Congruent or Polarized? Mining Opinion Dynamics towards Cryptocurrency on Twitter

Keywords: *Opinion Dynamics, Polarity, Cryptocurrency, Twitter*

Mintao Wei
mintao@uchicago.edu

CONTENT

1 Research Questions

2 Data and Methods

3 Results

4 Summary

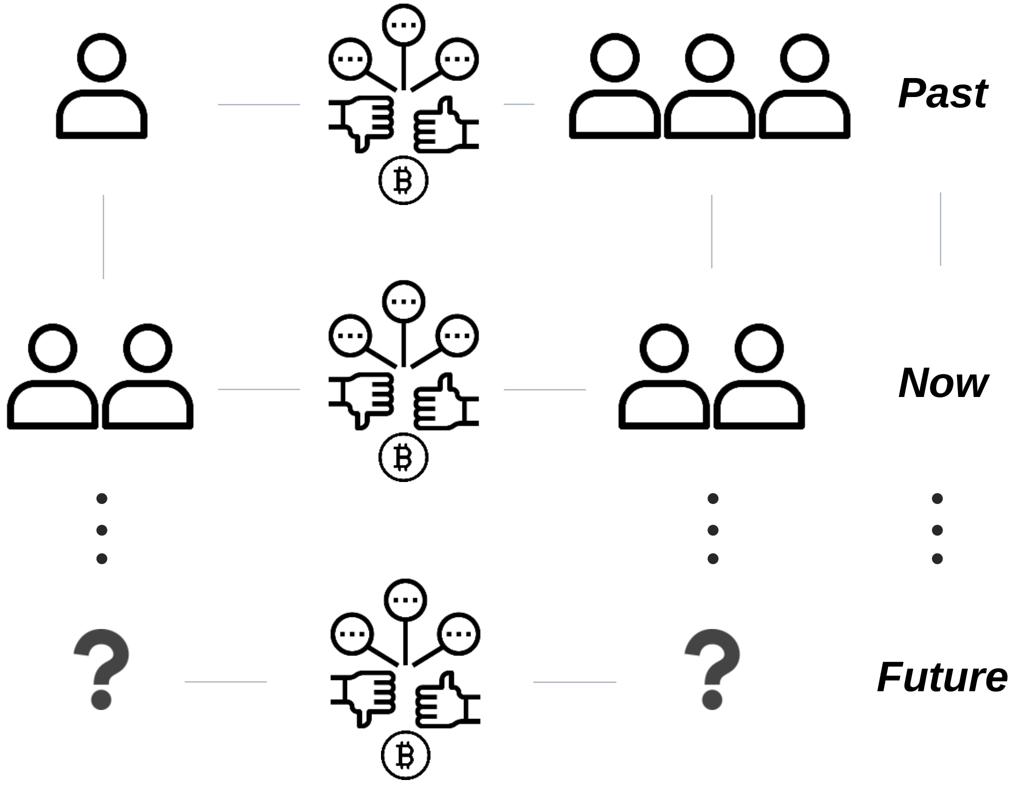
01

Research Questions

What are the research questions?

1. Research Questions

Framework



Questions

RQ1: How polarized are people's opinions towards cryptocurrency? (Overview)

RQ2: How does the polarity of cryptocurrency opinions on Twitter evolve over time?

RQ3: Can we effectively predict future opinion polarity on Twitter toward cryptocurrency?

02

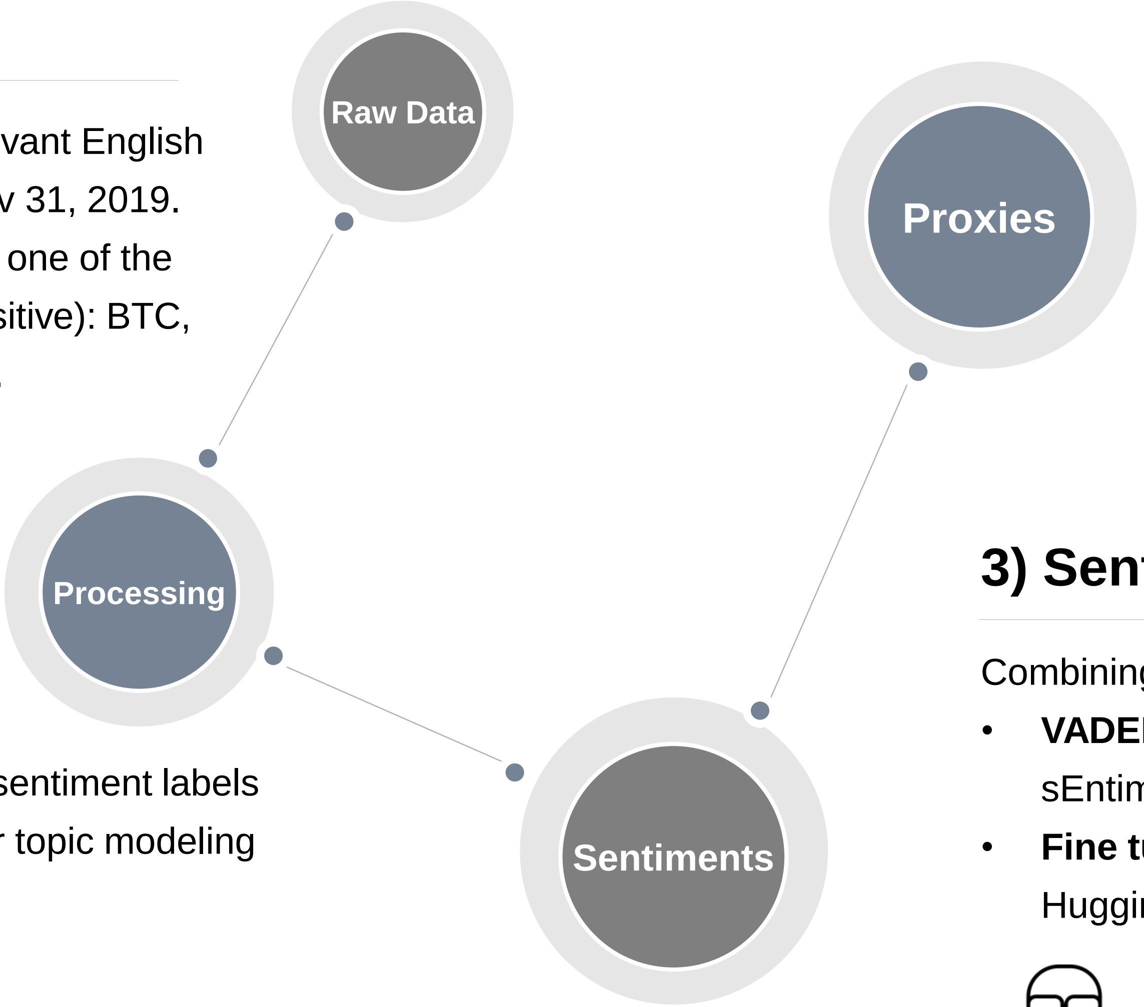
Data and Methods

What data and methods are used to answer the research questions?

2-1. Data

1) Data Source

The raw data is 4.75 million relevant English tweets from Aug 21, 2019 to Nov 31, 2019. All these tweets contain at least one of the following keywords (case insensitive): BTC, bitcoin, crypto, ETH, memecoin.



2) Processing

- Weigh the retweets
- Light textual processing for sentiment labels
- Heavy textual processing for topic modeling

3) Sentiment Classification

Combining 2 approaches to ensure robustness

- **VADER** (Valence Aware Dictionary and sEntiment Reasoner)
- **Fine tuned Pre-trained BERT** from Hugging Face

Hugging Face

4) Construct Proxies

$$\frac{\text{\# Positive/Negative Tweets}}{\text{\# All Crypto-related Tweets}}$$

2-2. Methods

RQ1: How polarized are people's opinions towards cryptocurrency?

- **Sentiment Classification and Data Analysis**
  - VADER
  - Fine-tuned pre-trained BERT
- **Representing Tweets**
  - TF-IDF word embedding + t-SNE & PCA
- **Topic Modeling**
  - LDA and Word Cloud
- **Girven Newman Community Detection on Hashtags**
  - Removing 'statistical insignificant' edges (Martinez-Romo et al., 2011)
  - Combine similar hashtags (difflib module)

RQ2: How does the polarity of cryptocurrency opinions on Twitter evolve over time?

- **Time Series Data Analysis**

RQ3: Can we effectively predict future opinion polarity on Twitter toward cryptocurrency?

- **Time Series Prediction**
  - ARMA
  - Exponential Smoothing (Simple, Double, Triple)
  - Facebook Prophet

03

Results

What are the findings?

3-1-1. Sentiment Data Analysis – RQ1 (Polarization Landscape)

Table 1: Overview of Cryptocurrency Sentiments and Opinions

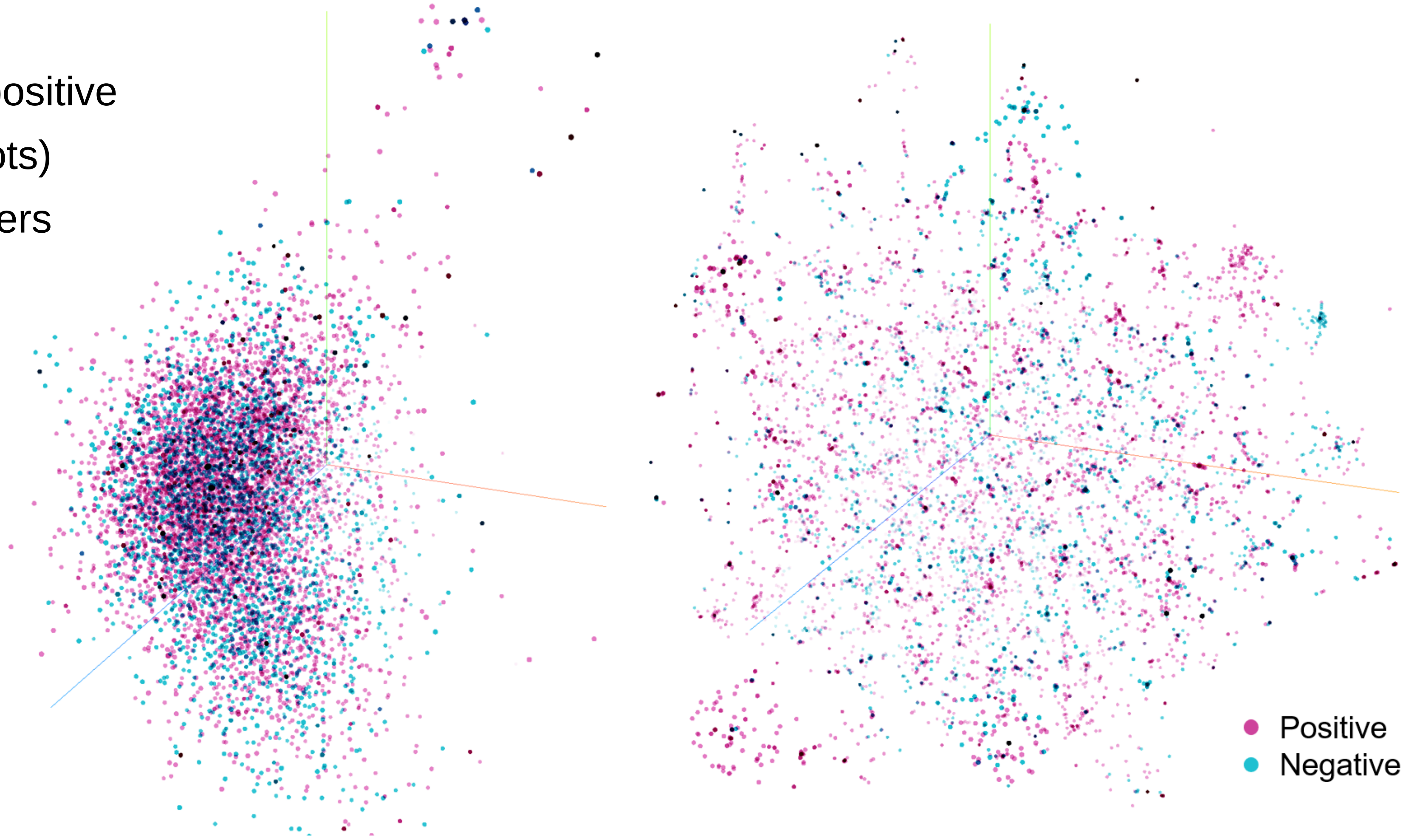
Sentiment Class	# Tweets(RT)	% Tweets(RT) (1)	Sentiment Intensity (2)	Avg. Likes per Tweet	Avg. Retweet per Tweet (3)
Negative	1,273,736	27%	0.38	3.44	0.70
Positive	3,476,861	73%	0.56	3.85	1.03
Total	4,750,597	100%	0.47	3.73	0.93
Polarity	2,203,125	46%	0.18	0.41	0.33

Note: Time Window: Aug 21, 2019 Ñov 30, 2019. The total number of tweets metrics exclude the original neutral tweets and include the retweet count. The sentiment intensity score produced by VADER ranges between 0 and 1 and has a larger value when the sentiment is strong.

3-1-2. Representing Tweets – RQ1 (Polarization Landscape)

Findings:

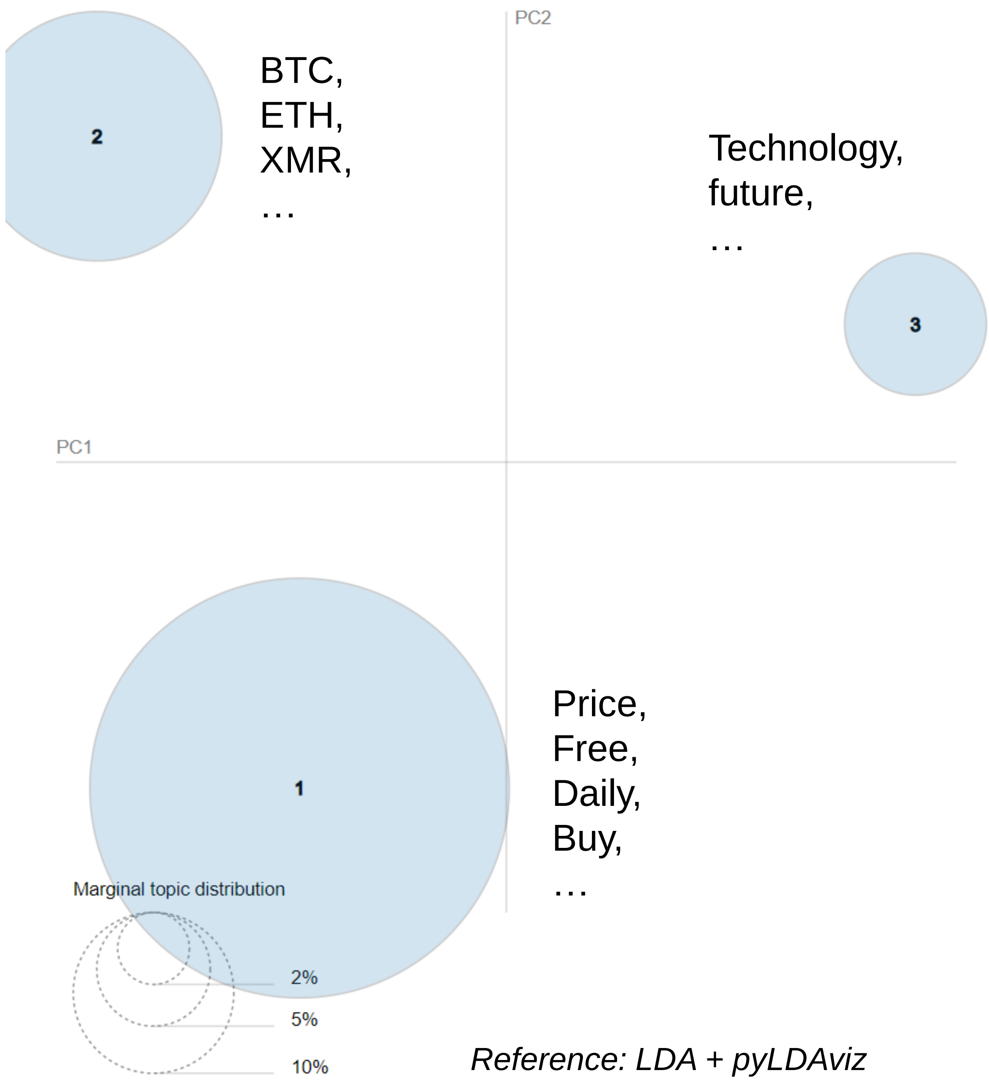
1. Again, more positive tweets (red dots)
2. No clear clusters

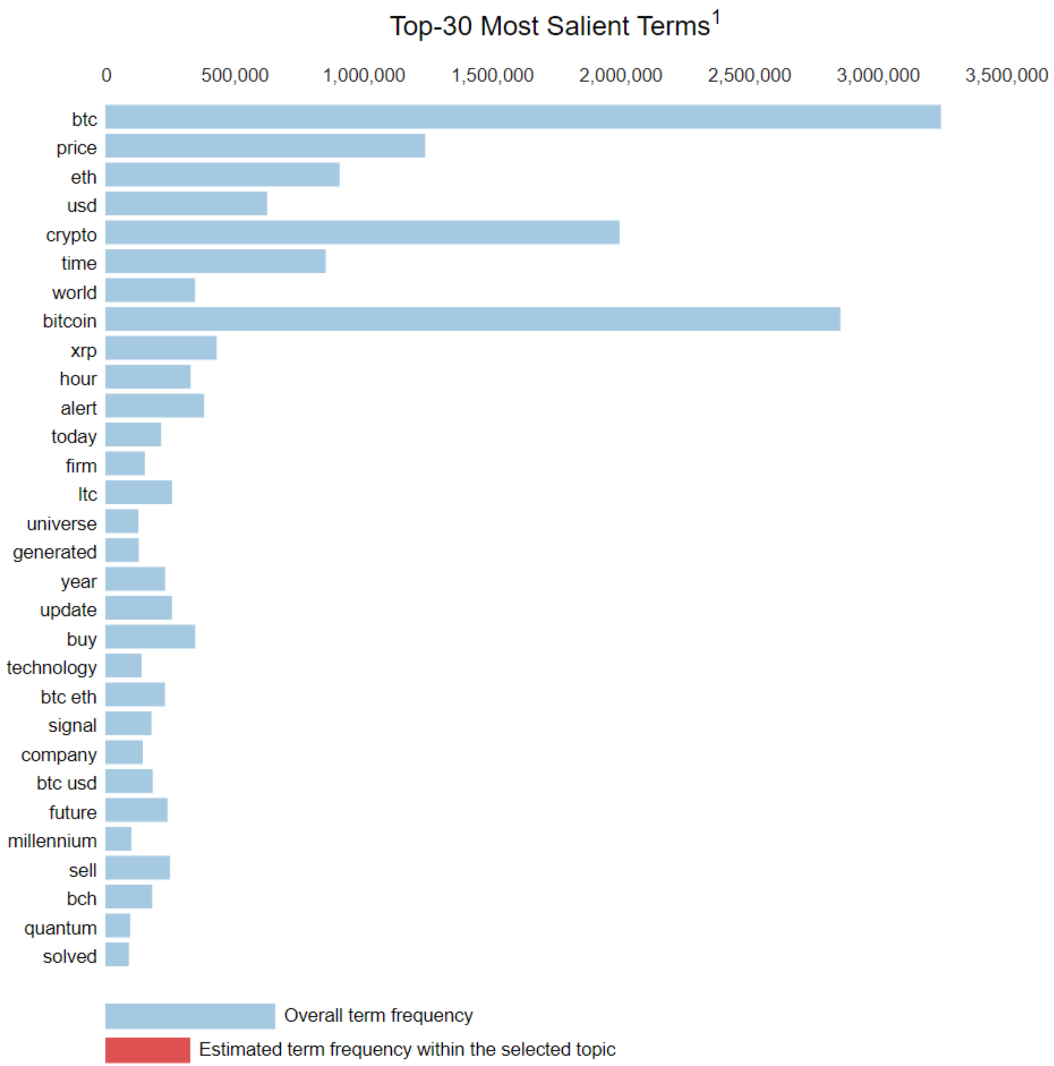


(a) PCA (b) t-SNE

Source: [Tensorflow Projector]

3-1-3. Topic Modeling – RQ1 (Polarization Landscape)





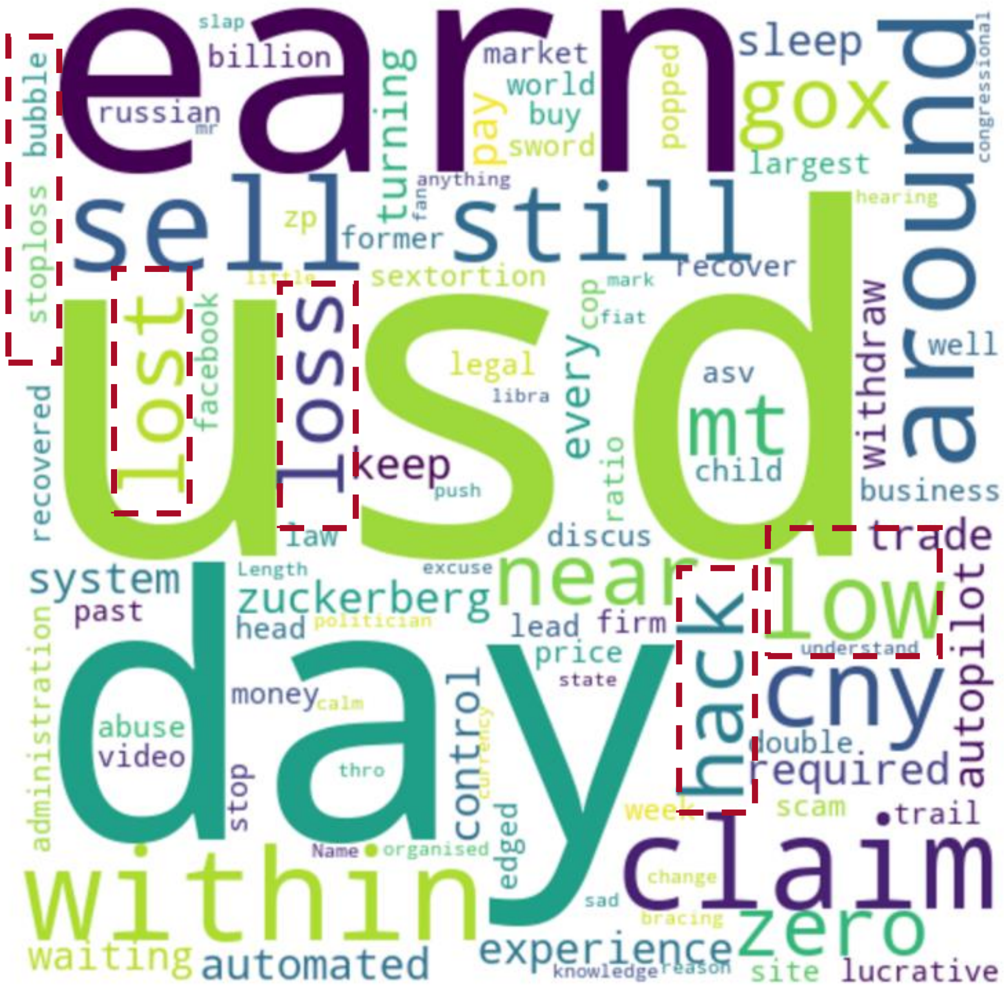
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Reference: LDA + pyLDAviz

3-1-4. Topic Modeling (cont.) – RQ1 (Polarization Landscape)

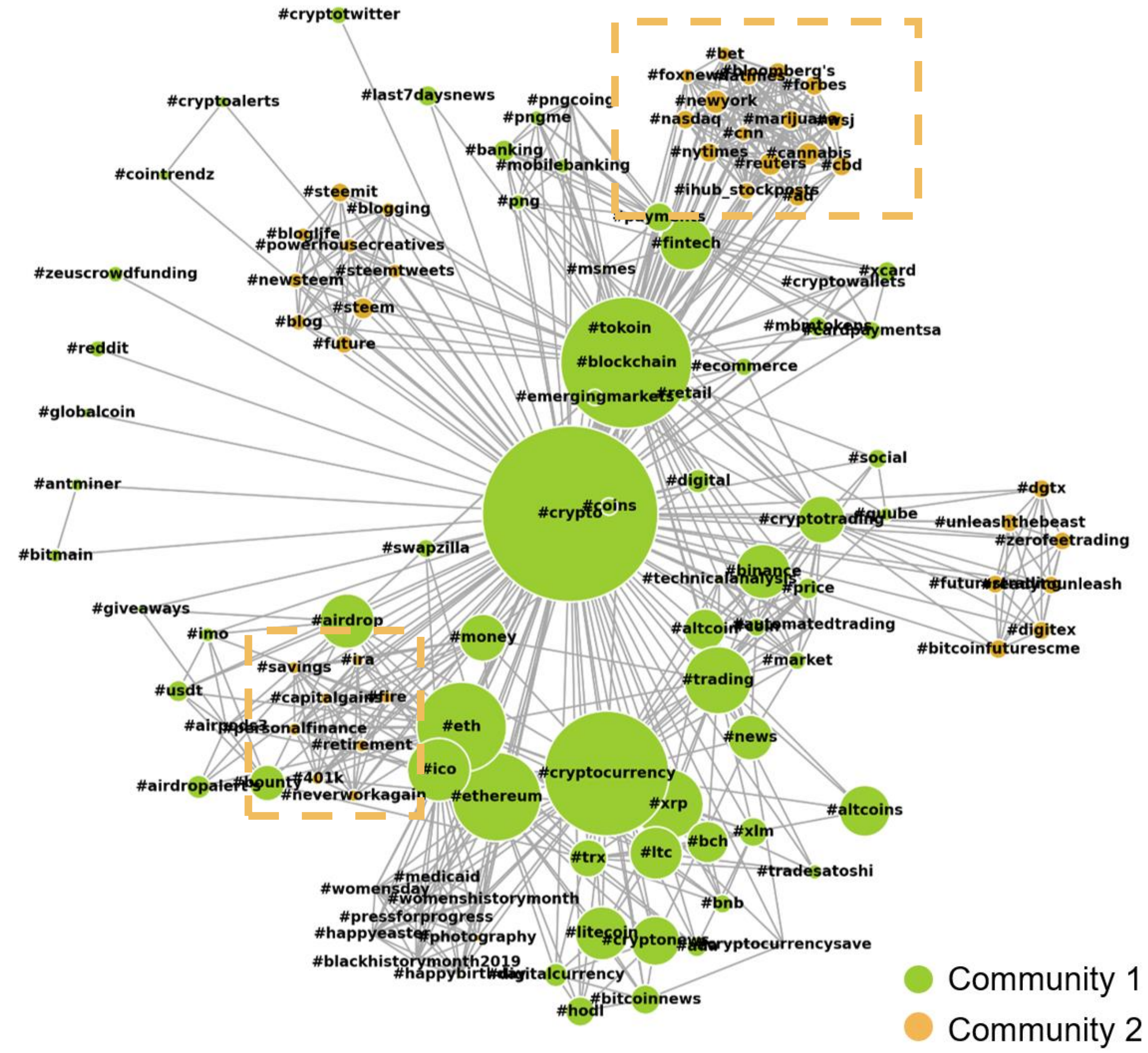
Positive Sentiments

Negative Sentiments

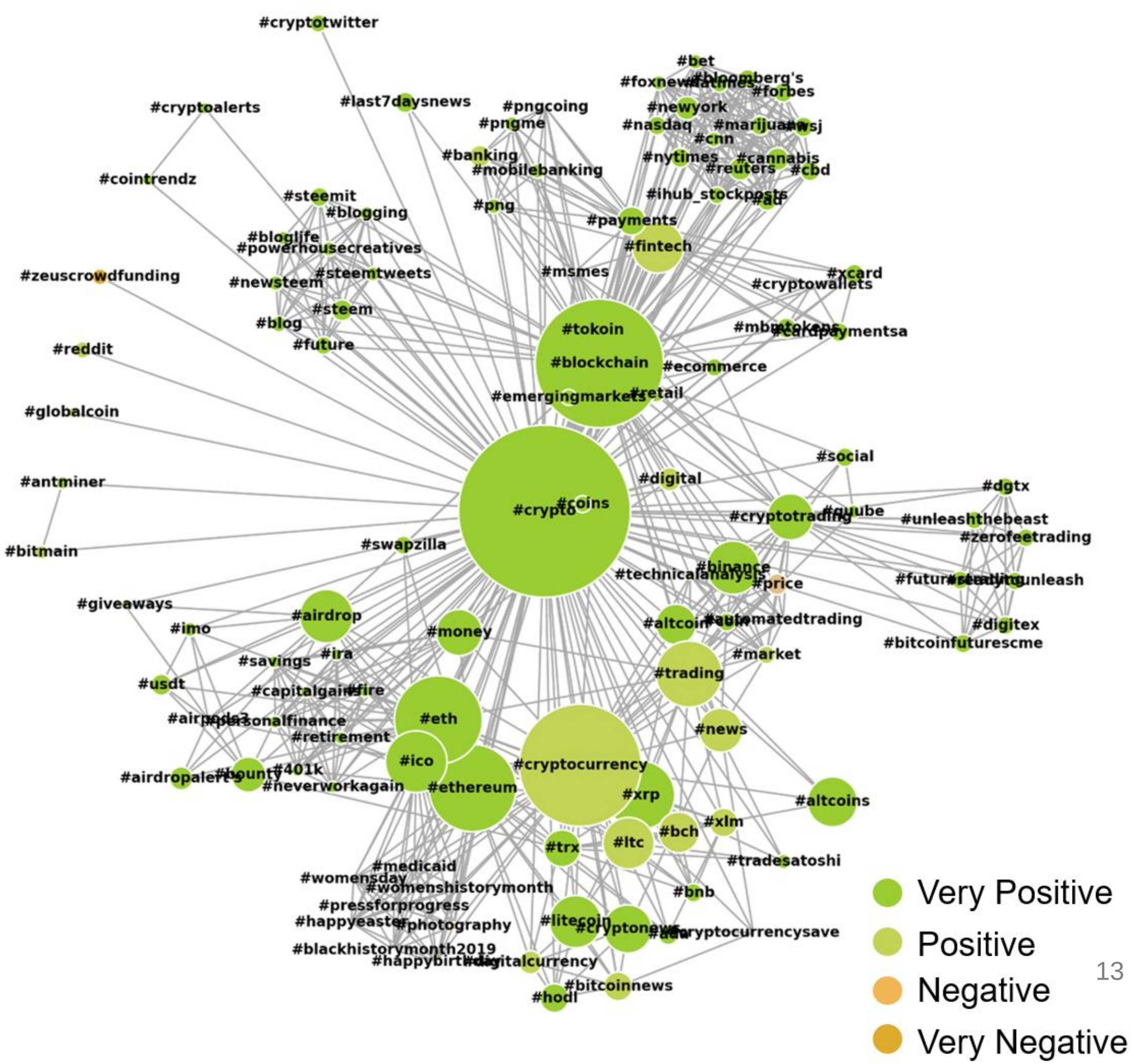


3-1-5. Newman Modularity Community Detection on Hashtags – RQ1 (Polarization Landscape)

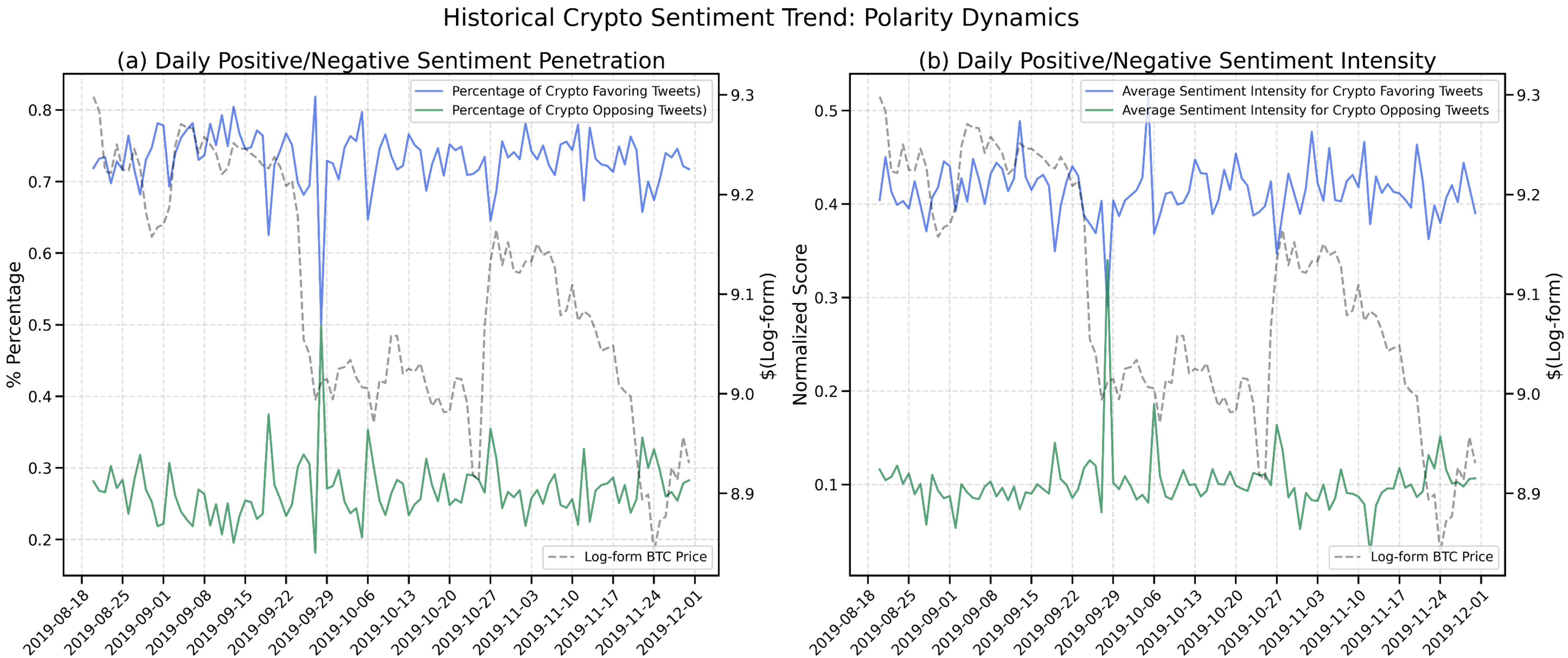
(a) Community Detection: Hashtag Co-occurrence Network



(b) Sentiment Classification: Hashtag Co-occurrence Network



3-2. Time Series Data Analysis – RQ2 (Historical Trend)



3-3. Time Series Prediction – RQ3 (Future Dynamics)

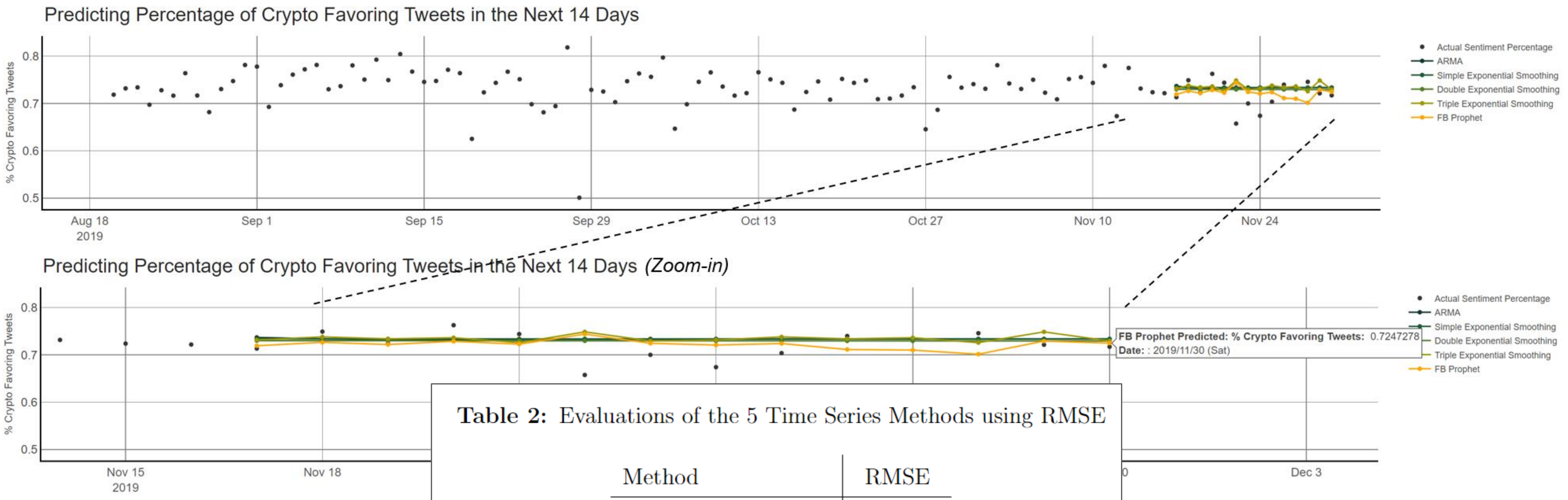


Table 2: Evaluations of the 5 Time Series Methods using RMSE

Method	RMSE
Double ES	0.029762
Simple ES (Baseline)	0.031296
ARMA(4,1)	0.031342
FB Prophet	0.034071
Triple ES	0.0344

04

Summary

What are the conclusions and research limitations?

4. Summary

RQ1: How polarized are people's opinions towards cryptocurrency?

- Opinion polarity is non-trivial (**73%** positive vs. **27%** negative)
- Positive tweets have ~**1.5x** higher sentiment intensity than negative tweets
- **Different topic words** for the two sentiment classes
- Cryptocurrency is a complex topic and comprises of **many sub-topics**

RQ2: How does the polarity of cryptocurrency opinions on Twitter evolve over time?

- Align with the previous findings
- Pattern valid across the time

RQ3: Can we effectively predict future opinion polarity on Twitter toward cryptocurrency?

- **Difficult to predict** using merely its own historical values

References

Reference

Bravo-Marquez, F., Gayo-Avello, D., Mendoza, M., & Poblete, B. (2012). Opinion Dynamics of Elections in Twitter. *2012 Eighth Latin American Web Congress*, 32–39. https://doi.org/10.1109/LA-WEB.2012.11

Hasan, S. Z., Ayub, H., Ellahi, A., & Saleem, M. (2022). A Moderated Mediation Model of Factors Influencing Intention to Adopt Cryptocurrency among University Students. *Human Behavior and Emerging Technologies*, *2022*, e9718920. https://doi.org/10.1155/2022/9718920

Lerman, K., & Ghosh, R. (2010). Information contagion: An empirical study of the spread of news on digg and twitter social networks. *Fourth International AAAI Conference on Weblogs and Social Media*.

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396. https://doi.org/10.1038/s41562-018-0353-0

Martinez-Romo, J., Araujo, L., Borge-Holthoefer, J., Arenas, A., Capitán, J. A., & Cuesta, J. A. (2011). Disentangling categorical relationships through a graph of co-occurrences. *Physical Review E*, *84*(4), 046108.

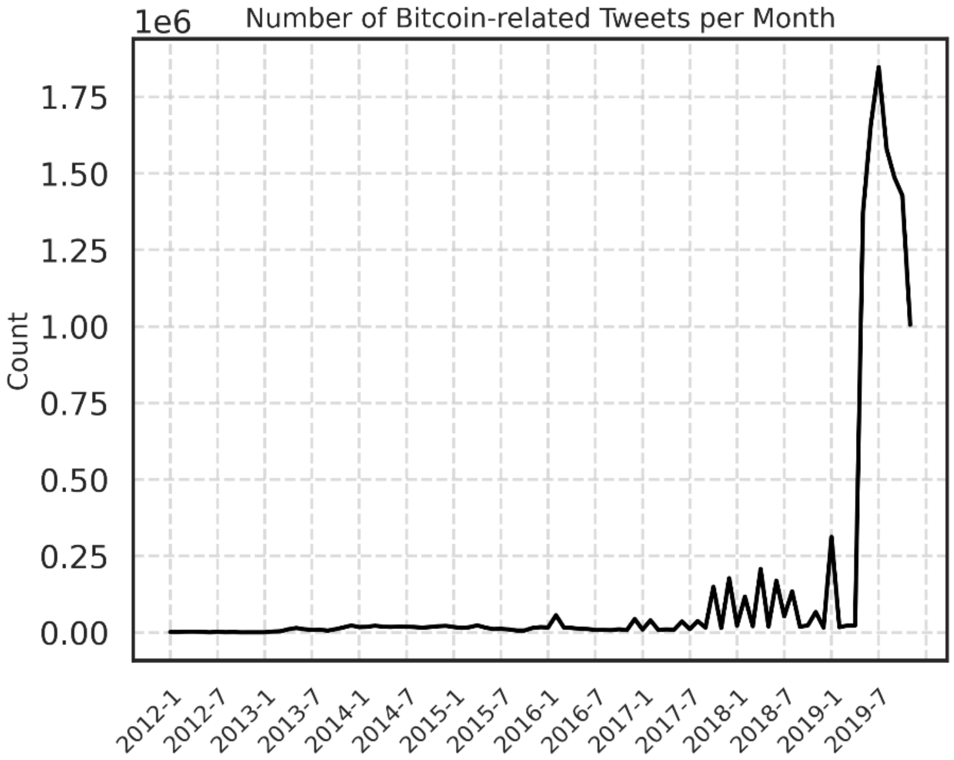
Thanks

Appendix

Initiatives

1) Interesting Information Cascades Going On

- **Viral spread** of those overnight-millionaire-in-bitcoin stories
- # bitcoin related tweets increases **~590 times** from 2012 to. While the entire Twitter platform only grows **~2 times** in terms of # tweets per day



2) Everlasting Debates on Cryptocurrencies

- Positive: Next generation, decentralization, hard asset, etc
- Negative: gambling, money-laundry, unstable value, etc

Source: Kaggle

Discussion

Limitation

- No ground truth sentiment labels
- Insufficient feature engineering in prediction

Future Direction

- Generate trustworthy ground truth sentiment labels
- Try machine learning / deep learning based model to predict the future