Profile Analysis for Cryptocurrency in Social Media

Husnu Saner Narman[*]

Alymbek Damir Uulu[*]

Jinwei Liu[+]

[*]Weisberg Division of Computer Science

Marshall University

[+]Institute for Simulation and Training

University of Central Florida

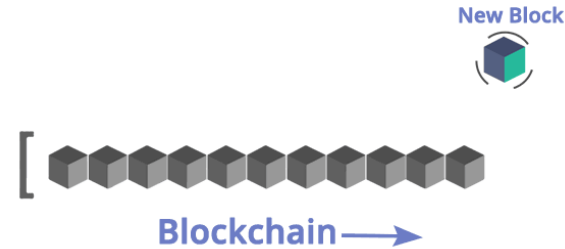narman@marshall.edu

http://mupfc.marshall.edu/~narman/

December 2018

Outline

- Introduction

- System Model

- Readability Indices
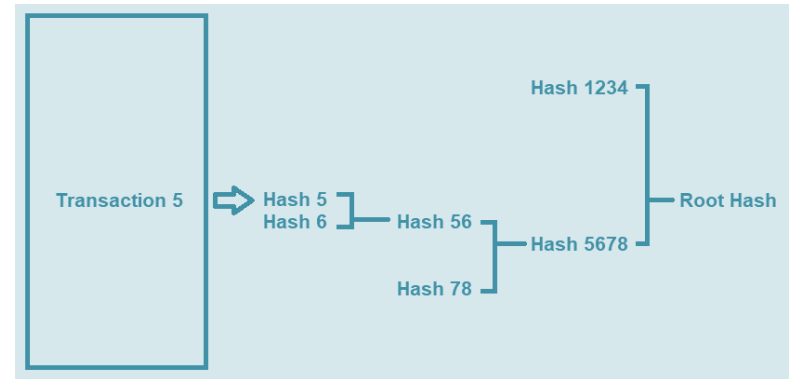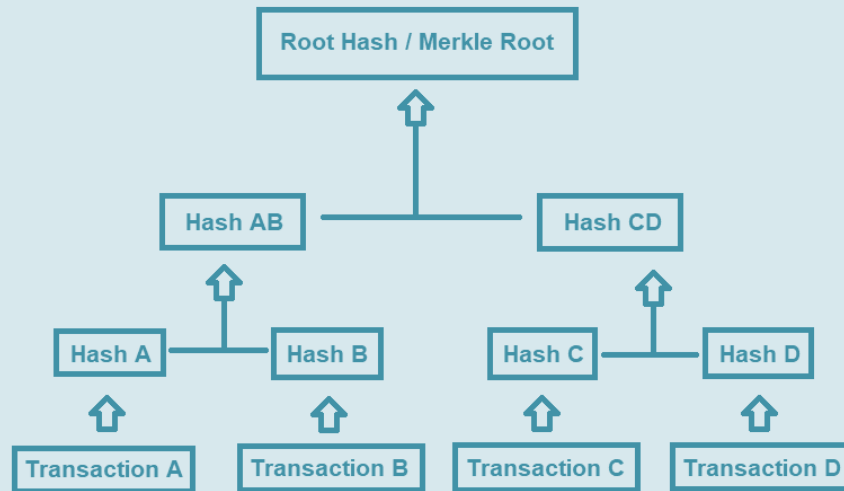
- Selected Cryptocurrencies

- Results

- Conclusion

Introduction

New Block

Blockchain →

- **Blockchain** ushers in a new era for the global financial system.

- One application is Cryptocurrency.

Introduction

System

Readability

Selected

Results

Conclusion

Cryptocurrency

- Verification of transactions

Root Hash / Merkle Root

Hash AB — Hash CD

Hash A — Hash B   Hash C — Hash D

Transaction A   Transaction B   Transaction C   Transaction D

Transaction 5 → Hash 5
Hash 6 — Hash 56
Hash 78 — Hash 5678 — Hash 1234 — Root Hash

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Understanding Public Opinion

- Protect new investors

- Interested parties

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Understanding Public Opinion

- Using social media activities of cryptocurrency related forums

- Twitter, Reddit, YouTube and etc.

Husnu S. Narman

Social Media Analysis for Bitcoin

- Strong interaction between the social media sentiment and the Bitcoin price
- Tendency for investors to overreact to the news on social media within a short period.

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Previous Works

There are several works which analyze cryptocurrency in terms of security, privacy, applications, usability, regulations, and technology.

- We are interested in **education levels** of the users who are active in social media which related to cryptocurrencies.

Husnu S. Narman

Objective

- Analyze the education levels of the users who are active in eight cryptocurrency subreddits (Bitcoin, Bitcoin Cash, Dash, Ether, Litecoin, Lumen, Monero, and Ripple) by using users' comments for each coin subreddit on Reddit.

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Data Gathering Model

- Reddit can have one or more subreddits for each cryptocurrency.

- Each cryptocurrency can have a number of posts in each subreddit

- Each post can have many comments from a number of users because users tend to respond to posts that match their interests

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion
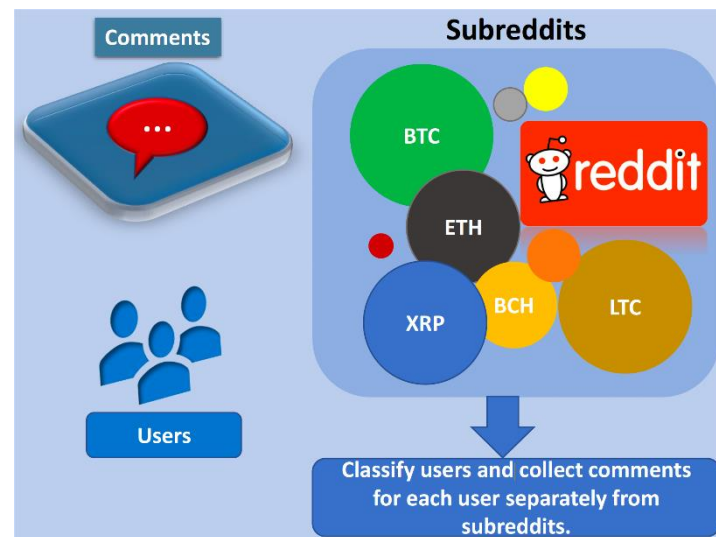
Data Gathering Model and Issues

- Ten to seventy top posts for each cryptocurrency to collect distinct usernames.

- It is possible that the user is interested in the coin, but has not invested in it.

- Informal structure of the comments (no or missing punctuation, shortened words and so forth), the obtained results approximate to the education levels of users.

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Readability Indices

Seven readability test techniques to test the collected comments to identify the education levels of the users.

A. Flesch-Kincaid Readability
B. Dale-Chall Readability
C. The Fog Scale (Gunning Fog)
D. Automated Readability Index
E. Simple Measure of Gobbledygook (SMOG)
F. Coleman-Liau Index
G. Linsear Write
H. Difficulty of Words (not a technique)
I. Average of Seven Readability Techniques (Average)

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Selected Cryptocurrencies
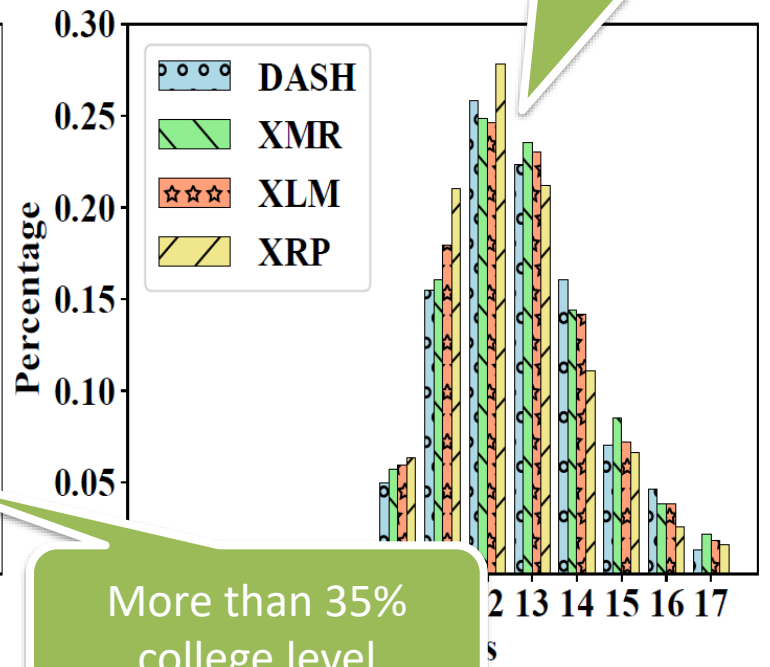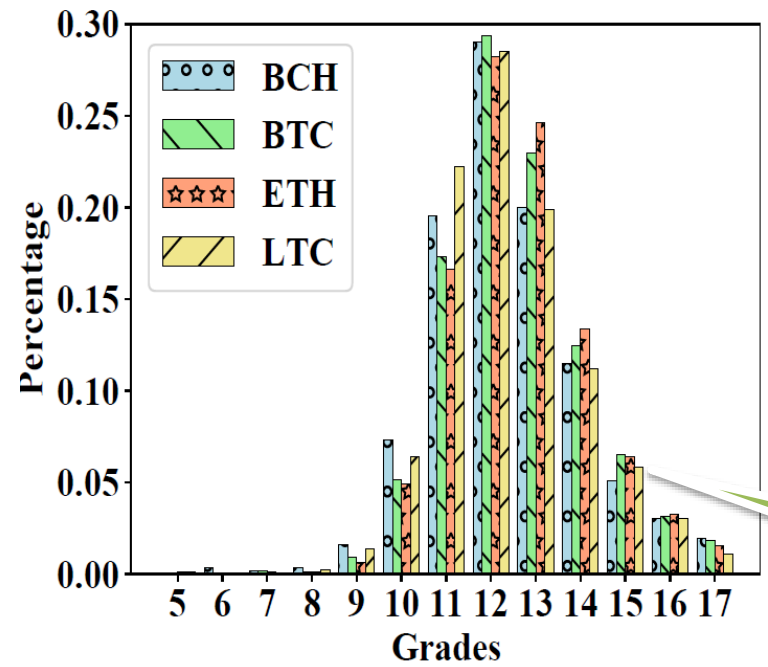
Eight Cryptocurrencies

A. Bitcoin (BTC)
B. BitcoinCash (BCH)
C. Etherium (ETH)
D. Litecoin (LTC)
E. Dash (DASH)
F. Monero (XMR)
G. XRP
H. Lumen (XLM)

Husnu S. Narman

Introduction

System

Readability

Selected

Results

Conclusion

Results

Result for FOG SCALE readability techniques

High School Level

More than 35% college level

Husnu S. Narman

Results

Introduction

System

Readability

Selected

Results

Conclusion

Average of all readability techniques

Middle School Level

Very low college level

Husnu S. Narman

Introduction

System
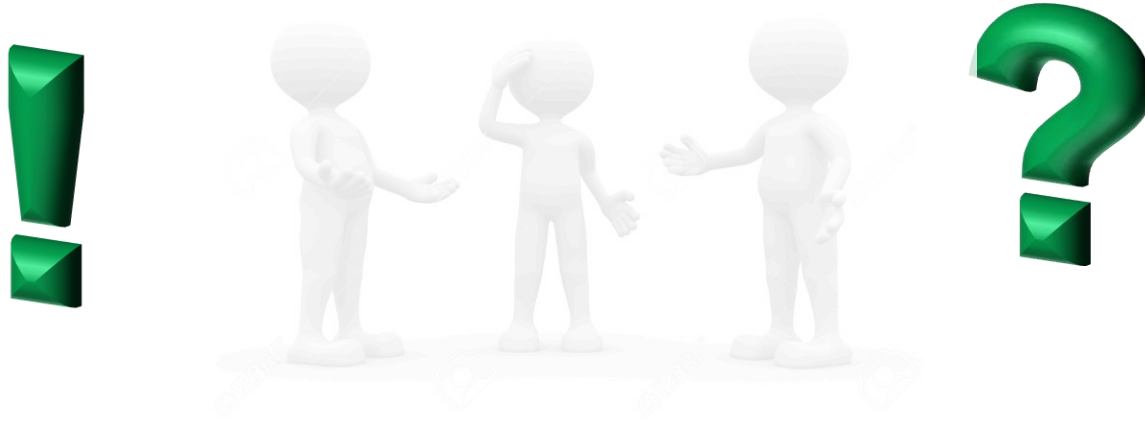
Readability

Selected
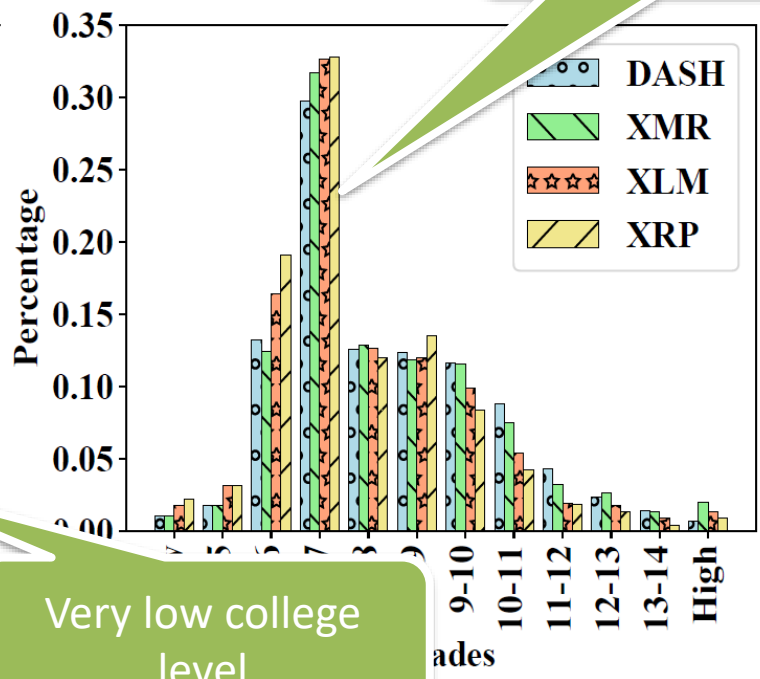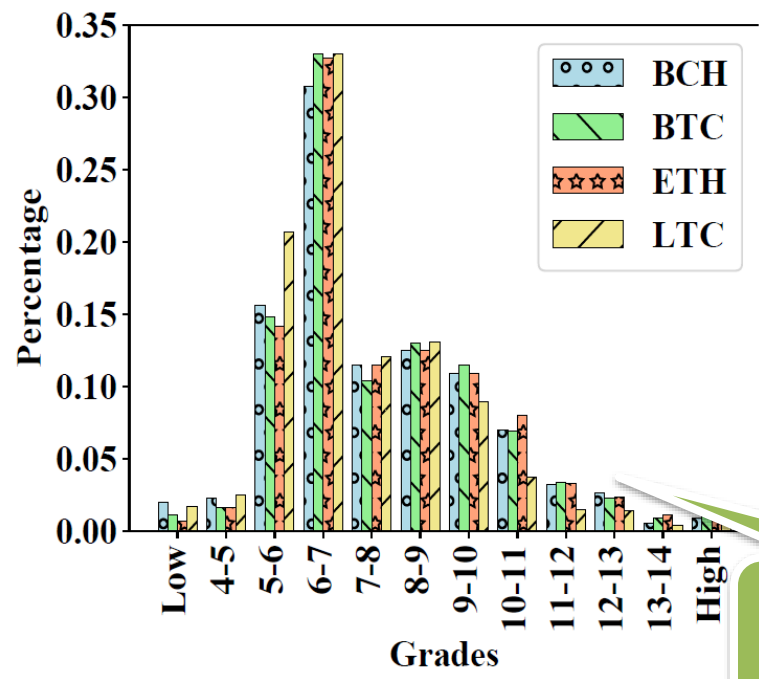
Results

Conclusion

Conclusion

There are differences between the obtained results from seven readability techniques.

The most significant difference is the Fog Scale results which show that 50% of users are grouped under 9th to 12th grades, and almost 35% of users are college students.

The average results of seven readability techniques show that the education levels of users are approximately 60% in middle school, nearly 30% in high school, and 10% in other levels.

16

Husnu S. Narman

Thank You

narman@marshall.edu

http://mupfc.marshall.edu/~narman/

Husnu S. Narman