



# **ASME** IDETC-CIE 2023

International Design Engineering Technical Conferences  
& Computers and Information in Engineering Conference

---

CONFERENCE AUGUST 20-23, 2023



**SIDI LAB**  
System Integration  
& Design Informatics Laboratory



**IDEAL**  
Integrated Design  
Automation Laboratory

# Evolutionary Co-Mention Network Analysis via Social Media Mining

IDETC/CIE 2023 - 115114

Phillip A. O. Gavino<sup>1</sup>, Yinshuang Xiao<sup>1</sup>, Yaxin Cui<sup>2</sup>, Wei Chen<sup>2</sup>, Zhenghui Sha<sup>1</sup>

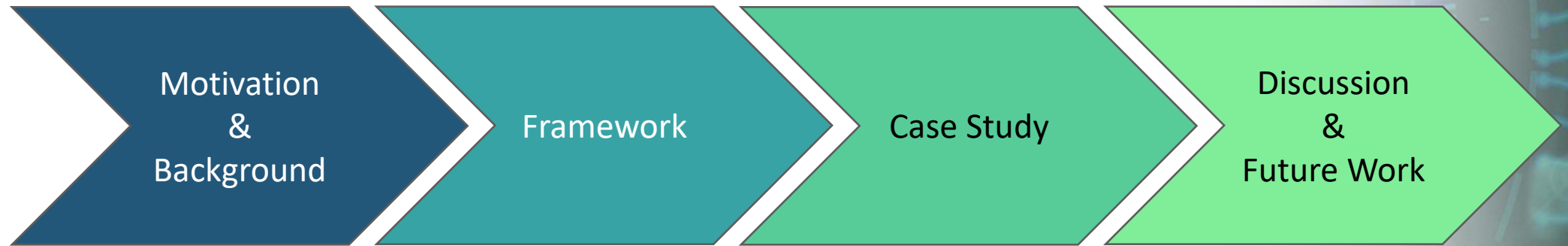
<sup>1</sup>Walker Department of Mechanical Engineering, University of Texas at Austin

<sup>2</sup>Department of Mechanical Engineering, Northwestern University





# Overview



# Background

Understanding competitive relationships of market systems is essential for product design



Quantitatively presenting competitive relationship

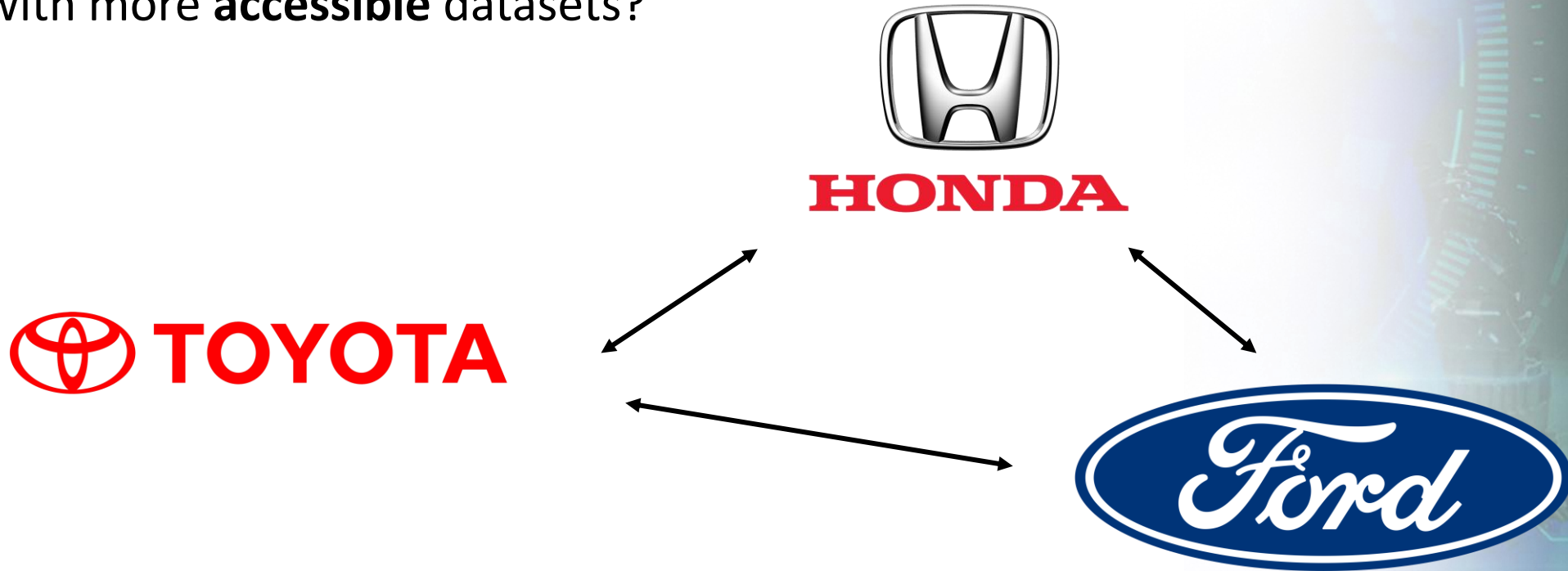


Most significant challenge is lack of data



# Motivating Question

How can we **quantitatively** represent/model/predict competitive relationships between brands/companies with more **accessible** datasets?



# Main Idea

## Social Media Data Sources



- Free large sources of data

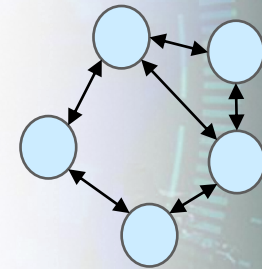
## Co-mention Relationships

$\text{Entity}_1 \leftrightarrow \text{Entity}_2$

Honda Civic  $\leftrightarrow$  Toyota Camry

- Two entities existing in the form of text

## Network Representation

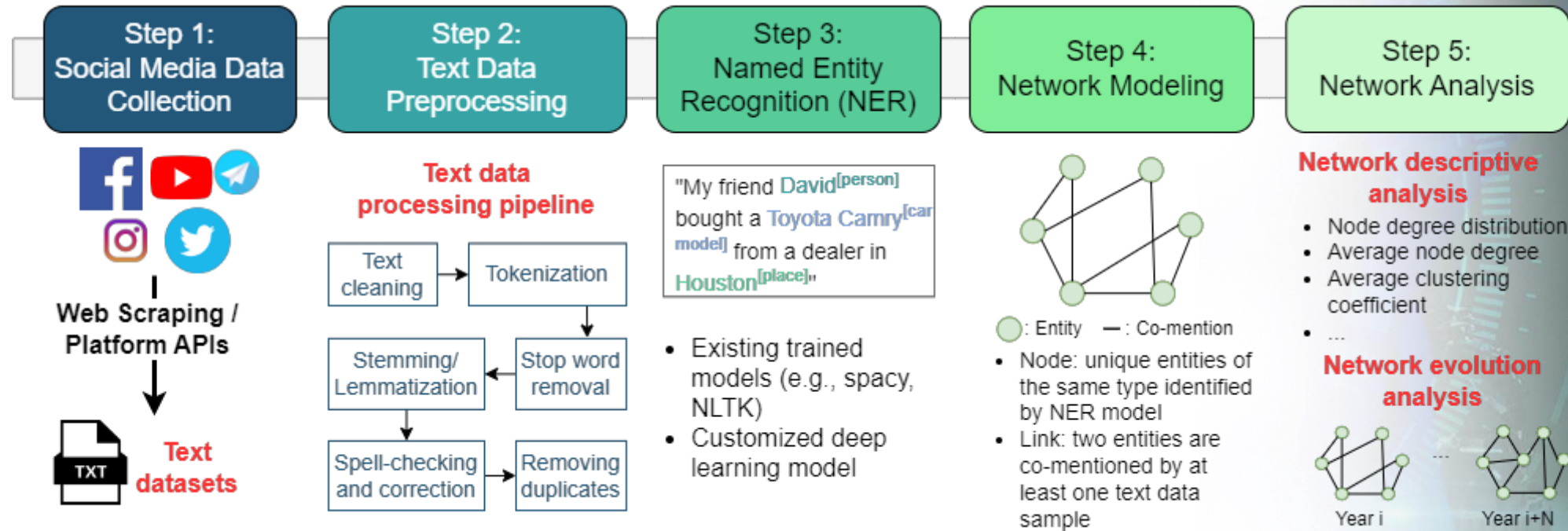


- Analyze relationships with metric results
- Interpret findings to the broader complex system





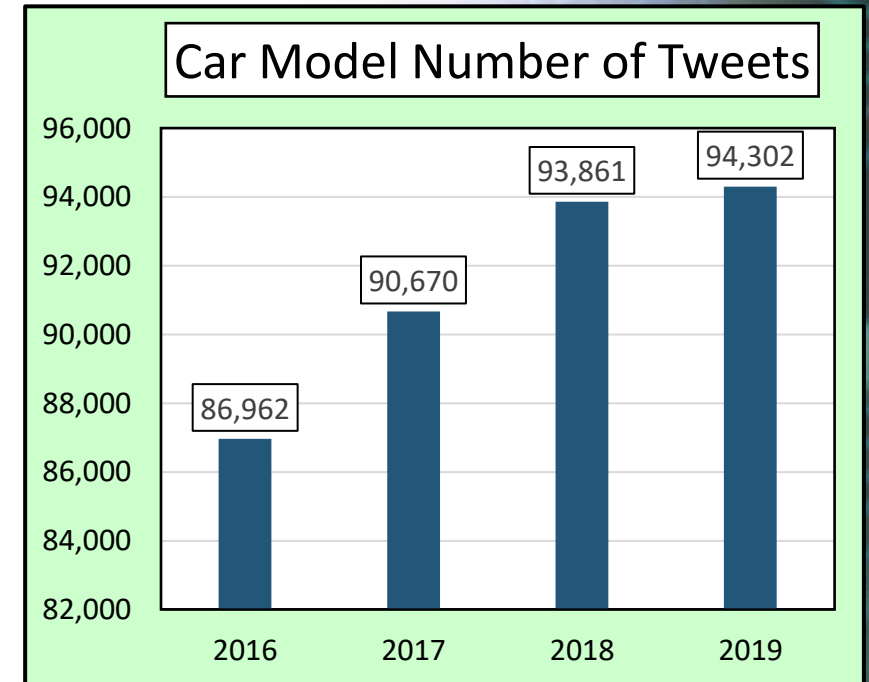
# Framework for Content-Based Data Mining



# Case Study: Data Preparation

## Data Scraping

- 949 Unique Car Models **from Cars.com**
- Tweets from 2016 – 2019
- Possible Maximum Number of Tweets Collected Per Year:
  - 240 Tweets for **each car model**





# Case Study: Data Preprocessing

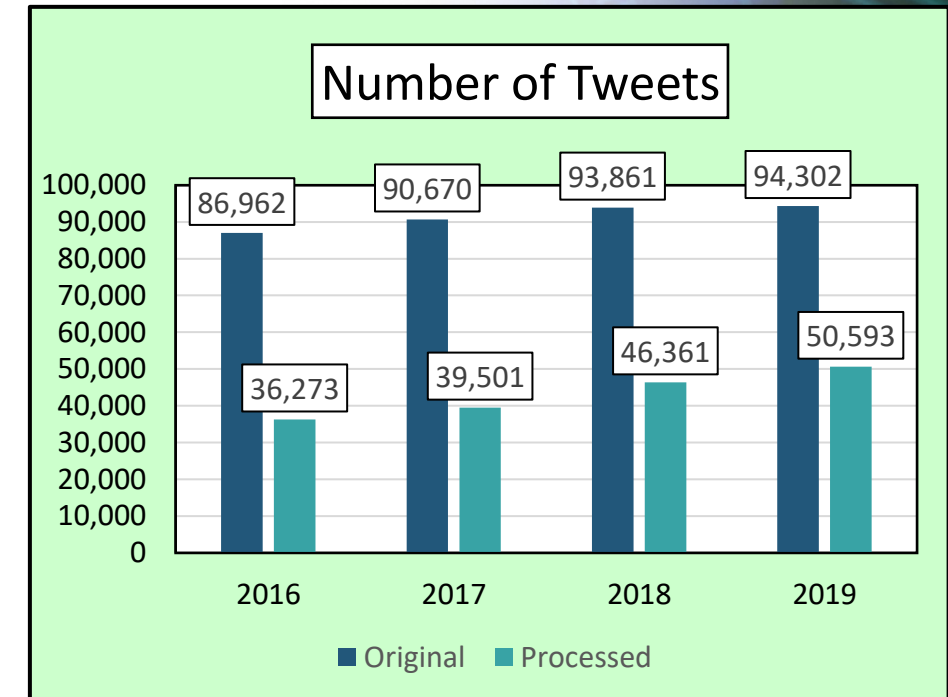
**Remove Duplicates**  
(Sequence Matching 60% threshold)

Raw

03 Audi A6 the (BEST  
OFFER) (San Diego) \$1500  
ift.tt/2FH6Yoo  
<https://t.co/AQ9cgwfoHR>

Processed

03	audi	a6
best	offer	san
diego	1500	



# Named Entity Recognition (NER) Process

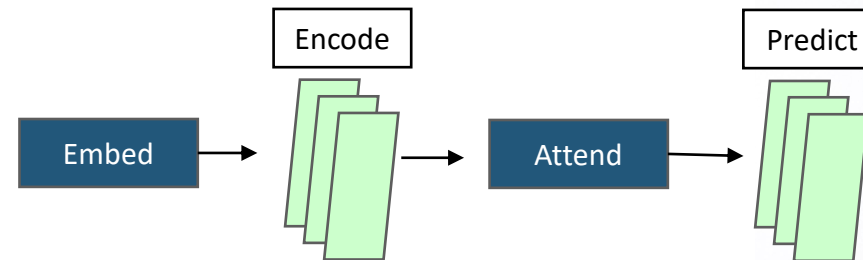
## Training Data

- Tweet 1: "lexus lc 500<sup>[CAR]</sup> save get porsche 911 gts<sup>[CAR]</sup>"
- Tweet 2: "say goodbye my old toyota rav4<sup>[CAR]</sup> thinking buy new ford f150<sup>[CAR]</sup> chevy silverado 1500<sup>[CAR]</sup>"
- Tweet 3: "my friends have toyota rav4<sup>[CAR]</sup> lexus lc 500 <sup>[CAR]</sup> want subaru wrx<sup>[CAR]</sup> bad raelene first need learn drive"

~2000 marked tweets

## NER Model

## Entity Predictions



# Case Study: Prediction Metrics

Accurate

my friend agreed  
toyota camry better  
acura ilx honda accord

toyota camry	acura ilx	honda accord
-----------------	-----------	-----------------

Inaccurate

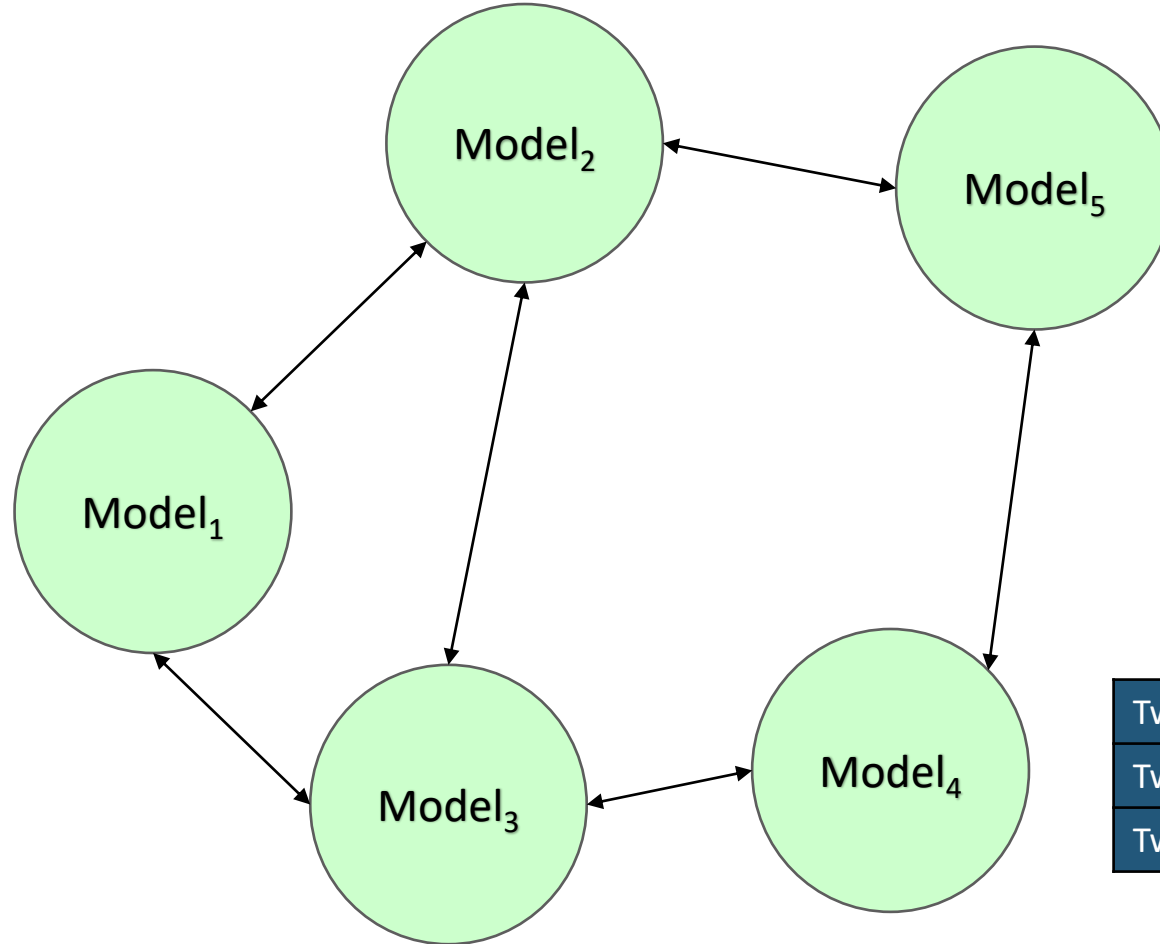
my friend agreed  
toyota camry better  
acura ilx honda accord

toyota camry	acura tlx	
-----------------	-----------	--

Year	F1-Score	Precision	Recall
2016	73.25%	80.42%	67.26%
2017	71.50%	77.67%	66.23%
2018	74.83%	74.03%	75.67%
2019	69.96%	74.37%	66.04%



# Case Study: Network Definition



Link:

- Co-mention

Node:

- Car Model

	Model <sub>1</sub>	Model <sub>2</sub>	Model <sub>3</sub>
Tweet <sub>1</sub>	Model <sub>1</sub>	Model <sub>2</sub>	Model <sub>3</sub>
Tweet <sub>2</sub>	Model <sub>3</sub>	Model <sub>4</sub>	
Tweet <sub>3</sub>	Model <sub>4</sub>	Model <sub>5</sub>	

Source	Target
Model <sub>1</sub>	Model <sub>2</sub>
Model <sub>1</sub>	Model <sub>3</sub>
Model <sub>2</sub>	Model <sub>3</sub>



# Case Study: Competition Analysis

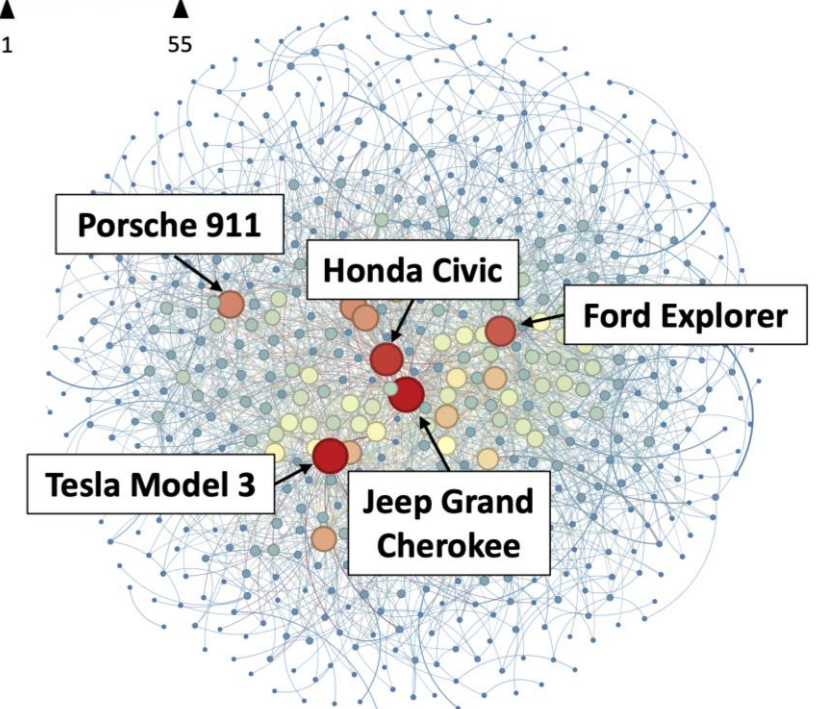
Top five car models with the largest unweighted degree.

2016		2017		2018		2019	
Model	UWD	Model	UWD	Model	UWD	Model	UWD
Honda Civic	19	Nissan Leaf	19	Ford Focus	50	Jeep Grand Cherokee	55
Mini Cooper	18	Jeep Wrangler	16	Toyota RAV4	37	Tesla Model 3	55
Audi R8	18	Honda Civic	15	Porsche 911	33	Honda Civic	51
Porsche 911	16	BMW X1	15	Honda Civic	32	Ford Explorer	47
Nissan Leaf	14	Mini Cooper	14	Nissan Leaf	31	Porsche 911	42

The most frequently co-mentioned car pairs by year.

Year	Linked Car Models	# of Co-mentions
2016	BMW X5 vs. Volvo XC90	6
2017	Tesla Model 3 vs. Chevrolet Bolt EV	8
2018	Buick Envision vs. Cadillac CT6	13
2019	Jeep Wrangler vs. Jeep Wrangler Unlimited	11

# of Connections



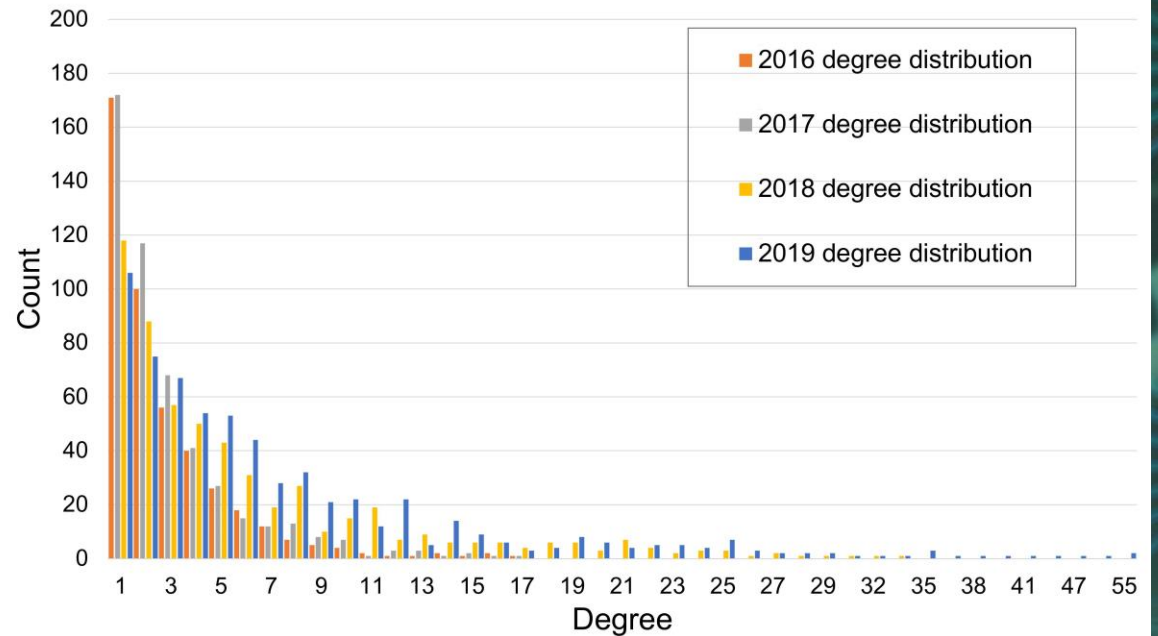
2019 Twitter Automobile Co-mention Network  
(# of Nodes: 640, # of Edges: 2289)



# Case Study: Network Analysis

Twitter co-mention network metrics by year.

Year	2016	2017	2018	2019
Density	0.007	0.006	0.011	0.012
Unweighted Avg. Deg.	2.913	3.037	6.079	7.466
Weighted Avg. Deg.	3.523	3.793	8.047	9.934
Avg. Local Cluster Coeff.	0.125	0.128	0.229	0.249



Twitter co-mention network degree distributions.





# Contribution & Limitations

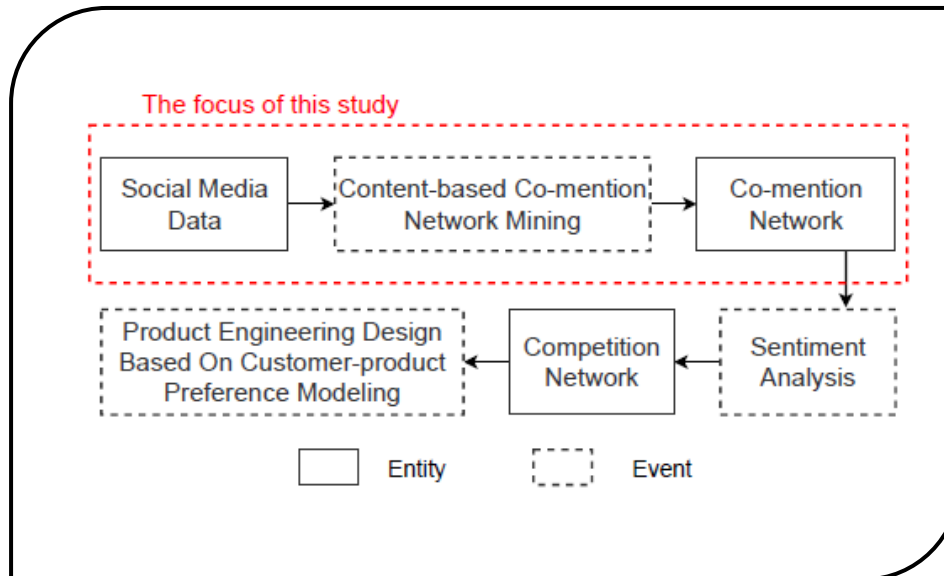
## Contributions

- Track changes in customer preferences
- Provide insights into customer perceptions and potential market competition structures

## Limitations

- Inability to standardize vehicle model names
  - Reduces the number of possible co-mentions
- Limited spam detection

# Summary & Future Work



## Future Work

- Sentiment analysis
- Predictive network model (e.g., ERGM, GNN)
- Design optimization/feature engineering





# Thank You!

## Acknowledgements

National Science Foundation (CMI #2005661 and #2203080)

Email: [phillip.gavino@utexas.edu](mailto:phillip.gavino@utexas.edu)

