



# INFORMATION CASCADES

Sandeep Soni

---

11/20/2023

# SCHEDULE CHANGES

---

- No lecture on Wednesday because of Thanksgiving recess
- No lecture next Monday; instead I'll hold office hours in class to help answer any questions
- Group presentations will be on 11/29
- Final project report due on or before 12/13

# GROUP PRESENTATIONS

---

- Please upload slides on canvas
- Keep the slides short and have less text; instead, make use of tables, graphs, etc
- Every group member has to present

# FINAL PROJECT REPORT

---

- See the canvas page for details.

# FINAL PROJECT REPORT

---

- Don't just dump results/methodological details as bullet points. Try to motivate your choices or explain your findings.
- Be skeptical of positive findings and rigorously test the conditions under which the findings hold

# FINAL PROJECT REPORT

---

- EDA is expected.
  - Method selection should be purposeful.
- Data cleaning/preprocessing is important
  - Write about it but don't take too much space

# FINAL PROJECT REPORT

---

- Write about the limitations of your approach and what are the steps to overcome the limitations
- If you see ethical issues at play in your project, write about them

# CONTENT ANALYSIS

---

- We have focused on intrinsic analysis of text (e.g., topic modeling, language modeling, etc)
- We have also used text to predict an outcome
- Many times, the outcome is some notion of popularity or impact

# POPULARITY

---

- Engagement metrics on many social media platforms indicate the popularity of the content



**Macaulay Culkin**  
@IncredibleCulk

Hey guys, wanna feel old?

I'm 40.

You're welcome.

5:13 PM · Aug 26, 2020

56K

525K

2.9M

11K

# IMPACT

---

- In some domains, the content is referred, indicating the importance of the content

**Bert: Pre-training of deep bidirectional transformers for language understanding**

J Devlin, MW Chang, K Lee, K Toutanova - arXiv preprint arXiv ..., 2018 - arxiv.org

... We introduce **BERT** and its detailed implementation in this ... For finetuning, the **BERT** model is first initialized with the pre-... A distinctive feature of **BERT** is its unified architecture across ...

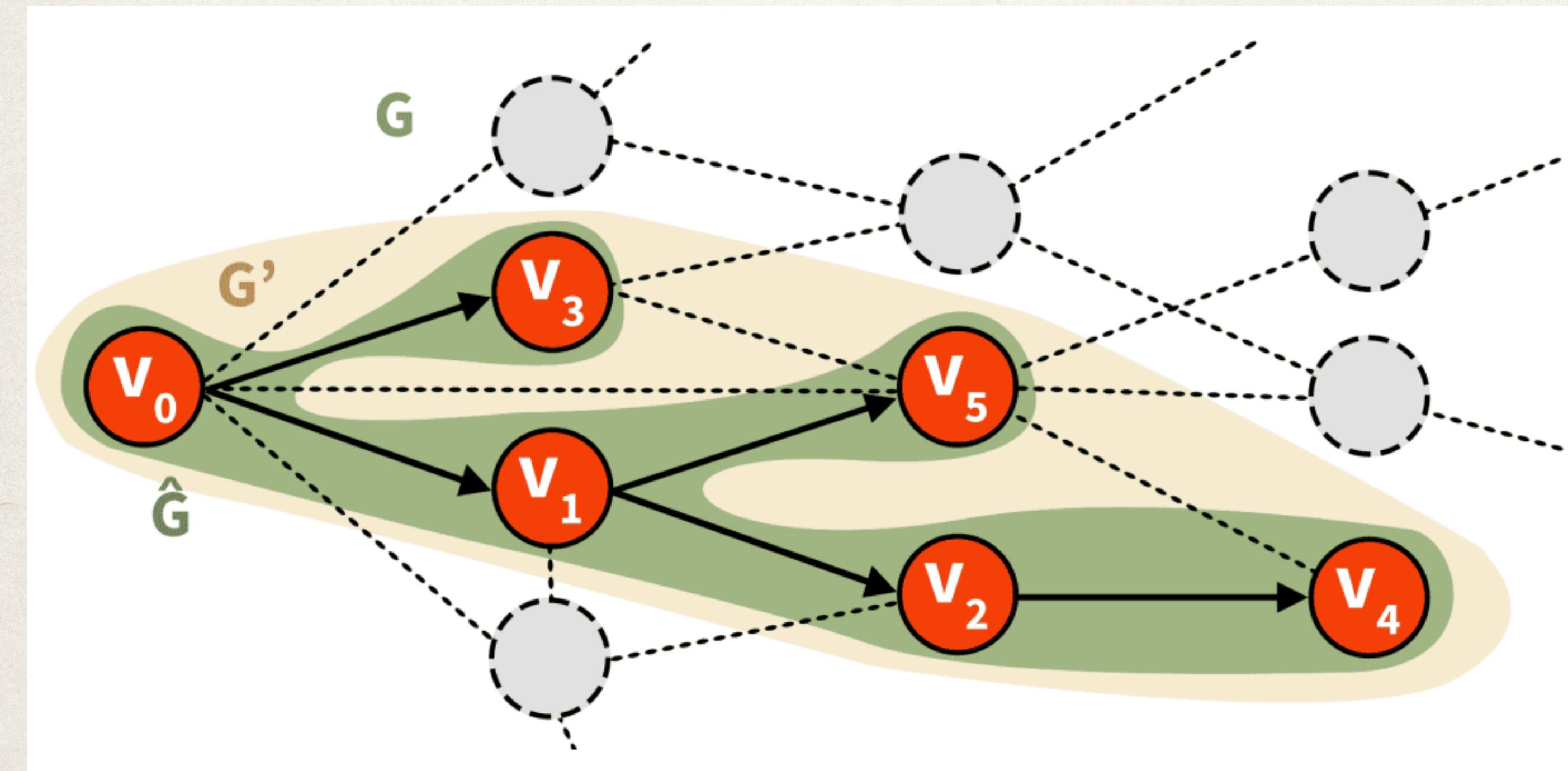
★ Save ⚡ Cite Cited by 83562 Related articles All 46 versions ☺

“What are factors for text to become very popular in short amount of time?”

# FACTORS AFFECTING POPULARITY

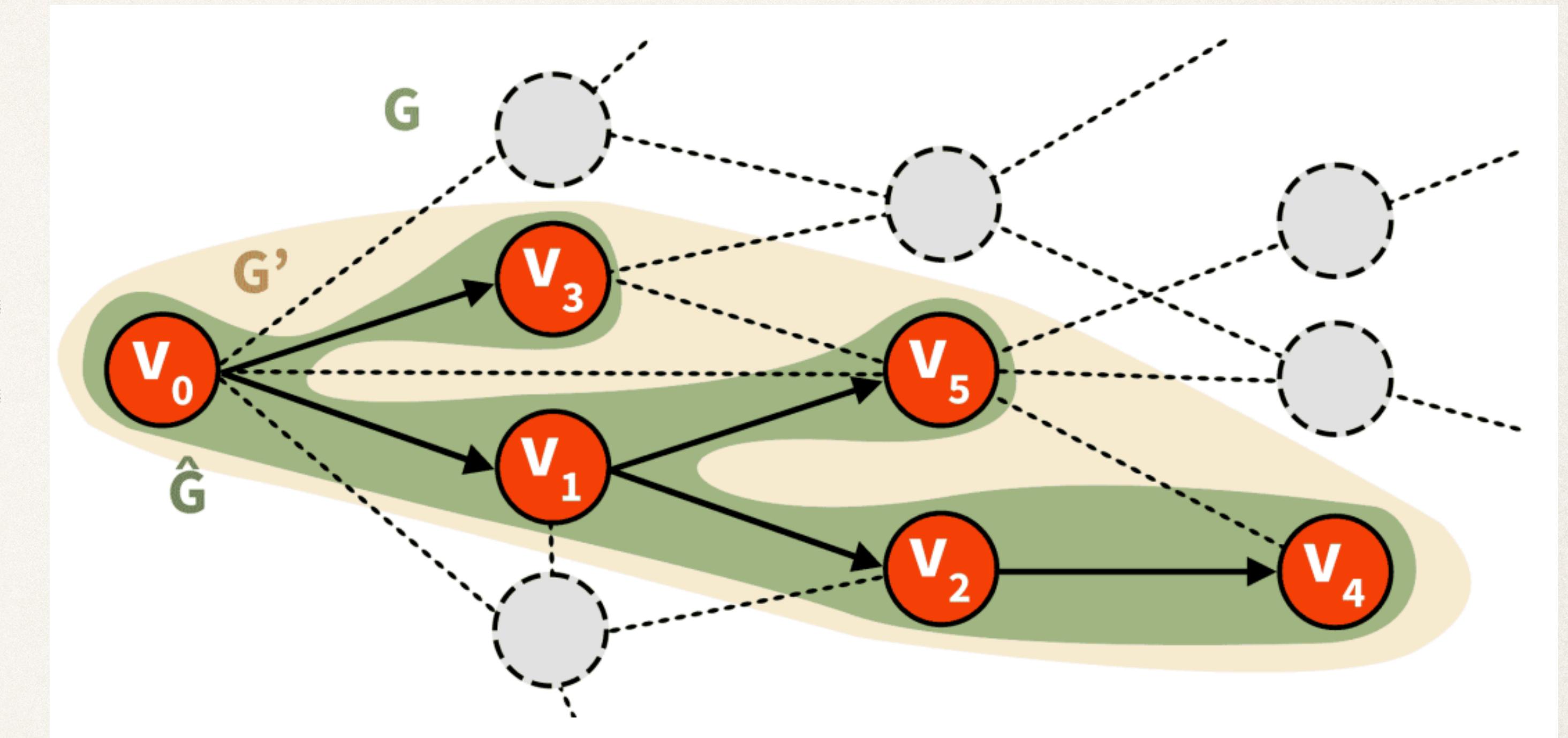
---

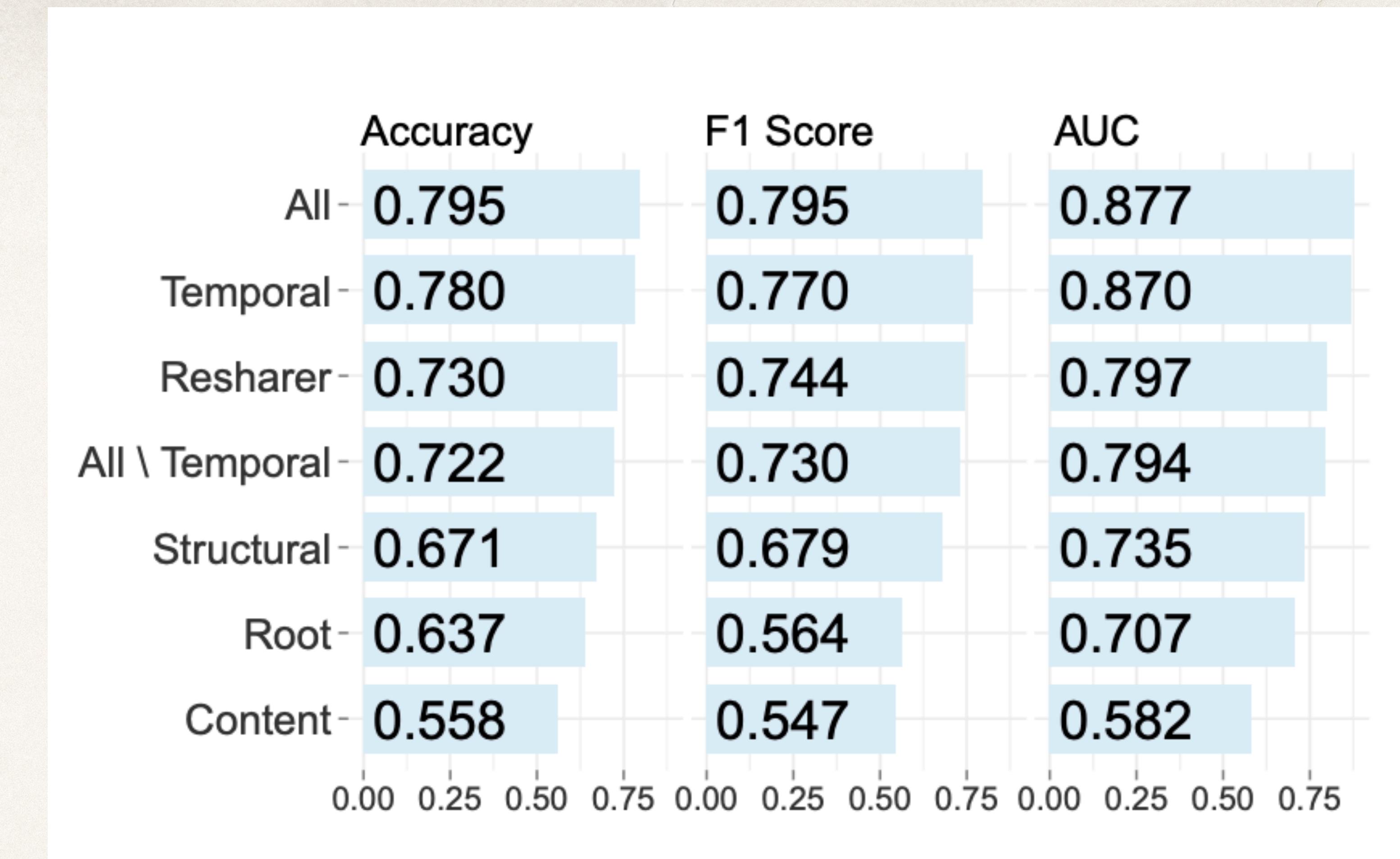
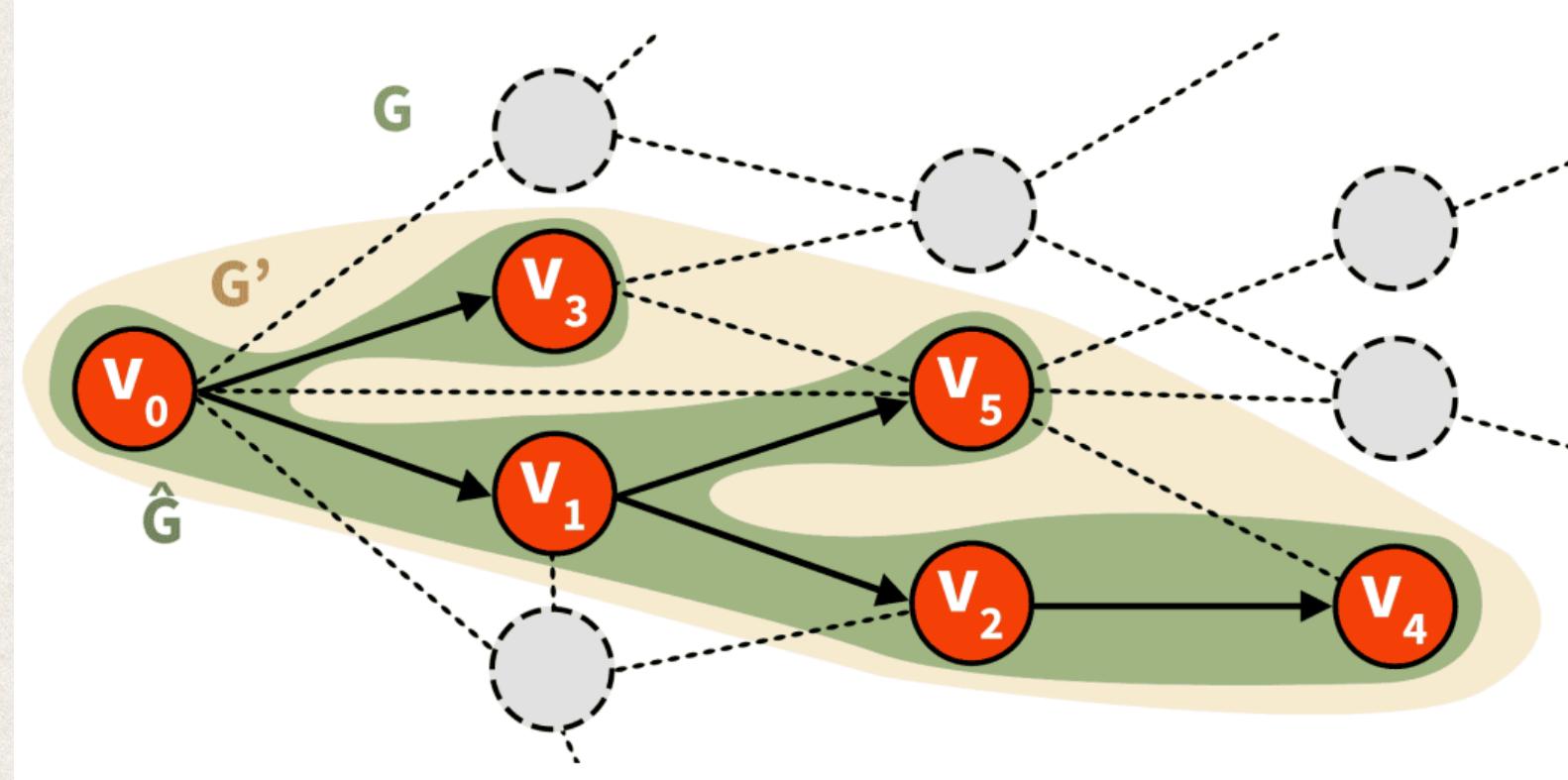
- Content



If we observe a Facebook post reshared  $k$  times, can we predict if the cascade will reach a certain size?

A cascade is a temporal sequence of events such as resharing of a post

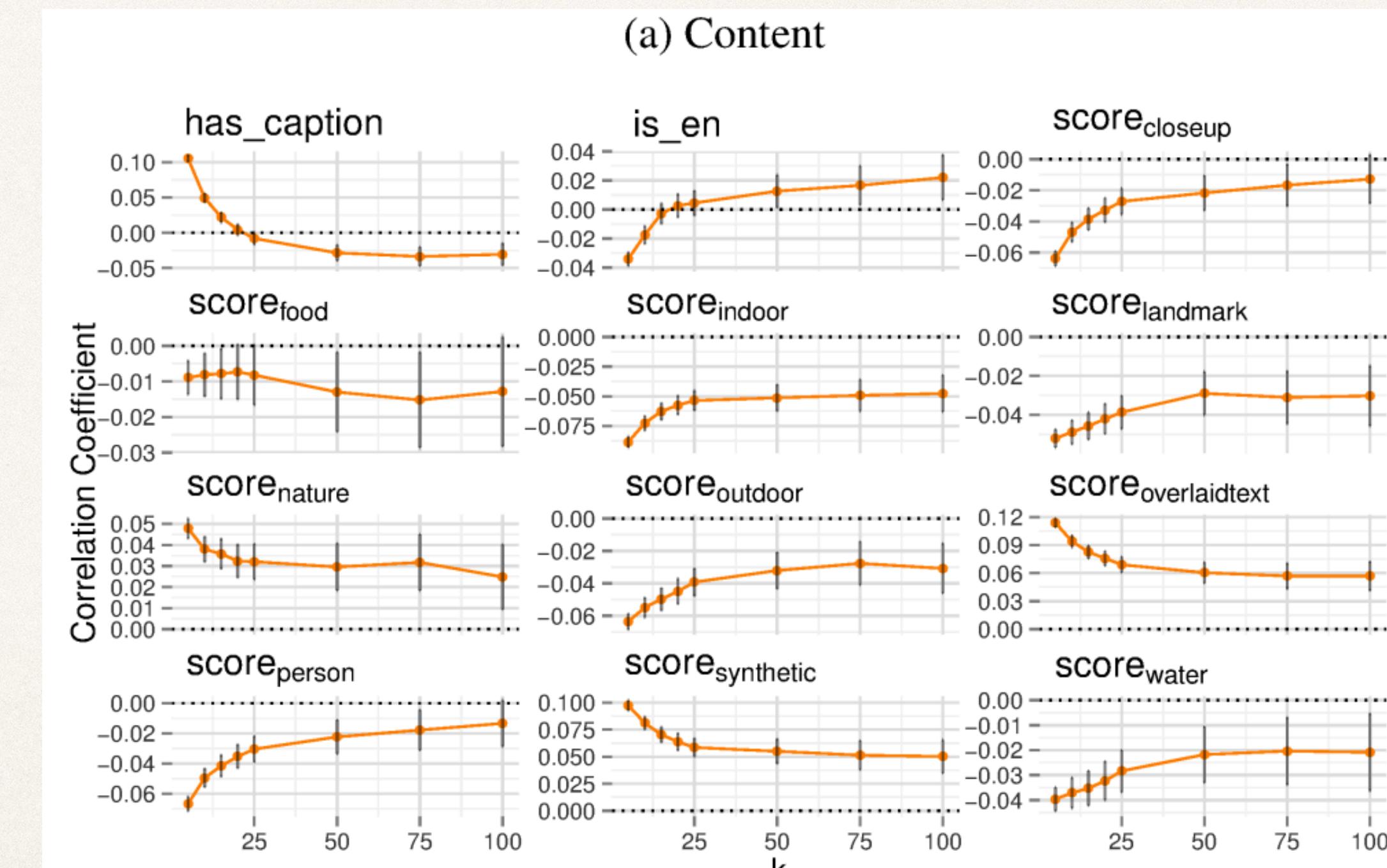


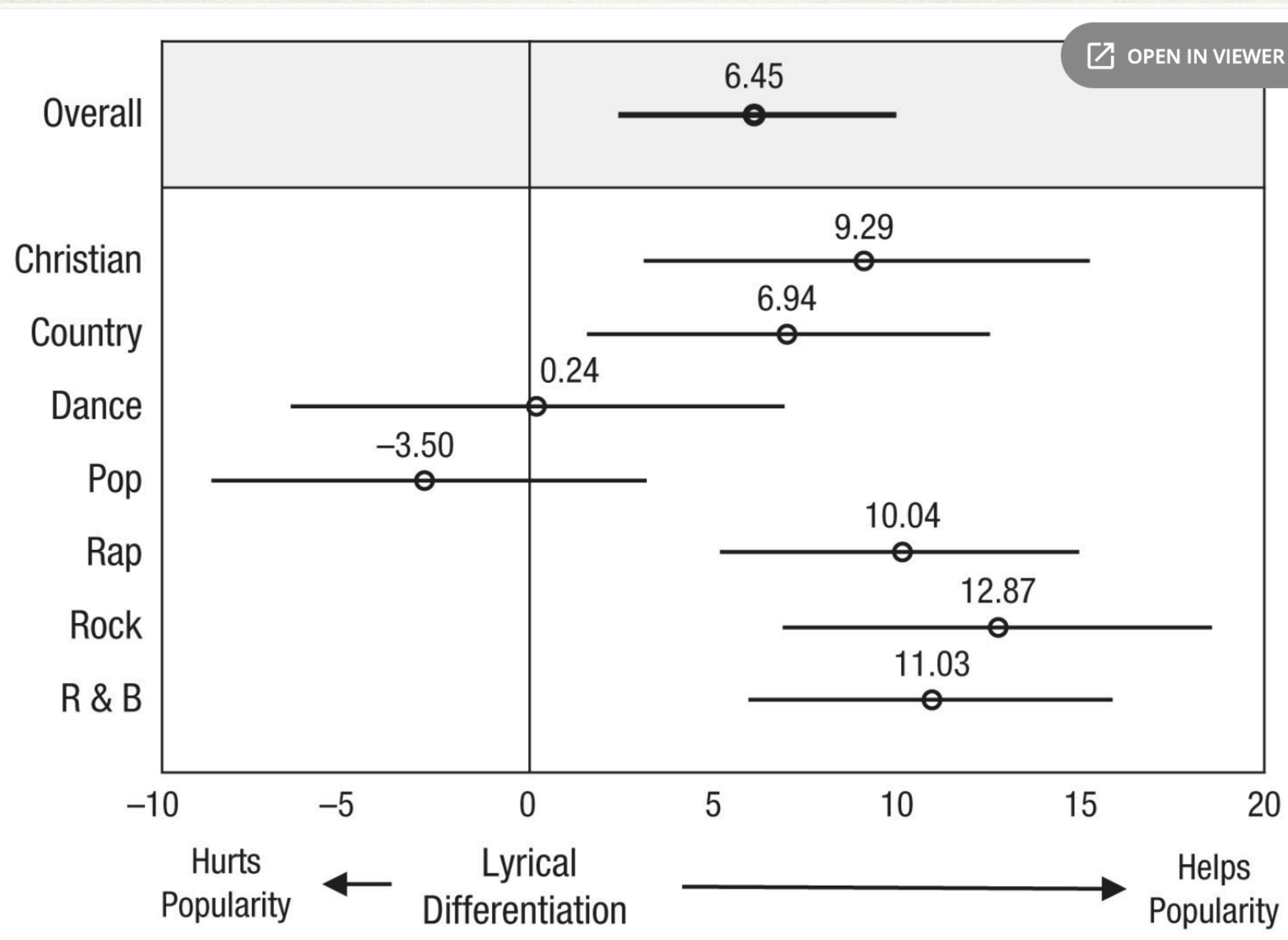


Content is an important factor in a post being reshared

## Content Features

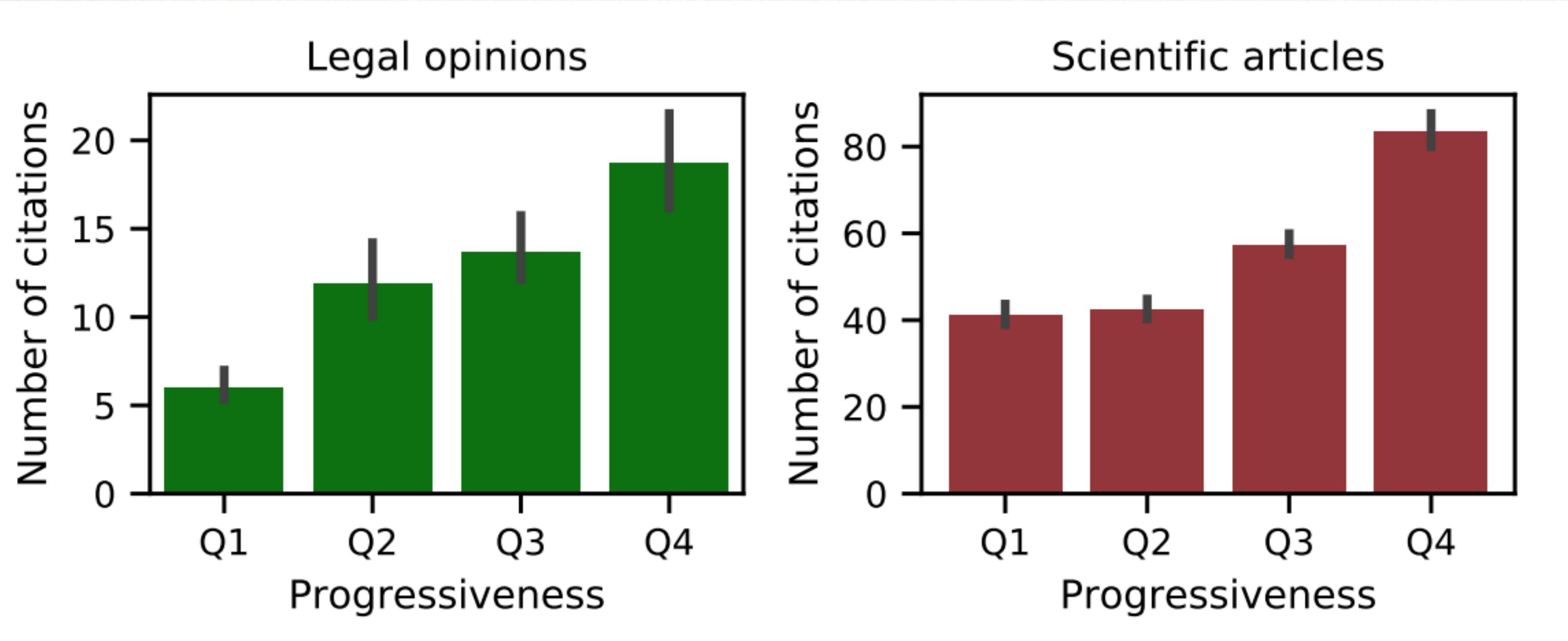
$score_{food/nature/...}$	The probability of the photo having a specific feature (food, overlaid text, landmark, nature, etc.)
$is\_en$	Whether the photo was posted by an English-speaking user or page
$has\_caption$	Whether the photo was posted with a caption
$liwc_{pos/neg/soc}$	Proportion of words in the caption that expressed positive or negative emotion, or sociality, if English





What features of a song lyric makes it popular?

Songs that are atypical of their genre tend to be more popular



Which papers/opinions have more impact?

Papers/opinions that are “ahead of their time”

# FACTORS AFFECTING POPULARITY

---

- Social connections

# FACTORS AFFECTING POPULARITY

---

- Social connections
  - We are generally part of a social network, so information flows from person to person
  - Position in a network and influence can determine popularity

# EPIDEMIC SPREAD

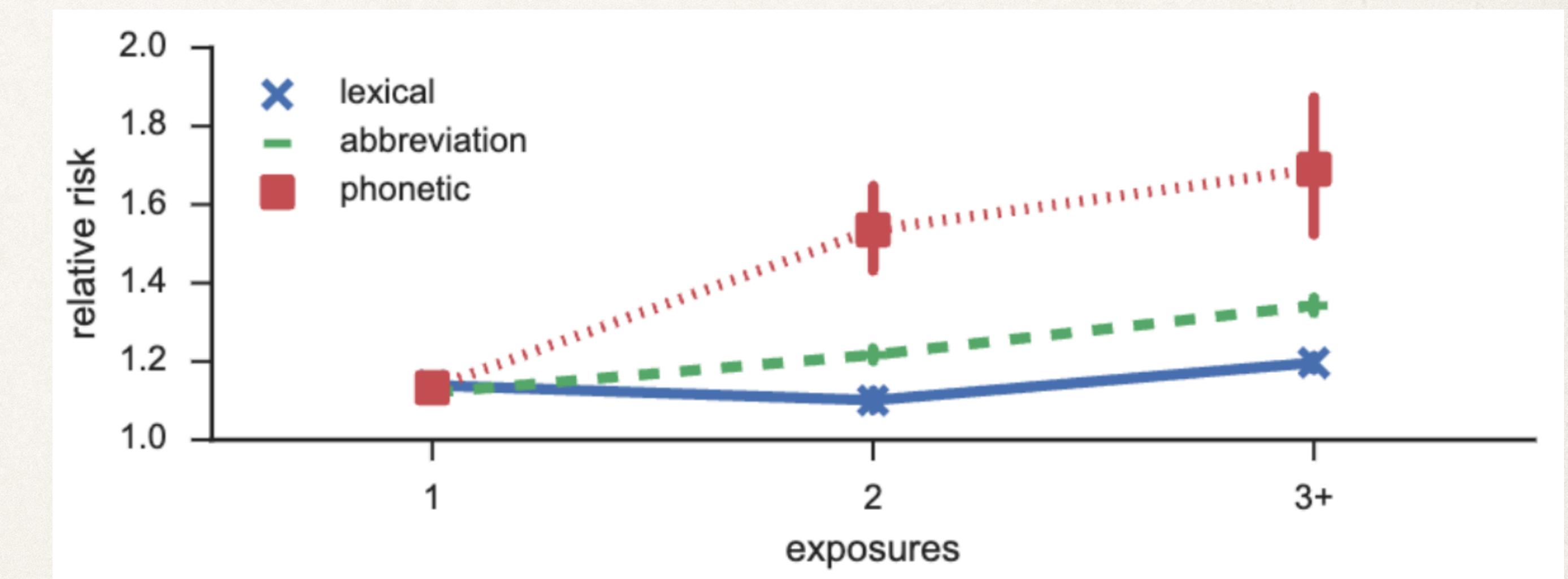
---

- Many diseases can be modeled as person-person interactions
  - Exposure: A person comes in contact with another person who carries a virus
  - Infection: The person becomes infected due to exposure

Can text popularity be thought of as a contagion?

# LANGUAGE CHANGE AS CONTAGION

- How many exposures are needed before someone adopts a new word?

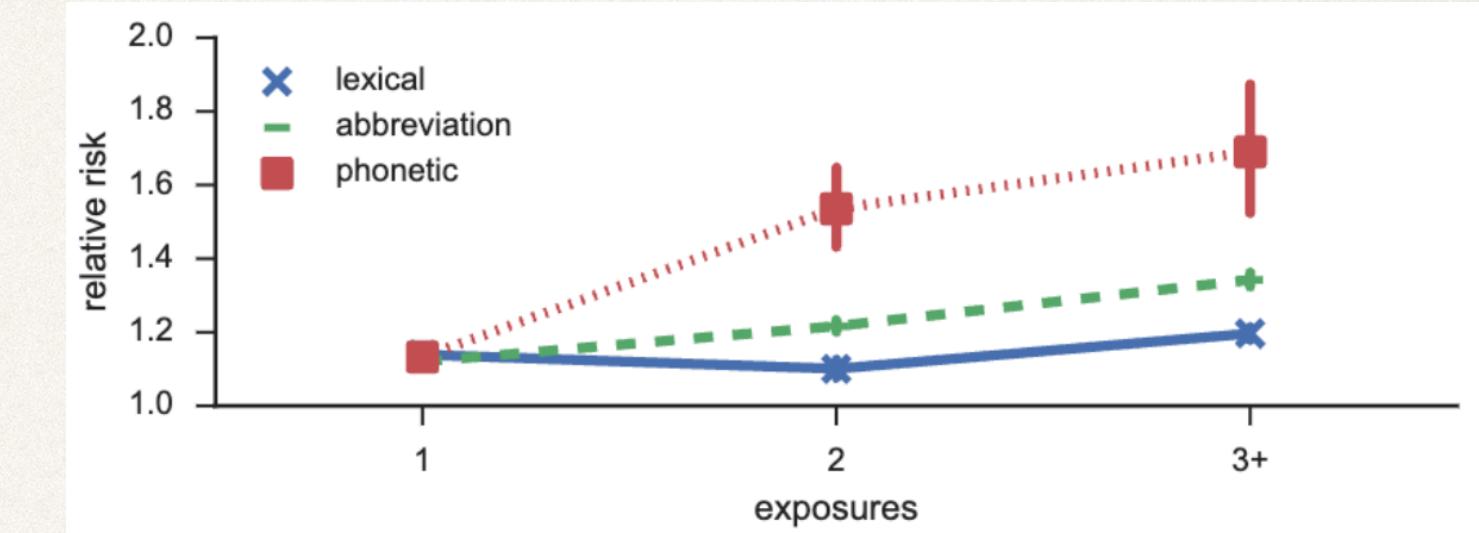


Goel et. al.

# LANGUAGE CHANGE AS CONTAGION

---

- If only one exposure is needed for someone to adopt a word for the first time, then the spread is called a simple contagion
- If multiple exposures are needed, then the spread is a complex contagion

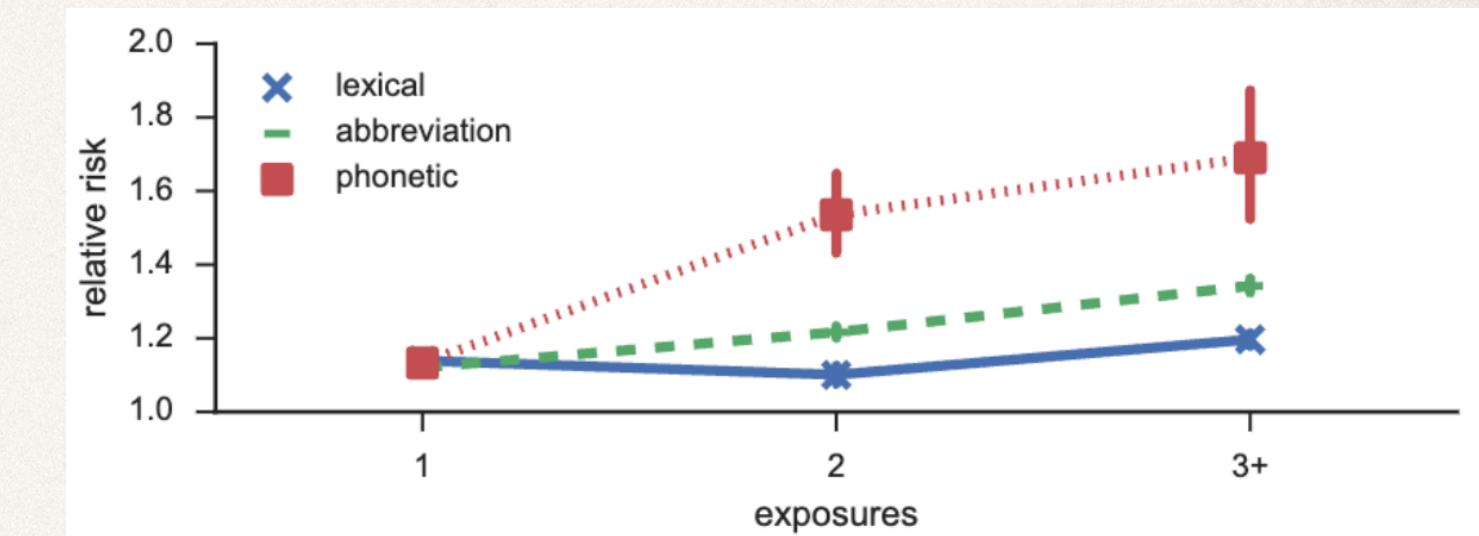


Goel et. al.

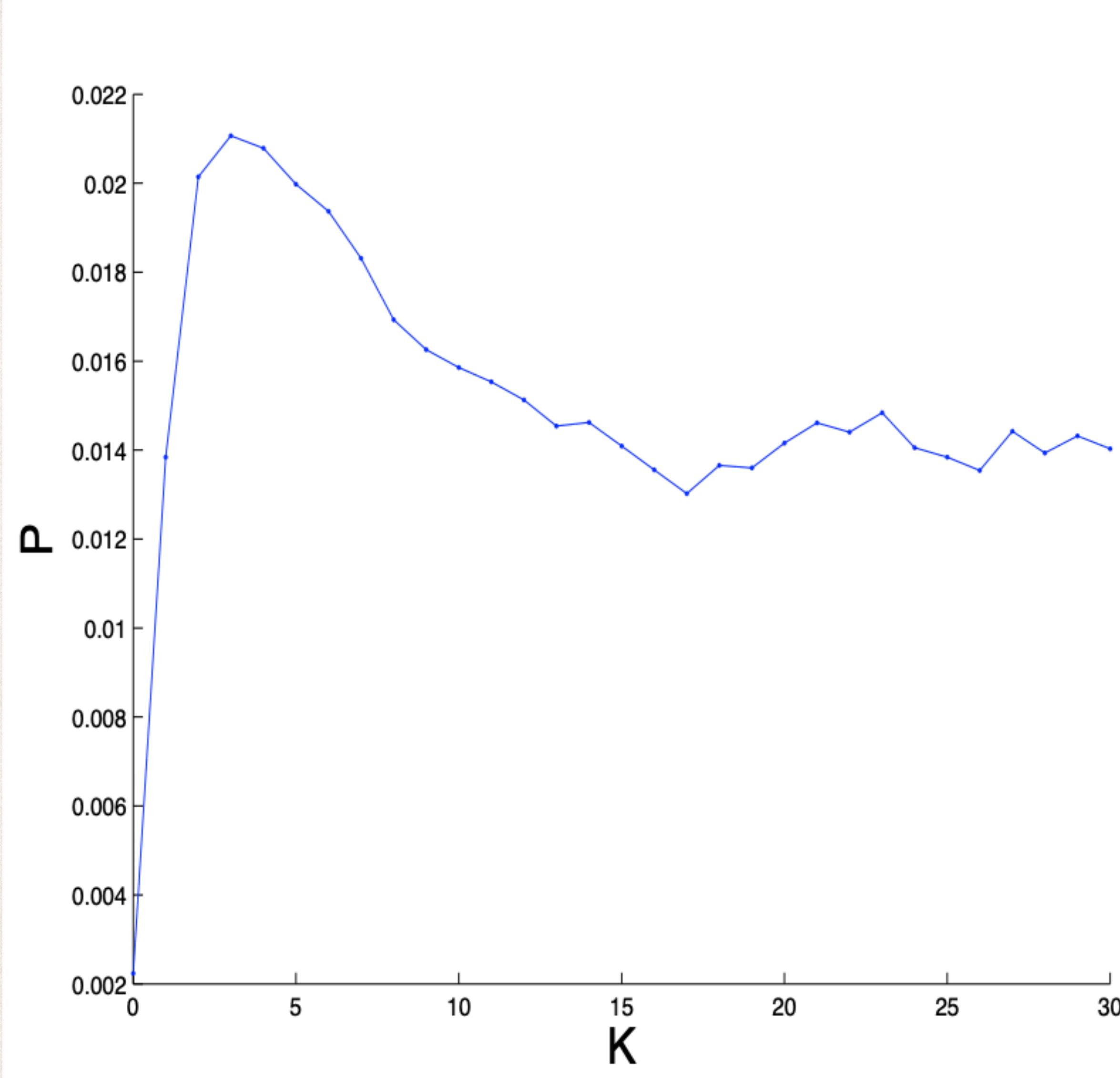
# LANGUAGE CHANGE AS CONTAGION

---

- Some words that are more familiar (e.g., *hella*) spread as a simple contagion
- Others spread that are new (e.g., *jawn*) require more exposures and hence follow complex contagion



Goel et. al.



Romero et. al. 2011

- Political hashtags on Twitter (e.g., #tcot) follow a complex contagion

# FACTORS AFFECTING POPULARITY

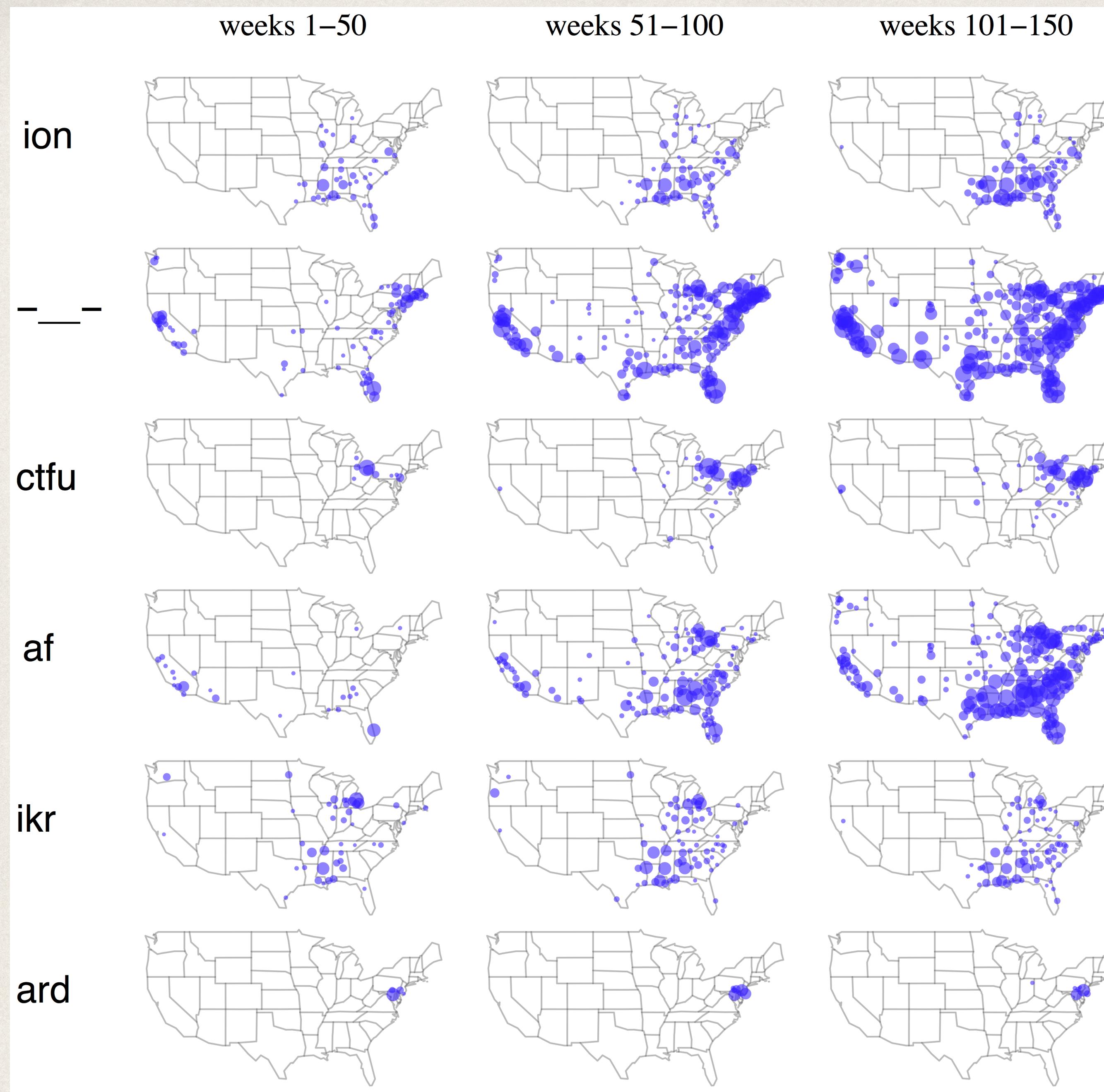
---

- Identity

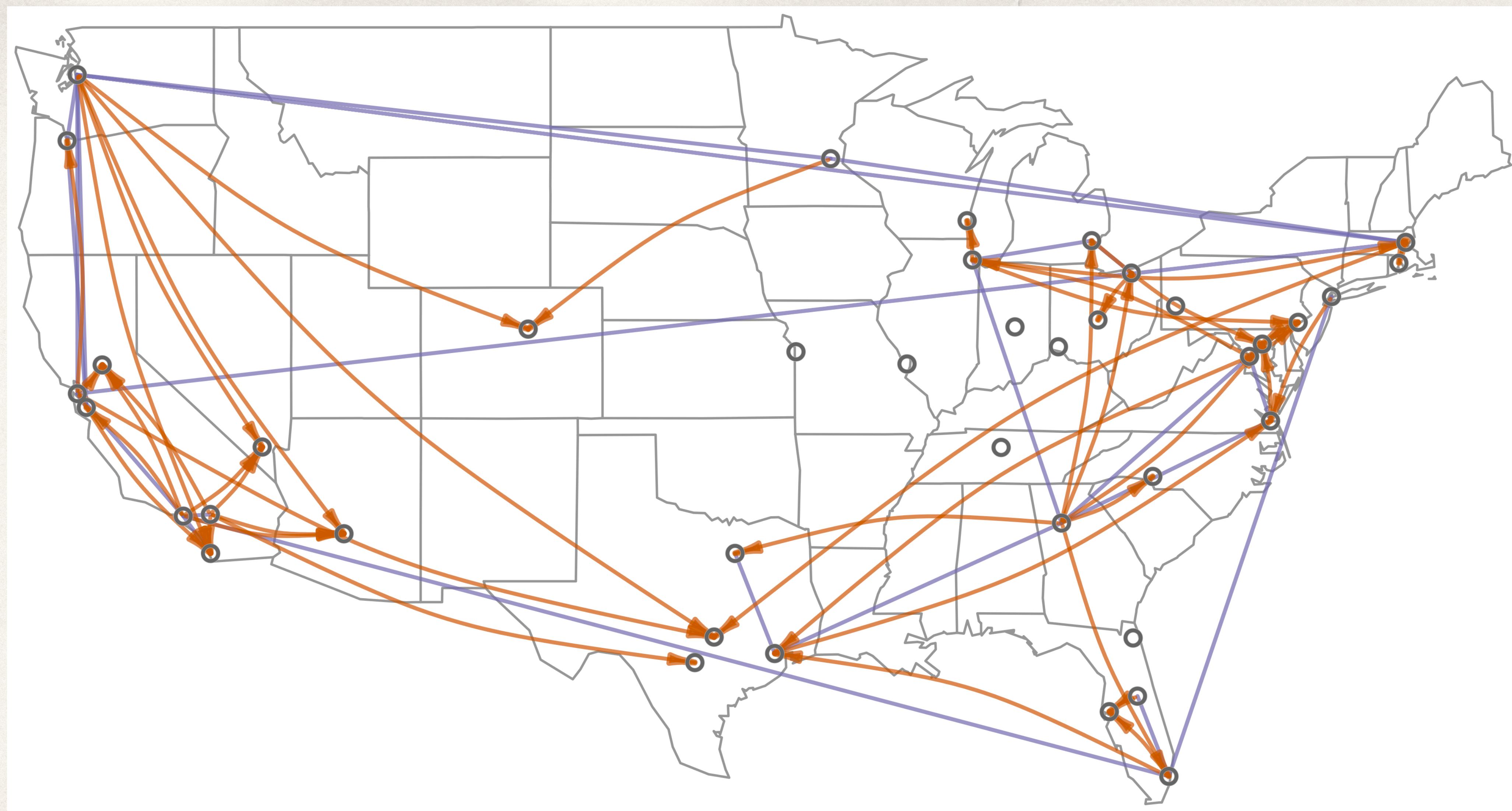
# FACTORS AFFECTING POPULARITY

---

- Identity
  - Adoption of items is often related to whether the item relates to your own identity and in response to dissimilar others.



Linguistic innovations (e.g., new words) show different diffusion patterns based on identifying characteristics of users such as their race, age, gender, and geography



Eisenstein et. al.

# TEXT AS DATA

---

- Measurement: Can we measure constructs that we're interested in using text as data?

# TEXT AS DATA

---

- Measurement: Can we measure constructs that we're interested in using text as data?

# TEXT AS DATA

---

- Representation: Can we transform text to find representations that are both mathematically amenable and have some grounding in the world?

# PROBLEMS

---

- Hypothesis testing
- Similarity and differences
- Classification and Prediction/Forecasting
- Language modeling

# MODELS

---

- Topic models
- Language models (N-gram, Transformers, RNN)
- Word2Vec

# TECHNOLOGIES

---



spaCy

PyTorch



HUGGING FACE



**THANK YOU!**