TikTok Engagement Traces Over Time and Health Risky Behaviors: Combining Data Linkage and Computational Methods

Xinyan (Eva) Zhao, UNC-Chapel Hill, Hussman School of Journalism and Media
Chau-Wai Wong, North Carolina State University, Department of Electrical and Computer Engineering

Theoretical Background

Social media information environment:

- Users are embedded within multiple, intersecting flows of content (curated flows, Thorson & Wells, 2015).
- A highly personalized information environment in the risk and health realm (Zhao et al. 2022).

TikTok:

- 834 million global users as of 2024.
- Offer immersive and interactive videos.

Our purpose:

- How does TikTok engagement affect users' downstream health behaviors?

Methodological Background

Self-report Issues:

- Accurately recalling media consumption is challenging for survey respondents (Prior, 2009; Ohme, 2020; Scharkow, 2016).
- Subjects can underreport health risky behaviors.

A Data Likage Approach:

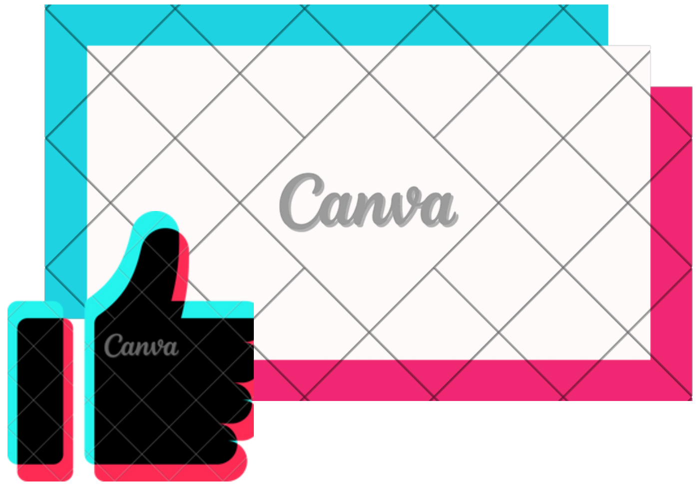
TikTok engagement traces

Self-reported survey responses

Theoretical Framework

The Selectivity Paradigm:

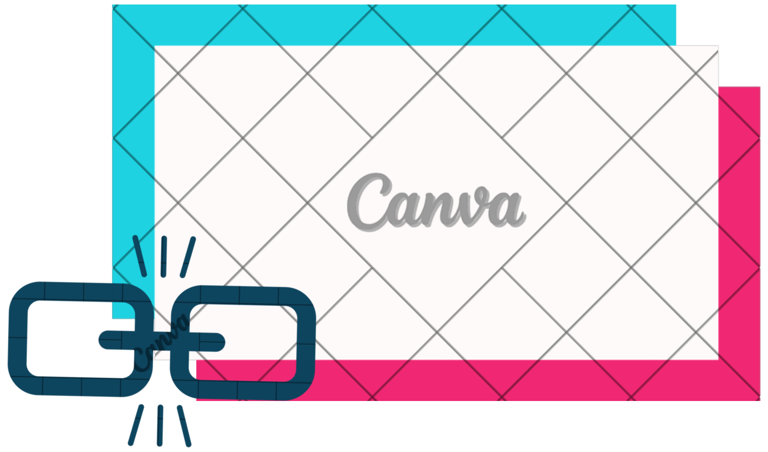
Individuals are limited in the amount of media content they can focus on, thus they choose the information based on personal needs and desires, and are influenced only by the information they select (Valkenburg, 2022).



Algorithmic Impact:

- Social media selective engagement shapes personalized content flows through a dynamic loop driven by platform algorithms.
- TikTok's For You feed algorithm (2020) recommends content starting from preferences as a new user and adjusting based on user interactions such as the videos one likes, to "ensure users see more of what they enjoy." (p. 1)
- Selective engagement can push more TikTok videos users engage with through algorithms over time.

Data Likage Approach



A user-centric approach that links data at the individual level:

- Participant consent needed.
- Authorize access through TikAPI.

A computational analysis of health-risk TikToks "liked" by participants (2020-23):

- Speech-to-text and large language models.
- Video summarization.

Self-reported drinking/vaping behaviors through a survey.

Key Hypotheses/RQ

RQ1: What is the temporal pattern of TikTok users liking videos containing smoking (a) and drinking (b) content?

H1: Controlling for demographic factors, the size of liked videos containing smoking (a) and drinking (b) content are associated with self-reported vaping (a) and drinking (b) behaviors.

RQ2: How is the size of liked videos containing additional health topics (i.e., vegetable and fruit consumption, exercise, marijuana) associated with self-reported vaping and drinking behaviors?

RQ3: What is the actual prevalence of smoking and drinking TikTok videos of different valence (positive vs. non-positive)? What is the behavioral impact of video valence?

RQ4: How do digital engagement traces at different time segments correlate with self-reported engagement and behavioral measures?

Methods

Survey Participants:

- Recruited through Prolific.
- Quota sampling: 1/4 Caucasian, 1/4 Hispanic, 1/4 Black, 1/4 Asian, Pacific Islander, American Indian, or Native American.
- Of the 1,102 survey respondents, 166 (15%) agreed to authorize TikTok access.

The Consented Sample:

- # "Liked" TikToks: M = 2,409, Med = 698, SD = 2,919
- # Health-risk TikToks: M = 131.30, Med = 29, SD = 181.50
- We randomly sampled and downloaded up to 200 health-risk videos for each participant.
- 13,724 videos were downloaded using a Python script, amounting to 93.01 GB of data.

Computational Analysis

Speech-to-Text (Deepgram, 2023)

- The neural engine transcribes speakers' sentences into texts and provides estimated topics with confidence levels.
- 74.8% contain human speech successfully transcribed.
- For each participant, we aggregated and normalized each topic's confidence levels to measure the size of liked drinking/smoking videos.

Valence Detection through GPT-4 (Ziem et al., 2024)

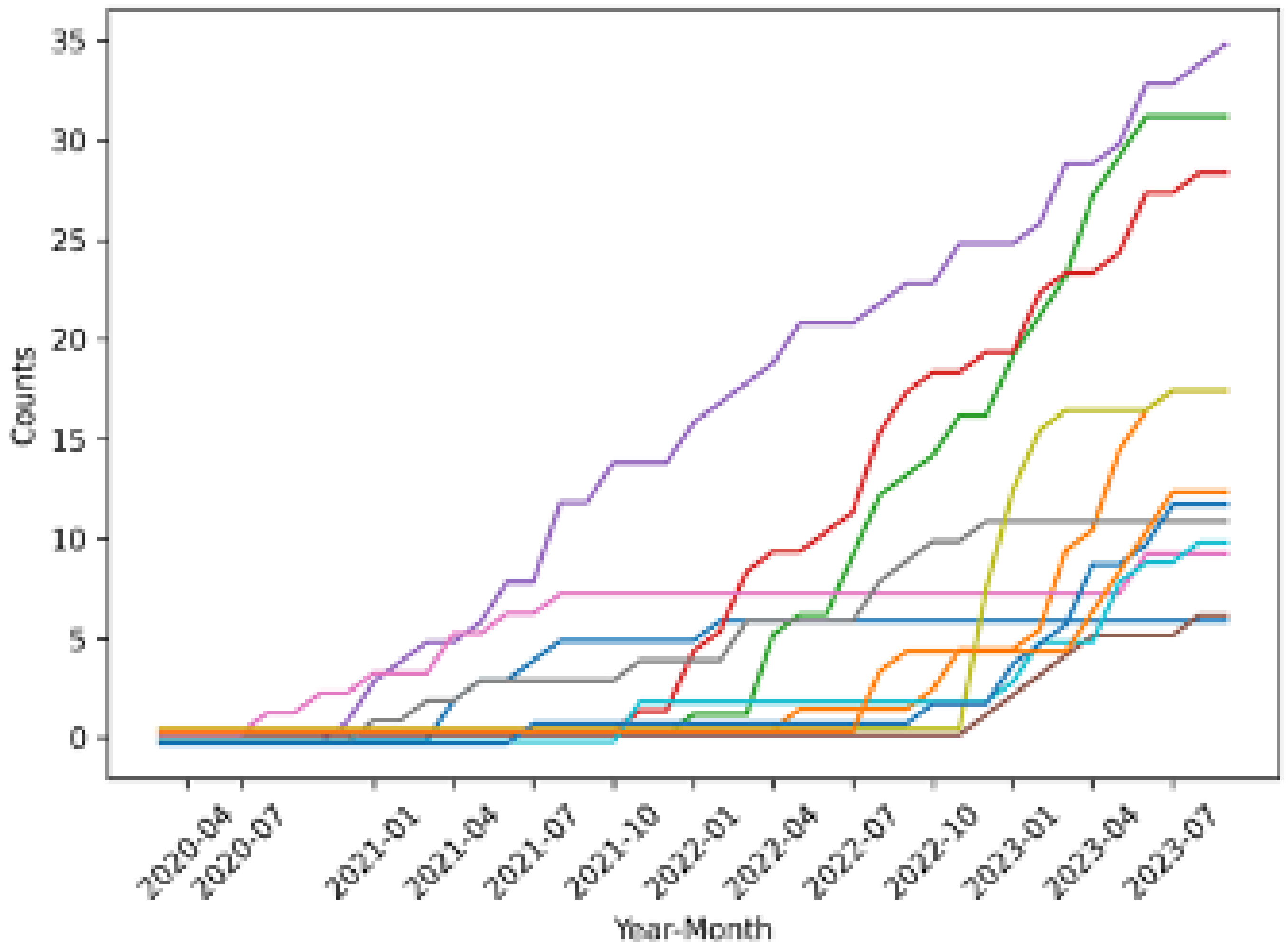
- For drinking videos, we measured the number of positive vs. nonpositive videos for each subject.

Video summarization (e.g., Iashin & Rahtu, 2020)

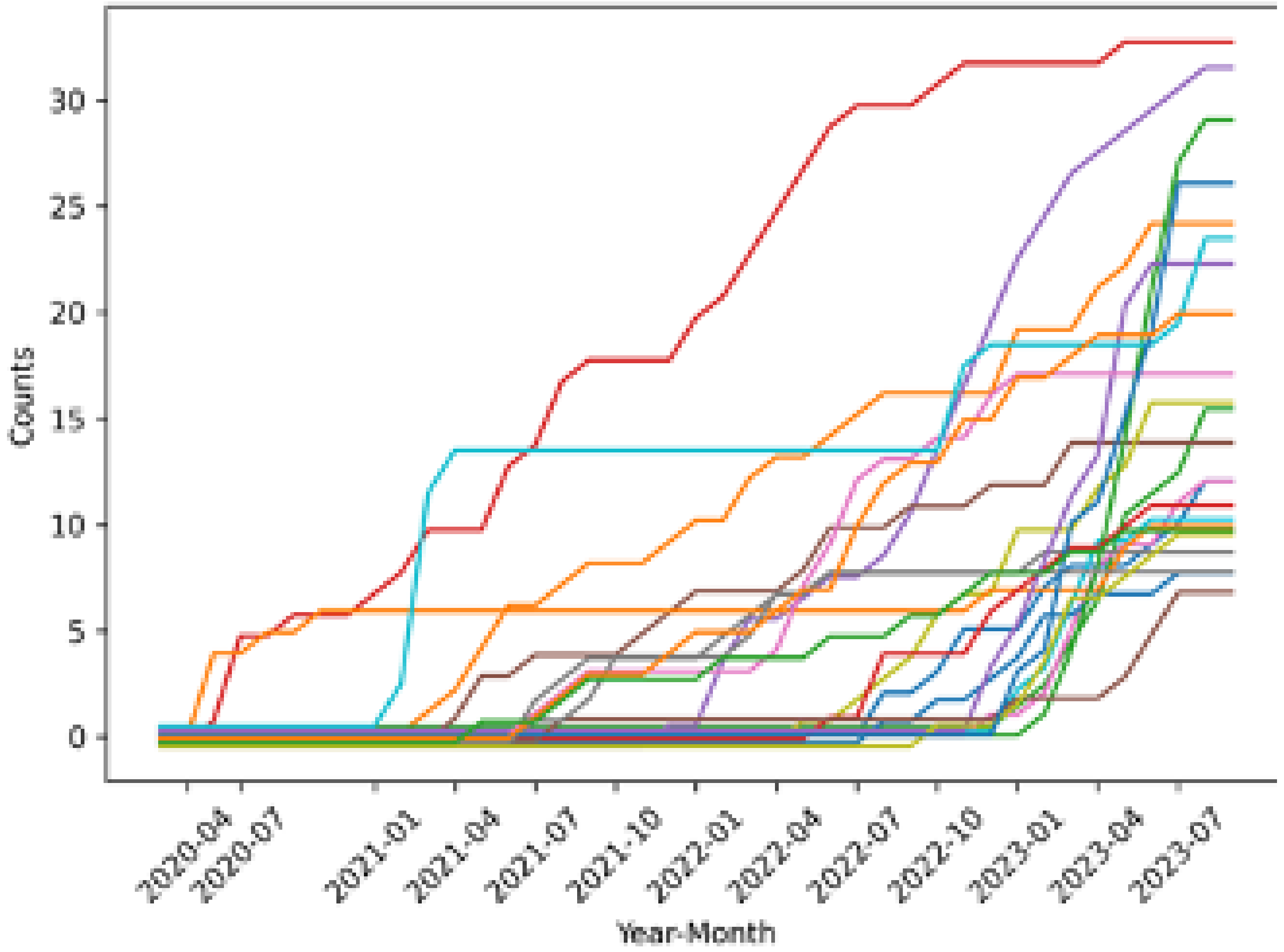
- Only output very general results, thus not used in analysis.

Trend Analysis Results

Significant trend of an increasing rate of liking drinking TikTok videos



Insignificant Results: A constant rate of liking drinking TikTok videos



An F-test was conducted between the null hypothesis of a linear line (i.e., viewing counts evenly spread over time) versus the alternative hypothesis of a quadratic line (viewing counts linearly change over time) for 35 participants who viewed more than 4 drinking videos across years.

Regression Results

	Current Drinking Behavior (Self-Reported)	Current Binge Drinking Behavior (Self-Reported)	Current Vaping Behavior (Self-reported)
Age	0.02 (0.09)	-0.06 (0.08)	0.01 (0.08)
Sex (male as the reference group)	0.03 (0.26)	-0.10 (0.24)	-0.37 (0.24)
Education	0.09 (0.09)	-0.01 (0.08)	-0.07 (0.08)
Ethnicity (non-white as the reference group)	0.07 (0.25)	-0.17 (0.23)	0.07 (0.23)
Total # of following	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Total # of hearts	-0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Former vaping behavior	0.63 (0.25) *	0.64 (0.23) **	1.58 (0.23) ***
Total # liked videos on drinking	6.03 (2.83) *	5.69 (2.66) *	2.22 (2.56)
Total # liked videos on fruit and vegetable consumption	-8.86 (4.26) *	-8.47 (4.00) *	5.28 (3.86)
Total # liked videos on smoking	-0.54 (1.50)	1.58 (1.40)	2.90 (1.36) *

Note. N = 166

- Self-reported drinking, binge drinking, vaping behaviors were positively affected by the size of liked TikToks on respective topics.
- Drinking and binge drinking behaviors were negatively affected by TikToks on fruit and vegetable consumption.
  - Neither liking pro- or anti-drinking content related to the self-reported behaviors.

Results: Digital Traces & Self Reports

Vaping-Related Measures

	Liking traces (09/2022-09/2023)	Liking traces (09/2021-09/2022)	Liking traces (04/2020-09/2021)	Reported Attention	Reported Engagement	Self-reported vaping behavior
Liking traces (09/2022-09/2023)	1.00	-.01	.98***	.11	.03	.19*
Liking traces (09/2021-09/2022)	-.01	1.00	.10	-.10	-.06	-.04
Liking traces 04/2020-09/2021)	.98***	.10	1.00	.08	-.01	.18*
Reported Attention to Vaping TikTok	.11	-.10	.08	1.00	.43***	.30***
Reported Vaping Content Engagement	.03	-.06	-.01	.43***	1.00	.44***
Self-reported Vaping behavior	.19*	-.04	.18*	.30***	.44***	1.00

- For vaping, liking traces in the recent one year did not correlate with self-reported TikTok engagement.
- Liking traces in the recent one year correlated with self-reported vaping and current drinking, but not binge drinking.

Takeaways

- Users who initially liked drinking-related videos tend to like more of such content over time, supporting **algorithmic impact**. The varied increasing rates over years suggest TikTok's algorithm might not have a unified impact on all users.
- The actual size of liked TikToks on drinking and smoking positively affected self-reported behaviors, supporting **selective engagement effects on downstream behaviors** using data linkage.
- For vaping, TikTok liking traces did not correlate with self-reported engagement. This stresses **self-report bias** and the importance of **objective measures of engagement.**

Implications for Data Linkage

Pros

- Digital traces and computational measures can be objective and reliable measures of social media engagement.
- Future research using multimodal analysis enabled by new models like GPT-4o?

Cons

- Complications in system design, social media API constraints/updates, and user privacy concerns.
- Only a small percentage of participants were willing to provide access to their TikTok data, limiting the generalizability of results and making this method costly.

Thank you!

TikTok Engagement Traces Over Time and Health Risky Behaviors: Combining Data Linkage and Computational Methods

Xinyan (Eva) Zhao, UNC-Chapel Hill, Hussman School of Journalism and Media
Chau-Wai Wong, North Carolina State University, Department of Electrical and Computer Engineering