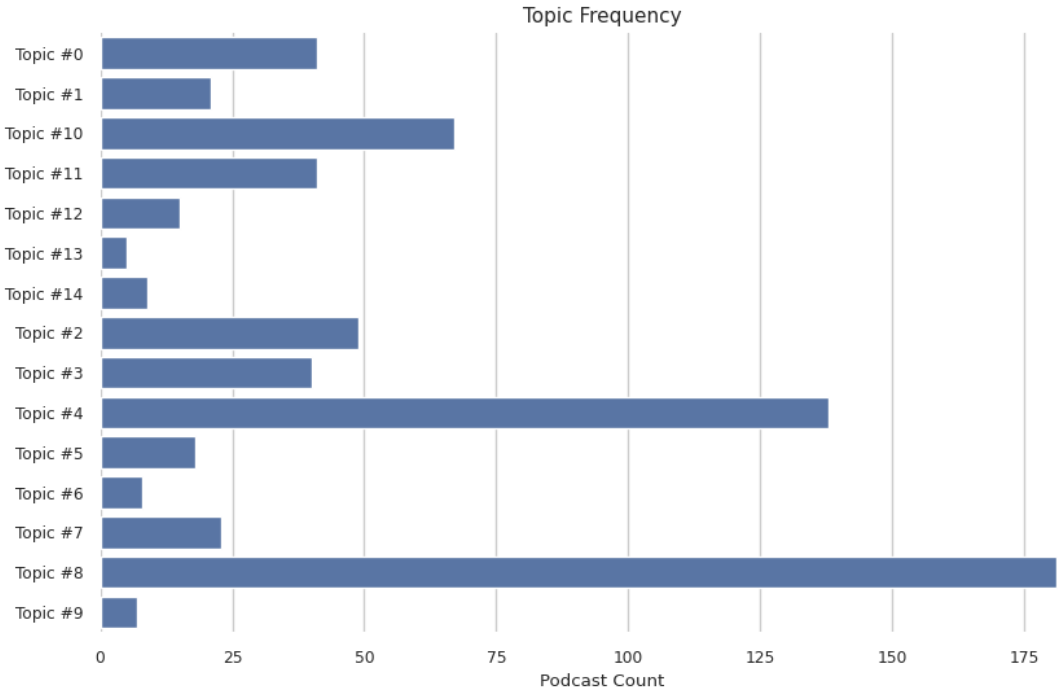
[81]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4      0.104340 -0.027336       1        1  31.413491
8      0.144224  0.004317       2        1  19.231978
2      0.133338  0.032928       3        1   7.987568
0      0.100490  0.227077       4        1   6.467036
11     0.076274 -0.186428       5        1   6.421300
10     0.077564 -0.047396       6        1   6.237074
3      0.113452  0.009453       7        1   5.121827
12    -0.064762  0.023112       8        1   3.230967
1      0.047838  0.041825       9        1   3.066545
7     -0.031854 -0.041285      10        1   2.710586
5     -0.013804 -0.129547      11        1   2.634422
14    -0.095643  0.108630      12        1   1.815744
6     -0.149124 -0.063730      13        1   1.587732
13    -0.223039  0.071771      14        1   1.190991
9     -0.219294 -0.023392      15        1   0.882738, topic_info=         Term
Freq        Total Category  logprob  loglift
2887     book  2293.000000  2293.000000  Default  30.0000  30.0000
10295    game  2101.000000  2101.000000  Default  29.0000  29.0000
28222   water  1728.000000  1728.000000  Default  28.0000  28.0000
24514    song  2418.000000  2418.000000  Default  27.0000  27.0000
6255    cream  1002.000000  1002.000000  Default  26.0000  26.0000
…       …           …           …       …       …       …
15043  lesson    13.171729    98.453628  Topic15  -6.0659   2.7184
17386    moon    15.416082   329.507716  Topic15  -5.9086   1.6677
16751  messag    14.539795   319.236941  Topic15  -5.9671   1.6409
14271     kid    14.382144   849.126475  Topic15  -5.9780   0.6517
4794    child    13.957235   768.183215  Topic15  -6.0080   0.7219
```

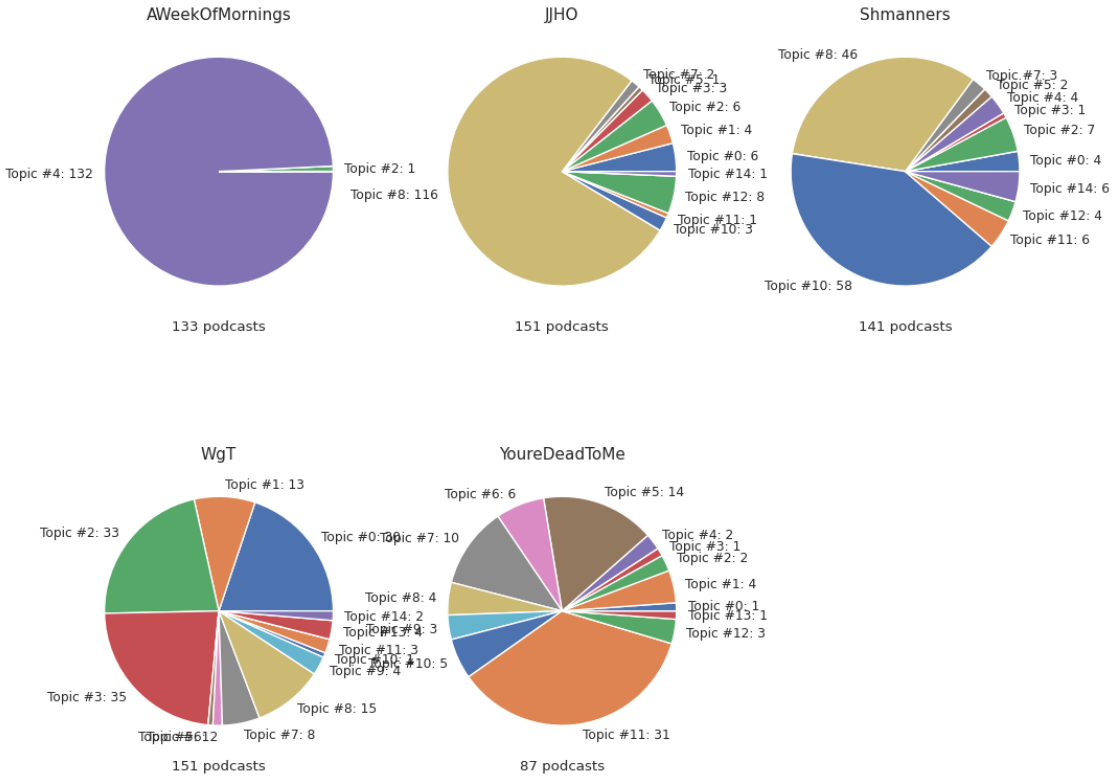
```
[1138 rows x 6 columns], token_table=       Topic      Freq    Term
term
161         1  0.330311  action
161         2  0.138680  action
161         3  0.045386  action
161         5  0.040343  action
161         6  0.037822  action
…       …        …      …
28990      10  0.016810    york
28990      11  0.002292    york
29124       5  0.867620   zhang
29124      11  0.117246   zhang
29180       8  0.925822  zoomer

[4461 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 9, 3, 1, 12, 11, 4, 13, 2, 8, 6, 15, 7, 14, 10])
```

[82]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: cream, ice, eat, butter, chocol, bread, meat, water, flavor, cake,
pizza, chicken, pie, soup, sauc
Topic #1: game, film, footbal, bar, mission, car, byron, shelley, vampir,
potato, team, paint, watch, basebal, sport
Topic #2: song, car, york, citi, disney, list, version, chair, danc, rock, roll,
band, kid, theme, wait
Topic #3: charact, audienc, film, song, star, list, watch, version, perform,
scene, televis, comedi, theme, actor, theater
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
connect, book, yesterday, bill, street, congressman
Topic #5: citi, empir, emperor, battl, peter, china, power, system, democraci,
armi, burton, water, comedi, worm, game
Topic #6: birthday, edit, caesar, poni, blame, celebr, watch, kyle, film, greg,
dragon, julius, nanci, court, answer
Topic #7: chocol, stone, beer, cocoa, clock, space, charact, power, water,
easter, hair, seri, age, milk, war
Topic #8: christma, book, dog, friend, season, slash, store, game, mom, holiday,
cat, box, husband, rule, dad
Topic #9: book, tank, tie, carniv, fish, bow, belt, sandler, burlesqu, jenga,
aquarium, gallon, jinger, star, neck
Topic #10: host, etiquett, check, slash, art, advic, hand, busi, exampl,
mcelroy, hair, pictur, appl, cover, parti
Topic #11: war, centuri, woman, comedi, franc, power, corner, radio, god, hors,
period, book, death, england, son
Topic #12: style, main, jazz, england, wood, town, tenni, radio, communiti,
news, sauc, game, fruit, ben, heat
```

Topic #13: rudi, smell, glitter, roller, comfort, josh, fan, gretchen, count, denis, cameron, ink, hotel, jeremi, jerom
Topic #14: tea, water, game, sale, market, system, drink, tast, bathroom, video, milk, experi, island, afternoon, cup





[83]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 15 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, presid, communiti, countri, morn, yesterday,
radio, bank, congressman, bill, street, trump, march
Topic #1: busi, box, hand, dog, slash, season, birthday, evid, cat, friend,
rule, store, support, justic, style
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, water, eat, salt,
richard, anni, birthday, pie, bowl
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, watch,
version, song, charact, weapon, season, mom
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, river, pino, war, song, presid
Topic #5: song, york, citi, list, album, danc, band, version, film, disney,
sound, vega, countri, coffe, chicago
Topic #6: game, host, system, bill, video, playstat, version, footbal, bar,
tenni, genesi, barker, golf, ball, control
Topic #7: water, bathroom, tast, drink, sarah, river, pie, glass, beer, kitchen,
sister, citi, differ, ami, ice
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, war, round, drink,
```

```
cacao, vanilla, sugar, york, coffe
Topic #9: book, kid, child, moon, seri, rabbit, dog, read, version, pooh,
lesson, chees, page, friend, style
Topic #10: war, charact, power, woman, centuri, comedi, star, film, corner, god,
stone, watch, empir, death, franc
Topic #11: bread, meat, eat, chicken, pizza, burger, sauc, sandwich, flavor,
beef, restaur, cut, taco, dip, soup
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, max, coffe, servic,
eat, drink, membership, saucer, dress
Topic #13: car, drive, baja, subaru, chip, river, transmiss, ride, door, truck,
watch, engin, york, film, road
Topic #14: audienc, song, particip, da, rock, perform, list, danc, experi, bar,
carolin, concert, shout, version, whoa
```

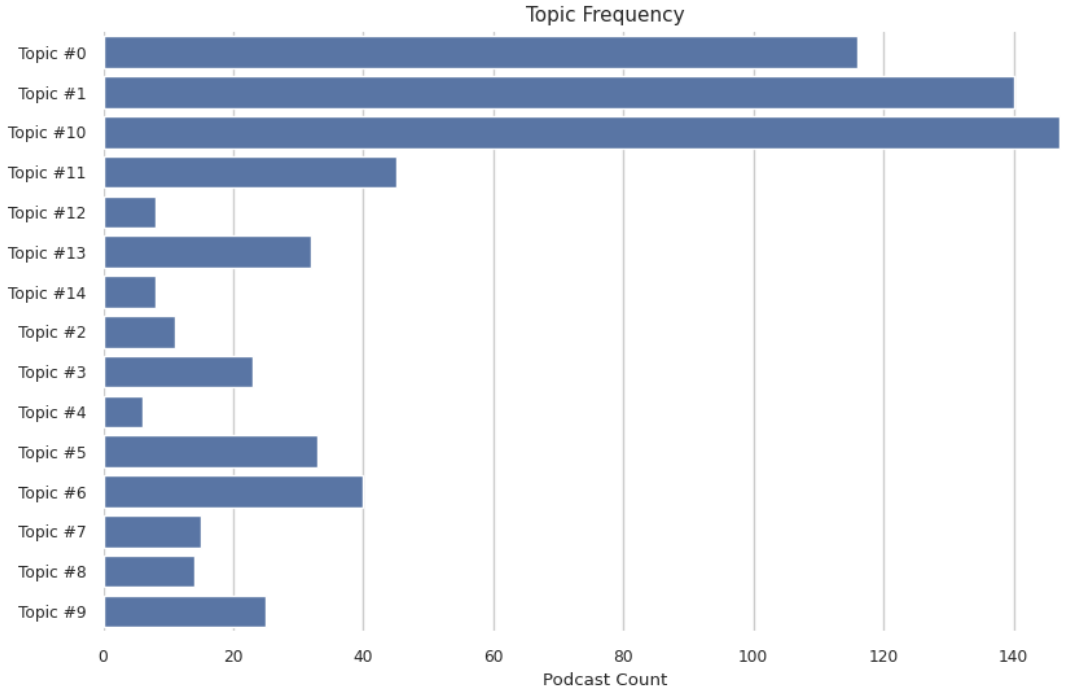
[83]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.045384  0.166788       1        1  16.374878
10    -0.065627  0.027565       2        1  14.373172
1     -0.004194  0.037024       3        1  13.535637
6     -0.077103 -0.019690       4        1   7.499705
5     -0.171202 -0.090961       5        1   7.299185
9     -0.039556  0.031469       6        1   6.613845
11     0.082821 -0.095723       7        1   6.502583
13    -0.087613 -0.004623       8        1   6.403681
7      0.085808  0.052768       9        1   4.332977
3     -0.020703 -0.108821      10        1   3.473983
14    -0.209495 -0.149620      11        1   3.187706
2      0.211042 -0.098559      12        1   2.989783
4     -0.045219  0.271441      13        1   2.762818
8      0.214873 -0.013185      14        1   2.632111
12     0.171548 -0.005875      15        1   2.017937, topic_info=
Term         Freq        Total Category  logprob  loglift
10295      game  3516.000000  3516.000000  Default  30.0000  30.0000
4933   christma  2481.000000  2481.000000  Default  29.0000  29.0000
2887       book  3427.000000  3427.000000  Default  28.0000  28.0000
26068       tea  2002.000000  2002.000000  Default  27.0000  27.0000
24514      song  3703.000000  3703.000000  Default  26.0000  26.0000
…           …           …           …      …        …        …
12551     hotel    54.490914   334.032533  Topic15  -5.4727   2.0899
13777      join    65.851903   666.027894  Topic15  -5.2834   1.5892
8651      enjoy    58.716016   458.400942  Topic15  -5.3981   1.8481
6255      cream    72.639273  2152.244554  Topic15  -5.1853   0.5143
25578   support    59.736830  1131.426330  Topic15  -5.3808   0.9618

[1359 rows x 6 columns], token_table=       Topic      Freq       Term
term
```

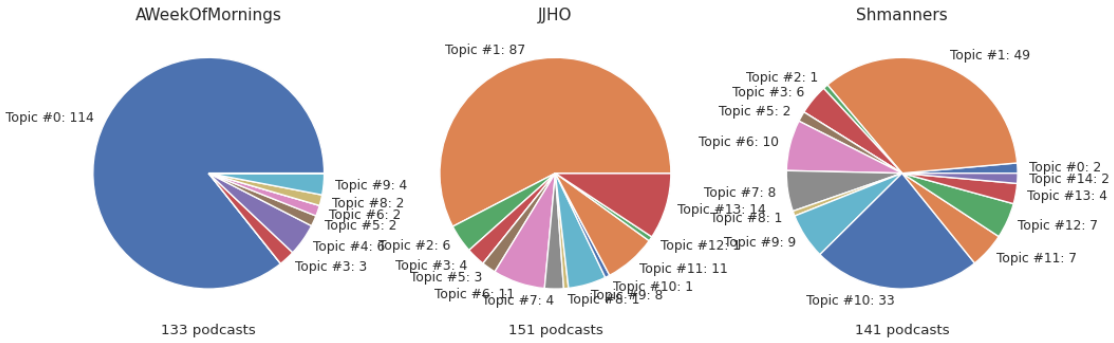
```
3          6  0.995797   aaronson
8         14  1.009503       abag
9          2  0.795536  abakhumid
66         4  0.940799    abscess
128       14  1.004410    achiott
…      …       …         …
29100      4  0.920803      zelda
29100      9  0.069495      zelda
29100     15  0.017374      zelda
29124      2  0.989709      zhang
29164      5  0.986738        zix

[4426 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 11, 2, 7, 6, 10, 12, 14, 8, 4, 15, 3, 5, 9, 13])
```

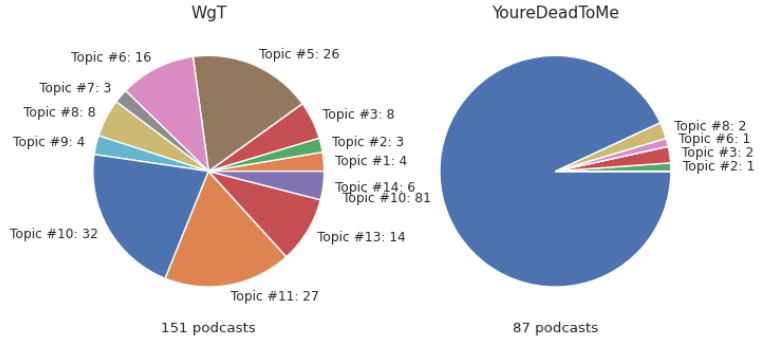
[84]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, presid, communiti, countri, morn, yesterday,
radio, bank, congressman, bill, street, trump, march
Topic #1: busi, box, hand, dog, slash, season, birthday, evid, cat, friend,
rule, store, support, justic, style
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, water, eat, salt,
richard, anni, birthday, pie, bowl
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, watch,
version, song, charact, weapon, season, mom
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, river, pino, war, song, presid
Topic #5: song, york, citi, list, album, danc, band, version, film, disney,
sound, vega, countri, coffe, chicago
Topic #6: game, host, system, bill, video, playstat, version, footbal, bar,
tenni, genesi, barker, golf, ball, control
Topic #7: water, bathroom, tast, drink, sarah, river, pie, glass, beer, kitchen,
sister, citi, differ, ami, ice
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, war, round, drink,
cacao, vanilla, sugar, york, coffe
Topic #9: book, kid, child, moon, seri, rabbit, dog, read, version, pooh,
lesson, chees, page, friend, style
Topic #10: war, charact, power, woman, centuri, comedi, star, film, corner, god,
stone, watch, empir, death, franc
Topic #11: bread, meat, eat, chicken, pizza, burger, sauc, sandwich, flavor,
beef, restaur, cut, taco, dip, soup
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, max, coffe, servic,
eat, drink, membership, saucer, dress
Topic #13: car, drive, baja, subaru, chip, river, transmiss, ride, door, truck,
watch, engin, york, film, road
```

Topic #14: audienc, song, particip, da, rock, perform, list, danc, experi, bar, carolin, concert, shout, version, whoa







[85]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 15 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: dog, car, docket, cat, letter, main, dracula, disput, justic, book,
mom, rule, evid, rudi, dad
Topic #1: wine, river, presid, state, yesterday, communiti, congressman, trump,
countri, bank, morn, hampton, vote, elect, bill
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, ghost, tree, carol, villain, song
Topic #3: bread, dip, breakfast, meat, soup, meal, pizza, eat, sauc, chicken,
sandwich, dinner, cook, burger, buffet
Topic #4: empir, emperor, war, franc, power, corner, armi, centuri, henri,
battl, greg, woman, citi, comedi, china
Topic #5: film, charact, star, list, disney, watch, action, tv, mission, scene,
car, jason, version, villain, theater
Topic #6: doordash, etiquett, host, parti, app, advic, book, art, hair, slash,
busi, deliveri, mcelroy, max, photographi
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond,
easter, kat, sugar, flavor, quaker, pot
Topic #8: cream, ice, cake, pie, bagel, rainer, vanilla, chocol, eat, birthday,
milk, water, cone, carlson, cooki
Topic #9: game, footbal, sport, basebal, playstat, island, golf, barker, ball,
tenni, monopoli, pursuit, trivia, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton,
camp, presid, pizza, spooner, russia, oak
Topic #11: water, appl, beer, cider, drink, bathroom, glass, coffe, flavor,
prohibit, juic, pie, river, tast, vanilla
Topic #12: tea, ceremoni, matcha, water, afternoon, cup, rudi, glitter,
breakfast, dinner, coffe, caffein, champagn, sugar, tank
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, vega, roll,
citi, boom, guitar, disney, jeff
Topic #14: byron, shelley, perci, vampir, book, jane, poetri, frankenstein,
mari, monster, corinn, florenc, woman, godwin, dracula
```

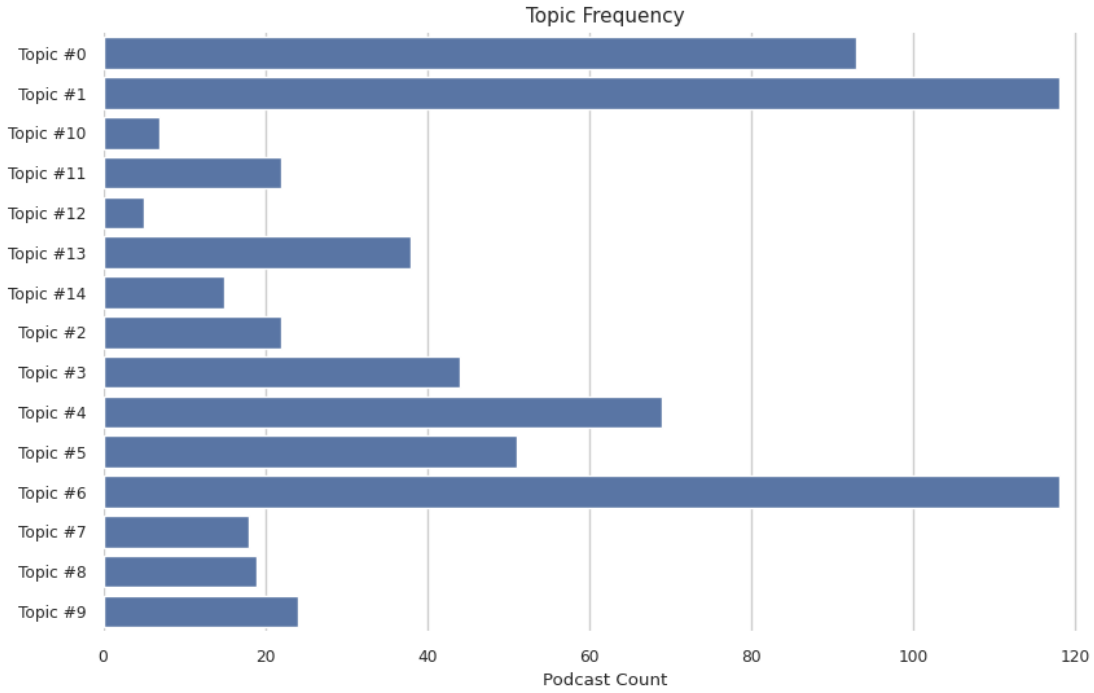
[85]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.072926  0.018577       1        1  17.509073
6     -0.049505  0.051603       2        1  12.796152
0      0.009450 -0.021745       3        1   9.961313
5     -0.080456 -0.086266       4        1   8.449790
```

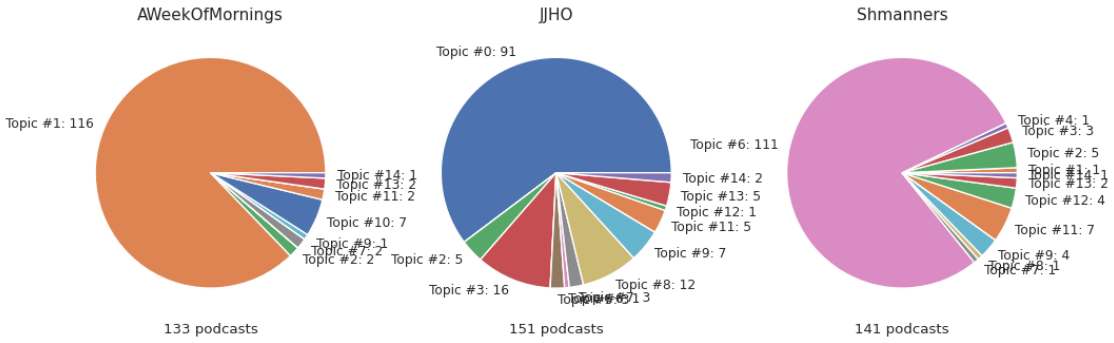
```
4     -0.177130  0.035349       5        1   8.358556
13    -0.088520 -0.052057       6        1   7.619303
3      0.187454 -0.036127       7        1   6.193442
11     0.099772  0.023894       8        1   5.624538
9     -0.026889  0.005919       9        1   5.349093
8      0.174703 -0.070048      10        1   3.897049
2      0.039638 -0.137527      11        1   3.777704
14    -0.188000 -0.053717      12        1   3.312463
7      0.169288 -0.084170      13        1   3.013434
10    -0.090709  0.069965      14        1   2.375121
12     0.093831  0.336349      15        1   1.762970, topic_info=
Term       Freq      Total Category  logprob  loglift
4933     christma  17.000000  17.000000  Default  30.0000  30.0000
26068         tea  12.000000  12.000000  Default  29.0000  29.0000
4877       chocol  15.000000  15.000000  Default  28.0000  28.0000
6255        cream  15.000000  15.000000  Default  27.0000  27.0000
12801         ice  12.000000  12.000000  Default  26.0000  26.0000
…             …         …         …       …       …        …
2488     birthday   0.587741   5.338482  Topic15  -5.5643   1.8318
26009        tast   0.572734   5.636072  Topic15  -5.5902   1.7516
8005        drink   0.574013   5.824895  Topic15  -5.5879   1.7209
16315         max   0.562632   5.353082  Topic15  -5.6080   1.7854
16642  membership   0.447820   2.626388  Topic15  -5.8362   2.2692

[1446 rows x 6 columns], token_table=       Topic      Freq      Term
term
119        8  1.114689  accuraci
161        1  0.283692    action
161        4  0.567385    action
165        4  0.612563     actor
249       11  1.090495    advent
…      …        …       …
28990      3  0.133446      york
28990      4  0.133446      york
28990      6  0.266892      york
29100      9  1.125909     zelda
29124      5  1.306986     zhang

[969 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 7, 1, 6, 5, 14, 4, 12, 10, 9, 3, 15, 8, 11, 13])
```

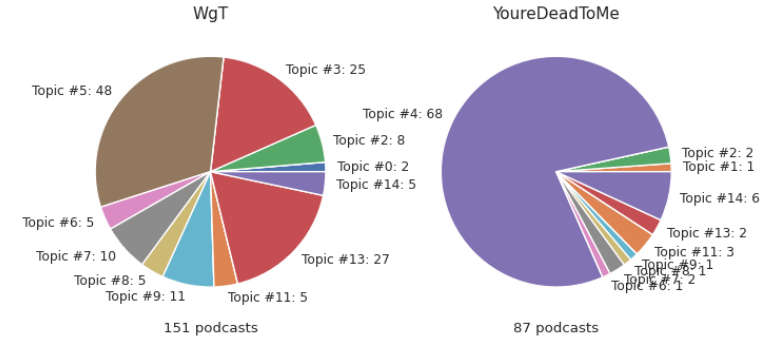
[86]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: dog, car, docket, cat, letter, main, dracula, disput, justic, book,
mom, rule, evid, rudi, dad
```

Topic #1: wine, river, presid, state, yesterday, communiti, congressman, trump, countri, bank, morn, hampton, vote, elect, bill
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah, winter, krampus, ghost, tree, carol, villain, song
Topic #3: bread, dip, breakfast, meat, soup, meal, pizza, eat, sauc, chicken, sandwich, dinner, cook, burger, buffet
Topic #4: empir, emperor, war, franc, power, corner, armi, centuri, henri, battl, greg, woman, citi, comedi, china
Topic #5: film, charact, star, list, disney, watch, action, tv, mission, scene, car, jason, version, villain, theater
Topic #6: doordash, etiquett, host, parti, app, advic, book, art, hair, slash, busi, deliveri, mcelroy, max, photographi
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond, easter, kat, sugar, flavor, quaker, pot
Topic #8: cream, ice, cake, pie, bagel, rainer, vanilla, chocol, eat, birthday, milk, water, cone, carlson, cooki
Topic #9: game, footbal, sport, basebal, playstat, island, golf, barker, ball, tenni, monopoli, pursuit, trivia, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton, camp, presid, pizza, spooner, russia, oak
Topic #11: water, appl, beer, cider, drink, bathroom, glass, coffe, flavor, prohibit, juic, pie, river, tast, vanilla
Topic #12: tea, ceremoni, matcha, water, afternoon, cup, rudi, glitter, breakfast, dinner, coffe, caffein, champagn, sugar, tank
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, vega, roll, citi, boom, guitar, disney, jeff
Topic #14: byron, shelley, perci, vampir, book, jane, poetri, frankenstein, mari, monster, corinn, florenc, woman, godwin, dracula


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count


AWeekOfMornings
Topic #1: 116
Topic #14: 1
Topic #13: 2
Topic #11: 2
Topic #10: 7
Topic #9: 1
Topic #7: 2
Topic #2: 2
133 podcasts
JJHO
Topic #0: 91
Topic #6: 111
Topic #14: 2
Topic #13: 5
Topic #12: 1
Topic #11: 5
Topic #9: 7
Topic #8: 12
Topic #2: 5
Topic #3: 16
151 podcasts
Shmanners
Topic #4: 1
Topic #3: 3
Topic #2: 5
Topic #12: 4
Topic #11: 7
Topic #9: 4
141 podcasts


WgT
Topic #3: 25
Topic #5: 48
Topic #2: 8
Topic #0: 2
Topic #14: 5
Topic #6: 5
Topic #7: 10
Topic #13: 27
Topic #8: 5
Topic #9: 11
Topic #11: 5
151 podcasts
YoureDeadToMe
Topic #4: 68
Topic #2: 2
Topic #1: 1
Topic #14: 6
Topic #13: 2
Topic #11: 3
87 podcasts

[87]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 15 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: dog, cat, car, evid, justic, rule, rudi, birthday, season, disput,
mom, turkey, docket, dracula, courtroom
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
peanut, bar, cacao, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, tradit, movi christma, santa claus, song, villain,
hanukkah, ghost
Topic #4: franc, henri, war, josephin, corner, england, revolut, centuri,
coloni, slaveri, woman, eleanor, power, stone, pyramid
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version,
york, car, theme song, theater, theme, perform
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, water, chocol,
vanilla, pie, rainer, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
bed, support, huga, work cancer, cancer cancer, support cancer, presid, hampton
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, advic,
deliveri, fashion, mcelroy, busi, code
Topic #9: bread, dip, meat, sauc, pizza, sandwich, breakfast, eat, soup,
chicken, meal, burger, flavor, butter, beef
Topic #10: byron, shelley, perci, vampir, jane, frankenstein, poetri, book,
mari, corinn, florenc, monster, switzerland, dracula, woman
Topic #11: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni,
afternoon, cup, beer, tea kind, coffe, drink, sugar, yeah tea
Topic #12: food bank, bank, march, hunger, march food, wine, food insecur,
communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: game, footbal, game show, golf, tenni, playstat, game game, sport,
croquet, basebal, ball, island, monopoli, barker, nintendo
Topic #14: empir, emperor, khan, armi, battl, china, shah, citi, tang, chinggi,
jahangir, genghi, peter, jahan, shah jahan
```

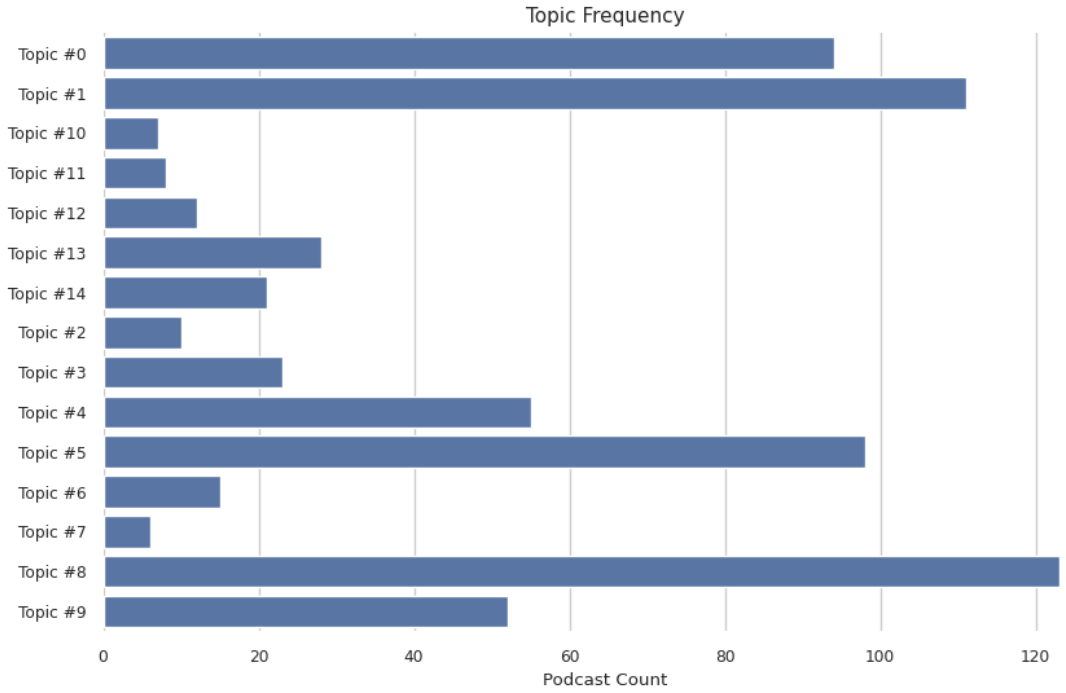
[87]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.030844 -0.217269       1        1  18.503381
5     -0.005781  0.052499       2        1  13.767723
8      0.026408  0.048432       3        1  13.664680
0     -0.066664  0.082443       4        1  10.972923
9     -0.168771  0.128881       5        1   7.523517
```

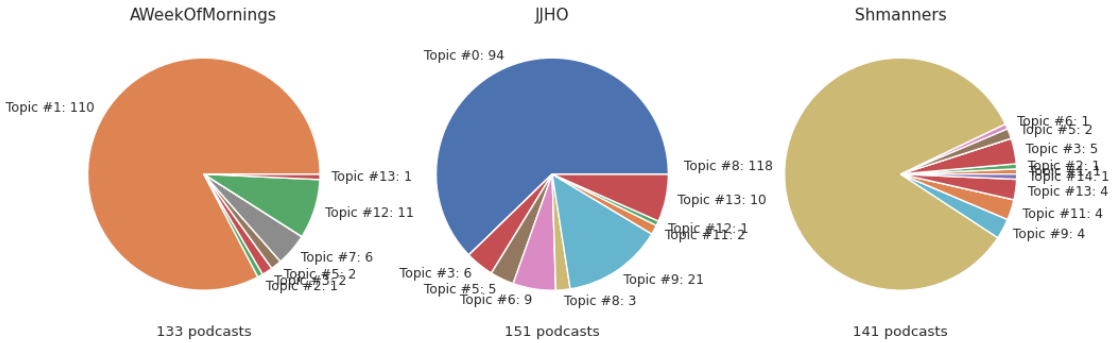
```
4      0.270876  0.034746       6        1   6.946919
13    -0.014461  0.016800       7        1   5.856305
14     0.282494  0.042448       8        1   3.821835
3     -0.080551  0.011444       9        1   3.653622
12    -0.050944 -0.261923      10        1   3.448468
6     -0.162678  0.128470      11        1   3.370680
2     -0.154423  0.025856      12        1   2.250537
10     0.198390  0.075301      13        1   2.240855
7     -0.000675 -0.268172      14        1   2.151306
11    -0.042375  0.100045      15        1   1.827249, topic_info=
Term      Freq     Total Category  logprob  loglift
84847    christma  8.000000  8.000000  Default  30.0000  30.0000
481205        tea  6.000000  6.000000  Default  29.0000  29.0000
83347      chocol  7.000000  7.000000  Default  28.0000  28.0000
240891  ice cream  6.000000  6.000000  Default  27.0000  27.0000
116310      cream  7.000000  7.000000  Default  26.0000  26.0000
…           …       …       …       …       …       …
363988      plant  0.332616  1.754217  Topic15  -7.0341   2.3396
309840       milk  0.339728  3.420490  Topic15  -7.0129   1.6930
202162    glitter  0.288997  1.249274  Topic15  -7.1747   2.5385
135957     dinner  0.308457  3.236864  Topic15  -7.1095   1.6516
51575   breakfast  0.285045  3.053488  Topic15  -7.1884   1.6310

[1543 rows x 6 columns], token_table=        Topic      Freq           Term
term
2599        2  0.469640         action
2739        2  1.832775    action movi
3021        2  0.498400          actor
5274        3  0.906895          advic
6048       15  0.589368      afternoon
…       …       …              …
553250      2  0.440785           york
553250      3  0.220392           york
553250      4  0.220392           york
554534      3  1.757948  zero deliveri
554602      8  1.639448          zhang

[570 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 6, 9, 1, 10, 5, 14, 15, 4, 13, 7, 3, 11, 8, 12])
```

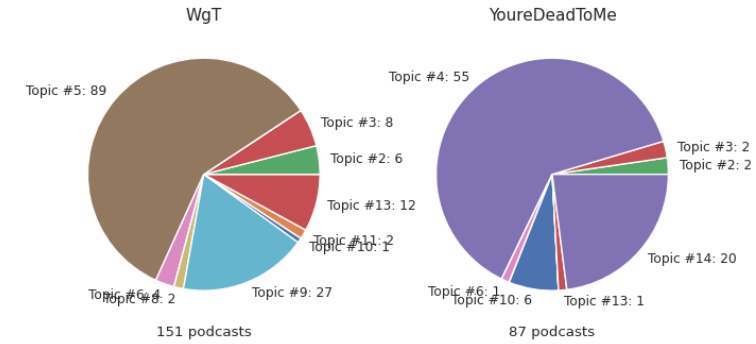
[88]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: dog, cat, car, evid, justic, rule, rudi, birthday, season, disput,
mom, turkey, docket, dracula, courtroom
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
```

morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter, peanut, bar, cacao, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma, scroog, ghost christma, tradit, movi christma, santa claus, song, villain, hanukkah, ghost
Topic #4: franc, henri, war, josephin, corner, england, revolut, centuri, coloni, slaveri, woman, eleanor, power, stone, pyramid
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version, york, car, theme song, theater, theme, perform
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, water, chocol, vanilla, pie, rainer, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi, bed, support, huga, work cancer, cancer cancer, support cancer, presid, hampton
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, advic, deliveri, fashion, mcelroy, busi, code
Topic #9: bread, dip, meat, sauc, pizza, sandwich, breakfast, eat, soup, chicken, meal, burger, flavor, butter, beef
Topic #10: byron, shelley, perci, vampir, jane, frankenstein, poetri, book, mari, corinn, florenc, monster, switzerland, dracula, woman
Topic #11: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni, afternoon, cup, beer, tea kind, coffe, drink, sugar, yeah tea
Topic #12: food bank, bank, march, hunger, march food, wine, food insecur, communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: game, footbal, game show, golf, tenni, playstat, game game, sport, croquet, basebal, ball, island, monopoli, barker, nintendo
Topic #14: empir, emperor, khan, armi, battl, china, shah, citi, tang, chinggi, jahangir, genghi, peter, jahan, shah jahan


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count


AWeekOfMornings
Topic #1: 110
Topic #13: 1
Topic #12: 11
Topic #7: 6
Topic #5: 2
Topic #3: 2
Topic #2: 1
133 podcasts
JJHO
Topic #0: 94
Topic #8: 118
Topic #13: 10
Topic #12: 1
Topic #11: 2
Topic #9: 21
Topic #8: 3
Topic #6: 9
Topic #5: 5
Topic #3: 6
151 podcasts
Shmanners
Topic #6: 1
Topic #5: 2
Topic #3: 5
Topic #2: 1
Topic #14: 1
Topic #13: 4
Topic #11: 4
Topic #9: 4
141 podcasts


WgT
Topic #5: 89
Topic #3: 8
Topic #2: 6
Topic #13: 12
Topic #11: 2
Topic #10: 1
Topic #9: 27
Topic #8: 2
Topic #6: 4
151 podcasts
YoureDeadToMe
Topic #4: 55
Topic #3: 2
Topic #2: 2
Topic #14: 20
Topic #13: 1
Topic #10: 6
Topic #6: 1
87 podcasts

[89]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 15 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: song, film, charact, list, game, disney, version, star, watch, movi
movi, york, car, theme song, theme, vega
Topic #1: wine, river, presid, state, yesterday, communiti, food bank,
congressman, countri, bank, trump, morn, hampton, bill, radio
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut
butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, villain, scroog, santa claus, song, ghost,
tradit, place christma
Topic #4: franc, emperor, henri, josephin, eleanor, power, tang, rebellion,
revolut, war, woman, louvertur, loui, slaveri, centuri
Topic #5: evid, dog, cat, car, justic, rule, rudi, season, turkey, birthday,
courtroom, slash, disput, jacki, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, water, vanilla, pie, rainer, cone
Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect
cancer, wine, pino, diagnosi, bed, support, huga, work cancer, connect cancer
connect, support cancer, cancer cancer
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, fashion,
advic, game, deliveri, max, mcelroy
Topic #9: bread, dip, meat, pizza, sauc, sandwich, eat, chicken, soup,
breakfast, meal, flavor, burger, butter, beef
Topic #10: shelley, byron, perci, vampir, jane, frankenstein, poetri, book,
mari, florenc, corinn, switzerland, monster, woman, dracula
Topic #11: tea, tea tea, afternoon tea, water, matcha, ceremoni, tea ceremoni,
afternoon, cup, tea tea tea, tea kind, coffe, beer, sugar, saucer
Topic #12: empir, khan, armi, battl, shah, citi, jahangir, jahan, chinggi, shah
jahan, genghi, peter, mogul, greec, fleet
Topic #13: pyramid, stone, tomb, stoneheng, sarah, egypt, imhotep, ancient,
archaeolog, site, maria, giza, mera, kingdom, copper
Topic #14: mayflow, coloni, ship, weston, misha, carver, popul, voyag, invest,
settlement, jane town, plymouth, england, merchant, dutch
```

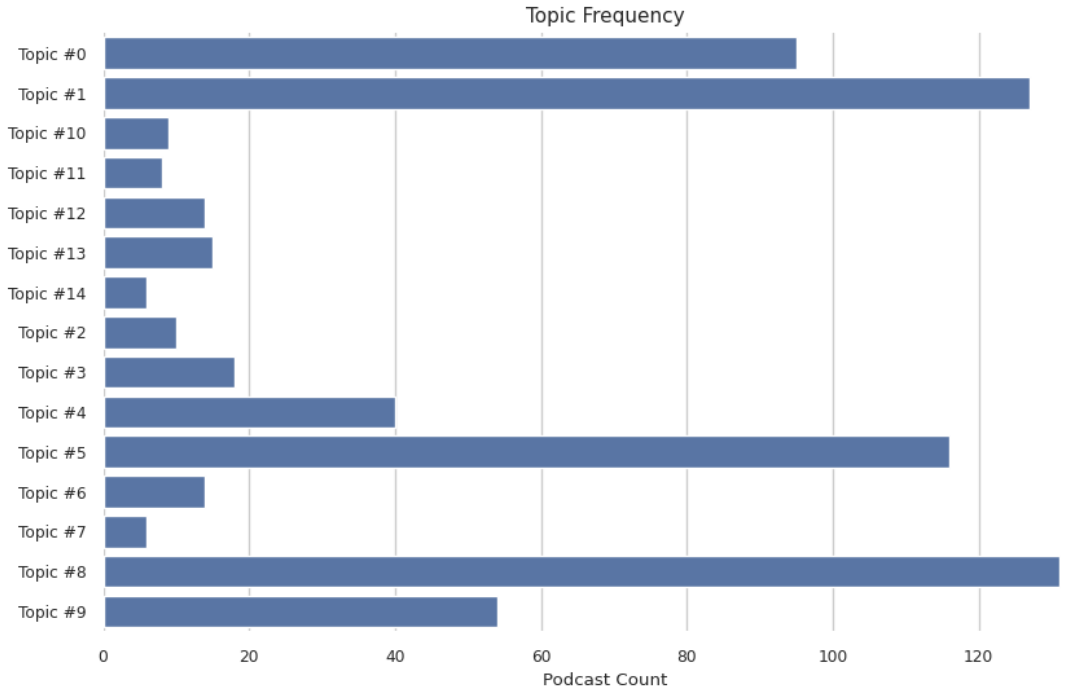
[89]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1      0.073367 -0.211352       1        1  22.208884
8      0.020007  0.007269       2        1  15.292786
5      0.116313  0.073294       3        1  14.227043
0      0.070944  0.024612       4        1  13.271453
```

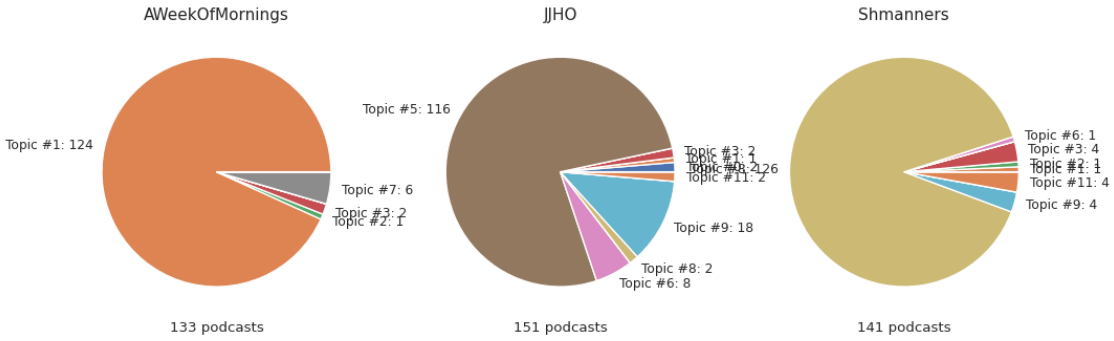
```
9      0.189320  0.115137       5        1   7.914820
4     -0.267849  0.032620       6        1   4.416608
3      0.116493  0.039867       7        1   3.565325
6      0.181048  0.116278       8        1   3.243354
12    -0.228745  0.009434       9        1   2.892949
13    -0.230744  0.049957      10        1   2.493299
10    -0.119192  0.042445      11        1   2.473484
7      0.055328 -0.387700      12        1   2.287882
2      0.157717  0.071849      13        1   2.162744
11     0.076787 -0.004761      14        1   1.986710
14    -0.210796  0.021052      15        1   1.562660, topic_info=
Term      Freq     Total Category  logprob  loglift
1141840        tea  6.000000  6.000000  Default  30.0000  30.0000
195860    christma  6.000000  6.000000  Default  29.0000  29.0000
192266      chocol  5.000000  5.000000  Default  28.0000  28.0000
566886   ice cream  5.000000  5.000000  Default  27.0000  27.0000
270440       cream  5.000000  5.000000  Default  26.0000  26.0000
…             …       …       …      …       …      …
1244935   virginia  0.300914  0.933641  Topic15  -7.3450   3.0265
481422    governor  0.294458  1.100564  Topic15  -7.3667   2.8403
1200930      trade  0.297795  1.737805  Topic15  -7.3554   2.3948
486799        greg  0.279805  1.907276  Topic15  -7.4177   2.2394
181807       charg  0.272943  1.325483  Topic15  -7.4425   2.5785

[1462 rows x 6 columns], token_table=         Topic      Freq           Term
term
5687         4  0.643923         action
6704         4  0.696973          actor
11712        2  0.551407          advic
13355       14  0.690050      afternoon
13682       14  0.941503  afternoon tea
…        …       …              …
1294001      1  0.927706      word week
1321637      4  1.133671      yeah movi
1337670      1  1.017397      yesterday
1339625      1  0.279989           york
1339625      4  0.279989           york

[462 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 1, 10, 5, 4, 7, 13, 14, 11, 8, 3, 12, 15])
```

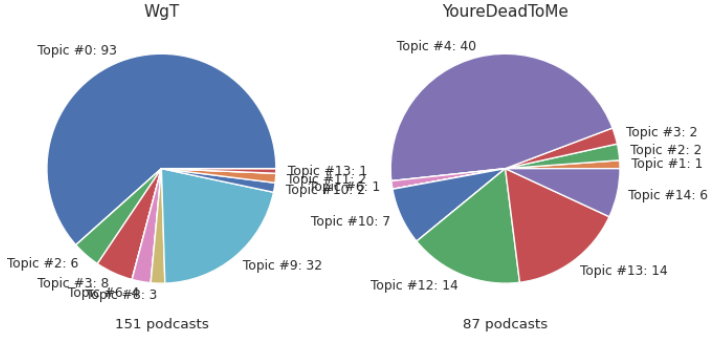
[90]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: song, film, charact, list, game, disney, version, star, watch, movi
movi, york, car, theme song, theme, vega
```

Topic #1: wine, river, presid, state, yesterday, communiti, food bank, congressman, countri, bank, trump, morn, hampton, bill, radio
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma, ghost christma, movi christma, villain, scroog, santa claus, song, ghost, tradit, place christma
Topic #4: franc, emperor, henri, josephin, eleanor, power, tang, rebellion, revolut, war, woman, louvertur, loui, slaveri, centuri
Topic #5: evid, dog, cat, car, justic, rule, rudi, season, turkey, birthday, courtroom, slash, disput, jacki, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice, cream ice cream, chocol, ice cream ice, water, vanilla, pie, rainer, cone
Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect cancer, wine, pino, diagnosi, bed, support, huga, work cancer, connect cancer connect, support cancer, cancer cancer
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, fashion, advic, game, deliveri, max, mcelroy
Topic #9: bread, dip, meat, pizza, sauc, sandwich, eat, chicken, soup, breakfast, meal, flavor, burger, butter, beef
Topic #10: shelley, byron, perci, vampir, jane, frankenstein, poetri, book, mari, florenc, corinn, switzerland, monster, woman, dracula
Topic #11: tea, tea tea, afternoon tea, water, matcha, ceremoni, tea ceremoni, afternoon, cup, tea tea tea, tea kind, coffe, beer, sugar, saucer
Topic #12: empir, khan, armi, battl, shah, citi, jahangir, jahan, chinggi, shah jahan, genghi, peter, mogul, greec, fleet
Topic #13: pyramid, stone, tomb, stoneheng, sarah, egypt, imhotep, ancient, archaeolog, site, maria, giza, mera, kingdom, copper
Topic #14: mayflow, coloni, ship, weston, misha, carver, popul, voyag, invest, settlement, jane town, plymouth, england, merchant, dutch


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count


AWeekOfMornings
Topic #1: 124
Topic #7: 6
Topic #3: 2
Topic #2: 1
133 podcasts
JJHO
Topic #5: 116
Topic #9: 18
Topic #8: 2
Topic #6: 8
151 podcasts
Shmanners
Topic #6: 1
Topic #3: 4
Topic #11: 4
Topic #9: 4
141 podcasts


WgT
Topic #0: 93
Topic #2: 6
Topic #3: 8
Topic #9: 32
151 podcasts
YoureDeadToMe
Topic #4: 40
Topic #3: 2
Topic #2: 2
Topic #1: 1
Topic #14: 6
Topic #10: 7
Topic #13: 14
Topic #12: 14
87 podcasts

[91]:
```
topic_count = 20
```

[92]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 20 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: eat, butter, bread, meat, cream, chocol, flavor, pizza, chicken, soup,
sauc, restaur, meal, cook, breakfast
Topic #1: game, car, chip, footbal, bar, potato, paint, mission, poni, snack,
kate, board, paul, shelley, bill
Topic #2: car, chair, citi, jason, pretzel, soul, chicago, drive, shape,
basebal, york, costum, sound, dream, wilco
Topic #3: song, list, film, charact, watch, version, audienc, star, theme,
topic, theater, rock, citi, perform, kid
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
connect, yesterday, bill, book, street, congressman
Topic #5: album, burton, bin, worm, scienc, democraci, china, jason, tang,
emperor, system, georg, compost, oscar, bee
Topic #6: birthday, edit, disney, caesar, celebr, blame, watch, jack, kyle,
julius, dragon, greg, host, rick, answer
Topic #7: stone, beer, charact, marvel, space, star, barnum, power, josephin,
seri, stoneheng, mind, lake, baker, captain
Topic #8: christma, game, parti, book, holiday, hand, monster, busi, hair,
slash, card, host, ghost, check, max
Topic #9: book, tank, tie, carniv, event, bow, sandler, burlesqu, fish, jewelri,
jenga, olymp, jinger, neck, star
Topic #10: art, citi, appl, book, slash, communiti, check, cider, advic,
etiquett, host, husband, bodi, societi, woman
Topic #11: woman, comedi, centuri, war, power, corner, franc, radio, god,
period, greg, age, england, son, death
Topic #12: ice, cream, cake, wood, tenni, sauc, england, news, fruit, birthday,
ketchup, parti, town, main, tradit
Topic #13: smell, rudi, fan, henri, count, denis, ink, jeremi, jerom, hotel,
cameron, toad, pie, eleanor, ballpoint
Topic #14: water, tea, game, sale, drink, market, tast, champagn, alan, plant,
bathroom, afternoon, island, flea, price
Topic #15: disney, ride, roller, milk, water, toilet, air, coaster, car,
disneyland, belt, line, cooki, pan, land
Topic #16: danc, pie, jeff, bank, salsa, crust, parti, mason, floor, trash, roy,
street, slide, fruit, march
Topic #17: empir, citi, war, peter, battl, armi, clock, water, power, gold,
khan, centuri, tapestri, radio, pan
Topic #18: chocol, york, milk, cocoa, bar, accent, car, season, championship,
aeta, watch, jane, bat, coffe, smoke
Topic #19: book, dog, friend, rule, slash, justic, cat, mom, evid, box, season,
husband, dad, store, letter
```

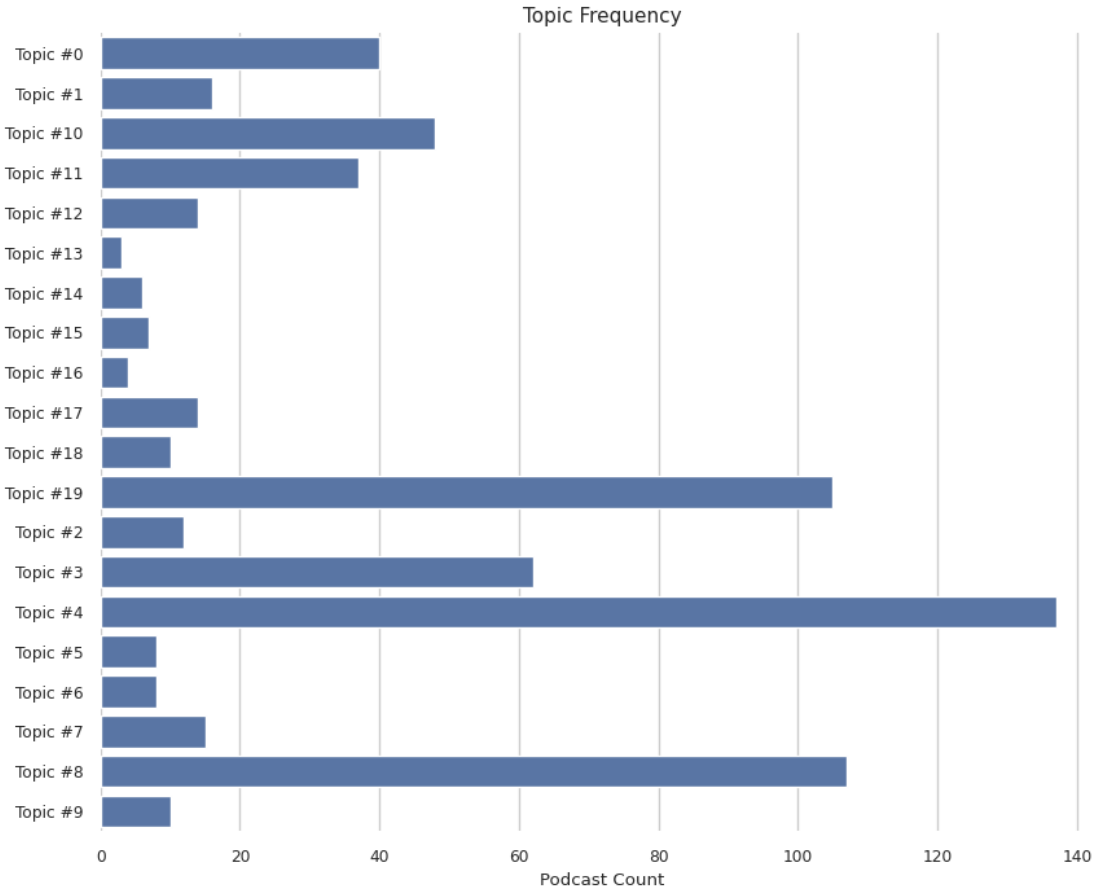
[92]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4     -0.126232  0.013215       1        1  30.291024
19    -0.204921 -0.027320       2        1  12.659513
8     -0.134815  0.000697       3        1   9.264715
3     -0.195556 -0.027133       4        1   9.191367
0     -0.113033 -0.203814       5        1   6.115281
11    -0.134302  0.149594       6        1   5.877554
10    -0.104620  0.049813       7        1   4.673109
12     0.047357 -0.076564       8        1   2.997336
7      0.037619  0.043633       9        1   2.270835
1      0.042645 -0.040952      10        1   2.219996
2      0.027805 -0.026090      11        1   2.078261
17    -0.016017  0.157337      12        1   2.010963
5      0.075780  0.078503      13        1   1.611803
6      0.094496  0.077273      14        1   1.510181
15     0.082070 -0.016122      15        1   1.419384
14     0.082197 -0.128156      16        1   1.332552
18     0.078135 -0.067326      17        1   1.331220
9      0.117665  0.083570      18        1   1.152239
16     0.162335 -0.074338      19        1   1.070501
13     0.181393  0.034179      20        1   0.922165, topic_info=
Term         Freq        Total Category  logprob  loglift
2887         book  2321.000000  2321.000000  Default  30.0000  30.0000
28222       water  1716.000000  1716.000000  Default  29.0000  29.0000
10295        game  2115.000000  2115.000000  Default  28.0000  28.0000
24514        song  2443.000000  2443.000000  Default  27.0000  27.0000
4877       chocol   838.000000   838.000000  Default  26.0000  26.0000
…            …          …           …       …       …       …
26093  technolog    20.599363   329.235924  Topic20  -5.6624   1.9147
6129        court    23.500906   846.836426  Topic20  -5.5306   1.1017
1691         ball    19.639978   352.940559  Topic20  -5.7101   1.7975
6808         date    18.607735   368.814811  Topic20  -5.7641   1.6995
2887         book    17.230459  2321.203440  Topic20  -5.8410  -0.2170

[1534 rows x 6 columns], token_table=       Topic      Freq    Term
term
98          1  0.239573  accent
98          2  0.089840  accent
98          4  0.049911  accent
98          6  0.074867  accent
98          8  0.134760  accent
…       …        …      …
29169      10  0.041315    zola
29169      16  0.123946    zola
29169      19  0.041315    zola
```

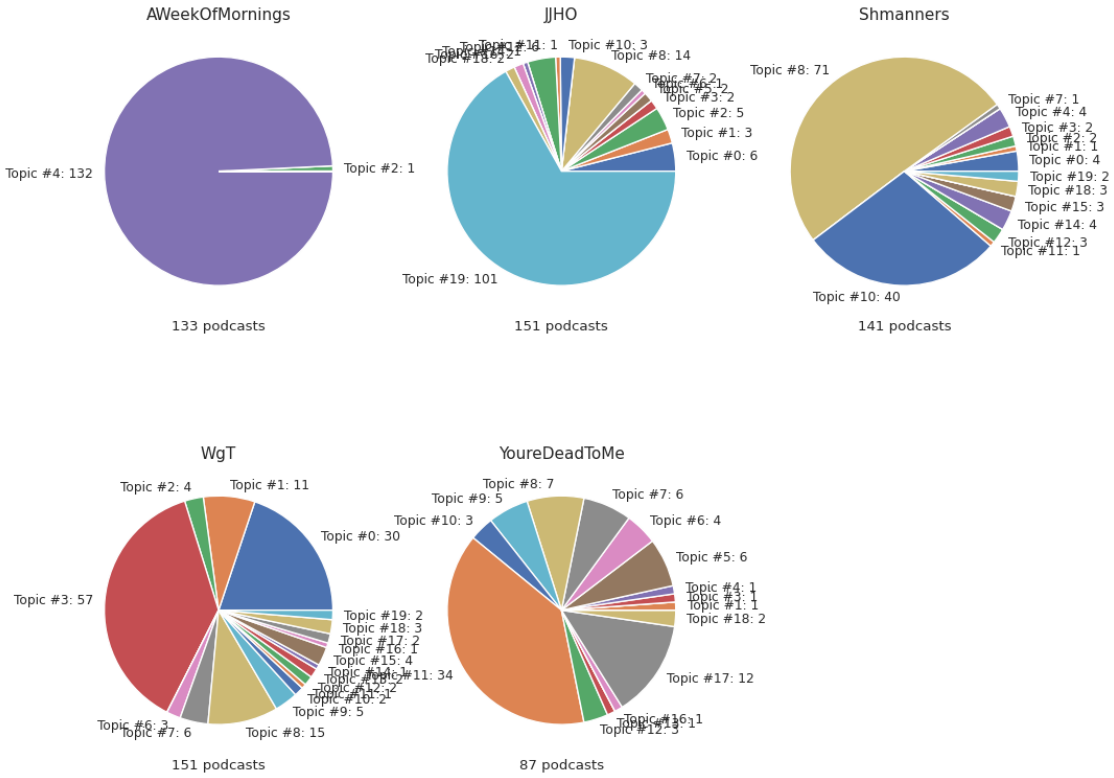
```
29180      8  0.832524  zoomer
29180     13  0.083252  zoomer

[6393 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 20, 9, 4, 1, 12, 11, 13, 8, 2, 3, 18, 6, 7, 16,
15, 19, 10, 17, 14])
```

[93]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: eat, butter, bread, meat, cream, chocol, flavor, pizza, chicken, soup,
sauc, restaur, meal, cook, breakfast
Topic #1: game, car, chip, footbal, bar, potato, paint, mission, poni, snack,
kate, board, paul, shelley, bill
Topic #2: car, chair, citi, jason, pretzel, soul, chicago, drive, shape,
basebal, york, costum, sound, dream, wilco
Topic #3: song, list, film, charact, watch, version, audienc, star, theme,
topic, theater, rock, citi, perform, kid
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
connect, yesterday, bill, book, street, congressman
Topic #5: album, burton, bin, worm, scienc, democraci, china, jason, tang,
emperor, system, georg, compost, oscar, bee
Topic #6: birthday, edit, disney, caesar, celebr, blame, watch, jack, kyle,
julius, dragon, greg, host, rick, answer
Topic #7: stone, beer, charact, marvel, space, star, barnum, power, josephin,
seri, stoneheng, mind, lake, baker, captain
Topic #8: christma, game, parti, book, holiday, hand, monster, busi, hair,
slash, card, host, ghost, check, max
Topic #9: book, tank, tie, carniv, event, bow, sandler, burlesqu, fish, jewelri,
jenga, olymp, jinger, neck, star
Topic #10: art, citi, appl, book, slash, communiti, check, cider, advic,
etiquett, host, husband, bodi, societi, woman
Topic #11: woman, comedi, centuri, war, power, corner, franc, radio, god,
period, greg, age, england, son, death
Topic #12: ice, cream, cake, wood, tenni, sauc, england, news, fruit, birthday,
ketchup, parti, town, main, tradit
Topic #13: smell, rudi, fan, henri, count, denis, ink, jeremi, jerom, hotel,
cameron, toad, pie, eleanor, ballpoint
Topic #14: water, tea, game, sale, drink, market, tast, champagn, alan, plant,
bathroom, afternoon, island, flea, price
Topic #15: disney, ride, roller, milk, water, toilet, air, coaster, car,
disneyland, belt, line, cooki, pan, land
Topic #16: danc, pie, jeff, bank, salsa, crust, parti, mason, floor, trash, roy,
street, slide, fruit, march
Topic #17: empir, citi, war, peter, battl, armi, clock, water, power, gold,
khan, centuri, tapestri, radio, pan
```

Topic #18: chocol, york, milk, cocoa, bar, accent, car, season, championship, aeta, watch, jane, bat, coffe, smoke
Topic #19: book, dog, friend, rule, slash, justic, cat, mom, evid, box, season, husband, dad, store, letter





[94]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 20 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, presid, state, countri, morn, vote, trump, elect, communiti,
yesterday, congressman, bill, hampton, support, star
Topic #1: bread, meat, eat, chicken, pizza, burger, sauc, sandwich, flavor,
beef, restaur, cut, taco, dip, soup
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, eat, salt, water,
richard, anni, birthday, pie, bowl
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, weapon,
version, watch, gift, mom, jacob, winter
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, oak, bedroom, presid, war
Topic #5: song, citi, album, danc, band, list, sound, rock, version, countri,
soundtrack, disney, record, boom, coffe
Topic #6: game, host, system, bill, playstat, video, footbal, version, bar,
genesi, tenni, barker, ball, golf, control
Topic #7: water, bathroom, tast, sarah, drink, pie, glass, kitchen, beer,
sister, differ, citi, ami, ice, butter
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink,
```

```
vanilla, sugar, war, coffe, tast
Topic #9: book, kid, child, moon, rabbit, read, seri, dog, pooh, lesson,
version, chees, page, friend, grover
Topic #10: war, centuri, power, woman, comedi, corner, franc, radio, citi,
empir, period, god, death, emperor, son
Topic #11: york, citi, film, list, chicago, watch, street, manhattan, park,
version, vega, bobo, accent, jackson, town
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, max, coffe, servic,
eat, drink, membership, dress, saucer
Topic #13: car, drive, baja, subaru, chip, transmiss, door, truck, engin,
outback, road, sophi, vehicl, ride, river
Topic #14: audienc, particip, song, da, perform, rock, list, experi, carolin,
bar, concert, danc, shout, whoa, artist
Topic #15: charact, star, film, watch, disney, list, scene, ride, version,
monster, seri, line, actor, god, action
Topic #16: river, wine, radio, festiv, countri, weekend, yesterday, state, farm,
station, bill, street, morn, rock, massachusett
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid,
russel, space, wait, wood, director
Topic #18: bank, march, communiti, wine, hunger, center, street, monday, team,
morn, congressman, insecur, state, store, mass
Topic #19: busi, box, slash, dog, hand, birthday, season, evid, cat, rule,
friend, store, justic, style, support
```

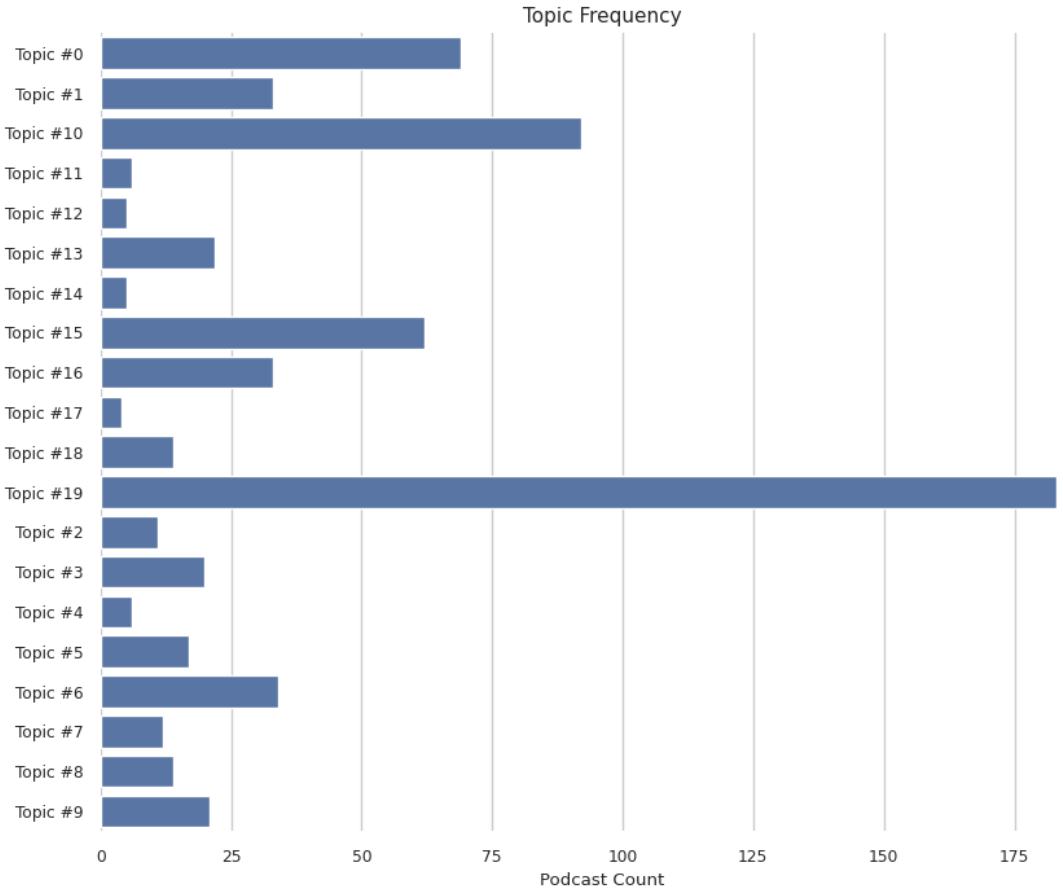
[94]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
19    -0.011806 -0.002051       1        1  14.506708
0      0.139289 -0.130712       2        1   9.341870
10     0.049328 -0.031053       3        1   8.486885
15     0.034968  0.133760       4        1   8.190150
16     0.148651 -0.120217       5        1   7.417464
6      0.013087  0.058612       6        1   6.214701
9      0.014734 -0.000690       7        1   5.377084
5      0.084889  0.172811       8        1   4.929864
1     -0.151470  0.029991       9        1   4.709365
13     0.028545  0.038244      10        1   4.513291
18     0.129634 -0.159036      11        1   4.344313
7     -0.113777 -0.036562      12        1   3.214027
11     0.056441  0.109256      13        1   2.997560
3     -0.059550  0.045536      14        1   2.925880
2     -0.227722 -0.063948      15        1   2.673472
4      0.160256 -0.177401      16        1   2.366501
8     -0.187703 -0.122249      17        1   2.268862
14     0.069291  0.219569      18        1   1.938175
17    -0.002255  0.092471      19        1   1.876046
12    -0.174829 -0.056331      20        1   1.707783, topic_info=
```

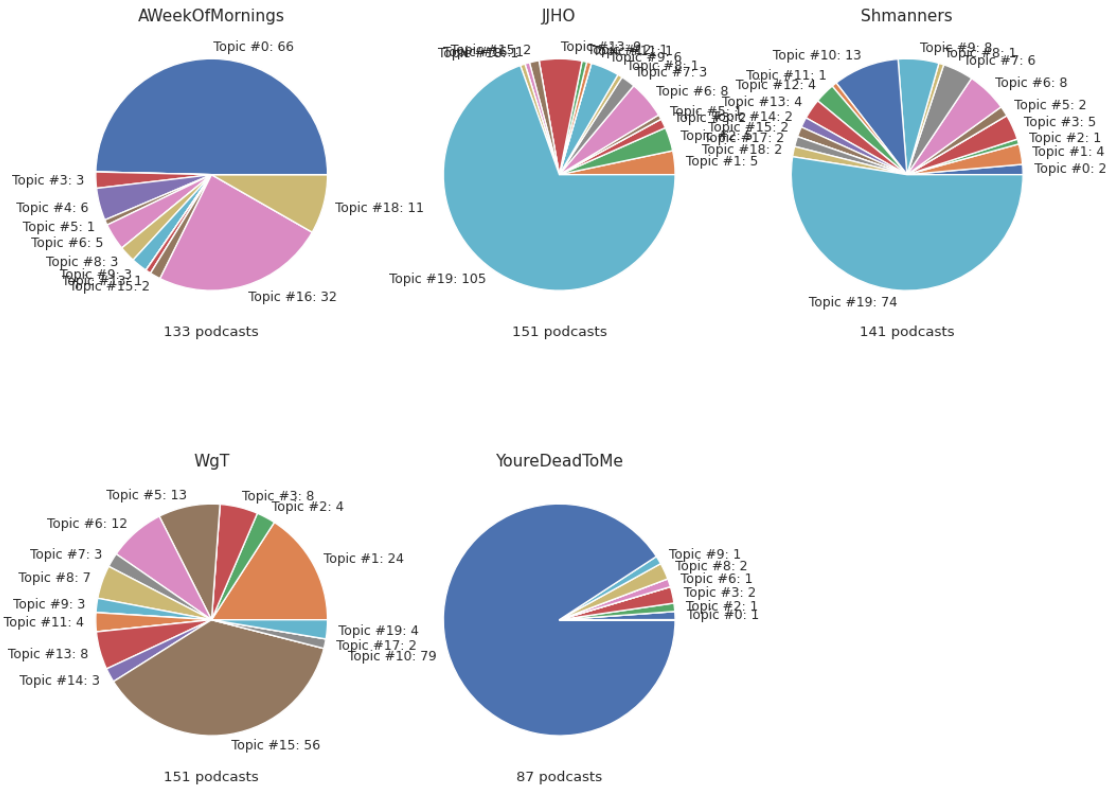
```
Term         Freq        Total Category  logprob  loglift
4933   christma  2223.000000  2223.000000  Default  30.0000  30.0000
10295      game  3192.000000  3192.000000  Default  29.0000  29.0000
2887       book  3100.000000  3100.000000  Default  28.0000  28.0000
26068       tea  1713.000000  1713.000000  Default  27.0000  27.0000
24514      song  3156.000000  3156.000000  Default  26.0000  26.0000
…           …           …           …      …       …       …
12551     hotel    47.127027   317.491776  Topic20  -5.4510   2.1624
8651      enjoy    50.528271   437.335818  Topic20  -5.3814   1.9118
13777      join    55.999144   686.899454  Topic20  -5.2786   1.5631
6255      cream    62.437838  1960.493420  Topic20  -5.1697   0.6232
25578   support    48.592966  1191.946576  Topic20  -5.4204   0.8701

[1868 rows x 6 columns], token_table=       Topic      Freq       Term
term
3           7  0.989012   aaronson
4           5  1.275609   aassenha
8          17  1.014788       abag
9           3  0.825545  abakhumid
43         11  0.929575     abject
…       …        …         …
29100      20  0.019402      zelda
29124       3  0.994783      zhang
29156       8  0.955407     zippel
29197      11  1.051536      zulal
29200       4  1.018934      zuvio

[6734 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[20, 1, 11, 16, 17, 7, 10, 6, 2, 14, 19, 8, 12, 4,
3, 5, 9, 15, 18, 13])
```

[95]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, presid, state, countri, morn, vote, trump, elect, communiti,
yesterday, congressman, bill, hampton, support, star
Topic #1: bread, meat, eat, chicken, pizza, burger, sauc, sandwich, flavor,
beef, restaur, cut, taco, dip, soup
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, eat, salt, water,
richard, anni, birthday, pie, bowl
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, weapon,
version, watch, gift, mom, jacob, winter
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, oak, bedroom, presid, war
Topic #5: song, citi, album, danc, band, list, sound, rock, version, countri,
soundtrack, disney, record, boom, coffe
```

Topic #6: game, host, system, bill, playstat, video, footbal, version, bar, genesi, tenni, barker, ball, golf, control
Topic #7: water, bathroom, tast, sarah, drink, pie, glass, kitchen, beer, sister, differ, citi, ami, ice, butter
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink, vanilla, sugar, war, coffe, tast
Topic #9: book, kid, child, moon, rabbit, read, seri, dog, pooh, lesson, version, chees, page, friend, grover
Topic #10: war, centuri, power, woman, comedi, corner, franc, radio, citi, empir, period, god, death, emperor, son
Topic #11: york, citi, film, list, chicago, watch, street, manhattan, park, version, vega, bobo, accent, jackson, town
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, max, coffe, servic, eat, drink, membership, dress, saucer
Topic #13: car, drive, baja, subaru, chip, transmiss, door, truck, engin, outback, road, sophi, vehicl, ride, river
Topic #14: audienc, particip, song, da, perform, rock, list, experi, carolin, bar, concert, danc, shout, whoa, artist
Topic #15: charact, star, film, watch, disney, list, scene, ride, version, monster, seri, line, actor, god, action
Topic #16: river, wine, radio, festiv, countri, weekend, yesterday, state, farm, station, bill, street, morn, rock, massachusett
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid, russel, space, wait, wood, director
Topic #18: bank, march, communiti, wine, hunger, center, street, monday, team, morn, congressman, insecur, state, store, mass
Topic #19: busi, box, slash, dog, hand, birthday, season, evid, cat, rule, friend, store, justic, style, support


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
25
50
75
100
125
150
175
Podcast Count



[96]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 20 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: cat, dog, docket, main, dracula, letter, mom, disput, evid, justic,
rule, birthday, season, dad, court
Topic #1: wine, river, state, presid, congressman, countri, yesterday, vote,
elect, bill, farm, trump, committe, hampton, telescop
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, ghost, tree, carol, villain, song
Topic #3: bread, dip, breakfast, meat, soup, meal, pizza, eat, sandwich, sauc,
chicken, dinner, cook, burger, buffet
Topic #4: empir, emperor, war, franc, power, corner, armi, henri, centuri,
battl, greg, woman, citi, china, comedi
Topic #5: film, charact, star, list, disney, watch, tv, action, scene, mission,
jason, jedi, trilog, actor, version
Topic #6: etiquett, doordash, host, parti, app, hair, advic, slash, art, busi,
mcelroy, max, deliveri, photographi, fund
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond,
```

```
easter, kat, sugar, flavor, quaker, pot
Topic #8: cream, ice, cake, pie, bagel, rainer, chocol, water, eat, vanilla,
milk, cone, birthday, carlson, snack
Topic #9: game, footbal, sport, playstat, basebal, island, golf, barker, ball,
tenni, pursuit, trivia, monopoli, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, presid, huga,
hampton, pizza, camp, spooner, oak, bedroom
Topic #11: appl, water, beer, cider, drink, flavor, glass, prohibit, coffe,
bathroom, juic, vanilla, tast, pie, fruit
Topic #12: tea, ceremoni, matcha, water, afternoon, cup, breakfast, dinner,
coffe, caffein, sugar, saucer, champagn, milk, astheticist
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, roll, vega,
disney, citi, boom, guitar, voic
Topic #14: rudi, glitter, tank, toad, gax, ha, patrick, blink, bridget, road,
partner, caesar, traffic, frisbe, powder
Topic #15: book, dedic, chees, kid, pooh, rabbit, horton, reader, moon, page,
milk, peter, grover, thackeri, child
Topic #16: bank, march, hunger, communiti, wine, insecur, confer, presid,
nutrit, center, congressman, ticket, counti, monday, state
Topic #17: virus, coronavirus, wine, muller, presid, vaccin, state, request,
virologist, quarantin, mask, distanc, trump, yesterday, morn
Topic #18: byron, shelley, perci, vampir, jane, frankenstein, poetri, mari,
corinn, florenc, monster, woman, godwin, franc, switzerland
Topic #19: car, drive, theater, subaru, baja, chip, transmiss, phone, josh,
citi, plane, road, wallet, river, york
```

[96]:

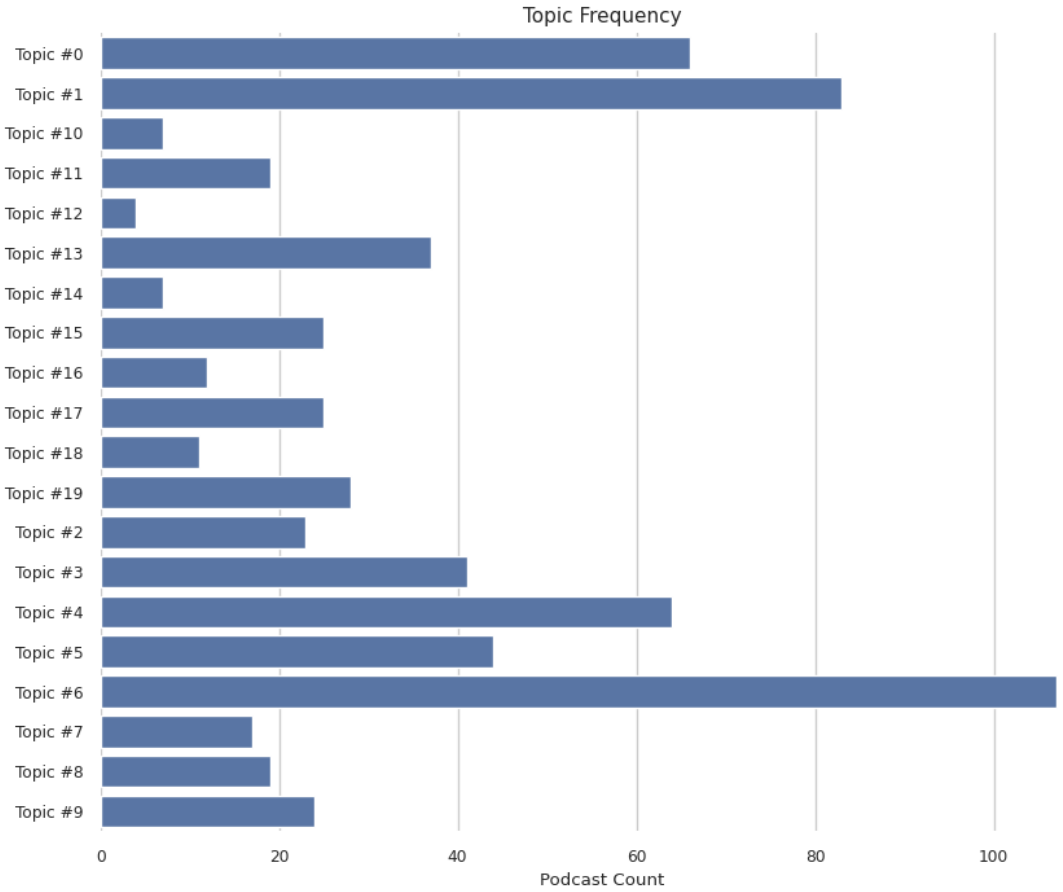
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.169598  0.093688       1        1  11.931616
6     -0.017065 -0.049057       2        1  10.498176
0      0.039472  0.011450       3        1   7.177937
4     -0.093405 -0.167540       4        1   6.912023
13    -0.014382 -0.100009       5        1   6.656756
5     -0.009177 -0.118380       6        1   6.342446
15    -0.036888 -0.071565       7        1   5.570130
19     0.032000  0.010688       8        1   5.416859
3      0.171616  0.092520       9        1   5.194146
17    -0.178012  0.116363      10        1   5.094506
9      0.025707 -0.036107      11        1   4.416121
16    -0.177323  0.131985      12        1   4.278300
11     0.114684  0.048925      13        1   3.920888
8      0.168863  0.069730      14        1   3.487838
2      0.057489 -0.010398      15        1   3.471326
7      0.127855  0.081088      16        1   2.534524
18    -0.051880 -0.273404      17        1   2.311715
10    -0.181367  0.125011      18        1   2.035672
```

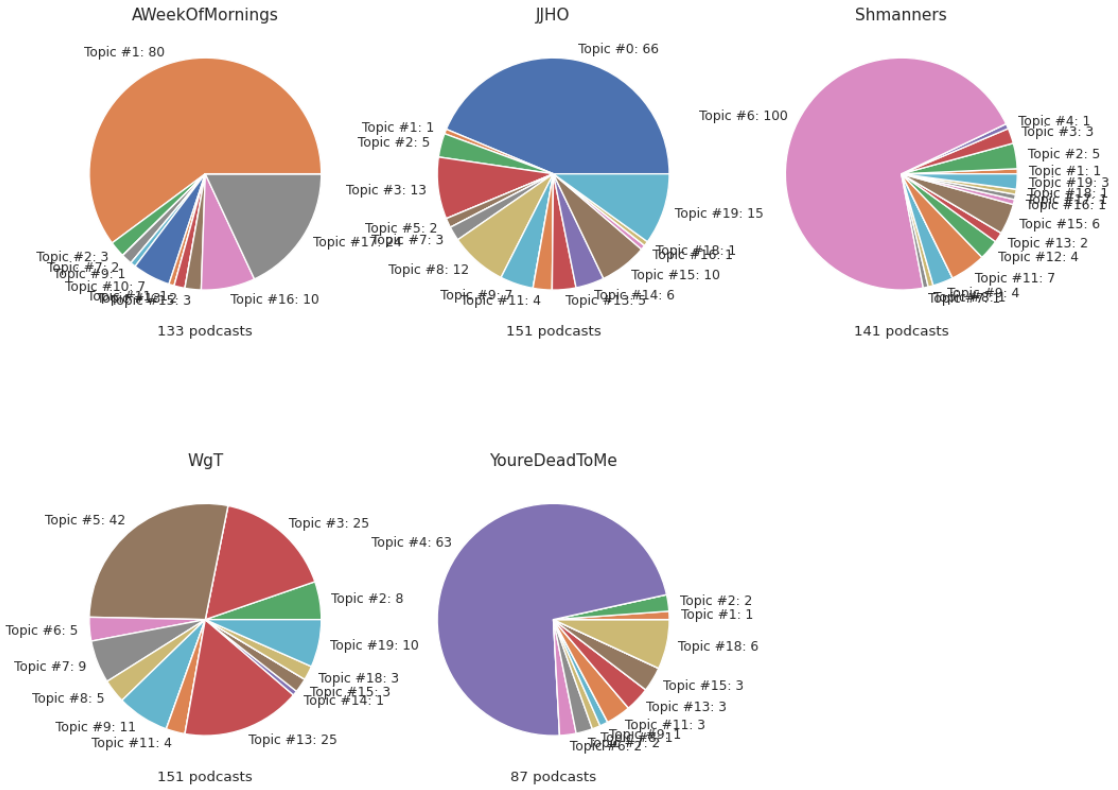
```
12     0.115451  0.013336      19        1   1.390721
14     0.075962  0.031677      20        1   1.358301, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  16.000000  16.000000  Default  30.0000  30.0000
26068       tea  10.000000  10.000000  Default  29.0000  29.0000
4877     chocol  13.000000  13.000000  Default  28.0000  28.0000
6255      cream  14.000000  14.000000  Default  27.0000  27.0000
12801       ice  11.000000  11.000000  Default  26.0000  26.0000
…          …        …         …       …      …        …
25007   station   0.410859   4.755147  Topic20  -5.6616   1.8502
4092       card   0.381805   4.175442  Topic20  -5.7349   1.9069
19542     peter   0.376245   4.357279  Topic20  -5.7496   1.8496
6657        dad   0.357054   6.014942  Topic20  -5.8019   1.4748
19075     parti   0.341776   8.008784  Topic20  -5.8457   1.1448

[1894 rows x 6 columns], token_table=       Topic      Freq      Term
term
98          8  0.367579    accent
119        13  1.292568  accuraci
161         1  0.285381    action
161         6  0.570761    action
161         8  0.285381    action
…       …        …         …
28990       6  0.131811      york
28990       8  0.131811      york
29100      11  1.283011     zelda
29124       4  1.513223     zhang
29169       2  0.938156      zola

[1062 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 7, 1, 5, 14, 6, 16, 20, 4, 18, 10, 17, 12, 9, 3,
8, 19, 11, 13, 15])
```

[97]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: cat, dog, docket, main, dracula, letter, mom, disput, evid, justic,
rule, birthday, season, dad, court
Topic #1: wine, river, state, presid, congressman, countri, yesterday, vote,
elect, bill, farm, trump, committe, hampton, telescop
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, ghost, tree, carol, villain, song
Topic #3: bread, dip, breakfast, meat, soup, meal, pizza, eat, sandwich, sauc,
chicken, dinner, cook, burger, buffet
Topic #4: empir, emperor, war, franc, power, corner, armi, henri, centuri,
battl, greg, woman, citi, china, comedi
```

Topic #5: film, charact, star, list, disney, watch, tv, action, scene, mission, jason, jedi, trilog, actor, version
Topic #6: etiquett, doordash, host, parti, app, hair, advic, slash, art, busi, mcelroy, max, deliveri, photographi, fund
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond, easter, kat, sugar, flavor, quaker, pot
Topic #8: cream, ice, cake, pie, bagel, rainer, chocol, water, eat, vanilla, milk, cone, birthday, carlson, snack
Topic #9: game, footbal, sport, playstat, basebal, island, golf, barker, ball, tenni, pursuit, trivia, monopoli, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, presid, huga, hampton, pizza, camp, spooner, oak, bedroom
Topic #11: appl, water, beer, cider, drink, flavor, glass, prohibit, coffe, bathroom, juic, vanilla, tast, pie, fruit
Topic #12: tea, ceremoni, matcha, water, afternoon, cup, breakfast, dinner, coffe, caffein, sugar, saucer, champagn, milk, astheticist
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, roll, vega, disney, citi, boom, guitar, voic
Topic #14: rudi, glitter, tank, toad, gax, ha, patrick, blink, bridget, road, partner, caesar, traffic, frisbe, powder
Topic #15: book, dedic, chees, kid, pooh, rabbit, horton, reader, moon, page, milk, peter, grover, thackeri, child
Topic #16: bank, march, hunger, communiti, wine, insecur, confer, presid, nutrit, center, congressman, ticket, counti, monday, state
Topic #17: virus, coronavirus, wine, muller, presid, vaccin, state, request, virologist, quarantin, mask, distanc, trump, yesterday, morn
Topic #18: byron, shelley, perci, vampir, jane, frankenstein, poetri, mari, corinn, florenc, monster, woman, godwin, franc, switzerland
Topic #19: car, drive, theater, subaru, baja, chip, transmiss, phone, josh, citi, plane, road, wallet, river, york


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
Podcast Count



[98]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 20 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: cat, dog, car, evid, justic, rule, birthday, season, disput, dracula,
turkey, docket, courtroom, mom, slash
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
peanut, bar, cacao, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, tradit, movi christma, santa claus, song, villain,
hanukkah, ghost
Topic #4: empir, battl, armi, greec, shah, democraci, fleet, salami, oracl,
jahangir, olymp, citi, greek, jahan, battl salami
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version,
york, car, theme song, theater, scene, theme
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, water,
vanilla, pie, rainer, bagel, cake cake, cone, eat
```

```
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, advic,
deliveri, fashion, mcelroy, busi, check
Topic #9: bread, dip, meat, sauc, pizza, sandwich, breakfast, eat, soup,
chicken, meal, burger, flavor, butter, beef
Topic #10: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni,
afternoon, cup, beer, tea kind, coffe, drink, sugar, yeah tea
Topic #11: byron, shelley, perci, vampir, jane, frankenstein, poetri, book,
mari, corinn, florenc, monster, switzerland, woman, godwin
Topic #12: food bank, bank, march, hunger, march food, wine, food insecur,
communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: josephin, danc, baker, josephin baker, franc, pari, dancer, cheetah,
venus, venus venus, danc floor, jazz, jeff, banana, chorus
Topic #14: henri, franc, eleanor, england, war, coloni, revolut, slaveri,
louvertur, woman, power, loui, richard, crusad, haiti
Topic #15: game, footbal, game show, golf, playstat, tenni, game game, sport,
croquet, basebal, ball, island, monopoli, barker, nintendo
Topic #16: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget,
road, ha, gax, guy gax, ha ha, frisbe, road toad
Topic #17: pyramid, stone, tomb, stoneheng, egypt, sarah, ancient, site,
imhotep, archaeolog, maria, giza, mera, kingdom, copper
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion,
zhang, guifei, rice, wu, empir, yang guifei, empress
Topic #19: khan, peter, chinggi, genghi, empir, genghi khan, phil, mongol,
chinggi khan, china, peter pan, silk, citi, temujin, jin
```

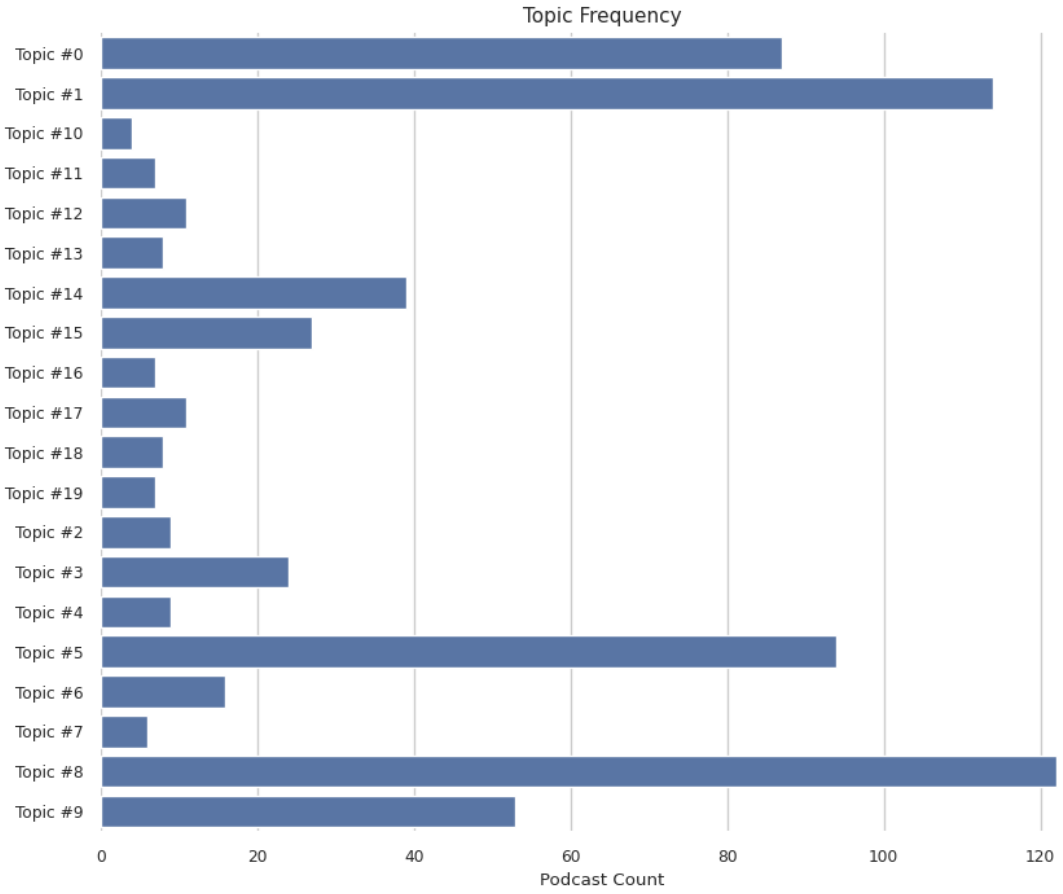
[98]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.075271  0.220270       1        1  18.439279
8     -0.032180 -0.042221       2        1  12.849986
5     -0.061011 -0.033020       3        1  12.514658
0     -0.117351 -0.090226       4        1   9.668203
9     -0.169362 -0.123440       5        1   7.115433
14     0.233636 -0.007547       6        1   5.269775
15    -0.093653 -0.023444       7        1   4.775885
3     -0.099505  0.016447       8        1   3.659423
6     -0.152271 -0.124076       9        1   3.365177
12    -0.090336  0.265559      10        1   3.150343
17     0.211502 -0.028034      11        1   2.380063
11     0.110848 -0.037753      12        1   2.274295
4      0.223679 -0.026758      13        1   2.193860
2     -0.139048  0.004224      14        1   2.165196
7     -0.063164  0.252444      15        1   2.120547
13     0.096984  0.021969      16        1   1.914488
18     0.206861  0.022255      17        1   1.646516
```

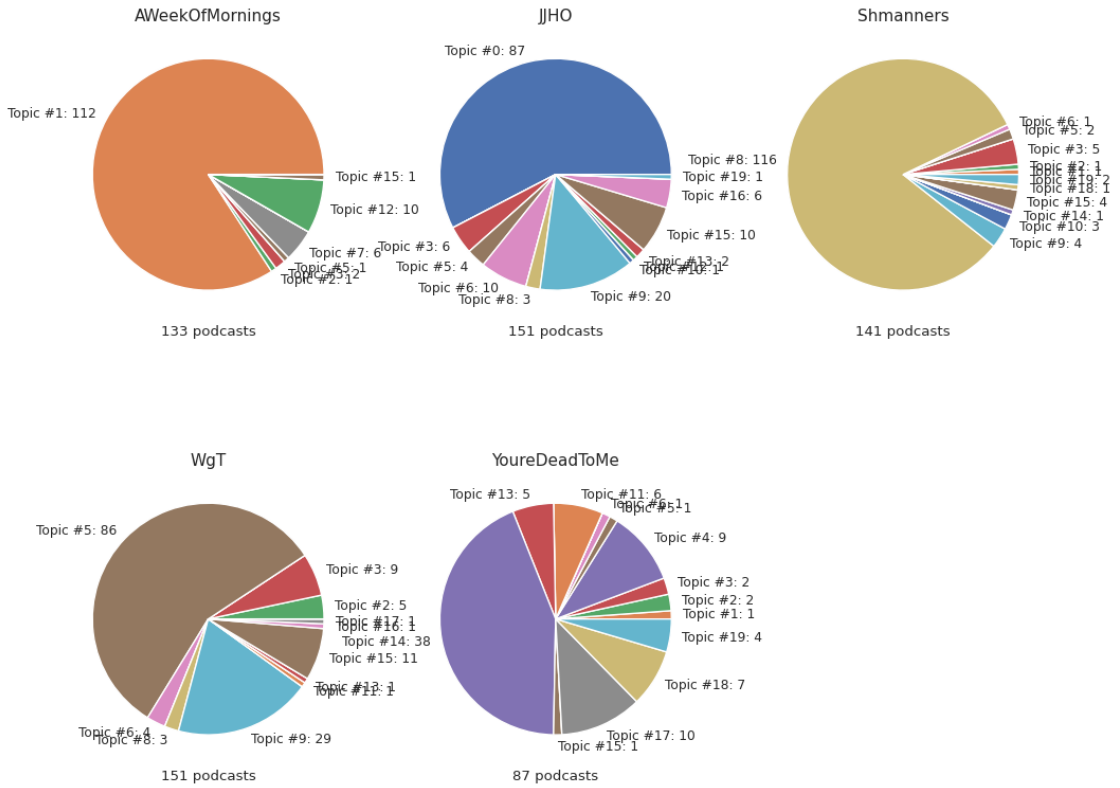
```
16    -0.091420 -0.104851      18        1   1.532885
19     0.182455 -0.033518      19        1   1.498429
10    -0.081393 -0.128281      20        1   1.465559, topic_info=
Term      Freq     Total Category  logprob  loglift
84847    christma  8.000000  8.000000  Default  30.0000  30.0000
83347      chocol  6.000000  6.000000  Default  29.0000  29.0000
481205        tea  5.000000  5.000000  Default  28.0000  28.0000
240891  ice cream  6.000000  6.000000  Default  27.0000  27.0000
116310      cream  7.000000  7.000000  Default  26.0000  26.0000
…           …       …       …       …      …      …
480200       tast  0.313950  2.957609  Topic20  -6.8713   1.9800
363988      plant  0.268453  1.664116  Topic20  -7.0278   2.3986
309840       milk  0.282658  3.315593  Topic20  -6.9763   1.7608
135957     dinner  0.255294  3.053156  Topic20  -7.0781   1.7414
51575   breakfast  0.228540  2.926335  Topic20  -7.1888   1.6731

[2003 rows x 6 columns], token_table=        Topic      Freq           Term
term
2599         3  0.491307         action
3021         3  0.503247          actor
5274         2  0.950290          advic
6048        20  0.657124      afternoon
6195        20  1.095064  afternoon tea
…       …       …             …
553250       1  0.224410           york
553250       2  0.224410           york
553250       3  0.224410           york
554534       2  1.861030  zero deliveri
554602      17  1.203980          zhang

[594 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 1, 10, 15, 16, 4, 7, 13, 18, 12, 5, 3, 8,
14, 19, 17, 20, 11])
```

[99]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: cat, dog, car, evid, justic, rule, birthday, season, disput, dracula,
turkey, docket, courtroom, mom, slash
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
peanut, bar, cacao, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, tradit, movi christma, santa claus, song, villain,
hanukkah, ghost
```

Topic #4: empir, battl, armi, greec, shah, democraci, fleet, salami, oracl, jahangir, olymp, citi, greek, jahan, battl salami
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version, york, car, theme song, theater, scene, theme
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, water, vanilla, pie, rainer, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi, bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, etiquett, host, app, hair, book, parti, slash, art, advic, deliveri, fashion, mcelroy, busi, check
Topic #9: bread, dip, meat, sauc, pizza, sandwich, breakfast, eat, soup, chicken, meal, burger, flavor, butter, beef
Topic #10: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni, afternoon, cup, beer, tea kind, coffe, drink, sugar, yeah tea
Topic #11: byron, shelley, perci, vampir, jane, frankenstein, poetri, book, mari, corinn, florenc, monster, switzerland, woman, godwin
Topic #12: food bank, bank, march, hunger, march food, wine, food insecur, communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: josephin, danc, baker, josephin baker, franc, pari, dancer, cheetah, venus, venus venus, danc floor, jazz, jeff, banana, chorus
Topic #14: henri, franc, eleanor, england, war, coloni, revolut, slaveri, louvertur, woman, power, loui, richard, crusad, haiti
Topic #15: game, footbal, game show, golf, playstat, tenni, game game, sport, croquet, basebal, ball, island, monopoli, barker, nintendo
Topic #16: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget, road, ha, gax, guy gax, ha ha, frisbe, road toad
Topic #17: pyramid, stone, tomb, stoneheng, egypt, sarah, ancient, site, imhotep, archaeolog, maria, giza, mera, kingdom, copper
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, zhang, guifei, rice, wu, empir, yang guifei, empress
Topic #19: khan, peter, chinggi, genghi, empir, genghi khan, phil, mongol, chinggi khan, china, peter pan, silk, citi, temujin, jin


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count



[100]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 20 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: song, film, charact, list, game, disney, version, star, watch, movi
movi, york, car, theme song, theme, vega
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
communiti, morn, hampton, vote, bill, radio, elect
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut
butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, villain, scroog, song, santa claus, ghost,
tradit, place christma
Topic #4: henri, battl, armi, eleanor, empir, franc, greec, war, olymp, crusad,
democraci, richard, king, chariot, fleet
Topic #5: evid, dog, cat, car, justic, rule, rudi, season, turkey, birthday,
courtroom, slash, disput, jacki, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, vanilla, water, pie, rainer, cone
```

```
Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect
cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer
connect, support cancer, cancer cancer
Topic #8: doordash, etiquett, host, app, parti, book, hair, slash, art, fashion,
advic, game, deliveri, max, mcelroy
Topic #9: bread, dip, meat, pizza, sauc, sandwich, chicken, soup, eat,
breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, book,
mari, florenc, corinn, switzerland, woman, monster, godwin
Topic #12: food bank, bank, march, march food, hunger, march food bank, wine,
food insecur, insecur, communiti, confer, food bank food, bank food, food food,
nutrit
Topic #13: josephin, baker, danc, josephin baker, franc, pari, cheetah, dancer,
venus, venus venus, banana, chorus, jazz, danc floor, danc danc
Topic #14: homo, languag, california man, stone, stone age, tim, ice age, site,
extinct, fossil, dna, fossil record, stoneheng, age, centuri
Topic #15: pyramid, tomb, stone, sarah, imhotep, egypt, ancient, maria,
archaeolog, giza, mera, kingdom, stoneheng, copper, mummi
Topic #16: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion,
guifei, zhang, wu, rice, yang guifei, empir, minist
Topic #17: coloni, mayflow, slaveri, louvertur, revolut, ship, haiti, weston,
popul, island, franc, misha, carver, freedom, war
Topic #18: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic,
butter, appl appl, juic, cider cider, drink, appl cider, glass
Topic #19: khan, empir, shah, peter, chinggi, jahangir, genghi, jahan, shah
jahan, mogul, genghi khan, mongol, akbar, phil, chinggi khan
```

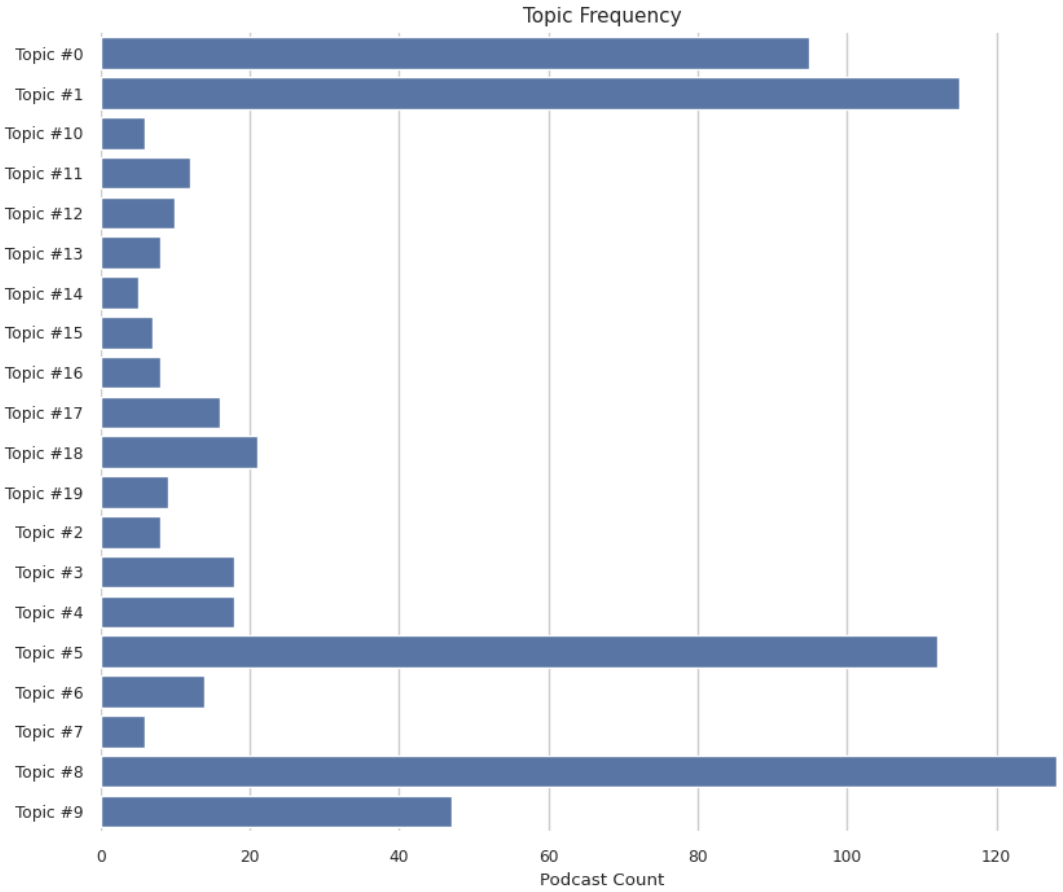
[100]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.104419 -0.228397       1        1  19.852530
8     -0.019025  0.053335       2        1  13.798862
5     -0.119547  0.117857       3        1  13.302265
0     -0.076534  0.063836       4        1  11.638653
9     -0.181358  0.156848       5        1   6.881171
18    -0.155671  0.039058       6        1   3.539838
3     -0.122088  0.015752       7        1   3.385444
6     -0.178103  0.129953       8        1   3.128130
12    -0.123406 -0.277014       9        1   3.038314
4      0.241999  0.028327      10        1   2.800267
11     0.124290  0.047948      11        1   2.543980
17     0.213793  0.000952      12        1   2.308898
7     -0.085482 -0.249323      13        1   2.117249
2     -0.158607  0.033095      14        1   1.967177
19     0.182506 -0.001956      15        1   1.852804
```

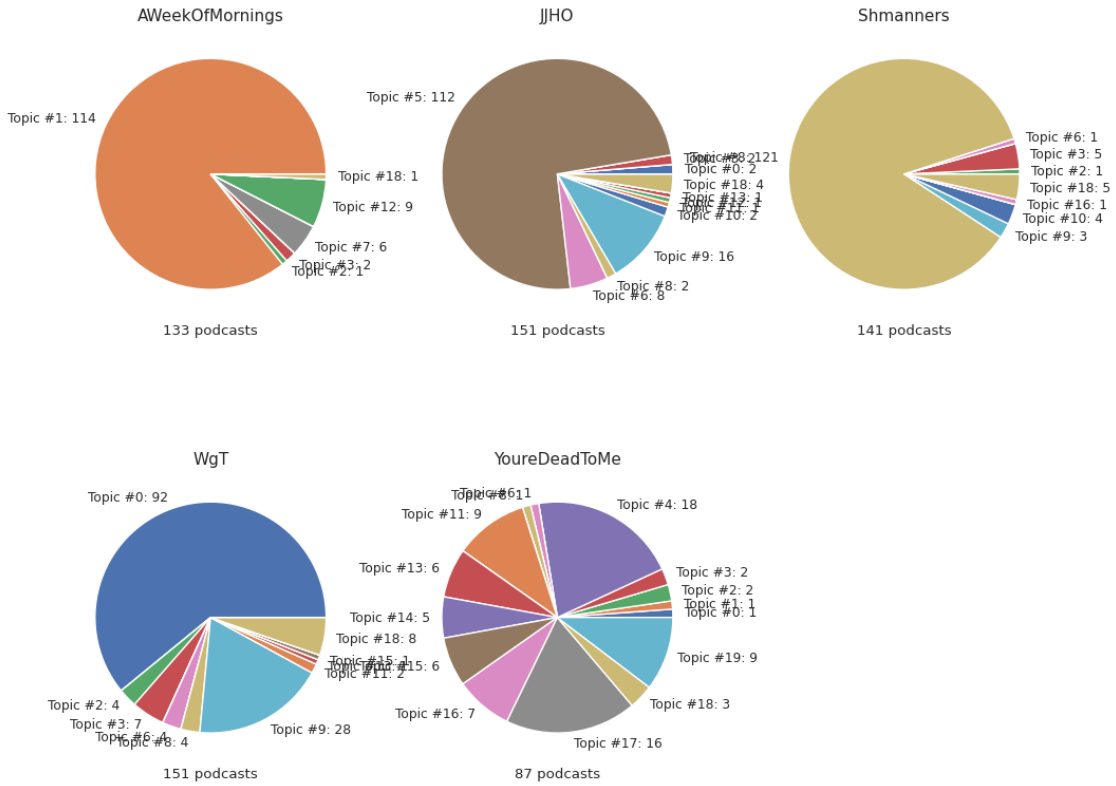
```
13     0.128931  0.006803      16        1   1.629437
16     0.163221 -0.039889      17        1   1.628329
10    -0.099141  0.072011      18        1   1.616365
15     0.191499  0.009842      19        1   1.531871
14     0.177143  0.020961      20        1   1.438416, topic_info=
Term      Freq     Total Category  logprob  loglift
1141840        tea  5.000000  5.000000  Default  30.0000  30.0000
195860     christma  6.000000  6.000000  Default  29.0000  29.0000
192266       chocol  5.000000  5.000000  Default  28.0000  28.0000
566886    ice cream  4.000000  4.000000  Default  27.0000  27.0000
270440        cream  5.000000  5.000000  Default  26.0000  26.0000
…              …       …       …       …      …       …
171827      centuri  0.320720  3.463404  Topic20  -7.1984   1.8622
502820         hair  0.292840  2.804069  Topic20  -7.2893   1.9824
458593         gene  0.240380  1.192298  Topic20  -7.4867   2.6402
141165   california  0.246459  1.892091  Topic20  -7.4618   2.2034
249678       corner  0.250560  2.979098  Topic20  -7.4453   1.7660

[1933 rows x 6 columns], token_table=         Topic      Freq           Term
term
5687          4  0.686855         action
6704          4  0.696780          actor
11712         2  0.589219          advic
13355        18  0.734557      afternoon
13682        18  1.067089  afternoon tea
…         …        …              …
1321637       4  1.248242      yeah movi
1337670       1  0.698567      yesterday
1339625       1  0.286995           york
1339625       4  0.286995           york
1342965      17  1.540921          zhang

[451 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 1, 10, 19, 4, 7, 13, 5, 12, 18, 8, 3, 20,
14, 17, 11, 16, 15])
```

[101]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: song, film, charact, list, game, disney, version, star, watch, movi
movi, york, car, theme song, theme, vega
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
communiti, morn, hampton, vote, bill, radio, elect
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut
butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
```

ghost christma, movi christma, villain, scroog, song, santa claus, ghost, tradit, place christma
Topic #4: henri, battl, armi, eleanor, empir, franc, greec, war, olymp, crusad, democraci, richard, king, chariot, fleet
Topic #5: evid, dog, cat, car, justic, rule, rudi, season, turkey, birthday, courtroom, slash, disput, jacki, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice, cream ice cream, chocol, ice cream ice, vanilla, water, pie, rainer, cone
Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer connect, support cancer, cancer cancer
Topic #8: doordash, etiquett, host, app, parti, book, hair, slash, art, fashion, advic, game, deliveri, max, mcelroy
Topic #9: bread, dip, meat, pizza, sauc, sandwich, chicken, soup, eat, breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni, afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, book, mari, florenc, corinn, switzerland, woman, monster, godwin
Topic #12: food bank, bank, march, march food, hunger, march food bank, wine, food insecur, insecur, communiti, confer, food bank food, bank food, food food, nutrit
Topic #13: josephin, baker, danc, josephin baker, franc, pari, cheetah, dancer, venus, venus venus, banana, chorus, jazz, danc floor, danc danc
Topic #14: homo, languag, california man, stone, stone age, tim, ice age, site, extinct, fossil, dna, fossil record, stoneheng, age, centuri
Topic #15: pyramid, tomb, stone, sarah, imhotep, egypt, ancient, maria, archaeolog, giza, mera, kingdom, stoneheng, copper, mummi
Topic #16: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, guifei, zhang, wu, rice, yang guifei, empir, minist
Topic #17: coloni, mayflow, slaveri, louvertur, revolut, ship, haiti, weston, popul, island, franc, misha, carver, freedom, war
Topic #18: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic, butter, appl appl, juic, cider cider, drink, appl cider, glass
Topic #19: khan, empir, shah, peter, chinggi, jahangir, genghi, jahan, shah jahan, mogul, genghi khan, mongol, akbar, phil, chinggi khan


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count



[102]:
```
topic_count = 25
```

[103]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 25 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: eat, bread, meat, butter, pizza, chicken, soup, sauc, cook, restaur,
meal, breakfast, sandwich, dip, dinner
Topic #1: game, footbal, bar, car, cooki, potato, team, mission, film, snack,
appl, basebal, sport, bill, board
Topic #2: car, citi, jason, pretzel, drive, soul, chicago, york, shape, dream,
jeff, engin, wilco, sound, busi
Topic #3: style, jazz, water, rammel, summer, town, graham, bed, product, pool,
watermelon, greec, coffe, tube, citi
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
yesterday, book, bill, street, congressman, connect
Topic #5: democraci, battl, burton, worm, bin, jason, michael, scienc, water,
citi, compost, game, greec, beetlejuic, dog
Topic #6: birthday, edit, caesar, poni, blame, answer, celebr, kyle, julius,
nanci, greg, chivalri, dragon, jack, disney
```

```
Topic #7: cancer, connect, beer, star, barnum, stoneheng, bed, pyramid, charact,
jedi, hair, stone, lake, saucer, blah
Topic #8: christma, holiday, monster, ghost, candi, santa, parti, game, golf,
tradit, club, spirit, version, vega, claus
Topic #9: book, tank, tie, carniv, fish, bow, sandler, star, jenga, aquarium,
jewelri, neck, rabbit, jinger, brother
Topic #10: slash, host, hand, check, art, busi, etiquett, advic, pictur, month,
exampl, book, code, husband, communiti
Topic #11: fashion, franc, revolut, power, war, ship, presid, game, postur,
slaveri, captain, plant, georg, pirat, system
Topic #12: news, tenni, england, town, game, sauc, child, birthday, communiti,
radio, bill, dad, heel, island, court
Topic #13: woman, smell, franc, henri, byron, vampir, centuri, book, shelley,
england, fan, marriag, husband, relationship, comedi
Topic #14: tea, water, game, drink, tast, champagn, system, video, afternoon,
bathroom, hole, playstat, comedi, genesi, island
Topic #15: ride, roller, milk, car, josephin, water, belt, tang, air, disney,
land, toilet, baker, line, emperor
Topic #16: danc, pie, jeff, crust, trash, mason, parti, floor, storag, slide,
lila, richard, roy, cw, water
Topic #17: war, centuri, corner, comedi, empir, power, radio, period, citi,
woman, greg, emperor, professor, god, age
Topic #18: chocol, york, milk, cocoa, accent, hors, bar, car, championship, bat,
jane, watch, season, war, cacao
Topic #19: book, friend, dog, rule, slash, justic, cat, evid, husband, season,
box, mom, store, dad, letter
Topic #20: cream, ice, cake, flavor, chocol, vanilla, market, rudi, banana,
salsa, sale, parti, game, eat, flea
Topic #21: chair, stone, peter, marvel, disney, pan, theater, captain, porch,
infin, age, sit, book, power, burlesqu
Topic #22: paint, chip, bog, car, door, sale, letter, richard, kate, ink, power,
sprayer, garag, season, club
Topic #23: song, list, charact, film, watch, audienc, version, theme, star, kid,
rock, perform, comedi, citi, topic
Topic #24: glitter, rudi, rick, watch, prohibit, aeta, valentin, drummer,
snowbal, alcohol, vote, fight, festiv, letter, hepburn
```

[103]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4      0.140839  0.011025       1        1  29.771894
19     0.231295 -0.032195       2        1  12.099043
23     0.228809 -0.025937       3        1   9.918425
10     0.179833  0.013621       4        1   8.137402
0      0.114642 -0.226154       5        1   4.966421
8      0.050662 -0.020212       6        1   4.446822
17     0.143837  0.125370       7        1   4.199399
```

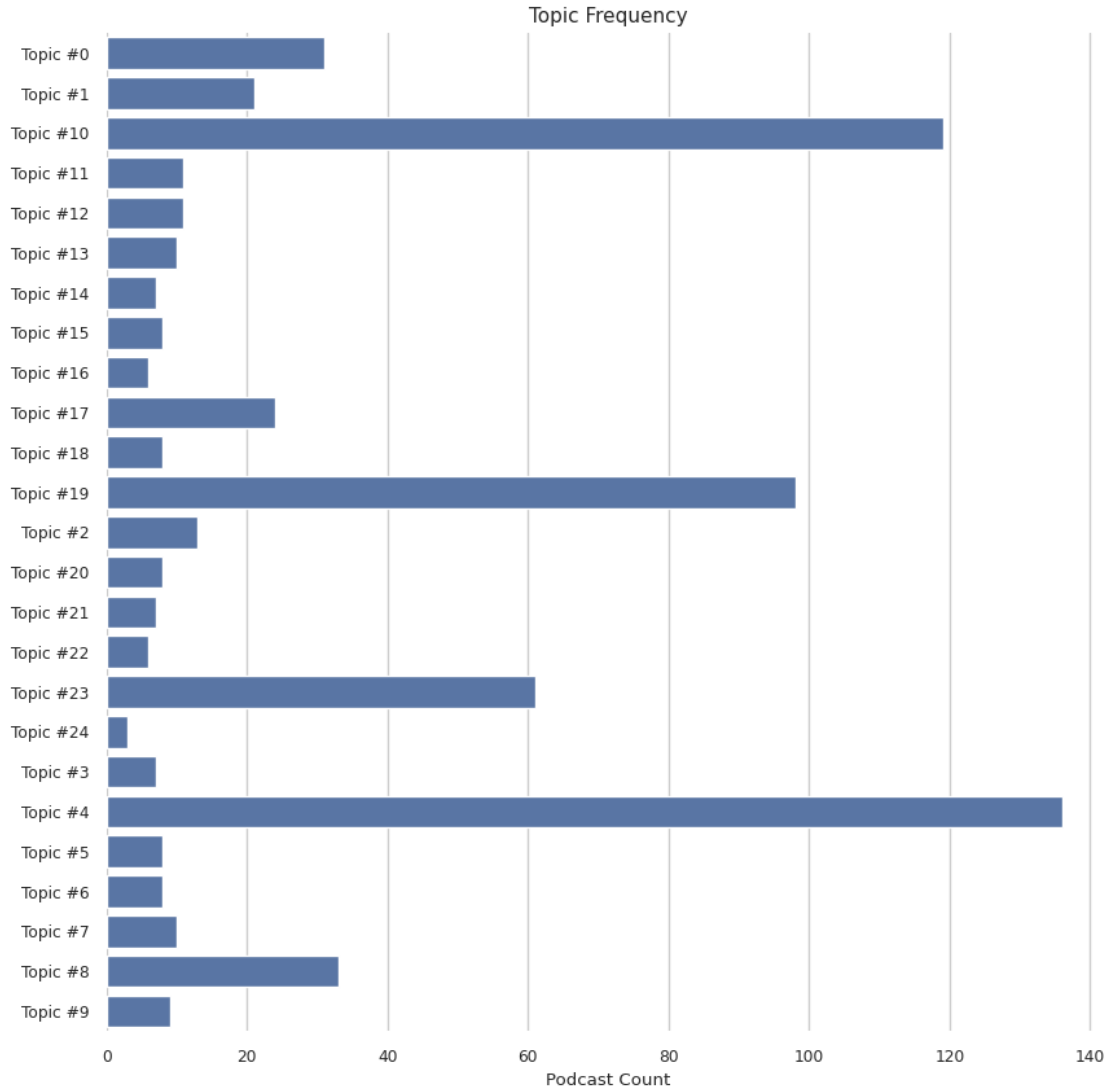
```
1      0.023466 -0.082029       8        1   2.690760
12     0.002842  0.036538       9        1   2.496635
2     -0.015929 -0.025008      10        1   1.902943
11    -0.012657  0.137851      11        1   1.828134
13     0.063984  0.169220      12        1   1.779715
7     -0.056621  0.008934      13        1   1.723428
14    -0.057214 -0.102947      14        1   1.524324
15    -0.094799  0.014553      15        1   1.335333
20    -0.097177 -0.203801      16        1   1.318046
5     -0.073929  0.036258      17        1   1.298873
6     -0.054871  0.072044      18        1   1.279353
3     -0.092336 -0.002905      19        1   1.271739
21    -0.053001  0.103207      20        1   1.259442
18    -0.057869 -0.037424      21        1   1.172606
9     -0.105750  0.074643      22        1   1.029393
16    -0.108064 -0.049329      23        1   0.937680
22    -0.117956 -0.001380      24        1   0.843027
24    -0.182036  0.006057      25        1   0.769160, topic_info=
Term         Freq        Total Category  logprob  loglift
2887       book  2320.000000  2320.000000  Default  30.0000  30.0000
6255      cream   945.000000   945.000000  Default  29.0000  29.0000
12801       ice   927.000000   927.000000  Default  28.0000  28.0000
10295      game  2130.000000  2130.000000  Default  27.0000  27.0000
4933   christma  1227.000000  1227.000000  Default  26.0000  26.0000
…          …           …           …      …       …       …
23678      ship    17.243801   410.326740  Topic25  -5.6588   1.6981
2748       bodi    16.479601   871.171898  Topic25  -5.7041   0.8999
19075     parti    16.723568  1148.998006  Topic25  -5.6894   0.6378
1789        bar    14.756844   753.174511  Topic25  -5.8145   0.9350
19085   partner    13.838681   356.188606  Topic25  -5.8788   1.6196

[1848 rows x 6 columns], token_table=       Topic      Freq     Term
term
57          1  0.572048    abort
57          9  0.364031    abort
57         13  0.020802    abort
57         20  0.010401    abort
57         23  0.010401    abort
…       …        …      …
29124      15  0.098114    zhang
29127       3  0.882295  zhongdu
29134       9  0.922885    zilla
29180      15  0.092035   zoomer
29180      24  0.828312   zoomer

[8076 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 20, 24, 11, 1, 9, 18, 2, 13, 3, 12, 14, 8, 15,
```

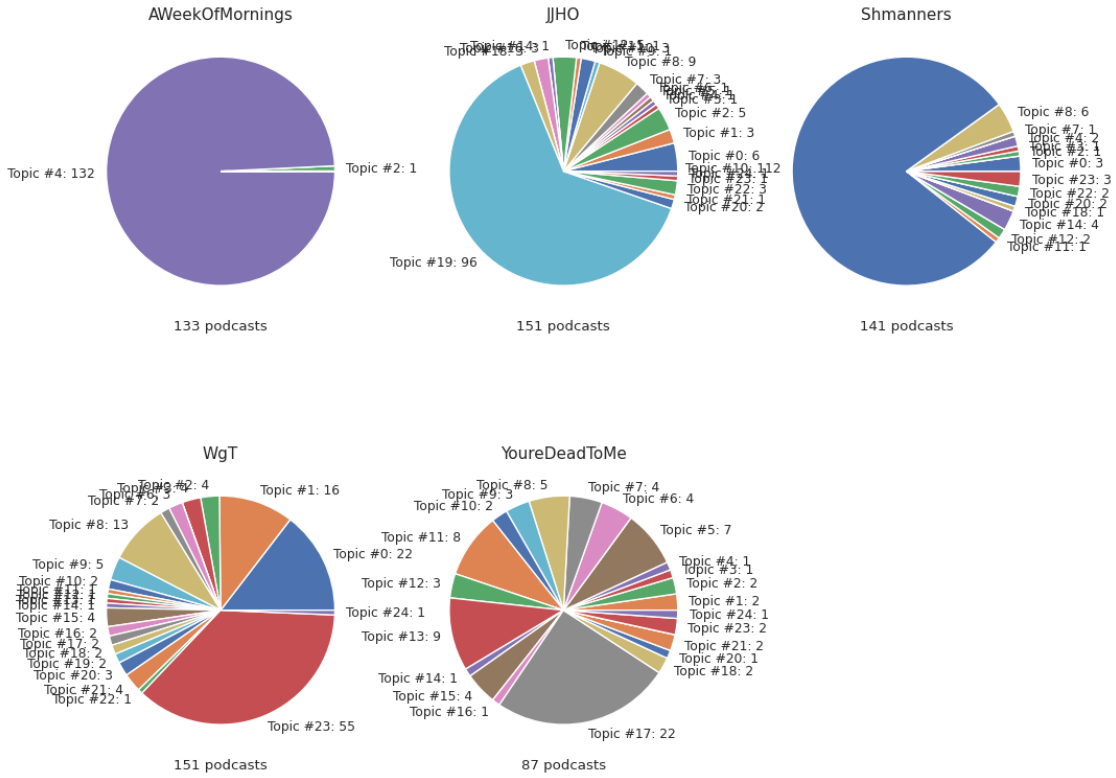
16, 21, 6, 7, 4, 22, 19, 10, 17, 23, 25])

[104]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: eat, bread, meat, butter, pizza, chicken, soup, sauc, cook, restaur,
meal, breakfast, sandwich, dip, dinner
Topic #1: game, footbal, bar, car, cooki, potato, team, mission, film, snack,
appl, basebal, sport, bill, board
Topic #2: car, citi, jason, pretzel, drive, soul, chicago, york, shape, dream,
jeff, engin, wilco, sound, busi
Topic #3: style, jazz, water, rammel, summer, town, graham, bed, product, pool,
watermelon, greec, coffe, tube, citi
Topic #4: wine, river, state, presid, countri, communiti, morn, song, radio,
yesterday, book, bill, street, congressman, connect
Topic #5: democraci, battl, burton, worm, bin, jason, michael, scienc, water,
citi, compost, game, greec, beetlejuic, dog
Topic #6: birthday, edit, caesar, poni, blame, answer, celebr, kyle, julius,
nanci, greg, chivalri, dragon, jack, disney
Topic #7: cancer, connect, beer, star, barnum, stoneheng, bed, pyramid, charact,
jedi, hair, stone, lake, saucer, blah
Topic #8: christma, holiday, monster, ghost, candi, santa, parti, game, golf,
tradit, club, spirit, version, vega, claus
Topic #9: book, tank, tie, carniv, fish, bow, sandler, star, jenga, aquarium,
jewelri, neck, rabbit, jinger, brother
Topic #10: slash, host, hand, check, art, busi, etiquett, advic, pictur, month,
exampl, book, code, husband, communiti
Topic #11: fashion, franc, revolut, power, war, ship, presid, game, postur,
slaveri, captain, plant, georg, pirat, system
Topic #12: news, tenni, england, town, game, sauc, child, birthday, communiti,
radio, bill, dad, heel, island, court
Topic #13: woman, smell, franc, henri, byron, vampir, centuri, book, shelley,
england, fan, marriag, husband, relationship, comedi
Topic #14: tea, water, game, drink, tast, champagn, system, video, afternoon,
bathroom, hole, playstat, comedi, genesi, island
Topic #15: ride, roller, milk, car, josephin, water, belt, tang, air, disney,
land, toilet, baker, line, emperor
Topic #16: danc, pie, jeff, crust, trash, mason, parti, floor, storag, slide,
lila, richard, roy, cw, water
Topic #17: war, centuri, corner, comedi, empir, power, radio, period, citi,
woman, greg, emperor, professor, god, age
Topic #18: chocol, york, milk, cocoa, accent, hors, bar, car, championship, bat,
jane, watch, season, war, cacao
Topic #19: book, friend, dog, rule, slash, justic, cat, evid, husband, season,
box, mom, store, dad, letter
Topic #20: cream, ice, cake, flavor, chocol, vanilla, market, rudi, banana,
```

salsa, sale, parti, game, eat, flea
Topic #21: chair, stone, peter, marvel, disney, pan, theater, captain, porch, infin, age, sit, book, power, burlesqu
Topic #22: paint, chip, bog, car, door, sale, letter, richard, kate, ink, power, sprayer, garag, season, club
Topic #23: song, list, charact, film, watch, audienc, version, theme, star, kid, rock, perform, comedi, citi, topic
Topic #24: glitter, rudi, rick, watch, prohibit, aeta, valentin, drummer, snowbal, alcohol, vote, fight, festiv, letter, hepburn





[105]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 25 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, river, state, presid, countri, morn, yesterday, radio,
communiti, bill, congressman, space, vote, professor, street
Topic #1: cake, birthday, pie, celebr, butter, greg, parti, eat, crust, brandi,
footbal, licoric, kid, plan, surpris
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, salt, water, richard,
anni, vanilla, bowl, isi, street
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, weapon,
version, gift, mom, jacob, winter, watch
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, presid, oak, bedroom, war
Topic #5: song, danc, album, band, sound, rock, list, version, countri,
soundtrack, record, jeff, georg, voic, rammel
Topic #6: game, host, system, bill, playstat, video, footbal, version, bar,
genesi, tenni, barker, ball, control, golf
Topic #7: water, bathroom, tast, drink, sarah, pie, glass, kitchen, beer,
sister, river, differ, ice, ami, butter
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink,
```

```
vanilla, war, coffe, sugar, tast
Topic #9: book, kid, child, moon, seri, rabbit, read, pooh, dog, lesson, chees,
version, page, friend, author
Topic #10: war, centuri, power, woman, comedi, corner, franc, radio, period,
empir, emperor, god, son, death, age
Topic #11: york, film, list, watch, manhattan, version, park, accent, citi,
bobo, jackson, street, hall, griffin, god
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, coffe, max, servic,
eat, drink, membership, saucer, dress
Topic #13: car, drive, baja, subaru, chip, transmiss, river, door, truck, engin,
outback, road, sophi, vehicl, insur
Topic #14: audienc, particip, song, da, perform, rock, list, experi, bar,
carolin, concert, shout, whoa, artist, version
Topic #15: charact, star, watch, film, list, scene, seri, actor, action,
version, god, background, fan, perform, stone
Topic #16: bread, meat, eat, chicken, burger, pizza, sauc, sandwich, flavor,
beef, taco, cut, dip, restaur, soup
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid,
russel, space, wait, deck, director
Topic #18: citi, song, chicago, vega, town, coffe, hollywood, duh, street,
state, philadelphia, los, california, pretzel, empir
Topic #19: busi, hand, card, soul, jason, handshak, hug, situat, blockchain,
code, market, phone, doordash, shake, pocket
Topic #20: bank, march, communiti, wine, hunger, presid, center, state, morn,
congressman, street, monday, support, mass, countri
Topic #21: style, jazz, main, pair, bed, theater, garbag, state, pizza, wear,
portland, stage, band, countri, trio
Topic #22: box, dog, slash, season, cat, evid, rule, justic, friend, support,
dad, store, court, fund, mom
Topic #23: monster, cooki, street, costum, ghost, oscar, candi, vampir, version,
kid, halloween, grover, count, parti, sexi
Topic #24: ride, disney, peter, line, pan, land, park, roller, coaster, wait,
disneyland, wheel, dan, version, adventur
```

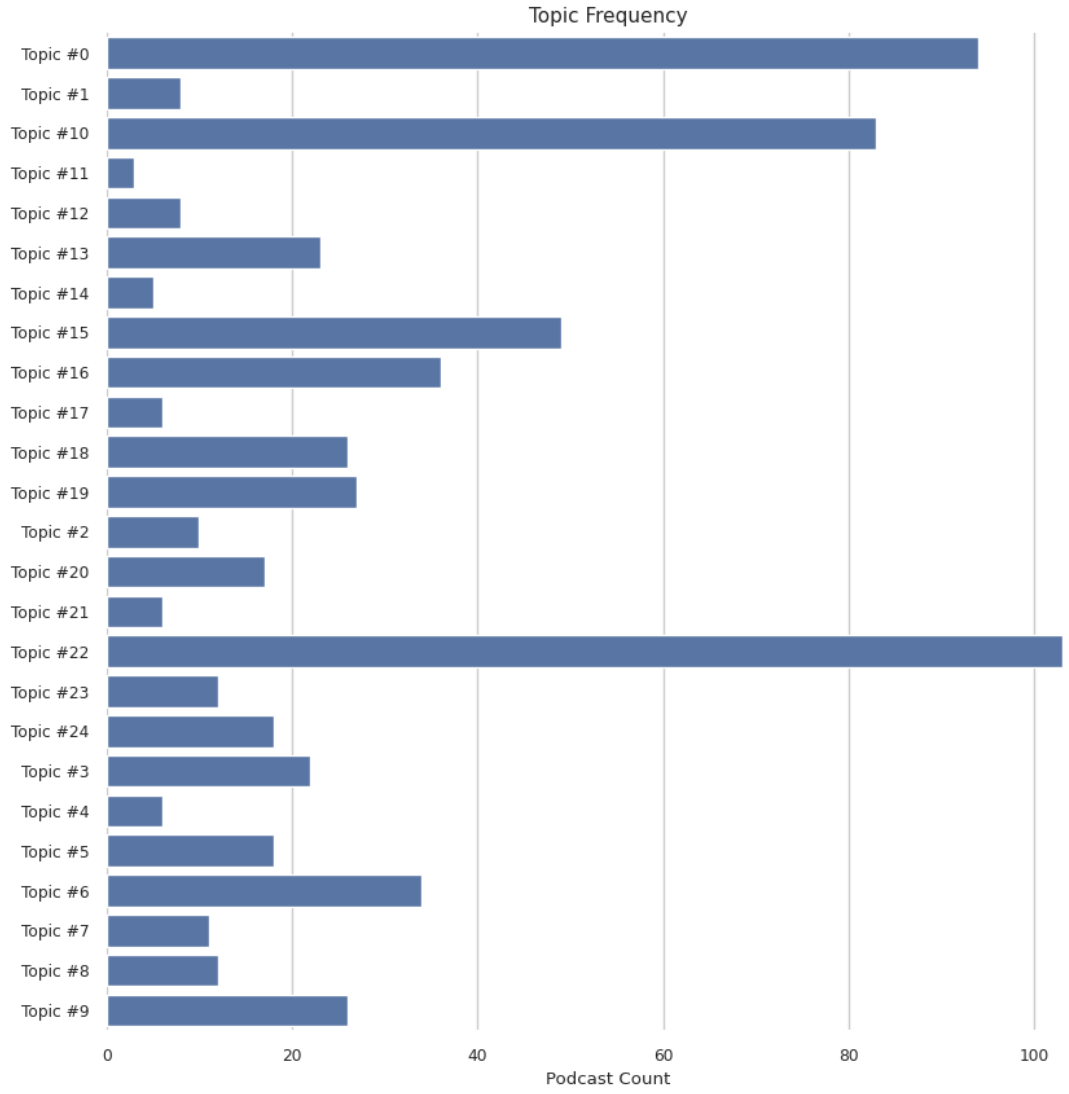
[105]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
0     -0.037812 -0.195375       1        1  10.418825
22     0.012279 -0.031361       2        1   9.315245
10    -0.032789 -0.086712       3        1   7.226655
15    -0.121034  0.067328       4        1   6.448586
6     -0.047807  0.035812       5        1   5.762715
20    -0.006092 -0.209826       6        1   5.457147
9     -0.021433 -0.029685       7        1   4.953641
19    -0.000124 -0.035525       8        1   4.614865
16     0.120738  0.096611       9        1   4.409479
13    -0.036530 -0.016951      10        1   4.346525
```

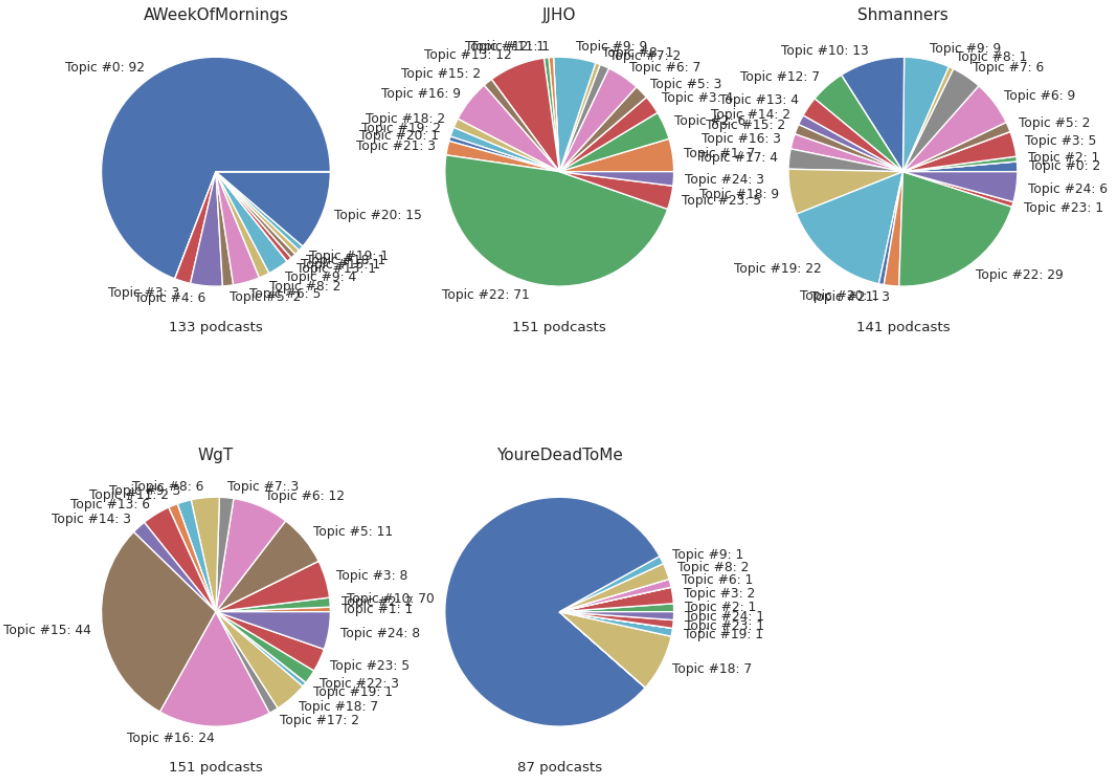
```
5     -0.165555  0.065367      11        1   4.265695
18    -0.032960 -0.057658      12        1   3.811356
23    -0.026001  0.105803      13        1   3.050773
7      0.098935  0.013675      14        1   2.898285
3      0.022351  0.079520      15        1   2.720407
24    -0.089499  0.142584      16        1   2.684317
2      0.220961  0.055821      17        1   2.390677
4     -0.025553 -0.238442      18        1   2.368383
11    -0.103195  0.026508      19        1   2.222805
8      0.221053 -0.001678      20        1   2.105203
21    -0.032997 -0.050857      21        1   1.974593
14    -0.171597  0.115651      22        1   1.683106
1      0.129571  0.072036      23        1   1.664461
17    -0.047318  0.057836      24        1   1.622497
12     0.172407  0.019518      25        1   1.583761, topic_info=
Term         Freq        Total Category  logprob  loglift
4933   christma  2119.000000  2119.000000  Default  30.0000  30.0000
10295      game  3126.000000  3126.000000  Default  29.0000  29.0000
2887       book  2930.000000  2930.000000  Default  28.0000  28.0000
26068       tea  1613.000000  1613.000000  Default  27.0000  27.0000
12801       ice  1875.000000  1875.000000  Default  26.0000  26.0000
…          …           …           …      …      …      …
8651      enjoy    47.366223   435.588317  Topic25  -5.3706   1.9266
13777      join    51.628583   668.528808  Topic25  -5.2844   1.5844
6255      cream    58.961494  1980.458852  Topic25  -5.1516   0.6312
12551     hotel    43.524289   340.905609  Topic25  -5.4552   2.0871
25578   support    46.627297  1150.234903  Topic25  -5.3863   0.9398

[2167 rows x 6 columns], token_table=       Topic      Freq      Term
term
3          7  1.061239  aaronson
4          1  1.167234  aassenha
8         20  0.962423      abag
43         6  1.049544    abject
54         6  1.049544   abordin
…      …        …       …
29100      5  0.921380     zelda
29100     14  0.061425     zelda
29100     25  0.020475     zelda
29197      6  0.943132     zulal
29200      4  1.010457     zuvio

[9175 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 23, 11, 16, 7, 21, 10, 20, 17, 14, 6, 19, 24, 8,
4, 25, 3, 5, 12, 9, 22, 15, 2, 18, 13])
```

[106]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, river, state, presid, countri, morn, yesterday, radio,
communiti, bill, congressman, space, vote, professor, street
Topic #1: cake, birthday, pie, celebr, butter, greg, parti, eat, crust, brandi,
footbal, licoric, kid, plan, surpris
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, salt, water, richard,
anni, vanilla, bowl, isi, street
Topic #3: christma, holiday, ghost, santa, tradit, claus, villain, list, weapon,
version, gift, mom, jacob, winter, watch
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, presid, oak, bedroom, war
Topic #5: song, danc, album, band, sound, rock, list, version, countri,
soundtrack, record, jeff, georg, voic, rammel
Topic #6: game, host, system, bill, playstat, video, footbal, version, bar,
genesi, tenni, barker, ball, control, golf
Topic #7: water, bathroom, tast, drink, sarah, pie, glass, kitchen, beer,
sister, river, differ, ice, ami, butter
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink,
vanilla, war, coffe, sugar, tast
Topic #9: book, kid, child, moon, seri, rabbit, read, pooh, dog, lesson, chees,
version, page, friend, author
Topic #10: war, centuri, power, woman, comedi, corner, franc, radio, period,
empir, emperor, god, son, death, age
Topic #11: york, film, list, watch, manhattan, version, park, accent, citi,
bobo, jackson, street, hall, griffin, god
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, coffe, max, servic,
eat, drink, membership, saucer, dress
Topic #13: car, drive, baja, subaru, chip, transmiss, river, door, truck, engin,
outback, road, sophi, vehicl, insur
Topic #14: audienc, particip, song, da, perform, rock, list, experi, bar,
carolin, concert, shout, whoa, artist, version
Topic #15: charact, star, watch, film, list, scene, seri, actor, action,
version, god, background, fan, perform, stone
Topic #16: bread, meat, eat, chicken, burger, pizza, sauc, sandwich, flavor,
beef, taco, cut, dip, restaur, soup
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid,
russel, space, wait, deck, director
Topic #18: citi, song, chicago, vega, town, coffe, hollywood, duh, street,
state, philadelphia, los, california, pretzel, empir
Topic #19: busi, hand, card, soul, jason, handshak, hug, situat, blockchain,
code, market, phone, doordash, shake, pocket
Topic #20: bank, march, communiti, wine, hunger, presid, center, state, morn,
congressman, street, monday, support, mass, countri
Topic #21: style, jazz, main, pair, bed, theater, garbag, state, pizza, wear,
```

portland, stage, band, countri, trio
Topic #22: box, dog, slash, season, cat, evid, rule, justic, friend, support, dad, store, court, fund, mom
Topic #23: monster, cooki, street, costum, ghost, oscar, candi, vampir, version, kid, halloween, grover, count, parti, sexi
Topic #24: ride, disney, peter, line, pan, land, park, roller, coaster, wait, disneyland, wheel, dan, version, adventur





[107]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 25 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: dog, docket, main, dracula, letter, book, cat, disput, mom, rule,
justic, evid, pizza, birthday, child
Topic #1: wine, river, bank, communiti, confer, congressman, farm, state, march,
yesterday, telescop, hunger, weekend, bill, committe
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, ghost, carol, tree, villain, song
Topic #3: breakfast, bread, meal, soup, meat, chicken, dinner, sandwich, eat,
buffet, cook, cooki, dessert, taco, burger
Topic #4: coloni, slaveri, war, revolut, louvertur, ship, mayflow, island,
haiti, popul, franc, weston, carver, freedom, trade
Topic #5: film, charact, star, disney, list, watch, tv, action, scene, mission,
book, jedi, version, jason, trilog
Topic #6: doordash, etiquett, app, parti, host, book, advic, art, hair,
deliveri, busi, slash, mcelroy, photographi, societi
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond, kat,
easter, sugar, quaker, pot, cadburi
```

```
Topic #8: cream, ice, cake, pie, bagel, vanilla, chocol, rainer, eat, birthday,
milk, cone, flavor, carlson, snack
Topic #9: game, footbal, sport, playstat, basebal, island, barker, golf, ball,
tenni, pursuit, trivia, monopoli, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton,
presid, camp, pizza, spooner, russia, oak
Topic #11: water, beer, bathroom, drink, logger, pie, tast, hair, river, swim,
glass, toilet, choir, smell, sarah
Topic #12: tea, ceremoni, matcha, cup, afternoon, water, coffe, caffein, sugar,
dinner, saucer, milk, astheticist, doordash, birthday
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, roll, vega,
boom, citi, guitar, disney, jeff
Topic #14: vote, elect, presid, trump, wine, state, countri, democraci,
congressman, book, yesterday, hampton, candi, offic, morn
Topic #15: car, drive, theater, subaru, baja, transmiss, chip, phone, citi,
york, plane, josh, wallet, road, butt
Topic #16: virus, coronavirus, wine, muller, vaccin, presid, river, state,
virologist, quarantin, mask, request, distanc, yesterday, morn
Topic #17: appl, cider, flavor, vanilla, butter, fruit, orang, juic, candi,
prohibit, seed, accuraci, coffe, drink, color
Topic #18: byron, shelley, perci, vampir, jane, book, poetri, frankenstein,
mari, corinn, florenc, monster, woman, godwin, switzerland
Topic #19: empir, emperor, armi, khan, battl, citi, china, tang, shah, peter,
chinggi, genghi, war, jahangir, greec
Topic #20: fund, max, membership, drive, member, support, cat, chivalri, join,
idiom, war, hors, bonus, phrase, month
Topic #21: dip, sauc, chip, barbecu, ranch, salsa, pizza, cracker, mustard,
ketchup, onion, queso, soy, spinach, condiment
Topic #22: rudi, glitter, tank, toad, gax, patrick, ha, blink, bridget, road,
partner, caesar, traffic, frisbe, powder
Topic #23: pyramid, stone, tomb, stoneheng, site, egypt, ancient, archaeolog,
sarah, imhotep, homo, languag, maria, corner, age
Topic #24: henri, franc, eleanor, josephin, richard, loui, crusad, england,
champagn, marriag, woman, joan, king, catherin, son
```

[107]:

PreparedData(topic_coordinates=

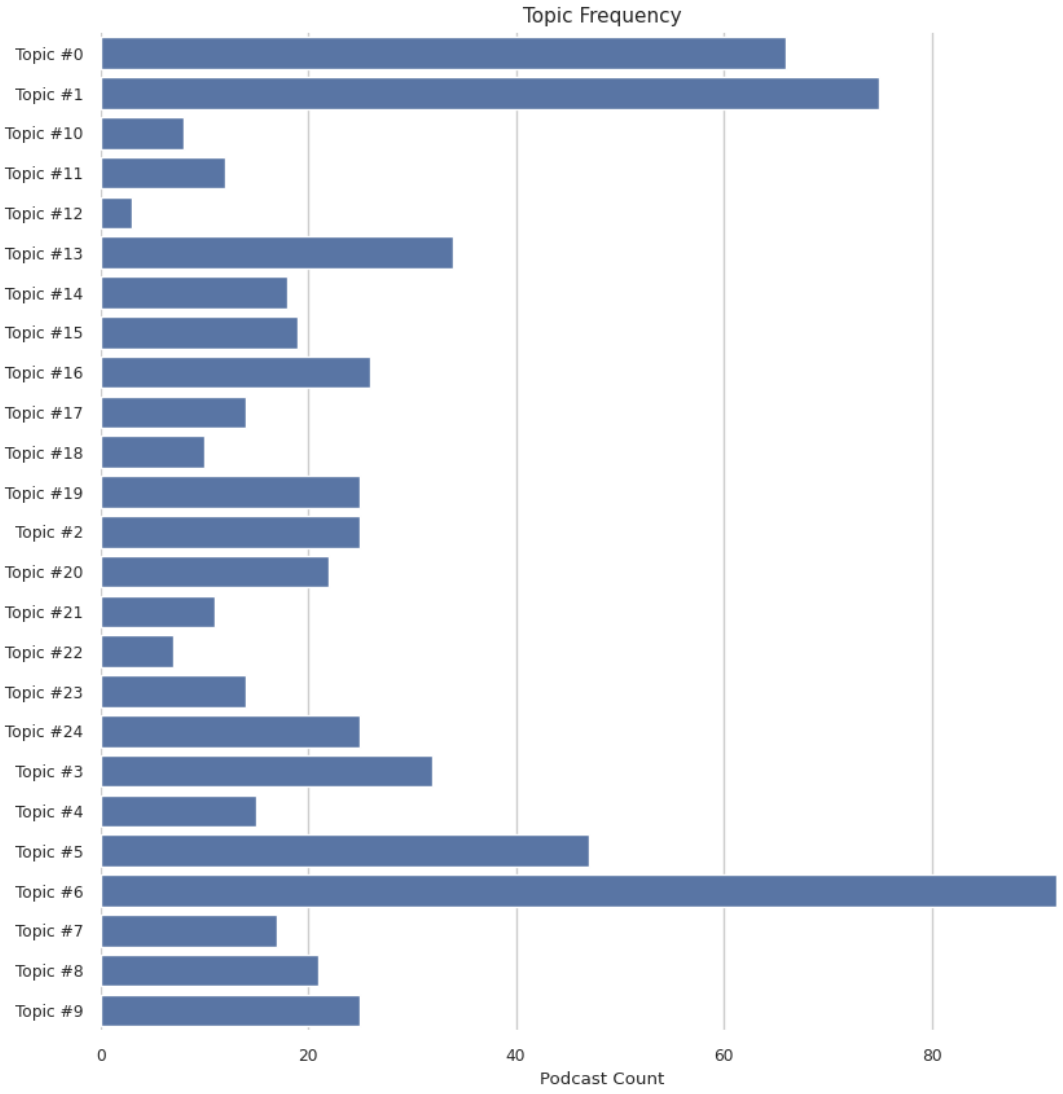
topic	x	y	topics	cluster	Freq
1	0.016818	0.190128	1	1	11.015195
6	0.027092	0.013531	2	1	9.400785
0	-0.070191	0.009866	3	1	6.478132
13	-0.017873	0.043957	4	1	6.189728
5	0.005051	0.005552	5	1	5.746152
16	0.008173	0.217318	6	1	4.626240
14	0.026855	0.208887	7	1	4.395785
3	-0.185345	-0.100807	8	1	4.242605
9	-0.030963	0.015015	9	1	4.085048

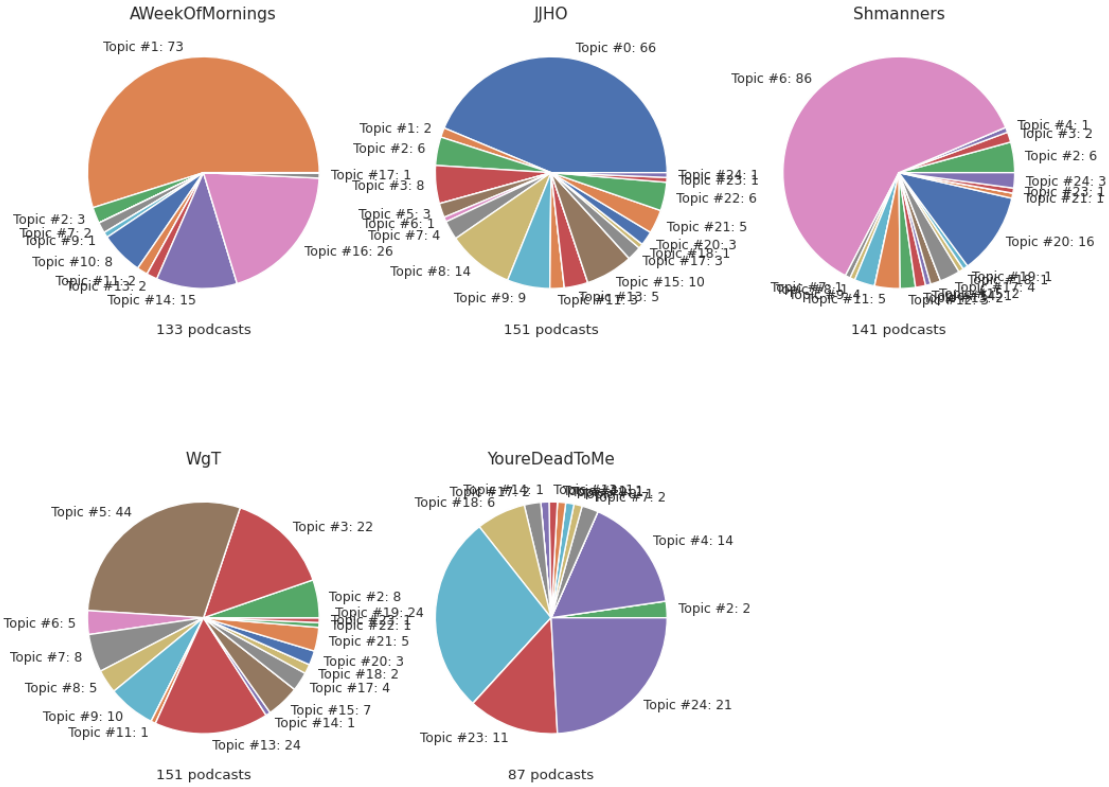
```
15    -0.055602  0.029013      10        1   3.905620
24     0.227593 -0.089500      11        1   3.815745
20    -0.017251 -0.008037      12        1   3.720854
2     -0.053363 -0.034706      13        1   3.405819
19     0.222393 -0.093768      14        1   3.323175
8     -0.134234 -0.094672      15        1   3.313493
11    -0.067870 -0.021788      16        1   3.224978
23     0.185133 -0.088310      17        1   2.822538
4      0.229208 -0.018118      18        1   2.641590
17    -0.096120 -0.061507      19        1   2.471707
7     -0.107055 -0.085324      20        1   2.406885
21    -0.182003 -0.076836      21        1   2.093016
18     0.163272 -0.085398      22        1   2.049758
10     0.007057  0.214949      23        1   1.982558
22    -0.067878  0.003011      24        1   1.524073
12    -0.032896 -0.092454      25        1   1.118519, topic_info=
Term       Freq      Total Category  logprob  loglift
4933    christma  16.000000  16.000000  Default  30.0000  30.0000
4877      chocol  13.000000  13.000000  Default  29.0000  29.0000
26068        tea   9.000000   9.000000  Default  28.0000  28.0000
6255       cream  13.000000  13.000000  Default  27.0000  27.0000
12801        ice  11.000000  11.000000  Default  26.0000  26.0000
…          …         …        …      …      …      …
19542      peter   0.345298   4.071594  Topic25  -5.6412   2.0258
20326    practic   0.311026   2.322589  Topic25  -5.7457   2.4826
3192   breakfast   0.336281   5.428301  Topic25  -5.6677   1.7117
14892       leaf   0.265111   1.257240  Topic25  -5.9055   2.9366
16315        max   0.268740   6.296801  Topic25  -5.8919   1.3391

[2198 rows x 6 columns], token_table=       Topic      Freq      Term
term
57         1  0.633853     abort
98        10  0.381033    accent
101        1  0.326625    access
101       12  0.326625    access
119       19  1.000817  accuraci
…      …        …       …
28990      5  0.140658      york
28990     10  0.140658      york
29100      9  1.327650     zelda
29124     14  1.283070     zhang
29169      2  0.907465      zola

[1132 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 7, 1, 14, 6, 17, 15, 4, 10, 16, 25, 21, 3, 20,
9, 12, 24, 5, 18, 8, 22, 19, 11, 23, 13])
```

[108]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: dog, docket, main, dracula, letter, book, cat, disput, mom, rule,
justic, evid, pizza, birthday, child
Topic #1: wine, river, bank, communiti, confer, congressman, farm, state, march,
yesterday, telescop, hunger, weekend, bill, committe
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, ghost, carol, tree, villain, song
Topic #3: breakfast, bread, meal, soup, meat, chicken, dinner, sandwich, eat,
buffet, cook, cooki, dessert, taco, burger
Topic #4: coloni, slaveri, war, revolut, louvertur, ship, mayflow, island,
haiti, popul, franc, weston, carver, freedom, trade
Topic #5: film, charact, star, disney, list, watch, tv, action, scene, mission,
book, jedi, version, jason, trilog
Topic #6: doordash, etiquett, app, parti, host, book, advic, art, hair,
deliveri, busi, slash, mcelroy, photographi, societi
Topic #7: chocol, cocoa, butter, candi, milk, peanut, bar, cacao, almond, kat,
easter, sugar, quaker, pot, cadburi
Topic #8: cream, ice, cake, pie, bagel, vanilla, chocol, rainer, eat, birthday,
milk, cone, flavor, carlson, snack
Topic #9: game, footbal, sport, playstat, basebal, island, barker, golf, ball,
tenni, pursuit, trivia, monopoli, bar, edit
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton,
presid, camp, pizza, spooner, russia, oak
Topic #11: water, beer, bathroom, drink, logger, pie, tast, hair, river, swim,
glass, toilet, choir, smell, sarah
Topic #12: tea, ceremoni, matcha, cup, afternoon, water, coffe, caffein, sugar,
dinner, saucer, milk, astheticist, doordash, birthday
Topic #13: song, danc, album, rock, band, list, soundtrack, audienc, roll, vega,
boom, citi, guitar, disney, jeff
Topic #14: vote, elect, presid, trump, wine, state, countri, democraci,
congressman, book, yesterday, hampton, candi, offic, morn
Topic #15: car, drive, theater, subaru, baja, transmiss, chip, phone, citi,
york, plane, josh, wallet, road, butt
Topic #16: virus, coronavirus, wine, muller, vaccin, presid, river, state,
virologist, quarantin, mask, request, distanc, yesterday, morn
Topic #17: appl, cider, flavor, vanilla, butter, fruit, orang, juic, candi,
prohibit, seed, accuraci, coffe, drink, color
Topic #18: byron, shelley, perci, vampir, jane, book, poetri, frankenstein,
mari, corinn, florenc, monster, woman, godwin, switzerland
Topic #19: empir, emperor, armi, khan, battl, citi, china, tang, shah, peter,
chinggi, genghi, war, jahangir, greec
Topic #20: fund, max, membership, drive, member, support, cat, chivalri, join,
idiom, war, hors, bonus, phrase, month
Topic #21: dip, sauc, chip, barbecu, ranch, salsa, pizza, cracker, mustard,
```

ketchup, onion, queso, soy, spinach, condiment
Topic #22: rudi, glitter, tank, toad, gax, patrick, ha, blink, bridget, road, partner, caesar, traffic, frisbe, powder
Topic #23: pyramid, stone, tomb, stoneheng, site, egypt, ancient, archaeolog, sarah, imhotep, homo, languag, maria, corner, age
Topic #24: henri, franc, eleanor, josephin, richard, loui, crusad, england, champagn, marriag, woman, joan, king, catherin, son





[109]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 25 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: car, dog, cat, evid, justic, rule, birthday, season, disput, dracula,
turkey, mom, docket, courtroom, main
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
peanut, cacao, bar, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, movi christma, tradit, santa claus, song, villain,
hanukkah, ghost
Topic #4: louvertur, revolut, slaveri, haiti, franc, island, napoleon, coloni,
war, freedom, virtu, plantat, presid, emperor, power
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version,
york, car, theater, scene, theme song, movi yeah
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, water,
vanilla, pie, rainer, bagel, cake cake, cone, eat
```

```
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, app, etiquett, book, host, hair, parti, deliveri, art,
advic, fashion, slash, code, mcelroy, busi
Topic #9: dip, bread, meat, sauc, pizza, sandwich, chicken, burger, beef,
flavor, eat, mustard, butter, barbecu, soup
Topic #10: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni,
afternoon, cup, tea kind, beer, coffe, drink, sugar, yeah tea
Topic #11: byron, shelley, perci, vampir, jane, frankenstein, poetri, book,
mari, corinn, florenc, woman, monster, switzerland, godwin
Topic #12: food bank, bank, march, march food, hunger, wine, food insecur,
communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: game, footbal, game show, golf, playstat, tenni, game game, sport,
ball, basebal, croquet, island, pizza, monopoli, barker
Topic #14: josephin, danc, baker, josephin baker, franc, pari, dancer, cheetah,
venus, venus venus, danc floor, jazz, banana, chorus, danc danc
Topic #15: battl, armi, empir, greec, olymp, democraci, fleet, salami, oracl,
greek, chariot, citi, battl salami, michael, water
Topic #16: homo, languag, california man, stone, stone age, tim, site, ice age,
extinct, fossil, dna, stoneheng, age, fossil record, centuri
Topic #17: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget,
road, gax, ha, guy gax, ha ha, frisbe, road toad
Topic #18: mayflow, coloni, ship, weston, carver, misha, voyag, popul,
settlement, invest, england, jane town, plymouth, merchant, dutch
Topic #19: henri, eleanor, franc, crusad, richard, loui, champagn, england, son,
king, marriag, pope, matilda, count champagn, woman
Topic #20: pyramid, tomb, stone, sarah, egypt, imhotep, ancient, maria,
archaeolog, giza, mera, kingdom, stoneheng, copper, mummi
Topic #21: khan, empir, shah, peter, chinggi, genghi, jahangir, jahan, shah
jahan, genghi khan, mogul, mongol, phil, akbar, chinggi khan
Topic #22: fund, max, max fund, membership, fund drive, chivalri, drive, member,
support, hors, join, cat, bonus, idiom, war
Topic #23: breakfast, cooki, dinner, buffet, meal, monster, candi, toast, eat,
dessert, bar, soup, breakfast breakfast, plate, cracker
Topic #24: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion,
zhang, guifei, rice, wu, yang guifei, empir, power
```

[109]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1      0.083085 -0.233295       1        1  18.112110
8      0.045834  0.030512       2        1  11.550853
5      0.082298  0.026924       3        1  11.184040
0      0.133641  0.068496       4        1   8.702469
13     0.112019 -0.004063       5        1   5.164293
9      0.172470  0.114023       6        1   5.137471
23     0.184545  0.118677       7        1   3.708062
```

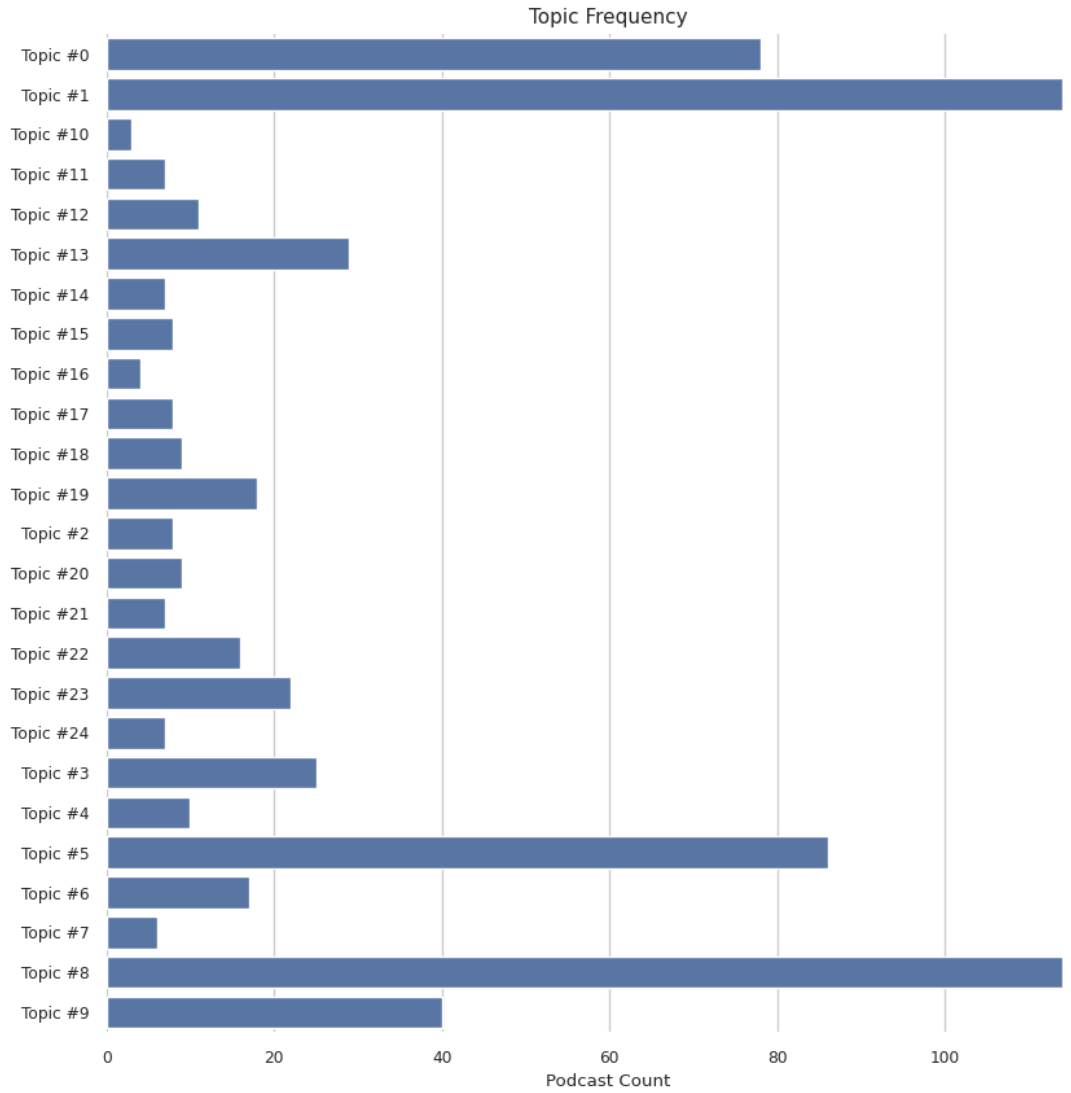
```
3      0.108013 -0.010918       8        1   3.562700
6      0.168314  0.114910       9        1   3.255830
12     0.107765 -0.261443      10        1   2.940995
22     0.086207  0.042022      11        1   2.930801
19    -0.188456  0.045878      12        1   2.523521
7      0.075043 -0.270500      13        1   2.080268
11    -0.115666  0.052002      14        1   1.964667
2      0.137210 -0.001594      15        1   1.900095
15    -0.181458  0.018961      16        1   1.843251
18    -0.198888  0.009060      17        1   1.807582
20    -0.175099  0.036510      18        1   1.757531
21    -0.188230  0.026609      19        1   1.637214
14    -0.108160 -0.029757      20        1   1.534103
4     -0.207427 -0.048373      21        1   1.525209
24    -0.170616 -0.010222      22        1   1.438321
16    -0.164337  0.028485      23        1   1.374765
17     0.109871  0.044004      24        1   1.235126
10     0.092021  0.093091      25        1   1.128721, topic_info=
Term      Freq     Total Category  logprob  loglift
84847    christma  8.000000  8.000000  Default  30.0000  30.0000
83347      chocol  6.000000  6.000000  Default  29.0000  29.0000
240891  ice cream  6.000000  6.000000  Default  28.0000  28.0000
116310      cream  7.000000  7.000000  Default  27.0000  27.0000
481205        tea  4.000000  4.000000  Default  26.0000  26.0000
…           …       …       …       …       …       …
363988      plant  0.213823  1.662809  Topic25  -6.9942   2.4330
76364    champagn  0.215380  2.151067  Topic25  -6.9869   2.1828
309840       milk  0.218554  3.174960  Topic25  -6.9723   1.8081
143407   doordash  0.166476  2.620092  Topic25  -7.2445   1.7280
201734      glass  0.165342  2.368841  Topic25  -7.2513   1.8219

[2490 rows x 6 columns], token_table=        Topic      Freq           Term
term
2599        3  0.509169         action
3021        3  0.526927          actor
5274        2  0.477800          advic
6195       25  1.288346  afternoon tea
6286        2  0.293955            age
…       …        …             …
553250      1  0.234776           york
553250      2  0.234776           york
553250      3  0.234776           york
554534      2  1.816814  zero deliveri
554602     22  1.342103          zhang

[594 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 1, 14, 10, 24, 4, 7, 13, 23, 20, 8, 12, 3,
```

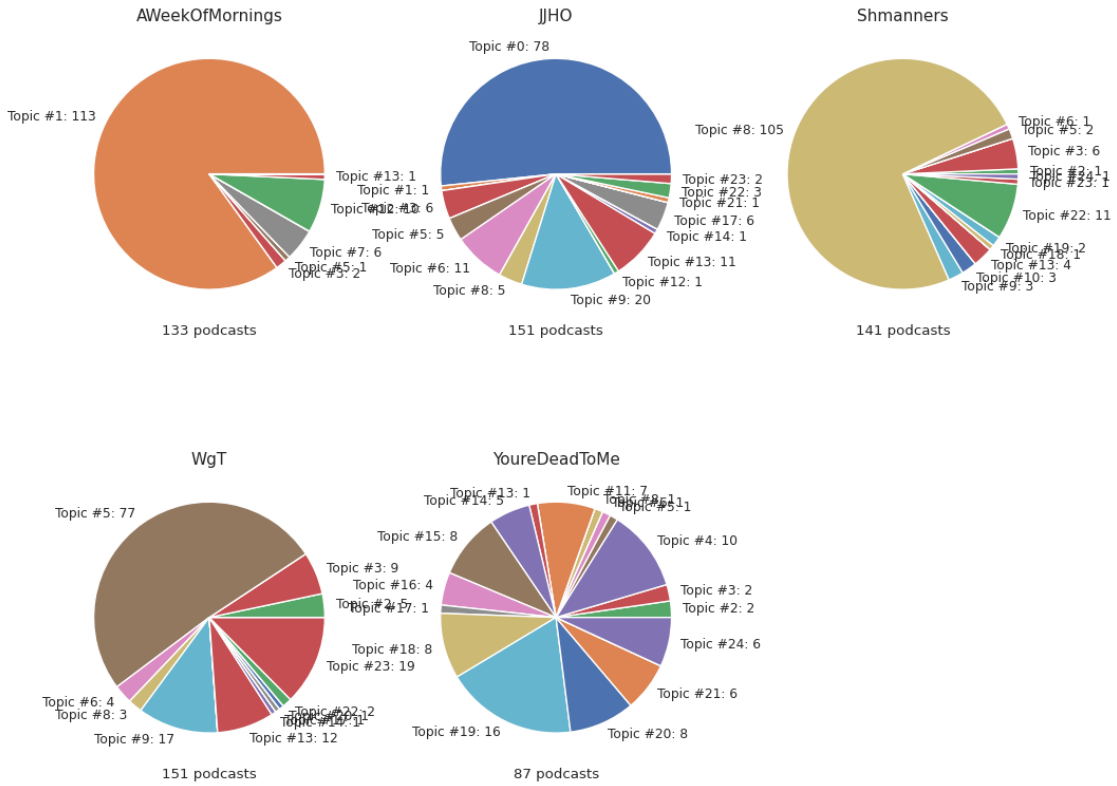
16, 19, 21, 22, 15, 5, 25, 17, 18, 11])

[110]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: car, dog, cat, evid, justic, rule, birthday, season, disput, dracula,
turkey, mom, docket, courtroom, main
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, radio, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
peanut, cacao, bar, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, movi christma, tradit, santa claus, song, villain,
hanukkah, ghost
Topic #4: louvertur, revolut, slaveri, haiti, franc, island, napoleon, coloni,
war, freedom, virtu, plantat, presid, emperor, power
Topic #5: song, film, charact, list, disney, star, movi movi, watch, version,
york, car, theater, scene, theme song, movi yeah
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, water,
vanilla, pie, rainer, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, app, etiquett, book, host, hair, parti, deliveri, art,
advic, fashion, slash, code, mcelroy, busi
Topic #9: dip, bread, meat, sauc, pizza, sandwich, chicken, burger, beef,
flavor, eat, mustard, butter, barbecu, soup
Topic #10: tea, tea tea, water, afternoon tea, matcha, ceremoni, tea ceremoni,
afternoon, cup, tea kind, beer, coffe, drink, sugar, yeah tea
Topic #11: byron, shelley, perci, vampir, jane, frankenstein, poetri, book,
mari, corinn, florenc, woman, monster, switzerland, godwin
Topic #12: food bank, bank, march, march food, hunger, wine, food insecur,
communiti, insecur, confer, bank food, nutrit, food food, center, counti
Topic #13: game, footbal, game show, golf, playstat, tenni, game game, sport,
ball, basebal, croquet, island, pizza, monopoli, barker
Topic #14: josephin, danc, baker, josephin baker, franc, pari, dancer, cheetah,
venus, venus venus, danc floor, jazz, banana, chorus, danc danc
Topic #15: battl, armi, empir, greec, olymp, democraci, fleet, salami, oracl,
greek, chariot, citi, battl salami, michael, water
Topic #16: homo, languag, california man, stone, stone age, tim, site, ice age,
extinct, fossil, dna, stoneheng, age, fossil record, centuri
Topic #17: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget,
road, gax, ha, guy gax, ha ha, frisbe, road toad
Topic #18: mayflow, coloni, ship, weston, carver, misha, voyag, popul,
settlement, invest, england, jane town, plymouth, merchant, dutch
Topic #19: henri, eleanor, franc, crusad, richard, loui, champagn, england, son,
king, marriag, pope, matilda, count champagn, woman
```

Topic #20: pyramid, tomb, stone, sarah, egypt, imhotep, ancient, maria, archaeolog, giza, mera, kingdom, stoneheng, copper, mummi
Topic #21: khan, empir, shah, peter, chinggi, genghi, jahangir, jahan, shah jahan, genghi khan, mogul, mongol, phil, akbar, chinggi khan
Topic #22: fund, max, max fund, membership, fund drive, chivalri, drive, member, support, hors, join, cat, bonus, idiom, war
Topic #23: breakfast, cooki, dinner, buffet, meal, monster, candi, toast, eat, dessert, bar, soup, breakfast breakfast, plate, cracker
Topic #24: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, zhang, guifei, rice, wu, yang guifei, empir, power





[111]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 25 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: game, footbal, golf, game show, tenni, playstat, sport, game game,
croquet, ball, pizza, basebal, barker, island, monopoli
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
communiti, morn, hampton, vote, bill, radio, elect
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut
butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, scroog, villain, santa claus, ghost, tradit,
place christma, krampus
Topic #4: emperor, henri, tang, eleanor, china, rebellion, franc, power, crusad,
richard, buddhism, loui, yang, england, dynasti
Topic #5: evid, car, dog, cat, birthday, turkey, rule, justic, season,
courtroom, slash, jacki, disput, mom, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, vanilla, water, rainer, pie, cone
```

```
Topic #7: cancer, cancer connect, connect, connect cancer, cancer connect
cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer
connect, support cancer, cancer cancer
Topic #8: doordash, app, etiquett, host, book, hair, fashion, parti, deliveri,
art, advic, slash, code, busi, mcelroy
Topic #9: bread, dip, meat, sauc, sandwich, pizza, chicken, soup, eat,
breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, mari,
florenc, book, corinn, woman, switzerland, godwin, dad
Topic #12: food bank, bank, march, march food, hunger, march food bank, wine,
food insecur, insecur, communiti, confer, food bank food, bank food, food food,
nutrit
Topic #13: louvertur, revolut, slaveri, haiti, franc, island, napoleon, coloni,
freedom, war, virtu, plantat, presid, marlena, toussaint
Topic #14: josephin, baker, josephin baker, danc, franc, pari, cheetah, venus,
dancer, venus venus, banana, chorus, jazz, chimpanze, grace
Topic #15: pyramid, stone, tomb, stoneheng, egypt, sarah, imhotep, ancient,
archaeolog, site, maria, giza, mera, kingdom, copper
Topic #16: battl, armi, empir, greec, olymp, fleet, democraci, salami, chariot,
greek, oracl, battl salami, citi, michael, ostrac
Topic #17: mayflow, coloni, ship, weston, misha, carver, voyag, popul, invest,
settlement, jane town, england, plymouth, merchant, dutch
Topic #18: khan, peter, chinggi, genghi, genghi khan, empir, phil, mongol,
chinggi khan, china, citi, silk, temujin, jin, peter pan
Topic #19: song, danc, album, band, rock, vega, list, boom boom, soundtrack,
audienc particip, rock roll, boom, jeff, boom boom boom, audienc
Topic #20: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget,
road, ha, ha ha ha, ha ha, gax, guy gax, road toad
Topic #21: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic,
butter, appl appl, cider cider, juic, drink, appl cider, glass
Topic #22: film, charact, star, list, movi movi, disney, watch, jason, scene,
version, car, movi yeah, action, mission, book
Topic #23: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal,
mahal, taj, delhi, ruler, india, timur, court
Topic #24: fund, max, max fund, membership, chivalri, fund drive, drive, hors,
member, support, join, max fund drive, cat, idiom, bonus
```

[111]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1      0.092614  0.229008       1        1  18.830239
8      0.030550 -0.087738       2        1  11.559934
5      0.132210 -0.129896       3        1  11.252194
22     0.086673 -0.067578       4        1   9.097538
9      0.167678 -0.082307       5        1   5.877924
```

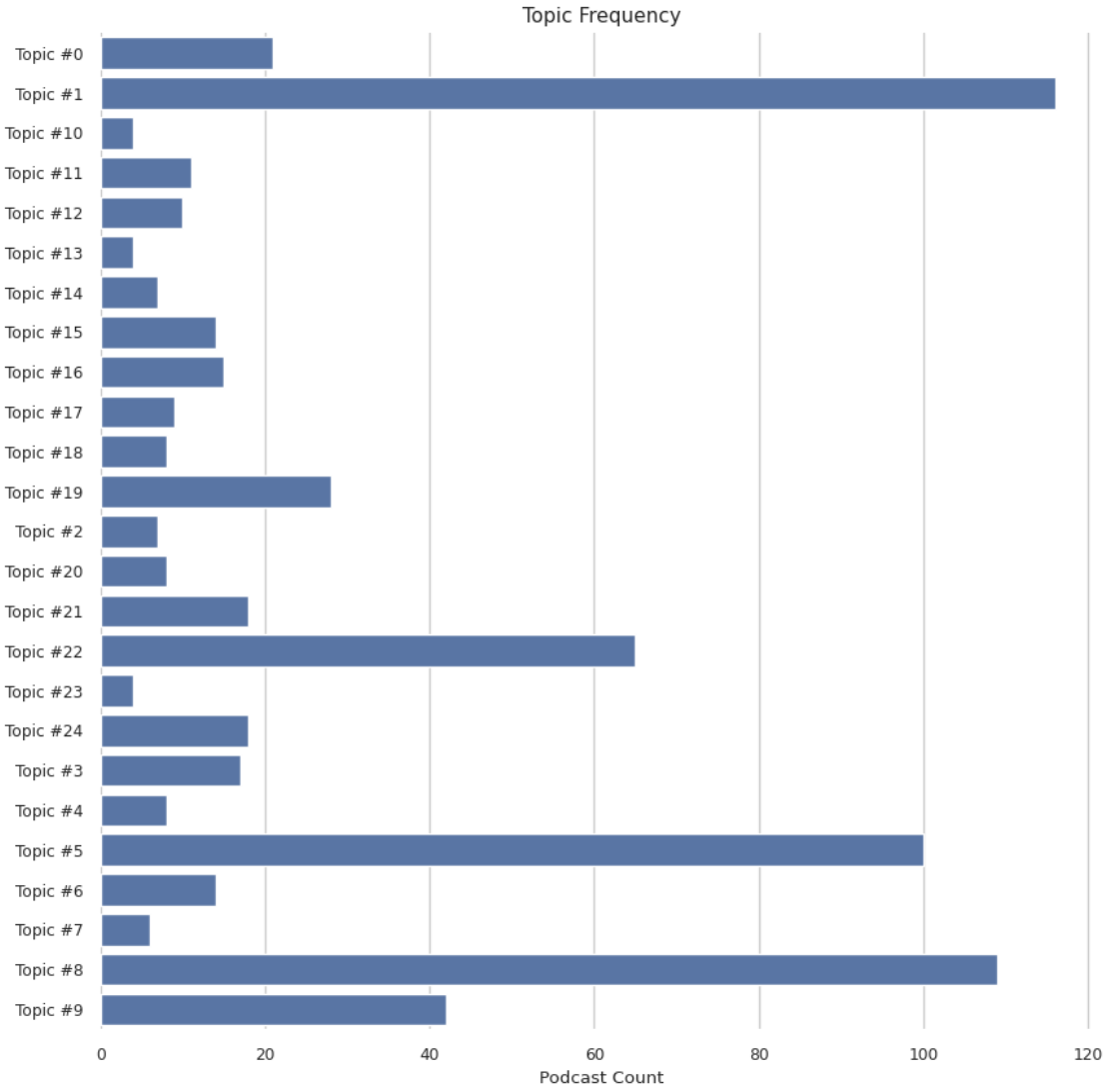
```
19     0.098939 -0.077112       6        1   4.997176
0      0.107785 -0.075362       7        1   3.340489
24     0.087883 -0.122379       8        1   3.122229
3      0.105820  0.049616       9        1   3.009320
6      0.159797 -0.046056      10        1   2.868106
21     0.139647  0.032416      11        1   2.789044
12     0.114477  0.279228      12        1   2.650168
11    -0.129431 -0.058716      13        1   2.183886
15    -0.207612  0.002399      14        1   2.131800
16    -0.218313  0.017995      15        1   2.034084
7      0.078656  0.212822      16        1   1.963419
2      0.130526  0.097999      17        1   1.787645
4     -0.241386 -0.017506      18        1   1.693404
17    -0.216153  0.012350      19        1   1.495255
20     0.100389 -0.120971      20        1   1.439336
18    -0.156578  0.004753      21        1   1.331663
10     0.090271 -0.067126      22        1   1.279386
14    -0.143583 -0.021390      23        1   1.171143
13    -0.227648  0.032464      24        1   1.118065
23    -0.183213  0.003088      25        1   0.976553, topic_info=
Term      Freq     Total Category  logprob  loglift
1141840        tea  4.000000  4.000000  Default  30.0000  30.0000
195860    christma  5.000000  5.000000  Default  29.0000  29.0000
192266      chocol  5.000000  5.000000  Default  28.0000  28.0000
566886   ice cream  4.000000  4.000000  Default  27.0000  27.0000
270440       cream  5.000000  5.000000  Default  26.0000  26.0000
…             …       …       …      …       …      …
264357       court  0.321451  2.844613  Topic25  -6.8088   2.4486
356977     emperor  0.256671  1.677795  Topic25  -7.0339   2.7515
943020      region  0.231503  1.395447  Topic25  -7.1371   2.8325
621212        khan  0.233878  1.708291  Topic25  -7.1269   2.6404
69613        battl  0.213976  2.312187  Topic25  -7.2158   2.2488

[2499 rows x 6 columns], token_table=         Topic      Freq           Term
term
5687         4  0.647065         action
6704         4  0.692120          actor
11712        2  0.608206          advic
13682       22  1.216786  afternoon tea
18624       25  1.491451          akbar
…        …       …              …
1294001      1  0.996273      word week
1321637      4  1.126766      yeah movi
1337670      1  0.715684      yesterday
1339625      1  0.287954           york
1339625      4  0.287954           york
```

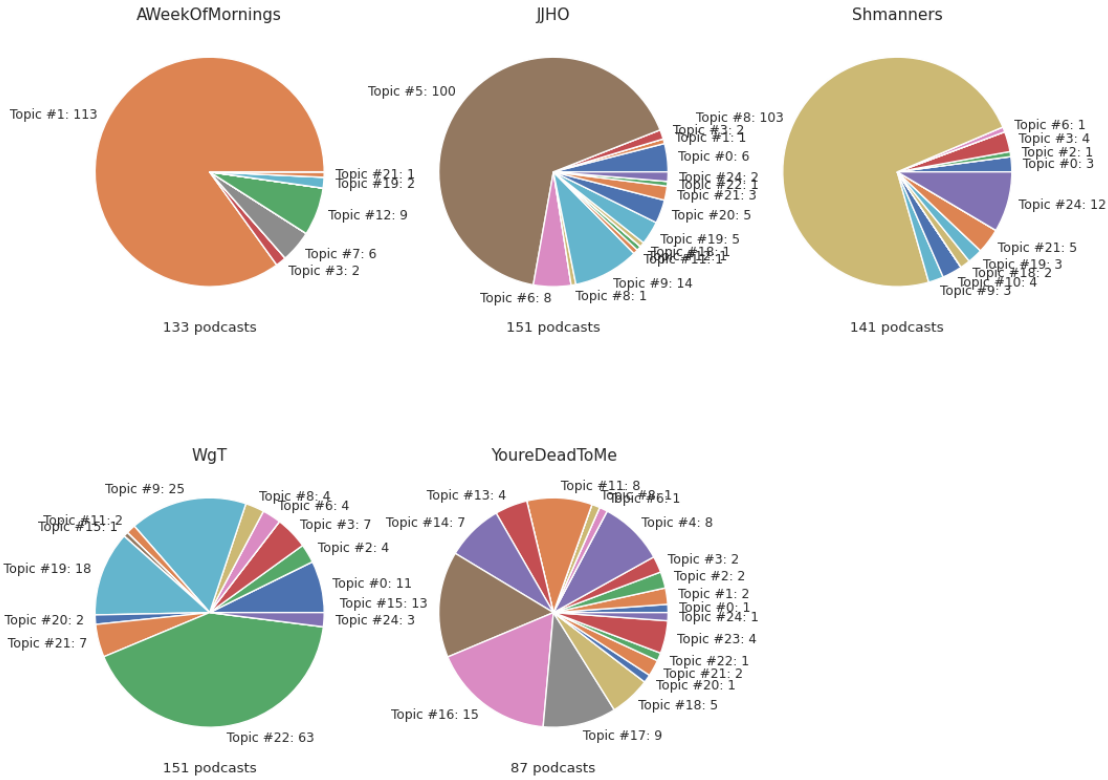
```
[451 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 23, 10, 20, 1, 25, 4, 7, 22, 13, 12, 16,
17, 8, 3, 5, 18, 21, 19, 11, 15, 14, 24])
```

[112]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: game, footbal, golf, game show, tenni, playstat, sport, game game,
croquet, ball, pizza, basebal, barker, island, monopoli
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
communiti, morn, hampton, vote, bill, radio, elect
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, cacao, bar, peanut
butter, kit kat, peanut, milk chocol, kat, cocoa butter, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, scroog, villain, santa claus, ghost, tradit,
place christma, krampus
Topic #4: emperor, henri, tang, eleanor, china, rebellion, franc, power, crusad,
richard, buddhism, loui, yang, england, dynasti
Topic #5: evid, car, dog, cat, birthday, turkey, rule, justic, season,
courtroom, slash, jacki, disput, mom, rise
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, vanilla, water, rainer, pie, cone
Topic #7: cancer, cancer connect, connect, connect cancer, cancer connect
cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer
connect, support cancer, cancer cancer
Topic #8: doordash, app, etiquett, host, book, hair, fashion, parti, deliveri,
art, advic, slash, code, busi, mcelroy
Topic #9: bread, dip, meat, sauc, sandwich, pizza, chicken, soup, eat,
breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, mari,
florenc, book, corinn, woman, switzerland, godwin, dad
Topic #12: food bank, bank, march, march food, hunger, march food bank, wine,
food insecur, insecur, communiti, confer, food bank food, bank food, food food,
nutrit
Topic #13: louvertur, revolut, slaveri, haiti, franc, island, napoleon, coloni,
freedom, war, virtu, plantat, presid, marlena, toussaint
Topic #14: josephin, baker, josephin baker, danc, franc, pari, cheetah, venus,
dancer, venus venus, banana, chorus, jazz, chimpanze, grace
Topic #15: pyramid, stone, tomb, stoneheng, egypt, sarah, imhotep, ancient,
archaeolog, site, maria, giza, mera, kingdom, copper
Topic #16: battl, armi, empir, greec, olymp, fleet, democraci, salami, chariot,
greek, oracl, battl salami, citi, michael, ostrac
Topic #17: mayflow, coloni, ship, weston, misha, carver, voyag, popul, invest,
settlement, jane town, england, plymouth, merchant, dutch
```

Topic #18: khan, peter, chinggi, genghi, genghi khan, empir, phil, mongol, chinggi khan, china, citi, silk, temujin, jin, peter pan
Topic #19: song, danc, album, band, rock, vega, list, boom boom, soundtrack, audienc particip, rock roll, boom, jeff, boom boom boom, audienc
Topic #20: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget, road, ha, ha ha ha, ha ha, gax, guy gax, road toad
Topic #21: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic, butter, appl appl, cider cider, juic, drink, appl cider, glass
Topic #22: film, charact, star, list, movi movi, disney, watch, jason, scene, version, car, movi yeah, action, mission, book
Topic #23: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal, mahal, taj, delhi, ruler, india, timur, court
Topic #24: fund, max, max fund, membership, chivalri, fund drive, drive, hors, member, support, join, max fund drive, cat, idiom, bonus





[113]:
```
topic_count = 30
```

[114]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'lda', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: lda , Number of Topics: 30 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: eat, meat, chicken, soup, bread, pizza, cook, meal, dip, sauc,
restaur, dinner, breakfast, burger, coffe
Topic #1: game, song, list, film, version, charact, watch, theme, topic, car,
kid, wait, audienc, star, god
Topic #2: car, citi, drive, song, york, bread, chicago, subaru, baja, vehicl,
flight, river, engin, street, transmiss
Topic #3: style, jazz, star, rammel, monologu, pulp, citi, scene, summer,
tarantino, fiction, theater, town, perform, hollywood
Topic #4: wine, river, state, presid, countri, morn, communiti, song, radio,
yesterday, book, bill, congressman, connect, support
Topic #5: bin, burton, seri, democraci, worm, scienc, watch, bee, beetlejuic,
lint, season, event, kendal, compost, action
Topic #6: edit, caesar, blame, julius, disney, kyle, roll, host, dragon, answer,
jack, line, pope, cow, figment
Topic #7: cancer, connect, beer, charact, hair, hand, star, car, bed, stoneheng,
```

josephin, stone, space, pyramid, busi
Topic #8: book, golf, vampir, byron, shelley, pizza, citi, vega, hotel, ball, prom, peter, ivan, parti, candi
Topic #9: book, tie, carniv, bow, sandler, pursuit, neck, poop, kid, moon, singer, hill, ibn, christma, column
Topic #10: book, dog, citi, cassi, cook, scott, spirit, candi, bodi, water, chees, beer, cowboy, pasta, medicin
Topic #11: war, woman, centuri, comedi, corner, power, radio, franc, god, period, greg, age, death, univers, author
Topic #12: tenni, news, sauc, england, game, town, ketchup, dog, valentin, birthday, heel, bill, communiti, island, fruit
Topic #13: smell, birthday, henri, celebr, fan, greg, franc, rudi, son, eleanor, jeremi, england, josh, power, denis
Topic #14: water, tea, game, system, tast, drink, bathroom, milk, afternoon, island, playstat, video, genesi, tarot, fruit
Topic #15: ride, roller, milk, disney, toilet, belt, water, land, disneyland, line, coaster, bagel, fuzz, park, air
Topic #16: danc, song, jeff, mason, storag, floor, roy, parti, art, cw, slide, york, fruit, harlem, renaiss
Topic #17: empir, battl, war, armi, clock, citi, water, gold, tapestri, power, frederick, tomb, court, centuri, ruler
Topic #18: york, accent, season, watch, bat, championship, aeta, jane, gift, smoke, david, car, basebal, seri, buffi
Topic #19: book, friend, dog, rule, slash, justic, cat, husband, evid, box, store, dad, mom, season, game
Topic #20: cream, ice, chocol, cake, pie, butter, flavor, candi, cooki, water, vanilla, eat, milk, fruit, banana
Topic #21: chair, stone, theater, peter, disney, marvel, pan, band, captain, iron, book, space, power, infin, soul
Topic #22: paint, chip, bog, car, door, richard, kate, ink, sprayer, letter, season, ballpoint, power, insur, gel
Topic #23: audienc, rock, song, particip, citi, emperor, china, peter, da, empir, tang, band, potter, phil, khan
Topic #24: glitter, rudi, watch, prohibit, rick, alcohol, jenga, jinger, toad, hepburn, parti, gift, beer, vote, drink
Topic #25: slash, host, check, busi, etiquett, art, pictur, parti, month, monster, advic, hand, max, exampl, code
Topic #26: christma, chocol, holiday, santa, cocoa, milk, winter, ghost, tradit, babi, club, hanukkah, tree, gift, bar
Topic #27: appl, cider, ham, lettuc, sandwich, juli, meat, snowbal, beef, fight, butter, deli, blood, roast, dracula
Topic #28: song, album, wood, band, paul, sound, record, porch, disney, soundtrack, phantom, guitar, georg, rock, heat
Topic #29: bank, communiti, countri, center, march, wine, space, mass, street, hunger, team, state, basebal, game, song

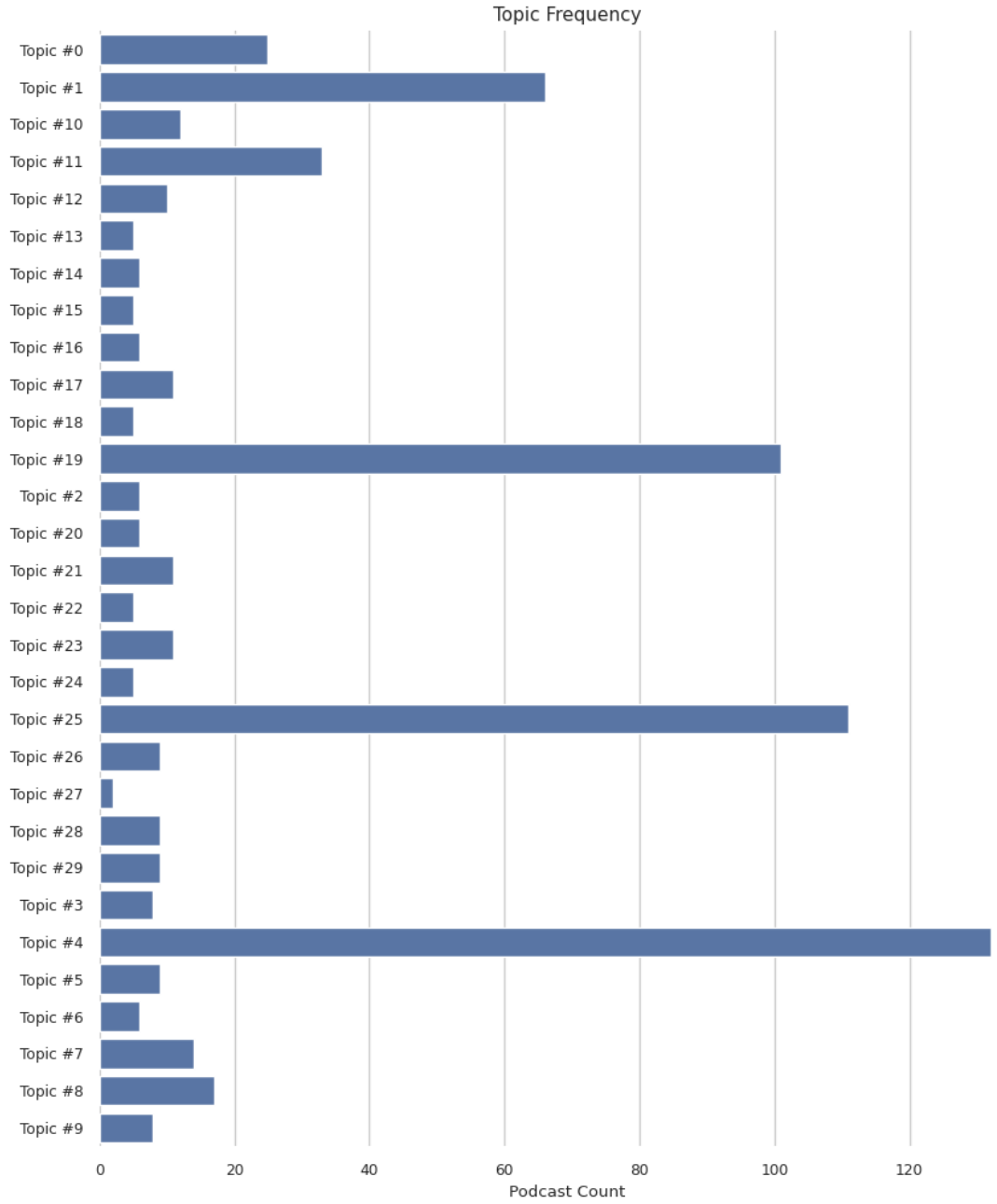
[114]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
4      0.164676 -0.018450       1        1  27.011666
19     0.237234  0.018472       2        1  11.697444
1      0.221607  0.013644       3        1  10.092052
25     0.213121 -0.016635       4        1   7.725746
11     0.185551 -0.094250       5        1   5.828725
29     0.087290 -0.017890       6        1   4.734932
0      0.109872  0.239825       7        1   3.947210
8      0.012910 -0.030530       8        1   2.286331
26    -0.015346  0.046432       9        1   2.228126
12    -0.007417 -0.007773      10        1   2.086772
7     -0.001589 -0.031506      11        1   1.916262
20    -0.032771  0.313168      12        1   1.818034
10    -0.040587  0.019372      13        1   1.770179
21     0.005044 -0.089559      14        1   1.440717
17     0.036855 -0.082831      15        1   1.330439
28    -0.040193 -0.051173      16        1   1.304806
3     -0.058979 -0.054496      17        1   1.237148
5     -0.045270 -0.040736      18        1   1.205157
14    -0.063013  0.116024      19        1   1.146967
2     -0.053093  0.032733      20        1   1.131914
23    -0.037548 -0.081864      21        1   1.105291
15    -0.114922 -0.015212      22        1   0.961291
18    -0.077654 -0.001621      23        1   0.895155
6     -0.092094 -0.072182      24        1   0.847288
13    -0.018131 -0.074404      25        1   0.839821
16    -0.109853 -0.039223      26        1   0.803170
9     -0.095303 -0.049979      27        1   0.793121
24    -0.116827  0.010487      28        1   0.725848
22    -0.093995 -0.027868      29        1   0.679160
27    -0.159574  0.088026      30        1   0.409226, topic_info=
Term         Freq        Total Category  logprob  loglift
24514      song  2394.000000  2394.000000  Default  30.0000  30.0000
2887       book  2315.000000  2315.000000  Default  29.0000  29.0000
4933   christma  1204.000000  1204.000000  Default  28.0000  28.0000
6255      cream   977.000000   977.000000  Default  27.0000  27.0000
28222     water  1728.000000  1728.000000  Default  26.0000  26.0000
…          …           …           …      …        …        …
5403       cold     8.154452    97.176445  Topic30  -5.7766   3.0207
23977    singer     9.087951   153.846796  Topic30  -5.6683   2.6697
11601      half    11.853750   965.614808  Topic30  -5.4026   1.0985
4317   catherin     8.827382   168.664967  Topic30  -5.6974   2.5486
27198    turkey     8.705569   378.414625  Topic30  -5.7112   1.7266

[2181 rows x 6 columns], token_table=       Topic      Freq        Term
```

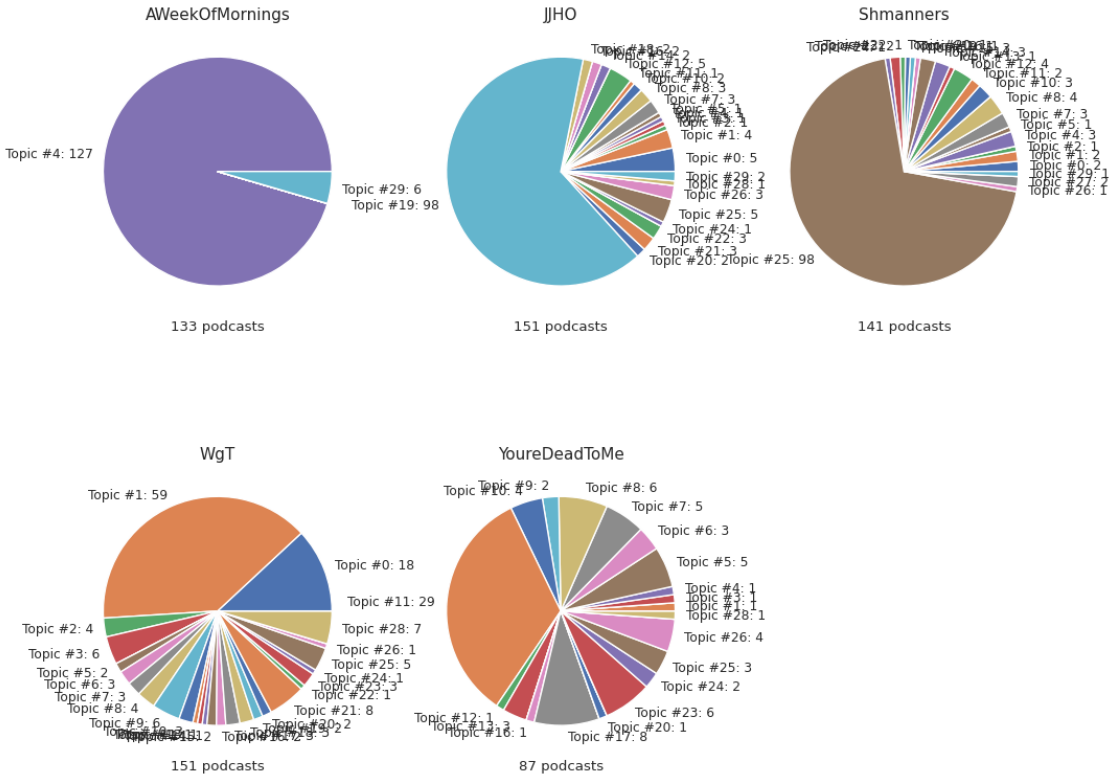
```
term
33        30  0.422215     abelsey
57         1  0.451831       abort
57         6  0.513445       abort
57        14  0.010269       abort
57        29  0.020538       abort
…      …       …          …
29155     28  0.128383     ziplock
29180     10  0.831751      zoomer
29180     22  0.083175      zoomer
29190      9  0.895707       zowel
29194     19  0.843702  zuckerberg

[9876 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[5, 20, 2, 26, 12, 30, 1, 9, 27, 13, 8, 21, 11, 22,
18, 29, 4, 6, 15, 3, 24, 16, 19, 7, 14, 17, 10, 25, 23, 28])
```

[115]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: eat, meat, chicken, soup, bread, pizza, cook, meal, dip, sauc,
restaur, dinner, breakfast, burger, coffe
Topic #1: game, song, list, film, version, charact, watch, theme, topic, car,
kid, wait, audienc, star, god
Topic #2: car, citi, drive, song, york, bread, chicago, subaru, baja, vehicl,
flight, river, engin, street, transmiss
Topic #3: style, jazz, star, rammel, monologu, pulp, citi, scene, summer,
tarantino, fiction, theater, town, perform, hollywood
Topic #4: wine, river, state, presid, countri, morn, communiti, song, radio,
yesterday, book, bill, congressman, connect, support
Topic #5: bin, burton, seri, democraci, worm, scienc, watch, bee, beetlejuic,
lint, season, event, kendal, compost, action
Topic #6: edit, caesar, blame, julius, disney, kyle, roll, host, dragon, answer,
jack, line, pope, cow, figment
Topic #7: cancer, connect, beer, charact, hair, hand, star, car, bed, stoneheng,
josephin, stone, space, pyramid, busi
Topic #8: book, golf, vampir, byron, shelley, pizza, citi, vega, hotel, ball,
prom, peter, ivan, parti, candi
Topic #9: book, tie, carniv, bow, sandler, pursuit, neck, poop, kid, moon,
singer, hill, ibn, christma, column
Topic #10: book, dog, citi, cassi, cook, scott, spirit, candi, bodi, water,
chees, beer, cowboy, pasta, medicin
Topic #11: war, woman, centuri, comedi, corner, power, radio, franc, god,
period, greg, age, death, univers, author
Topic #12: tenni, news, sauc, england, game, town, ketchup, dog, valentin,
birthday, heel, bill, communiti, island, fruit
```

Topic #13: smell, birthday, henri, celebr, fan, greg, franc, rudi, son, eleanor, jeremi, england, josh, power, denis
Topic #14: water, tea, game, system, tast, drink, bathroom, milk, afternoon, island, playstat, video, genesi, tarot, fruit
Topic #15: ride, roller, milk, disney, toilet, belt, water, land, disneyland, line, coaster, bagel, fuzz, park, air
Topic #16: danc, song, jeff, mason, storag, floor, roy, parti, art, cw, slide, york, fruit, harlem, renaiss
Topic #17: empir, battl, war, armi, clock, citi, water, gold, tapestri, power, frederick, tomb, court, centuri, ruler
Topic #18: york, accent, season, watch, bat, championship, aeta, jane, gift, smoke, david, car, basebal, seri, buffi
Topic #19: book, friend, dog, rule, slash, justic, cat, husband, evid, box, store, dad, mom, season, game
Topic #20: cream, ice, chocol, cake, pie, butter, flavor, candi, cooki, water, vanilla, eat, milk, fruit, banana
Topic #21: chair, stone, theater, peter, disney, marvel, pan, band, captain, iron, book, space, power, infin, soul
Topic #22: paint, chip, bog, car, door, richard, kate, ink, sprayer, letter, season, ballpoint, power, insur, gel
Topic #23: audienc, rock, song, particip, citi, emperor, china, peter, da, empir, tang, band, potter, phil, khan
Topic #24: glitter, rudi, watch, prohibit, rick, alcohol, jenga, jinger, toad, hepburn, parti, gift, beer, vote, drink
Topic #25: slash, host, check, busi, etiquett, art, pictur, parti, month, monster, advic, hand, max, exampl, code
Topic #26: christma, chocol, holiday, santa, cocoa, milk, winter, ghost, tradit, babi, club, hanukkah, tree, gift, bar
Topic #27: appl, cider, ham, lettuc, sandwich, juli, meat, snowbal, beef, fight, butter, deli, blood, roast, dracula
Topic #28: song, album, wood, band, paul, sound, record, porch, disney, soundtrack, phantom, guitar, georg, rock, heat
Topic #29: bank, communiti, countri, center, march, wine, space, mass, street, hunger, team, state, basebal, game, song


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #20
Topic #21
Topic #22
Topic #23
Topic #24
Topic #25
Topic #26
Topic #27
Topic #28
Topic #29
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count



[116]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'count')
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: count , Model: nmf , Number of Topics: 30 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: wine, presid, state, countri, morn, vote, trump, elect, yesterday,
communiti, congressman, hampton, bill, support, news
Topic #1: cake, birthday, celebr, greg, parti, eat, brandi, licoric, pie, plan,
kid, footbal, chang, august, surpris
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, salt, water, richard,
anni, bowl, vanilla, isi, recip
Topic #3: christma, holiday, ghost, santa, tradit, claus, list, villain, weapon,
version, gift, jacob, mom, marley, watch
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, oak, bedroom, presid, war
Topic #5: song, version, soundtrack, holiday, list, disney, album, rammel,
beauti, theme, voic, coffe, duh, tennesse, mention
Topic #6: game, host, system, bill, playstat, video, footbal, version, bar,
genesi, barker, tenni, ball, control, golf
Topic #7: water, bathroom, tast, sarah, drink, glass, beer, kitchen, sister,
differ, ami, ice, migrain, basement, experi
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink,
```

```
vanilla, coffe, war, sugar, tast
Topic #9: book, kid, child, moon, rabbit, read, seri, pooh, lesson, dog, chees,
page, version, friend, art
Topic #10: war, centuri, woman, power, comedi, corner, franc, radio, period,
empir, son, emperor, god, death, england
Topic #11: york, film, list, watch, manhattan, version, citi, park, accent,
jackson, bobo, street, hall, griffin, god
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, coffe, max, servic,
eat, membership, drink, saucer, dress
Topic #13: car, drive, baja, subaru, chip, transmiss, door, truck, engin,
outback, sophi, road, vehicl, river, insur
Topic #14: audienc, particip, song, da, rock, perform, list, experi, carolin,
bar, concert, shout, whoa, artist, danc
Topic #15: charact, star, watch, film, list, scene, seri, actor, action,
version, god, background, season, fan, perform
Topic #16: river, wine, radio, festiv, state, yesterday, weekend, station, farm,
street, bill, morn, massachusett, countri, area
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid,
russel, wait, space, deck, director
Topic #18: citi, chicago, vega, town, hollywood, coffe, pretzel, street, state,
duh, los, philadelphia, empir, battl, shape
Topic #19: busi, hand, card, soul, jason, handshak, hug, situat, code,
blockchain, phone, market, shake, doordash, pocket
Topic #20: bank, march, communiti, wine, hunger, center, monday, street, team,
congressman, morn, insecur, mass, state, store
Topic #21: style, jazz, main, pair, bed, theater, garbag, state, wear, pizza,
portland, stage, colleg, trio, countri
Topic #22: box, dog, season, cat, slash, evid, rule, justic, friend, store, dad,
support, fund, court, mom
Topic #23: pie, crust, butter, water, kumar, laila, freezer, smell, bowl, mom,
lila, freez, recip, roll, ice
Topic #24: ride, disney, peter, line, pan, land, park, roller, coaster, wait,
disneyland, wheel, dan, version, adventur
Topic #25: monster, cooki, street, costum, ghost, oscar, candi, vampir, version,
kid, halloween, grover, count, parti, sexi
Topic #26: danc, jeff, cw, floor, song, slide, parti, wed, dancer, state,
husband, colleg, cha, celebr, bird
Topic #27: bread, meat, eat, chicken, burger, pizza, sauc, sandwich, beef,
flavor, taco, cut, dip, restaur, soup
Topic #28: album, band, sound, rock, record, wilco, michael, countri, anthoni,
van, jeff, guitar, bass, paul, georg
Topic #29: stone, power, infin, space, soul, realiti, marvel, vision, mind,
mine, mcu, energi, war, captain, loki
```

[116]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
```

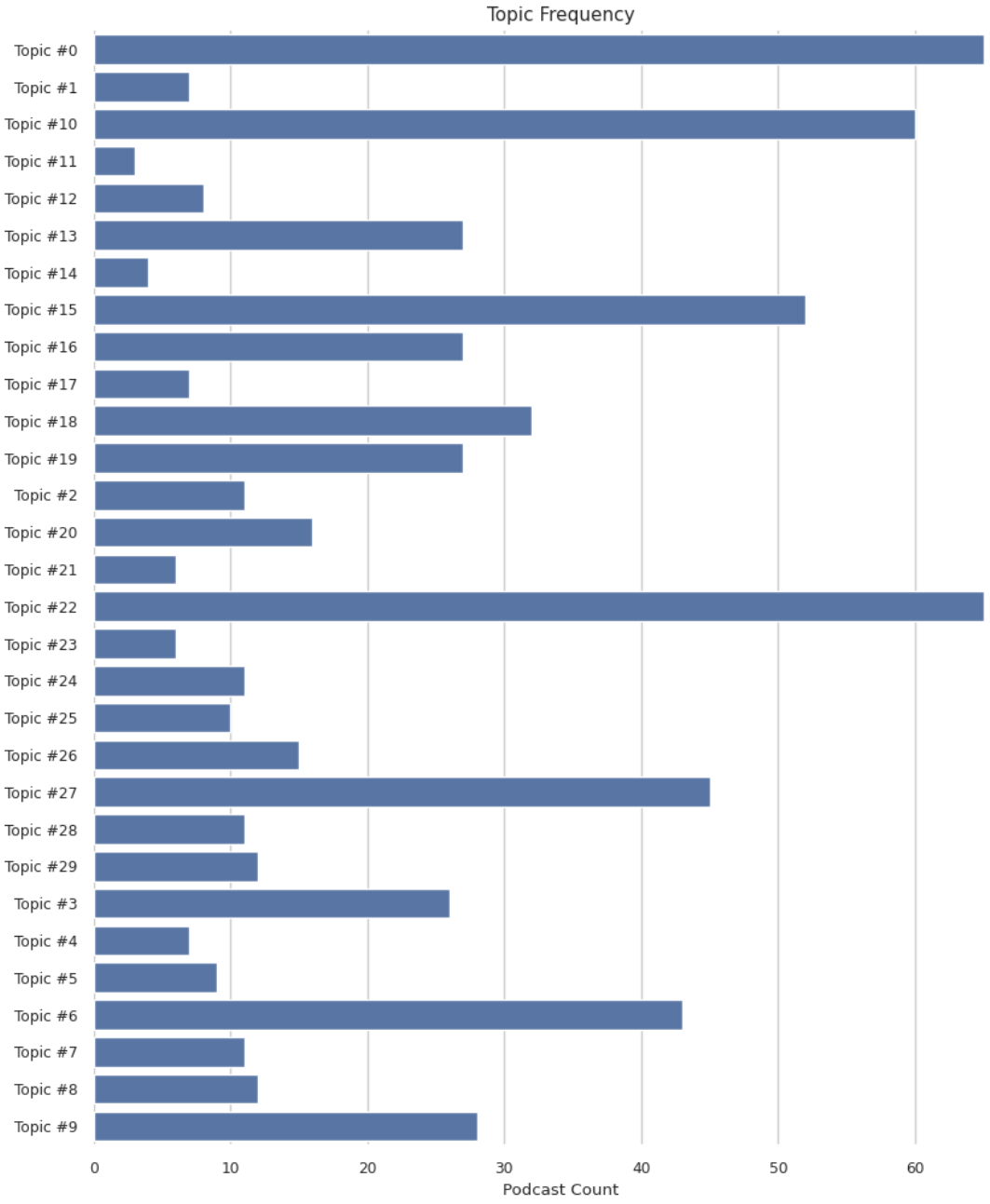
```
0     -0.114436  0.169650       1        1  7.659720
22     0.029544  0.031523       2        1  6.796961
15    -0.075104 -0.122226       3        1  5.755704
6     -0.019176 -0.054960       4        1  5.615437
10    -0.067299  0.053691       5        1  5.435973
16    -0.117353  0.163750       6        1  5.298809
9     -0.025333  0.003847       7        1  4.904215
27     0.167129 -0.035435       8        1  4.767890
20    -0.086973  0.191629       9        1  4.324635
19    -0.005278  0.010201      10        1  4.317963
13    -0.024351 -0.006941      11        1  3.894763
18    -0.048227  0.025615      12        1  3.596173
5     -0.071857 -0.161974      13        1  3.333880
28    -0.108003 -0.052642      14        1  3.079125
3      0.049044 -0.060986      15        1  2.672305
7      0.096930  0.025554      16        1  2.368563
2      0.218629  0.045560      17        1  2.352001
26    -0.075324 -0.017416      18        1  2.288290
4     -0.111894  0.196451      19        1  2.271911
25     0.032334 -0.109298      20        1  2.266206
29    -0.101981 -0.006078      21        1  2.225413
8      0.186105  0.092850      22        1  2.071486
24    -0.016645 -0.156785      23        1  1.911192
11    -0.060692 -0.066705      24        1  1.719221
21    -0.018302  0.033077      25        1  1.603889
12     0.162475  0.032326      26        1  1.562901
17    -0.008853 -0.072618      27        1  1.523729
14    -0.095485 -0.178039      28        1  1.517157
1      0.123058 -0.008242      29        1  1.469639
23     0.187318  0.034620      30        1  1.394850, topic_info=          Term
Freq        Total Category  logprob  loglift
24514      song  3240.000000  3240.000000  Default  30.0000  30.0000
4933   christma  2136.000000  2136.000000  Default  29.0000  29.0000
2887       book  3070.000000  3070.000000  Default  28.0000  28.0000
10295      game  3122.000000  3122.000000  Default  27.0000  27.0000
26068       tea  1622.000000  1622.000000  Default  26.0000  26.0000
…          …           …           …       …       …       …
25198     store    39.294959   910.463863  Topic30  -5.4304   1.1295
8291        eat    37.927092  1176.762368  Topic30  -5.4658   0.8375
19641     phone    33.494937   711.555210  Topic30  -5.5901   1.2163
13955    justic    30.483341   491.655434  Topic30  -5.6843   1.4918
25778    system    30.777140   764.093185  Topic30  -5.6747   1.0605

[2536 rows x 6 columns], token_table=       Topic      Freq      Term
term
3          7  1.016373  aaronson
4          6  0.821956  aassenha
```

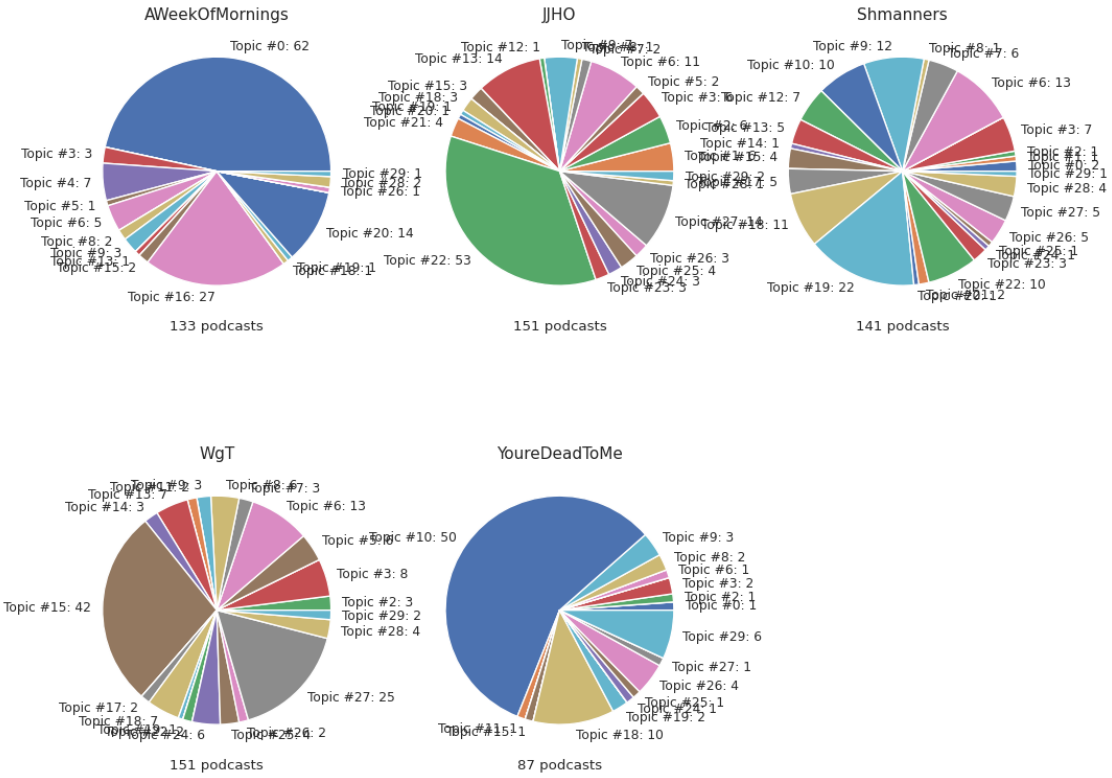
```
8          22  0.958290      abag
43          9  0.904121    abject
54          9  0.904121   abordin
…       …        …        …
29100      16  0.061975     zelda
29176       9  0.958640    zoolal
29178       9  0.958640    zoolol
29197       9  1.022550     zulal
29200       3  1.005826     zuvio

[11675 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[1, 23, 16, 7, 11, 17, 10, 28, 21, 20, 14, 19, 6,
29, 4, 8, 3, 27, 5, 26, 30, 9, 25, 12, 22, 13, 18, 15, 2, 24])
```

[117]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: wine, presid, state, countri, morn, vote, trump, elect, yesterday,
communiti, congressman, hampton, bill, support, news
Topic #1: cake, birthday, celebr, greg, parti, eat, brandi, licoric, pie, plan,
kid, footbal, chang, august, surpris
Topic #2: ice, cream, cake, chocol, centuri, milk, cone, salt, water, richard,
anni, bowl, vanilla, isi, recip
Topic #3: christma, holiday, ghost, santa, tradit, claus, list, villain, weapon,
version, gift, jacob, mom, marley, watch
Topic #4: cancer, connect, support, wine, bed, month, communiti, organ,
diagnosi, hampton, pino, oak, bedroom, presid, war
Topic #5: song, version, soundtrack, holiday, list, disney, album, rammel,
beauti, theme, voic, coffe, duh, tennesse, mention
Topic #6: game, host, system, bill, playstat, video, footbal, version, bar,
genesi, barker, tenni, ball, control, golf
Topic #7: water, bathroom, tast, sarah, drink, glass, beer, kitchen, sister,
differ, ami, ice, migrain, basement, experi
Topic #8: chocol, cocoa, milk, bar, butter, flavor, candi, round, cacao, drink,
vanilla, coffe, war, sugar, tast
Topic #9: book, kid, child, moon, rabbit, read, seri, pooh, lesson, dog, chees,
page, version, friend, art
Topic #10: war, centuri, woman, power, comedi, corner, franc, radio, period,
empir, son, emperor, god, death, england
Topic #11: york, film, list, watch, manhattan, version, citi, park, accent,
jackson, bobo, street, hall, griffin, god
Topic #12: tea, afternoon, water, cup, dinner, milk, sugar, coffe, max, servic,
eat, membership, drink, saucer, dress
Topic #13: car, drive, baja, subaru, chip, transmiss, door, truck, engin,
outback, sophi, road, vehicl, river, insur
Topic #14: audienc, particip, song, da, rock, perform, list, experi, carolin,
```

bar, concert, shout, whoa, artist, danc
Topic #15: charact, star, watch, film, list, scene, seri, actor, action,
version, god, background, season, fan, perform
Topic #16: river, wine, radio, festiv, state, yesterday, weekend, station, farm,
street, bill, morn, massachusett, countri, area
Topic #17: chair, sit, porch, reclin, boy, comfort, seat, ground, bodi, kid,
russel, wait, space, deck, director
Topic #18: citi, chicago, vega, town, hollywood, coffe, pretzel, street, state,
duh, los, philadelphia, empir, battl, shape
Topic #19: busi, hand, card, soul, jason, handshak, hug, situat, code,
blockchain, phone, market, shake, doordash, pocket
Topic #20: bank, march, communiti, wine, hunger, center, monday, street, team,
congressman, morn, insecur, mass, state, store
Topic #21: style, jazz, main, pair, bed, theater, garbag, state, wear, pizza,
portland, stage, colleg, trio, countri
Topic #22: box, dog, season, cat, slash, evid, rule, justic, friend, store, dad,
support, fund, court, mom
Topic #23: pie, crust, butter, water, kumar, laila, freezer, smell, bowl, mom,
lila, freez, recip, roll, ice
Topic #24: ride, disney, peter, line, pan, land, park, roller, coaster, wait,
disneyland, wheel, dan, version, adventur
Topic #25: monster, cooki, street, costum, ghost, oscar, candi, vampir, version,
kid, halloween, grover, count, parti, sexi
Topic #26: danc, jeff, cw, floor, song, slide, parti, wed, dancer, state,
husband, colleg, cha, celebr, bird
Topic #27: bread, meat, eat, chicken, burger, pizza, sauc, sandwich, beef,
flavor, taco, cut, dip, restaur, soup
Topic #28: album, band, sound, rock, record, wilco, michael, countri, anthoni,
van, jeff, guitar, bass, paul, georg
Topic #29: stone, power, infin, space, soul, realiti, marvel, vision, mind,
mine, mcu, energi, war, captain, loki


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #20
Topic #21
Topic #22
Topic #23
Topic #24
Topic #25
Topic #26
Topic #27
Topic #28
Topic #29
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
10
20
30
40
50
60
Podcast Count



[118]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 1))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 30 , tidf ngram range: (1, 1)
--------------------------------------------------------------------------------
Topic #0: dog, docket, dracula, main, letter, cat, disput, book, rule, justic,
evid, birthday, mom, dad, husband
Topic #1: wine, river, bank, communiti, confer, congressman, farm, march, state,
telescop, hunger, yesterday, weekend, massachusett, bill
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, carol, tree, villain, ghost, maci
Topic #3: dip, sauc, chip, ranch, cracker, salsa, barbecu, onion, snack, queso,
spinach, soy, mustard, restaur, hummus
Topic #4: emperor, empir, khan, china, tang, chinggi, shah, genghi, peter,
jahangir, jahan, buddhism, mongol, phil, dynasti
Topic #5: film, charact, star, list, tv, watch, action, mission, jedi, trilog,
scene, actor, monologu, seri, televis
Topic #6: doordash, etiquett, app, host, book, parti, advic, art, deliveri,
hair, busi, slash, mcelroy, photographi, societi
Topic #7: chocol, cocoa, butter, milk, peanut, candi, bar, cacao, kat, pot,
quaker, almond, sugar, cadburi, kit
```

```
Topic #8: cream, ice, cake, pie, bagel, chocol, vanilla, rainer, eat, milk,
cone, birthday, carlson, snack, flavor
Topic #9: game, footbal, sport, playstat, island, basebal, barker, golf, ball,
tenni, pursuit, monopoli, trivia, nintendo, pub
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton,
presid, camp, pizza, spooner, russia, oak
Topic #11: song, danc, vega, list, citi, york, soundtrack, tennesse, boom,
carter, duh, audienc, hit, slide, chicago
Topic #12: tea, ceremoni, matcha, cup, afternoon, water, coffe, caffein, sugar,
saucer, astheticist, champagn, milk, doordash, birthday
Topic #13: rock, album, band, roll, guitar, paul, concert, song, record, georg,
singer, wilco, jagger, toni, revolv
Topic #14: vote, elect, presid, trump, wine, state, countri, democraci,
congressman, book, yesterday, hampton, morn, offic, joe
Topic #15: car, drive, theater, subaru, baja, transmiss, chip, phone, josh,
road, wallet, plane, butt, seat, truck
Topic #16: water, beer, bathroom, drink, logger, pie, tast, river, hair, swim,
glass, choir, toilet, smell, sarah
Topic #17: virus, coronavirus, wine, muller, vaccin, presid, river, state,
virologist, quarantin, mask, request, distanc, yesterday, morn
Topic #18: byron, shelley, perci, vampir, jane, book, poetri, frankenstein,
mari, corinn, florenc, godwin, woman, switzerland, comedi
Topic #19: appl, cider, flavor, vanilla, fruit, orang, butter, prohibit, juic,
seed, accuraci, coffe, drink, chapman, candi
Topic #20: fund, max, membership, drive, member, chivalri, support, cat, join,
idiom, war, hors, bonus, phrase, month
Topic #21: breakfast, dinner, meal, buffet, cooki, toast, eat, dessert, janet,
plate, cracker, candi, topic, soup, snack
Topic #22: rudi, glitter, tank, toad, gax, patrick, ha, blink, bridget, road,
partner, caesar, traffic, frisbe, powder
Topic #23: costum, candi, monster, ghost, halloween, jason, parti, spirit, sexi,
easter, trick, hoot, spooki, cooki, skeleton
Topic #24: josephin, danc, baker, franc, pari, jazz, dancer, cheetah, fashion,
venus, banana, chorus, napoleon, jeff, perform
Topic #25: henri, franc, england, war, coloni, eleanor, revolut, slaveri, woman,
louvertur, richard, ship, power, loui, centuri
Topic #26: pyramid, stone, tomb, stoneheng, site, egypt, ancient, archaeolog,
sarah, imhotep, maria, homo, languag, corner, kingdom
Topic #27: bread, meat, sandwich, chicken, pizza, soup, cook, burger, beef,
mayonnais, butter, eat, lettuc, tomato, recip
Topic #28: disney, peter, ride, book, pan, cronk, walt, cooki, disneyland, edit,
jungl, charact, cinderella, lana, potter
Topic #29: battl, armi, empir, greec, olymp, democraci, fleet, chariot, greek,
salami, oracl, citi, war, michael, ostrac
```

[118]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
```

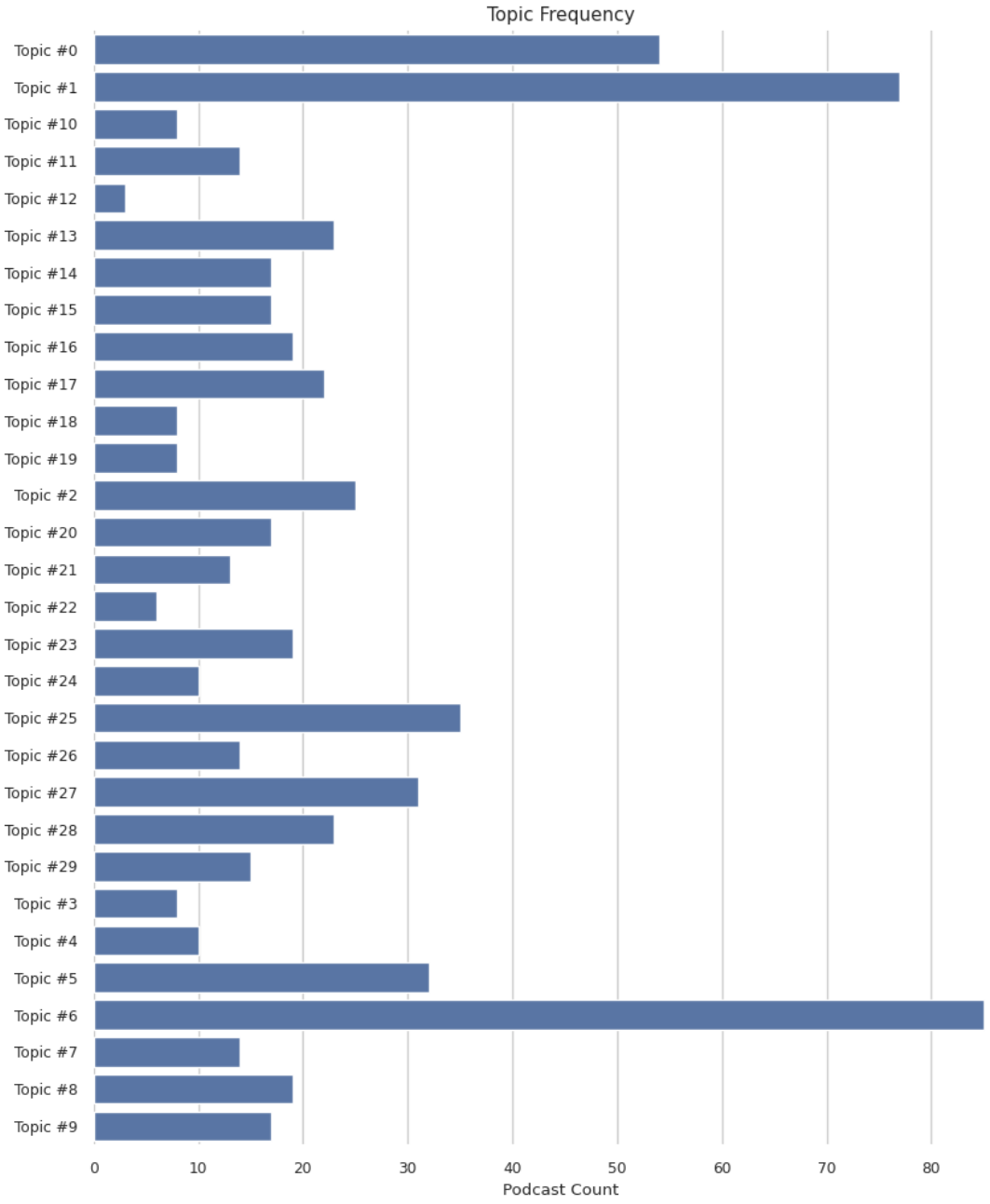
```
topic
1      0.029263  0.185095       1        1  10.503440
6      0.026398  0.028336       2        1   8.112020
0     -0.063068  0.022721       3        1   5.337463
25     0.234451 -0.060126       4        1   5.011505
5      0.009671  0.029244       5        1   4.305474
27    -0.170330 -0.107775       6        1   4.253242
13    -0.002922  0.090093       7        1   4.215128
17     0.020782  0.207537       8        1   3.996614
14     0.037201  0.208178       9        1   3.751159
16    -0.063089 -0.011812      10        1   3.540839
28    -0.001208 -0.037353      11        1   3.458227
15    -0.050291  0.032919      12        1   3.276103
2     -0.048300 -0.020094      13        1   3.220365
9     -0.030524  0.033626      14        1   3.124531
8     -0.133822 -0.097130      15        1   3.102517
20    -0.012166  0.005173      16        1   3.041394
23    -0.062680 -0.010502      17        1   2.767374
29     0.220239 -0.122686      18        1   2.742049
11    -0.029647  0.044700      19        1   2.669723
26     0.182866 -0.112859      20        1   2.527743
21    -0.161195 -0.085528      21        1   2.409465
24     0.147937  0.012605      22        1   2.104795
7     -0.100813 -0.081015      23        1   2.010433
10     0.021814  0.202673      24        1   1.889405
19    -0.098527 -0.041622      25        1   1.868063
18     0.153957 -0.066153      26        1   1.686624
4      0.216130 -0.108441      27        1   1.470537
3     -0.169470 -0.083831      28        1   1.319629
22    -0.058502  0.010838      29        1   1.293033
12    -0.044156 -0.066811      30        1   0.991108, topic_info=
Term       Freq      Total Category  logprob  loglift
4933   christma  15.000000  15.000000  Default  30.0000  30.0000
4877     chocol  12.000000  12.000000  Default  29.0000  29.0000
6255      cream  13.000000  13.000000  Default  28.0000  28.0000
26068       tea   8.000000   8.000000  Default  27.0000  27.0000
12801       ice  11.000000  11.000000  Default  26.0000  26.0000
…          …         …         …       …       …       …
7431     dinner   0.315001   5.692071  Topic30  -5.6121   1.7199
28428     wheel   0.245757   1.572542  Topic30  -5.8603   2.7580
19542     peter   0.253379   5.673834  Topic30  -5.8298   1.5054
13542     japan   0.230824   1.175465  Topic30  -5.9230   2.9863
16315       max   0.242229   5.685385  Topic30  -5.8748   1.4583

[2565 rows x 6 columns], token_table=       Topic      Freq      Term
term
57         1  0.649786     abort
```

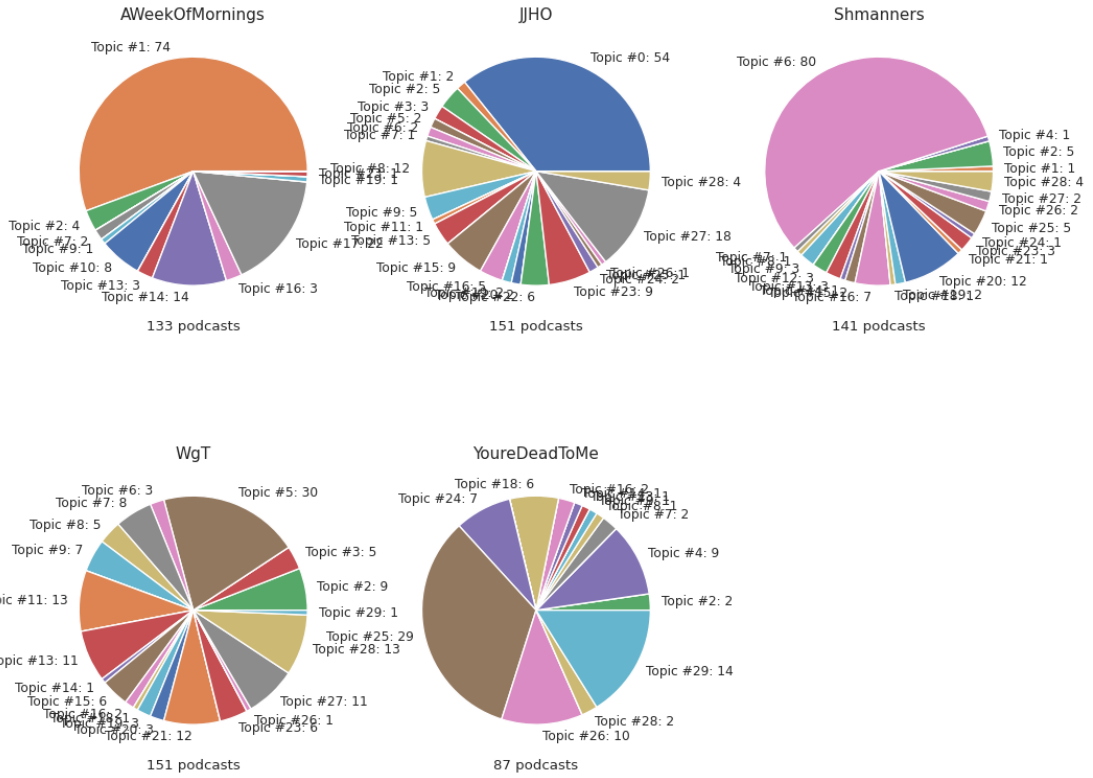
```
98        12  0.403791    accent
101        1  0.341397    access
101       16  0.341397    access
119       25  1.152892  accuraci
…      …       …       …
28990     12  0.149333      york
28990     19  0.149333      york
28990     22  0.149333      york
29100     14  1.561219     zelda
29122     18  1.184352      zeus

[1160 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 7, 1, 26, 6, 28, 14, 18, 15, 17, 29, 16, 3, 10,
9, 21, 24, 30, 12, 27, 22, 25, 8, 11, 20, 19, 5, 4, 23, 13])
```

[119]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: dog, docket, dracula, main, letter, cat, disput, book, rule, justic,
evid, birthday, mom, dad, husband
Topic #1: wine, river, bank, communiti, confer, congressman, farm, march, state,
telescop, hunger, yesterday, weekend, massachusett, bill
Topic #2: christma, holiday, santa, claus, scroog, tradit, gift, hanukkah,
winter, krampus, carol, tree, villain, ghost, maci
Topic #3: dip, sauc, chip, ranch, cracker, salsa, barbecu, onion, snack, queso,
spinach, soy, mustard, restaur, hummus
Topic #4: emperor, empir, khan, china, tang, chinggi, shah, genghi, peter,
jahangir, jahan, buddhism, mongol, phil, dynasti
Topic #5: film, charact, star, list, tv, watch, action, mission, jedi, trilog,
scene, actor, monologu, seri, televis
Topic #6: doordash, etiquett, app, host, book, parti, advic, art, deliveri,
hair, busi, slash, mcelroy, photographi, societi
Topic #7: chocol, cocoa, butter, milk, peanut, candi, bar, cacao, kat, pot,
quaker, almond, sugar, cadburi, kit
Topic #8: cream, ice, cake, pie, bagel, chocol, vanilla, rainer, eat, milk,
cone, birthday, carlson, snack, flavor
Topic #9: game, footbal, sport, playstat, island, basebal, barker, golf, ball,
tenni, pursuit, monopoli, trivia, nintendo, pub
Topic #10: cancer, connect, wine, pino, support, bed, diagnosi, huga, hampton,
presid, camp, pizza, spooner, russia, oak
Topic #11: song, danc, vega, list, citi, york, soundtrack, tennesse, boom,
carter, duh, audienc, hit, slide, chicago
Topic #12: tea, ceremoni, matcha, cup, afternoon, water, coffe, caffein, sugar,
saucer, astheticist, champagn, milk, doordash, birthday
Topic #13: rock, album, band, roll, guitar, paul, concert, song, record, georg,
singer, wilco, jagger, toni, revolv
```

Topic #14: vote, elect, presid, trump, wine, state, countri, democraci, congressman, book, yesterday, hampton, morn, offic, joe
Topic #15: car, drive, theater, subaru, baja, transmiss, chip, phone, josh, road, wallet, plane, butt, seat, truck
Topic #16: water, beer, bathroom, drink, logger, pie, tast, river, hair, swim, glass, choir, toilet, smell, sarah
Topic #17: virus, coronavirus, wine, muller, vaccin, presid, river, state, virologist, quarantin, mask, request, distanc, yesterday, morn
Topic #18: byron, shelley, perci, vampir, jane, book, poetri, frankenstein, mari, corinn, florenc, godwin, woman, switzerland, comedi
Topic #19: appl, cider, flavor, vanilla, fruit, orang, butter, prohibit, juic, seed, accuraci, coffe, drink, chapman, candi
Topic #20: fund, max, membership, drive, member, chivalri, support, cat, join, idiom, war, hors, bonus, phrase, month
Topic #21: breakfast, dinner, meal, buffet, cooki, toast, eat, dessert, janet, plate, cracker, candi, topic, soup, snack
Topic #22: rudi, glitter, tank, toad, gax, patrick, ha, blink, bridget, road, partner, caesar, traffic, frisbe, powder
Topic #23: costum, candi, monster, ghost, halloween, jason, parti, spirit, sexi, easter, trick, hoot, spooki, cooki, skeleton
Topic #24: josephin, danc, baker, franc, pari, jazz, dancer, cheetah, fashion, venus, banana, chorus, napoleon, jeff, perform
Topic #25: henri, franc, england, war, coloni, eleanor, revolut, slaveri, woman, louvertur, richard, ship, power, loui, centuri
Topic #26: pyramid, stone, tomb, stoneheng, site, egypt, ancient, archaeolog, sarah, imhotep, maria, homo, languag, corner, kingdom
Topic #27: bread, meat, sandwich, chicken, pizza, soup, cook, burger, beef, mayonnais, butter, eat, lettuc, tomato, recip
Topic #28: disney, peter, ride, book, pan, cronk, walt, cooki, disneyland, edit, jungl, charact, cinderella, lana, potter
Topic #29: battl, armi, empir, greec, olymp, democraci, fleet, chariot, greek, salami, oracl, citi, war, michael, ostrac


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #20
Topic #21
Topic #22
Topic #23
Topic #24
Topic #25
Topic #26
Topic #27
Topic #28
Topic #29
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
10
20
30
40
50
60
70
80
Podcast Count



[120]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 2))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 30 , tidf ngram range: (1, 2)
--------------------------------------------------------------------------------
Topic #0: birthday, evid, cat, turkey, courtroom, season, justic, rise, rule,
slash, favor, truth, jacki, court, butcherbox
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, virus, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
cacao, peanut, bar, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, tradit, movi christma, santa claus, villain, hanukkah,
krampus, winter
Topic #4: battl, armi, empir, democraci, fleet, salami, oracl, battl salami,
greec, citi, ostrac, water, delphi, darius, michael
Topic #5: film, charact, star, movi movi, list, watch, mission, scene, action,
car, movi yeah, jason, monologu, jedi, trilog
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, vanilla,
pie, rainer, water, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
```

bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, etiquett, app, host, hair, parti, art, slash, book, advic,
fashion, deliveri, mcelroy, busi, code
Topic #9: bread, meat, sandwich, chicken, beef, burger, lettuc, butter,
mayonnais, ham, cook, breast, slice, tomato, eat
Topic #10: byron, shelley, perci, vampir, jane, poetri, frankenstein, book,
mari, florenc, corinn, woman, switzerland, godwin, dad
Topic #11: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea kind, yeah tea, coffe, caffein, saucer, sugar
Topic #12: food bank, bank, march, march food, hunger, wine, food insecur,
communiti, insecur, confer, bank food, nutrit, food food, center, congressman
Topic #13: game, footbal, game show, playstat, tenni, game game, sport, croquet,
basebal, island, ball, barker, monopoli, nintendo, genesi
Topic #14: josephin, baker, josephin baker, danc, franc, pari, cheetah, venus,
dancer, venus venus, jazz, banana, chorus, barnum, chimpanze
Topic #15: song, album, danc, rock, band, vega, list, rock roll, boom boom,
soundtrack, boom, roll, jeff, guitar, song song
Topic #16: pyramid, stone, tomb, stoneheng, egypt, site, sarah, ancient,
imhotep, archaeolog, maria, giza, kingdom, mera, copper
Topic #17: pizza, dracula, docket, letter, main, dog, golf, book, josh, car,
york, friend, blood, disput, poop
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion,
zhang, guifei, rice, wu, yang guifei, empir, power
Topic #19: disney, peter pan, ride, peter, pan, cronk, walt, book, disney world,
disneyland, jungl, edit, roller, charact, song
Topic #20: henri, eleanor, franc, crusad, richard, loui, champagn, england, son,
king, marriag, pope, matilda, count champagn, woman
Topic #21: coloni, slaveri, louvertur, revolut, mayflow, ship, haiti, franc,
war, popul, island, weston, freedom, carver, trade
Topic #22: olymp, chariot, olympia, event, keon, discus, ceremoni, chariot race,
zeus, greec, greek, race, gold, statu, cliff
Topic #23: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget,
road, gax, guy gax, ha, ha ha, frisbe, road toad
Topic #24: breakfast, buffet, dinner, meal, toast, soup, dessert, eat, plate,
cooki, breakfast breakfast, dinner time, janet, bar, cracker
Topic #25: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic,
drink, appl appl, juic, cider cider, butter, glass, orang
Topic #26: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal,
mahal, taj, delhi, ruler, india, timur, court
Topic #27: khan, chinggi, peter, genghi, empir, genghi khan, phil, mongol,
chinggi khan, china, citi, silk, temujin, jin, zhongdu
Topic #28: dip, sauc, chip, dip dip, salsa, cracker, ranch, barbecu, mustard,
queso, onion, soy, spinach, vehicl dip, trader joe
Topic #29: monster, cooki, costum, candi, cooki monster, halloween, ghost,
oscar, jason, grover, monster yeah, sexi, easter, monster monster, elmo

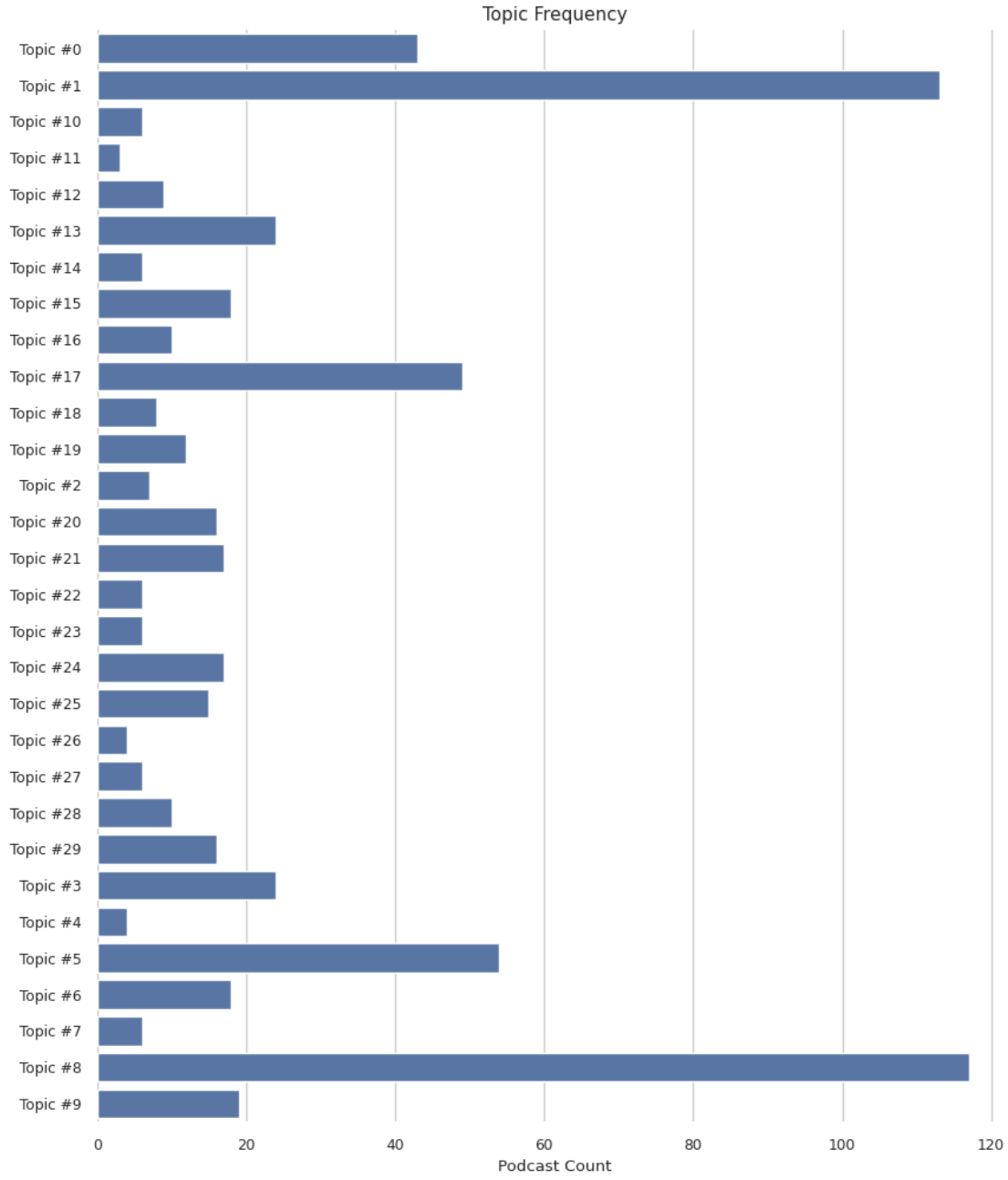
[120]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1      0.065963  0.158799       1        1  16.941689
8      0.033087  0.053052       2        1  11.151200
5      0.065083  0.114965       3        1   7.406083
17     0.121359 -0.006799       4        1   5.509097
0      0.106595 -0.001494       5        1   5.395967
13     0.086127  0.102549       6        1   4.272953
15     0.080192  0.131136       7        1   3.574331
3      0.087980  0.001692       8        1   3.387125
9      0.134724 -0.209921       9        1   3.156977
6      0.145249 -0.155956      10        1   3.044286
12     0.082401  0.148391      11        1   2.835946
29     0.122451  0.052720      12        1   2.799776
25     0.110797 -0.049410      13        1   2.680011
24     0.154453 -0.148358      14        1   2.645208
21    -0.227963  0.002494      15        1   2.456370
20    -0.202666  0.002882      16        1   2.237666
16    -0.198560 -0.051335      17        1   2.092155
7      0.049362  0.168489      18        1   2.012546
19     0.057989  0.076924      19        1   1.921722
28     0.144819 -0.150531      20        1   1.882718
2      0.107401 -0.095055      21        1   1.615524
18    -0.197096 -0.065979      22        1   1.455223
4     -0.190806 -0.039730      23        1   1.399916
27    -0.179096 -0.044184      24        1   1.365608
10    -0.154050  0.039642      25        1   1.358074
22    -0.200659 -0.052817      26        1   1.255766
14    -0.130084  0.092659      27        1   1.239236
23     0.069281  0.019089      28        1   1.161110
26    -0.205939 -0.060297      29        1   0.943875
11     0.061607 -0.033618      30        1   0.801842, topic_info=
Term      Freq     Total Category  logprob  loglift
84847    christma  8.000000  8.000000  Default  30.0000  30.0000
83347      chocol  5.000000  5.000000  Default  29.0000  29.0000
240891  ice cream  6.000000  6.000000  Default  28.0000  28.0000
116310      cream  6.000000  6.000000  Default  27.0000  27.0000
64655      cancer  4.000000  4.000000  Default  26.0000  26.0000
…             …       …       …       …        …        …
480200       tast  0.167545  2.771443  Topic30  -6.8962   2.0201
145930      drink  0.165150  2.828527  Topic30  -6.9106   1.9854
76364    champagn  0.136262  2.106674  Topic30  -7.1028   2.0877
143407   doordash  0.135348  2.373555  Topic30  -7.1096   1.9617
301740        max  0.128620  2.754833  Topic30  -7.1606   1.7618

[2969 rows x 6 columns], token_table=        Topic      Freq         Term
```

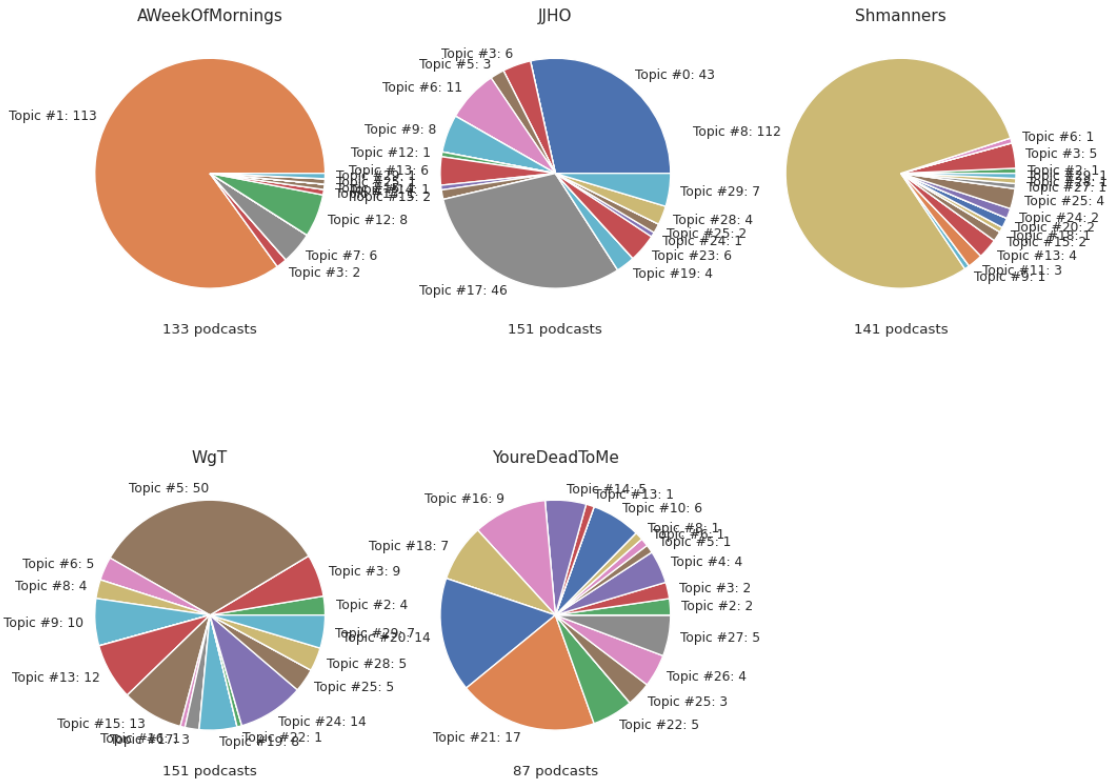
```
term
986          4  0.731202       accent
2599         3  0.508758       action
2739         3  1.822335  action movi
3021         3  0.544300        actor
5274         2  0.500280        advic
…        …        …           …
552419       1  0.878518    yesterday
553250       1  0.248089         york
553250       3  0.248089         york
553250       4  0.248089         york
554602      22  1.334231        zhang

[568 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 9, 6, 18, 1, 14, 16, 4, 10, 7, 13, 30, 26, 25,
22, 21, 17, 8, 20, 29, 3, 19, 5, 28, 11, 23, 15, 24, 27, 12])
```

[121]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: birthday, evid, cat, turkey, courtroom, season, justic, rise, rule,
slash, favor, truth, jacki, court, butcherbox
Topic #1: wine, river, presid, state, yesterday, countri, trump, congressman,
morn, communiti, hampton, vote, bill, virus, word week
Topic #2: chocol, cocoa, chocol chocol, milk, butter, candi, peanut butter,
cacao, peanut, bar, milk chocol, kit kat, cocoa butter, kat, histori chocol
Topic #3: christma, holiday, christma movi, santa, claus, christma christma,
scroog, ghost christma, tradit, movi christma, santa claus, villain, hanukkah,
krampus, winter
Topic #4: battl, armi, empir, democraci, fleet, salami, oracl, battl salami,
greec, citi, ostrac, water, delphi, darius, michael
Topic #5: film, charact, star, movi movi, list, watch, mission, scene, action,
car, movi yeah, jason, monologu, jedi, trilog
Topic #6: ice cream, cream, ice, cake, cream cake, cream ice, chocol, vanilla,
pie, rainer, water, bagel, cake cake, cone, eat
Topic #7: cancer, cancer connect, connect, connect cancer, wine, pino, diagnosi,
bed, support, huga, work cancer, support cancer, cancer cancer, presid, hampton
Topic #8: doordash, etiquett, app, host, hair, parti, art, slash, book, advic,
fashion, deliveri, mcelroy, busi, code
Topic #9: bread, meat, sandwich, chicken, beef, burger, lettuc, butter,
mayonnais, ham, cook, breast, slice, tomato, eat
Topic #10: byron, shelley, perci, vampir, jane, poetri, frankenstein, book,
mari, florenc, corinn, woman, switzerland, godwin, dad
Topic #11: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea kind, yeah tea, coffe, caffein, saucer, sugar
Topic #12: food bank, bank, march, march food, hunger, wine, food insecur,
```

communiti, insecur, confer, bank food, nutrit, food food, center, congressman
Topic #13: game, footbal, game show, playstat, tenni, game game, sport, croquet, basebal, island, ball, barker, monopoli, nintendo, genesi
Topic #14: josephin, baker, josephin baker, danc, franc, pari, cheetah, venus, dancer, venus venus, jazz, banana, chorus, barnum, chimpanze
Topic #15: song, album, danc, rock, band, vega, list, rock roll, boom boom, soundtrack, boom, roll, jeff, guitar, song song
Topic #16: pyramid, stone, tomb, stoneheng, egypt, site, sarah, ancient, imhotep, archaeolog, maria, giza, kingdom, mera, copper
Topic #17: pizza, dracula, docket, letter, main, dog, golf, book, josh, car, york, friend, blood, disput, poop
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, zhang, guifei, rice, wu, yang guifei, empir, power
Topic #19: disney, peter pan, ride, peter, pan, cronk, walt, book, disney world, disneyland, jungl, edit, roller, charact, song
Topic #20: henri, eleanor, franc, crusad, richard, loui, champagn, england, son, king, marriag, pope, matilda, count champagn, woman
Topic #21: coloni, slaveri, louvertur, revolut, mayflow, ship, haiti, franc, war, popul, island, weston, freedom, carver, trade
Topic #22: olymp, chariot, olympia, event, keon, discus, ceremoni, chariot race, zeus, greec, greek, race, gold, statu, cliff
Topic #23: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget, road, gax, guy gax, ha, ha ha, frisbe, road toad
Topic #24: breakfast, buffet, dinner, meal, toast, soup, dessert, eat, plate, cooki, breakfast breakfast, dinner time, janet, bar, cracker
Topic #25: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic, drink, appl appl, juic, cider cider, butter, glass, orang
Topic #26: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal, mahal, taj, delhi, ruler, india, timur, court
Topic #27: khan, chinggi, peter, genghi, empir, genghi khan, phil, mongol, chinggi khan, china, citi, silk, temujin, jin, zhongdu
Topic #28: dip, sauc, chip, dip dip, salsa, cracker, ranch, barbecu, mustard, queso, onion, soy, spinach, vehicl dip, trader joe
Topic #29: monster, cooki, costum, candi, cooki monster, halloween, ghost, oscar, jason, grover, monster yeah, sexi, easter, monster monster, elmo


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #20
Topic #21
Topic #22
Topic #23
Topic #24
Topic #25
Topic #26
Topic #27
Topic #28
Topic #29
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
120
Podcast Count



[122]:
```
vectorizer, data, model = topic_analyse(documents, topic_count, 'nmf', 'tfidf',␣
↪(1, 3))
pyLDAvis.lda_model.prepare(model, data, vectorizer)
```

```
--------------------------------------------------------------------------------
Vectorizer: tfidf , Model: nmf , Number of Topics: 30 , tidf ngram range: (1, 3)
--------------------------------------------------------------------------------
Topic #0: song, album, danc, band, rock, vega, list, soundtrack, boom boom, rock
roll, boom, disney, audienc particip, boom boom boom, roll
Topic #1: wine, river, presid, state, yesterday, countri, congressman, trump,
communiti, morn, hampton, vote, bill, word week, radio
Topic #2: chocol, cocoa, chocol chocol, milk, butter, cacao, candi, bar, kit
kat, milk chocol, peanut butter, kat, cocoa butter, peanut, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, scroog, villain, santa claus, tradit, ghost,
place christma, krampus
Topic #4: war, centuri, woman, gretchen, corner, comedi, sophi, catherin, power,
greg, radio, comedi corner, god, caesar, roman
Topic #5: evid, car, birthday, turkey, justic, rule, cat, dog, season,
courtroom, slash, rise, jacki, mom, favor
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, vanilla, water, pie, rainer, cone
```

Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer connect, support cancer, cancer cancer
Topic #8: doordash, app, etiquett, host, hair, fashion, deliveri, parti, art, advic, code, slash, busi, mcelroy, doordash app
Topic #9: dip, bread, meat, sauc, sandwich, pizza, chicken, soup, eat, breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni, afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, mari, florenc, corinn, switzerland, dad, godwin, franc, william godwin
Topic #12: food bank, bank, march, march food, hunger, march food bank, wine, food insecur, insecur, communiti, confer, food bank food, bank food, food food, nutrit
Topic #13: khan, peter, chinggi, genghi, genghi khan, empir, phil, mongol, chinggi khan, china, citi, silk, temujin, jin, peter pan
Topic #14: pyramid, tomb, stone, imhotep, sarah, egypt, maria, ancient, giza, archaeolog, mera, kingdom, stoneheng, copper, mummi
Topic #15: josephin, baker, josephin baker, danc, franc, cheetah, pari, venus, dancer, venus venus, banana, chorus, jazz, chimpanze, grace
Topic #16: battl, armi, empir, olymp, greec, fleet, salami, chariot, democraci, greek, oracl, battl salami, michael, olympia, ostrac
Topic #17: mayflow, coloni, ship, weston, misha, carver, popul, voyag, invest, jane town, settlement, plymouth, england, dutch, merchant
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, guifei, wu, rice, zhang, yang guifei, minist, empress
Topic #19: game, footbal, golf, game show, tenni, playstat, sport, game game, croquet, ball, pizza, basebal, barker, monopoli, island
Topic #20: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget, road, ha, ha ha ha, ha ha, gax, guy gax, road toad
Topic #21: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic, butter, appl appl, cider cider, juic, drink, appl cider, orang
Topic #22: henri, eleanor, franc, crusad, richard, loui, champagn, england, count champagn, matilda, son, marri, marriag, geoffrey, annul
Topic #23: homo, languag, california man, stone, stone age, tim, ice age, extinct, site, fossil, dna, fossil record, stoneheng, age, languag languag
Topic #24: book, book book, chees, end book, pooh, horton, dedic, rabbit, kid, letter, book kind, grover, reader, milk, book yeah
Topic #25: fund, max, max fund, membership, chivalri, fund drive, drive, member, hors, support, join, max fund drive, cat, idiom, bonus
Topic #26: film, charact, star, movi movi, list, watch, disney, jason, scene, car, mission, movi yeah, action, monologu, york
Topic #27: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal, mahal, taj, delhi, ruler, india, timur, taj mahal
Topic #28: louvertur, revolut, haiti, slaveri, franc, island, napoleon, coloni, freedom, virtu, plantat, emperor, marlena, toussaint, home franc
Topic #29: monster, cooki, costum, candi, cooki monster, halloween, oscar, ghost, monster yeah, grover, sexi, bar, monster monster, elmo, vampir

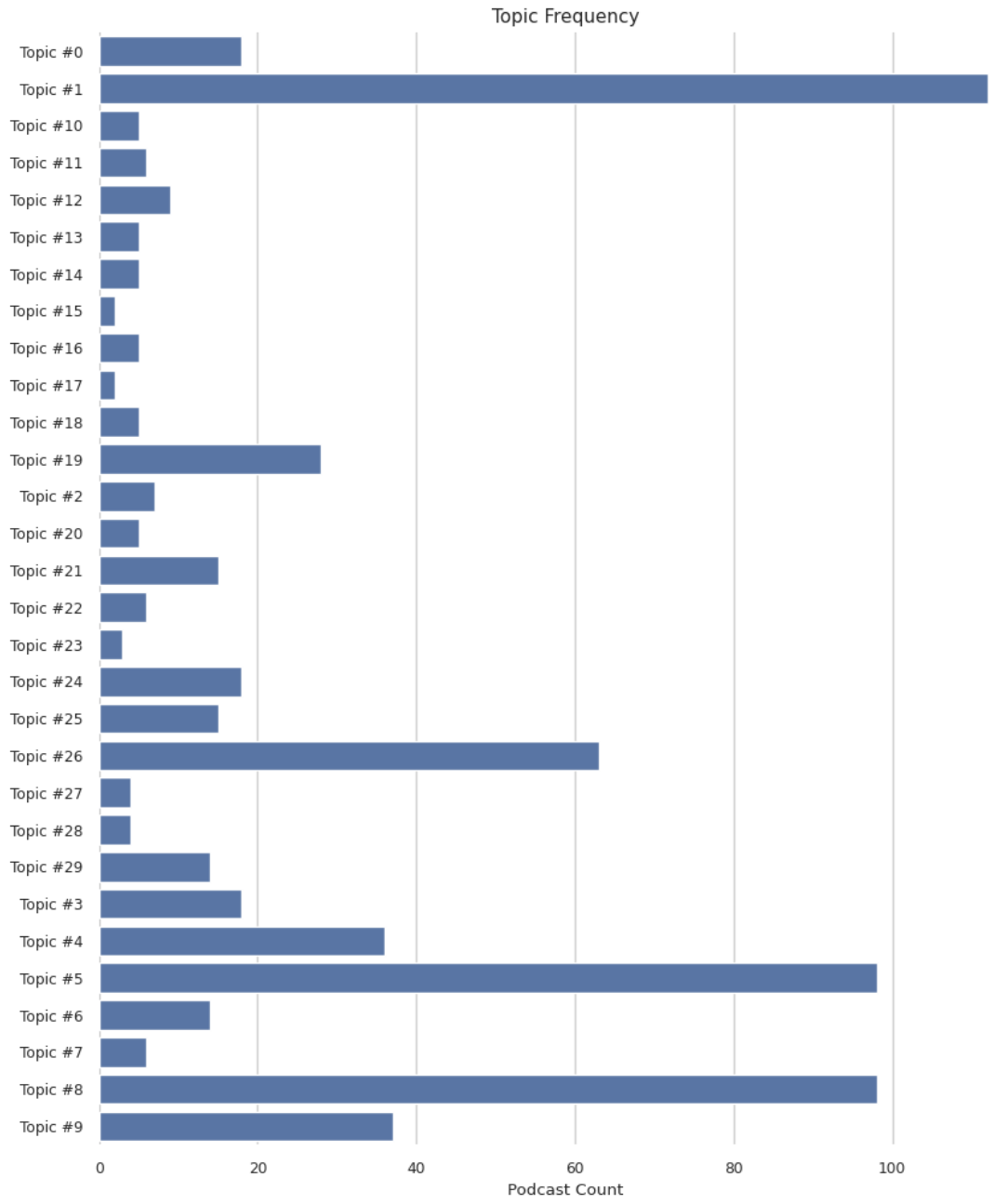
[122]:
```
PreparedData(topic_coordinates=              x         y  topics  cluster
Freq
topic
1     -0.035321  0.178868       1        1  17.587652
5     -0.127864 -0.090038       2        1  10.506386
8     -0.030290 -0.110453       3        1   9.695025
26    -0.078295 -0.079886       4        1   8.003986
9     -0.162145 -0.005985       5        1   5.070938
4      0.193914 -0.040171       6        1   4.309739
19    -0.095558 -0.105348       7        1   4.100883
24    -0.084104 -0.043591       8        1   3.744964
0     -0.078253 -0.125474       9        1   3.342527
3     -0.090753  0.095046      10        1   2.870088
29    -0.150438  0.029591      11        1   2.752652
6     -0.137476  0.034736      12        1   2.724145
12    -0.056506  0.243670      13        1   2.723017
25    -0.079728 -0.135730      14        1   2.685381
21    -0.123269  0.077126      15        1   2.321345
7     -0.012280  0.139730      16        1   1.960699
2     -0.115941  0.191351      17        1   1.522455
16     0.157737  0.025492      18        1   1.459610
10    -0.069952 -0.062424      19        1   1.338127
22     0.107270 -0.045019      20        1   1.293318
20    -0.094850 -0.111316      21        1   1.231956
28     0.163214  0.011109      22        1   1.157847
11     0.061345 -0.092231      23        1   1.144780
14     0.180322  0.033012      24        1   1.128774
13     0.123915  0.009200      25        1   0.968158
17     0.072764 -0.004859      26        1   0.927113
23     0.111796 -0.027702      27        1   0.920577
15     0.096689 -0.084058      28        1   0.855371
18     0.177165  0.098128      29        1   0.827446
27     0.176894 -0.002774      30        1   0.825043, topic_info=
Term      Freq     Total Category  logprob  loglift
1141840        tea  4.000000  4.000000  Default  30.0000  30.0000
195860    christma  5.000000  5.000000  Default  29.0000  29.0000
192266      chocol  4.000000  4.000000  Default  28.0000  28.0000
566886   ice cream  4.000000  4.000000  Default  27.0000  27.0000
270440       cream  5.000000  5.000000  Default  26.0000  26.0000
…              …       …       …       …       …       …
264357       court  0.290513  2.795939  Topic30  -6.7414   2.5332
356977     emperor  0.238866  2.258996  Topic30  -6.9372   2.5507
621212        khan  0.223177  1.417763  Topic30  -7.0051   2.9486
943020      region  0.214619  1.459754  Topic30  -7.0442   2.8803
69613        battl  0.206712  2.236025  Topic30  -7.0818   2.4164

[2995 rows x 6 columns], token_table=         Topic      Freq           Term
```

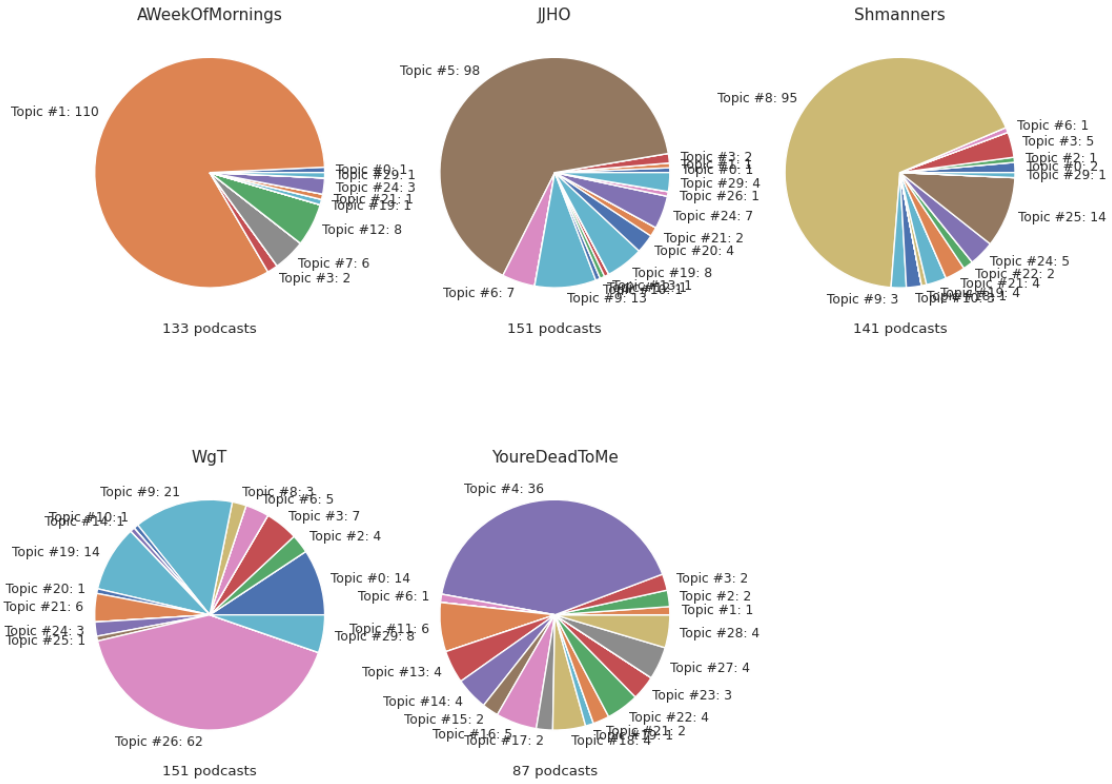
```
term
5687         4  0.632761         action
6704         4  0.687490          actor
11712        3  0.621807          advic
13682       19  1.152669  afternoon tea
18624       30  1.560167          akbar
…         …       …              …
1294001      1  1.024757      word week
1321637      4  1.069155      yeah movi
1337670      1  0.728646      yesterday
1339625      1  0.298497           york
1339625      4  0.298497           york

[413 rows x 3 columns], R=30, lambda_step=0.01, plot_opts={'xlab': 'PC1',
'ylab': 'PC2'}, topic_order=[2, 6, 9, 27, 10, 5, 20, 25, 1, 4, 30, 7, 13, 26,
22, 8, 3, 17, 11, 23, 21, 29, 12, 15, 14, 18, 24, 16, 19, 28])
```

[123]:
```
# visualisation
topic_df, topic_columns = prep_topic_df(model, data, vectorizer)
plot_topic_frequencies(topic_df)
plot_all_pod_topic_freq(topic_df, topic_columns)
```

```
Topic #0: song, album, danc, band, rock, vega, list, soundtrack, boom boom, rock
roll, boom, disney, audienc particip, boom boom boom, roll
Topic #1: wine, river, presid, state, yesterday, countri, congressman, trump,
communiti, morn, hampton, vote, bill, word week, radio
Topic #2: chocol, cocoa, chocol chocol, milk, butter, cacao, candi, bar, kit
kat, milk chocol, peanut butter, kat, cocoa butter, peanut, quaker
Topic #3: christma, christma movi, holiday, santa, claus, christma christma,
ghost christma, movi christma, scroog, villain, santa claus, tradit, ghost,
place christma, krampus
Topic #4: war, centuri, woman, gretchen, corner, comedi, sophi, catherin, power,
greg, radio, comedi corner, god, caesar, roman
Topic #5: evid, car, birthday, turkey, justic, rule, cat, dog, season,
courtroom, slash, rise, jacki, mom, favor
Topic #6: ice cream, cream, ice, cake, cream cake, ice cream cake, cream ice,
cream ice cream, chocol, ice cream ice, vanilla, water, pie, rainer, cone
Topic #7: cancer, cancer connect, connect, cancer connect cancer, connect
cancer, wine, pino, diagnosi, bed, support, work cancer, huga, connect cancer
connect, support cancer, cancer cancer
Topic #8: doordash, app, etiquett, host, hair, fashion, deliveri, parti, art,
advic, code, slash, busi, mcelroy, doordash app
Topic #9: dip, bread, meat, sauc, sandwich, pizza, chicken, soup, eat,
breakfast, meal, burger, beef, mustard, potato
Topic #10: tea, tea tea, afternoon tea, matcha, water, ceremoni, tea ceremoni,
afternoon, cup, tea tea tea, tea kind, saucer, caffein, sugar, yeah tea
Topic #11: shelley, byron, perci, vampir, jane, frankenstein, poetri, mari,
florenc, corinn, switzerland, dad, godwin, franc, william godwin
```

Topic #12: food bank, bank, march, march food, hunger, march food bank, wine, food insecur, insecur, communiti, confer, food bank food, bank food, food food, nutrit
Topic #13: khan, peter, chinggi, genghi, genghi khan, empir, phil, mongol, chinggi khan, china, citi, silk, temujin, jin, peter pan
Topic #14: pyramid, tomb, stone, imhotep, sarah, egypt, maria, ancient, giza, archaeolog, mera, kingdom, stoneheng, copper, mummi
Topic #15: josephin, baker, josephin baker, danc, franc, cheetah, pari, venus, dancer, venus venus, banana, chorus, jazz, chimpanze, grace
Topic #16: battl, armi, empir, olymp, greec, fleet, salami, chariot, democraci, greek, oracl, battl salami, michael, olympia, ostrac
Topic #17: mayflow, coloni, ship, weston, misha, carver, popul, voyag, invest, jane town, settlement, plymouth, england, dutch, merchant
Topic #18: emperor, tang, china, buddhism, yang, evelyn, dynasti, rebellion, guifei, wu, rice, zhang, yang guifei, minist, empress
Topic #19: game, footbal, golf, game show, tenni, playstat, sport, game game, croquet, ball, pizza, basebal, barker, monopoli, island
Topic #20: rudi, glitter, rudi place, toad road, toad, tank, patrick, bridget, road, ha, ha ha ha, ha ha, gax, guy gax, road toad
Topic #21: appl, cider, beer, water, prohibit, flavor, vanilla, appl juic, butter, appl appl, cider cider, juic, drink, appl cider, orang
Topic #22: henri, eleanor, franc, crusad, richard, loui, champagn, england, count champagn, matilda, son, marri, marriag, geoffrey, annul
Topic #23: homo, languag, california man, stone, stone age, tim, ice age, extinct, site, fossil, dna, fossil record, stoneheng, age, languag languag
Topic #24: book, book book, chees, end book, pooh, horton, dedic, rabbit, kid, letter, book kind, grover, reader, milk, book yeah
Topic #25: fund, max, max fund, membership, chivalri, fund drive, drive, member, hors, support, join, max fund drive, cat, idiom, bonus
Topic #26: film, charact, star, movi movi, list, watch, disney, jason, scene, car, mission, movi yeah, action, monologu, york
Topic #27: shah, jahangir, jahan, shah jahan, empir, mogul, akbar, mughal, mahal, taj, delhi, ruler, india, timur, taj mahal
Topic #28: louvertur, revolut, haiti, slaveri, franc, island, napoleon, coloni, freedom, virtu, plantat, emperor, marlena, toussaint, home franc
Topic #29: monster, cooki, costum, candi, cooki monster, halloween, oscar, ghost, monster yeah, grover, sexi, bar, monster monster, elmo, vampir


Topic Frequency
Topic #0
Topic #1
Topic #10
Topic #11
Topic #12
Topic #13
Topic #14
Topic #15
Topic #16
Topic #17
Topic #18
Topic #19
Topic #2
Topic #20
Topic #21
Topic #22
Topic #23
Topic #24
Topic #25
Topic #26
Topic #27
Topic #28
Topic #29
Topic #3
Topic #4
Topic #5
Topic #6
Topic #7
Topic #8
Topic #9
0
20
40
60
80
100
Podcast Count



If you have made it this far down in this document **thank you**. It was an interesting project and given more time would have many angles to dig deeper.