Research Question: Can Online Discussions Predict the Lifecycle of a Food Trend?

Kadeeja Zumreen, Sriya Kondury, Sayalee Chivate

30 August 2025

1 Background and Research Questions

In today's digital age, food trends emerge and spread at an unprecedented pace, driven largely by social media and online communities. Platforms such as TikTok, Instagram, Reddit, and YouTube have changed the way recipes, ingredients, and diets capture attention and gain momentum. Viral recipes like baked feta pasta or the rise of niche products such as mushroom coffee and matcha show how quickly online conversations can influence consumer behavior and decisions (Saura et al., 2020). These online trends have significant real-world impacts, affecting grocery store inventory, inspiring new products, and shaping conversations around health and culture. For businesses and researchers, understanding the lifecycle of these trends – when they will start, peak, and fade can be extremely valuable.

Previous studies have used online data to analyze food conversations. Research has shown that social media data can predict real world outcomes like public health statistics (Fried et al., 2014) and reveal public opinion on broad topics like organic food (Singh and Glińska-Neweś, 2022),or the popularity of different cuisines (Khulbe and Pathak, 2019). These studies confirm that online discussions are a rich source of data for analyzing food trends.

However, a key gap exists in current research, it describes what people are talking about, but not how those conversations change over time to drive a trend's rise and fall. This project will fill that gap by focusing on the dynamics of food trends as they evolve. We aim to analyze food trend discussions over time across one or two platforms, using a combination of sentiment analysis and topic modeling. This approach will allow us to capture both the mood and themes of conversation driving food trends. **Our overarching goal is to understand how the sentiment and themes of online discussions on social media reflect and predict the lifecycle of modern food trends.**

1.1 What factors (online conversations, news coverage, celebrity endorsements) drive the rise and decline of food trends?

Social media is a treasure trove of conversations in the online space, but the lifecycle of food trends is based on more than casual conversation. Conversations may be amplified by traditional media, the endorsement of celebrities and influencers, health or sustainability narratives, and or spontaneous cultural moments (for example, a viral TikTok video or a restaurant shift in innovation). For this reason, we will investigate how these factors coexist and which have the most defining influence in order to accelerate, peak, and ultimately saturate specific food trends.

1.2 Can we build a model to predict a trend's peak popularity and its longevity?

One challenge in analyzing online food trends is that their growth trajectories are often highly nonlinear: some trends experience explosive viral peaks followed by steep declines, some develop gradually and persist over time while some trends never become viral and in a matter of few days, cease to exist. By applying machine learning and time series analysis to social media data, we will explore whether models can be trained to forecast not only when a trend will reach its peak visibility but also how long it will remain culturally relevant. We will also evaluate whether incorporating external data sources (e.g., market sales, Google search data, or restaurant adoption) improves predictive accuracy.

1.3 Where do food trends originate, and how do they spread geographically across online communities and markets?

Food trends may begin development from specific cultural origins - like urban restaurant scenes, influencer communities, or demographic groups - before breaking into new territories and larger markets. Through methods such as geotagged posts, network analysis and cultural mapping, we will explore how these trends are disseminated and whether they follow trails that transcend geographical and demographical boundaries. Tracking the trend's pathway will clarify the origin for trends and the unique aspect digital communication plays in extending or distorting the spread of food trends in comparison to food fads of the past.

2 Data

2.1 Social Media Data

We plan to use multiple social media datasets to collect discussions revolving around food trends. The Stanford SNAP Collection (`https://snap.stanford.edu/data/index.html#socnets`) provides large-scale Reddit data, including subreddit submissions and comment networks. Although this data may be less current, it provides us with a structured and well-documented starting point in our exploratory analysis. Using this data, we can begin testing sentiment and topic modeling approaches and identifying patterns in online discussions.

2.2 YouTube Data

The Trending YouTube Video Statistics dataset (`https://www.kaggle.com/datasets/datasnaek/youtube-new`) provides a daily record of the top trending videos on YouTube across multiple countries (including the U.S., U.K., Germany, France, Canada, South Korea, India, Russia, and Japan). Each file contains up to 200 trending videos per day, and this includes metadata like video title, channel title, time of publishing, views, likes, dislikes, description, and comment count.

Since YouTube defines "trending" based on the number of user interactions (views, shares, likes, comments), rather than just the raw view counts, this dataset is very useful for identifying viral content. By filtering specifically for food-related videos, we can track the speed by which new trends emerge and fade across multiple regions.

2.3 Future Data Sources

Depending on the scope of this project, we may supplement these datasets with more recent social media discussions. Reddit's API provides access to current posts and comments in food-related subreddits (eg: r/food, r/easyrecipes, r/cooking). We may also consider using the Youtube API directly to gather the most recent data and capture comment level sentiment. These future data sources would allow us to compare our older structured datasets with more up to date conversations.

3 Methods

For exploratory data analysis, we will use Python libraries like pandas and matplotlib to track how often food-related posts and videos appear over time. These will help us analyze the growth, peak, and decline stages of specific food trends.

To understand how users feel about these trends, we will apply sentiment analysis with existing tools such as VADER or TextBlob. These tools will classify text into categories like positive, neutral, and negative, which allows us to observe any shifts in enthusiasm as a trend evolves.

To identify any recurring themes, we plan to use topic modeling methods such as Latent Dirichlet Allocation (LDA). This will help group posts and video descriptions into broader categories like recipes and health concerns, which can give us a clear sense of what conversations drive different stages of the trend lifecycle.

Finally to explore predictive modeling, we will test whether early patterns in sentiment and discussion topics can predict the likely peak or duration of a food trend. Using machine learning methods like regression or random forest can be used to estimate how long a trend is likely to stay relevant.

References

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., and Bell, D. (2014). Analyzing the language of food on social media. *arXiv preprint arXiv:1409.2195*.

Khulbe, D. and Pathak, M. (2019). Modeling food popularity dependencies using social media data. *arXiv preprint arXiv:1906.12331*.

Saura, J. R., Reyes-Menendez, A., and Thomas, S. B. (2020). Gaining a deeper understanding of nutrition using social networks and user-generated content. *Internet Interventions*, 20:100312.

Singh, A. and Glińska-Neweś, A. (2022). Modeling the public attitude towards organic foods: a big data and text mining approach. *Journal of Big Data*, 9(1):2.