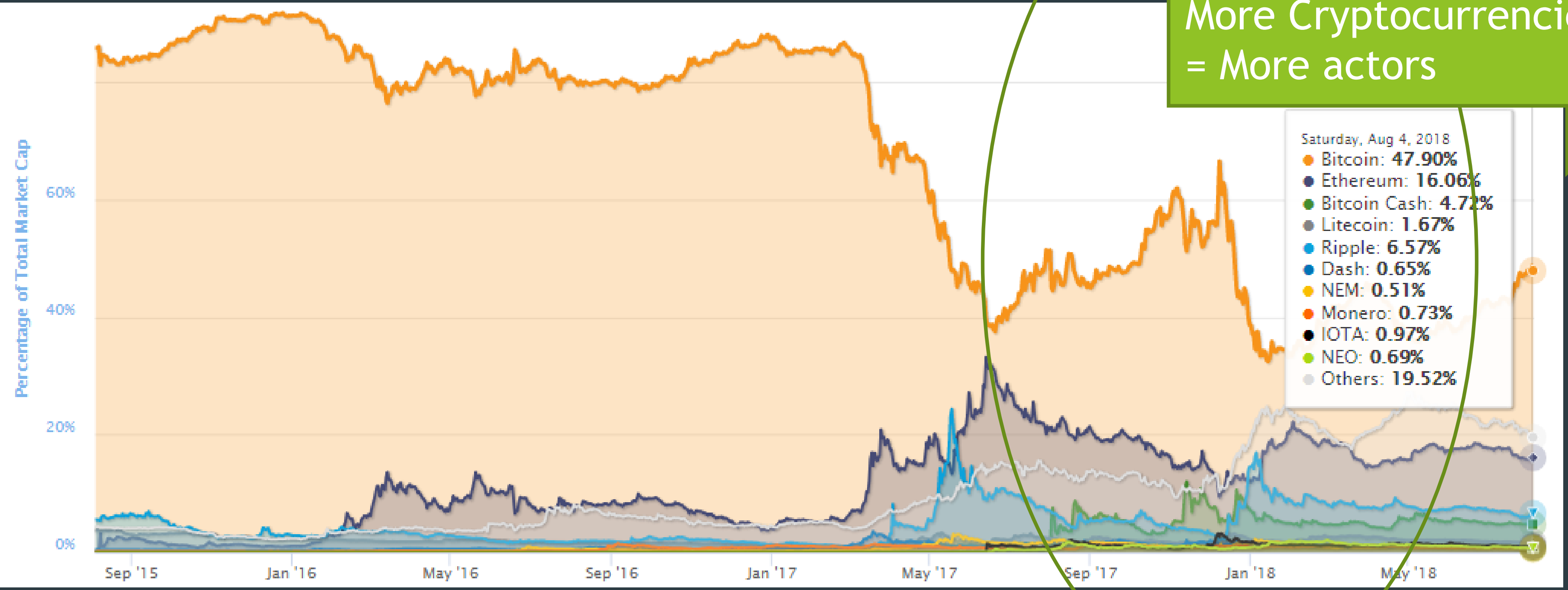
Social network analysis of Cryptocurrencies on Twitter

Use - network analysis techniques

To - Study interaction between Cryptocurrencies

In - Twitter

Motivation



More Cryptocurrencies
= More actors

Motivation(cont.)

- There are - Many Cryptocurrencies in the market
- Attract – Many People interested in different Coins
- Express – Thoughts via Twitter
- So…

"Can we use Twitter to study social interaction about cryptocurrencies?"

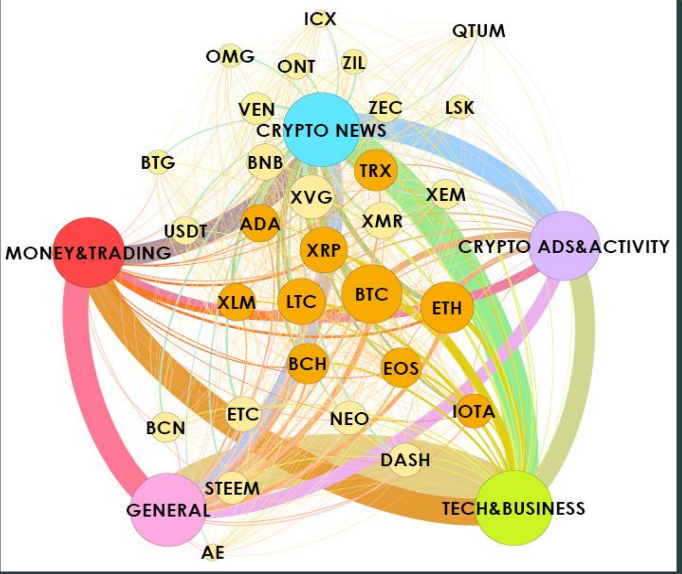
Research Question

- What is the relationship between cryptocurrencies?
- And How do they interact to each other?

Why is it important?

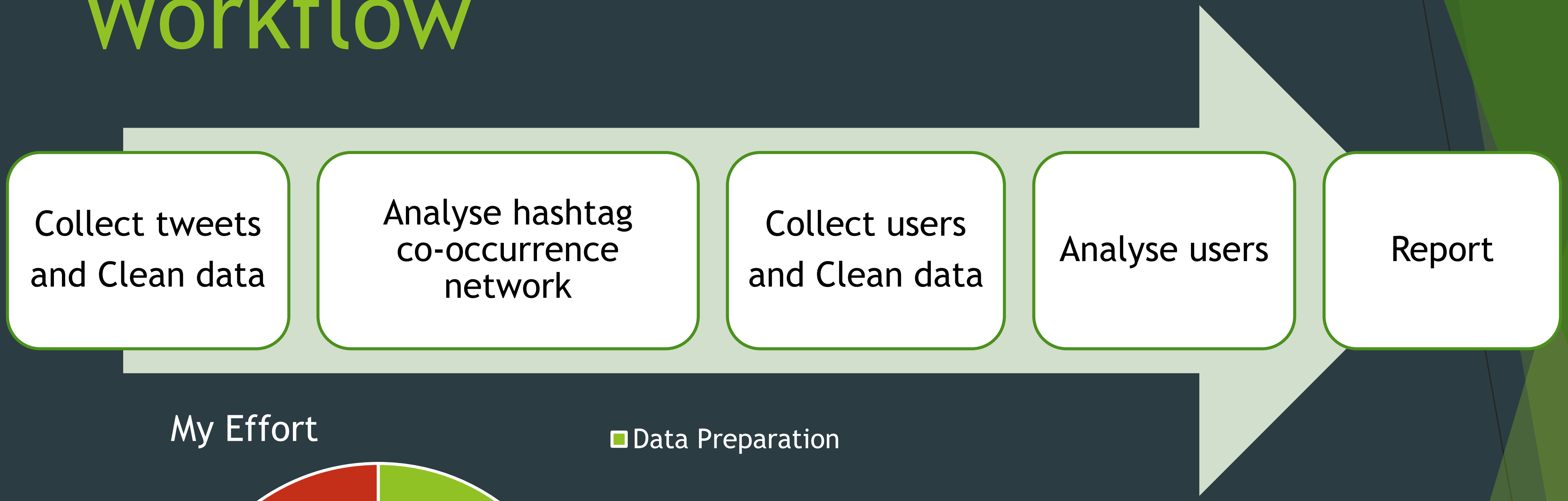
- It fulfils research gaps…
  - Most research focused on price prediction and modelling but “No one studied user interactions in Social network”
- Benefits investors
  - Make market more transparent
  - Dependency between cryptocurrencies will be revealed
  - User behaviours will be revealed
  - **Investor can use insights to create trading strategies.**

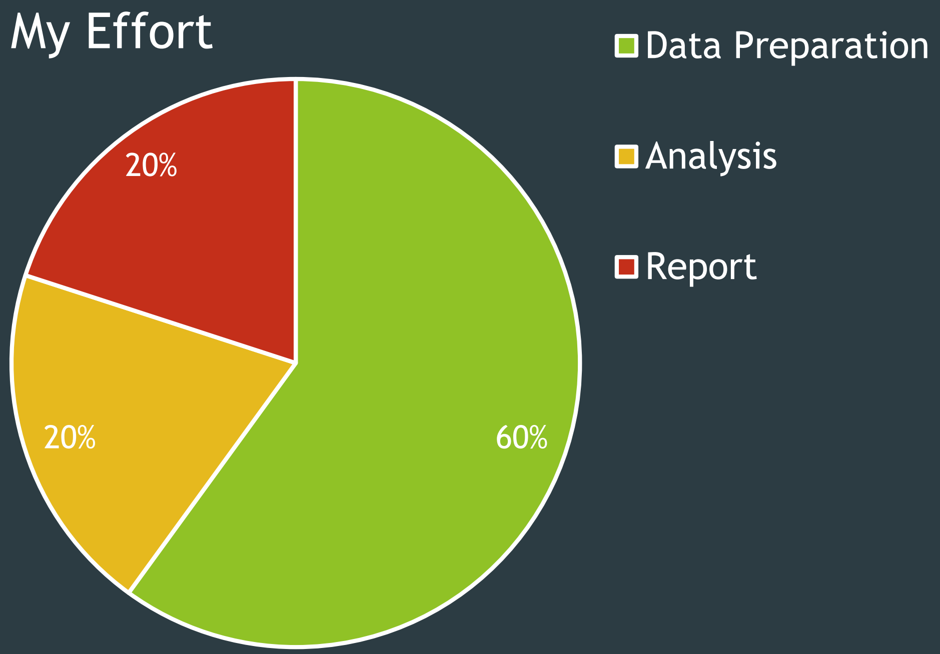
Design



- Analyse 2 networks
  1. Hashtag Co-occurrence Network
     - It contains Hashtag that co-occurred with Cryptocurrencies hashtags
     - To study: “How each coin is talked about in social network?”
  2. User Network
     - Users who actively tweet about Cryptocurrencies
     - To study: “How users who like different coins interact to each other?”

Workflow


Collect tweets and Clean data
Analyse hashtag co-occurrence network
Collect users and Clean data
Analyse users
Report


My Effort
20%
20%
60%
Data Preparation
Analysis
Report

Collect Tweets and Prepare data → Analyse hashtag co-occurrence network → Collect users and Prepare data → Analyse users → Report

- Collect data: Tweepy API, Search for Tweets containing hashtags of top-30 cryptocurrencies in the market Ex.
  - 45 days (13 May - 26 June 2018)

bitcoin	btc	
ethereum	eth	ether
ripple	xrp	
bitcoincas	bch	
eosio	eos	
litecoin	ltc	

- Clean data:
  - **Exclude Bots** : Botometer API + Manual
  - **Exclude Retweet** & Tweets with Non-English hashtag
  - **Keep only important hashtag**
    - To choose the Threshold, we need Robustness test! -> Top-4000 hashtags
- Find co-occurring hashtags
  - **Exclude hashtags that co-occurred by chance.**
    - Keep only P(A,B) > P(A)P(B)

Collect Tweets and Clean data → Analyse hashtag co-occurrence network → Collect users and Clean data → Analyse users → Report

- Raw data

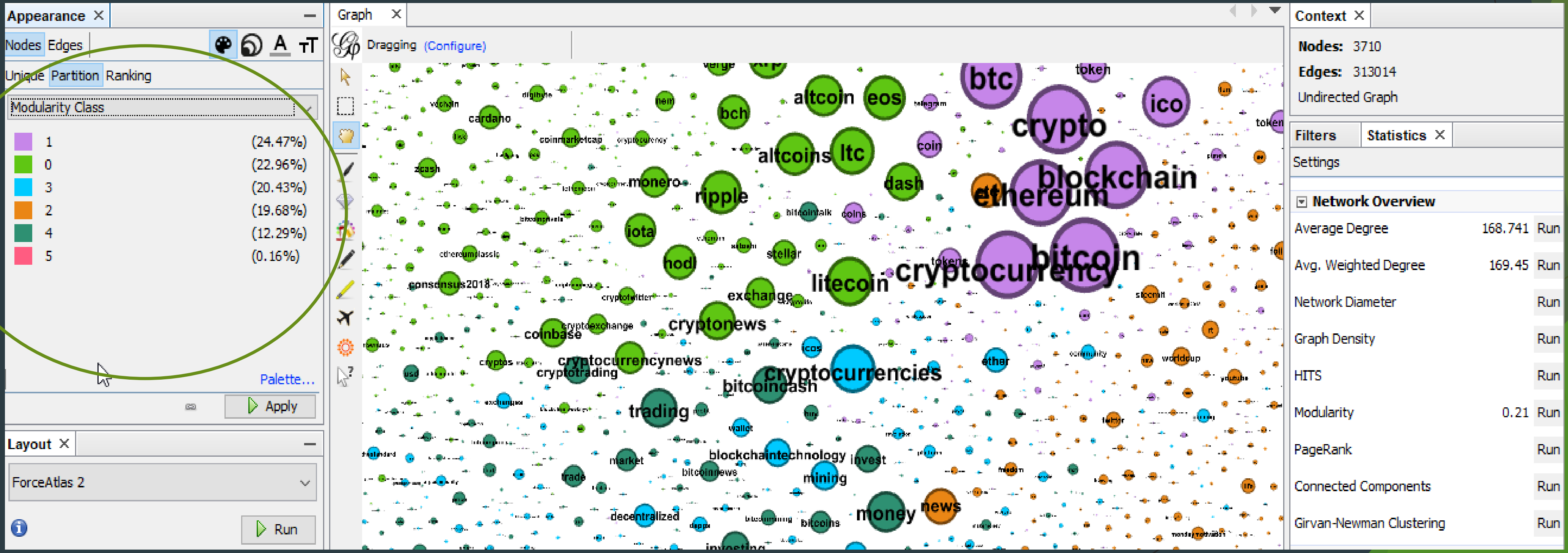
Tweets	~ 6 millions
Users	~ 600 k

- Cleaned data

Users	~ 140 k	
Total Hashtags	~ 100 k	
Hashtags > threshold	~ 4 k	← Node
Co-occurring Hashtags Pairs	~ 320 k	
Co-occurring Hashtags Pairs (Excluding by chance)	~ 310 k	← Edges

Collect Tweets and Clean data → **Analyse hashtag co-occurrence network** → Collect users and Clean data → Analyse users → Report

- Build hashtag co-occurrence network
  - Node = hashtag and Size = how often it co-occurred with others
- Perform Community Detection – Group related hashtags to “Topic”



Collect Tweets and Clean data → **Analyse hashtag co-occurrence network** → Collect users and Clean data → Analyse users → Report

- **Perform Community Detection** – to cluster groups of hashtags strongly related

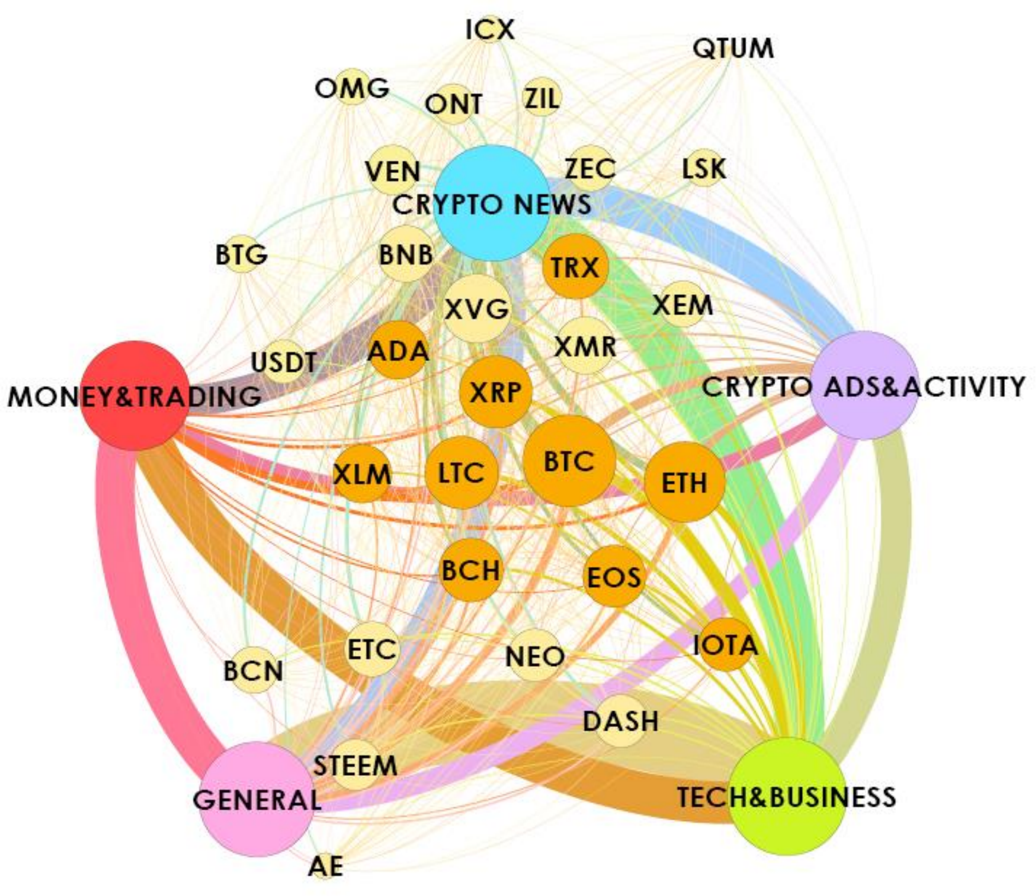
1_crypto_news	2_crypto_ads	3_general	4_bussiness_tech	5_trading
altcoins	crypto	news	cryptocurrencies	money
altcoin	cryptocurrency	usa	fintech	trading
cryptonews	blockchain	love	technology	investing
hodl	ico	entrepreneur	business	cryptotrading
cryptocurrencynews	token	worldcup	tech	invest
exchange	airdrop	follow	finance	investment
coinbase	coin	rt	mining	forex
consensus2018	tokens	success	blockchaintechnology	market
bittrex	free	music	ai	trade
cryptoexchange	tokensale	art	iot	stocks
cryptos	erc20	trump	future	gold
coinmarketcap	coins	twitter	currency	wallstreet
dgb	giveaway	retweet	decentralized	cash
moon	bounty	youtube	startup	investors
cryptotwitter	telegram	facebook	china	trader
dogecoin	airdrops	video	icos	usd
digibyte	airdropalert	life	marketing	investor

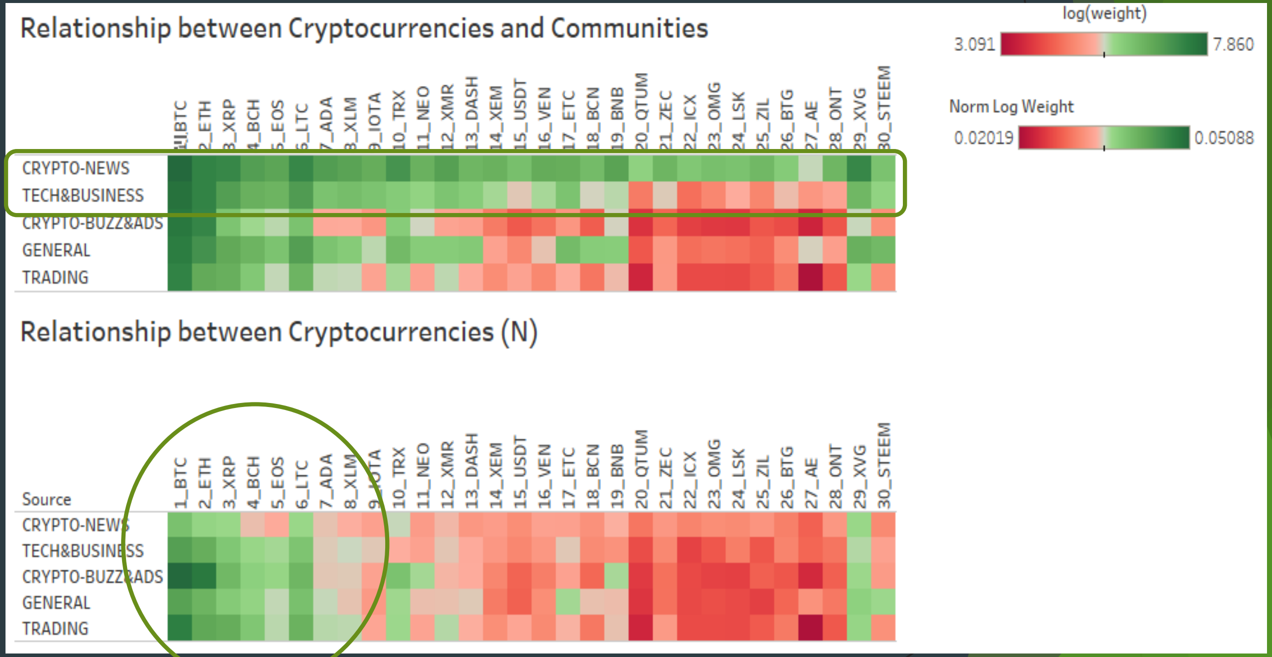
Ex. Cryptocurrencies hashtags

bitcoin	ethereum	ripple	bitcoincash
btc	eth	ripple	bitcoincash
btcusd	ether	ripplenet	bitcoin_cash
hitbtc	ethereum	xrp	bch
btcnews	myetherwallet	xrpbtc	bchpls
freebtc	etherum	xrpthesta	bchforeveryone
btcprice	etherium	xrpcommı	bchusd
btcturk	ethusd	xrparmy	ibch
bitebtc	ethereummining	xrpusd	bchfaucet
btcmining	etherc	xrpsymbol	
btctrading	etherdelta	instaxrp	
wearebtcp	ethbtc	xrpcommunitty	
btcpics	ethlend	xrpnyc	
btcfx	ethereummark		

Collect Tweets and Clean data → **Analyse hashtag co-occurrence network** → Collect users and Clean data → Analyse users → Report

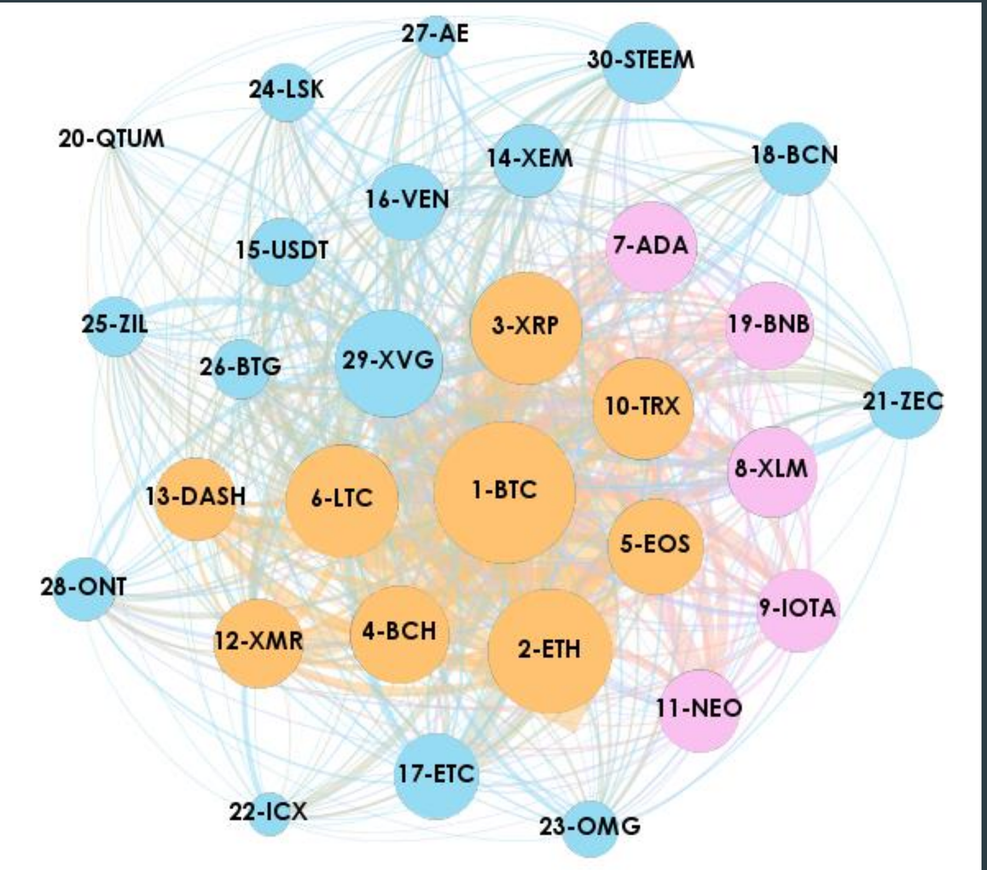
- Merge nodes to Cryptocurrencies and Topics (Communities)
- **“How likely that coin X will be mentioned in topic Y ?”**
- The higher rank a coin is the more prominent (big node) it seems to be in social network and likely to be mentioned in various topics

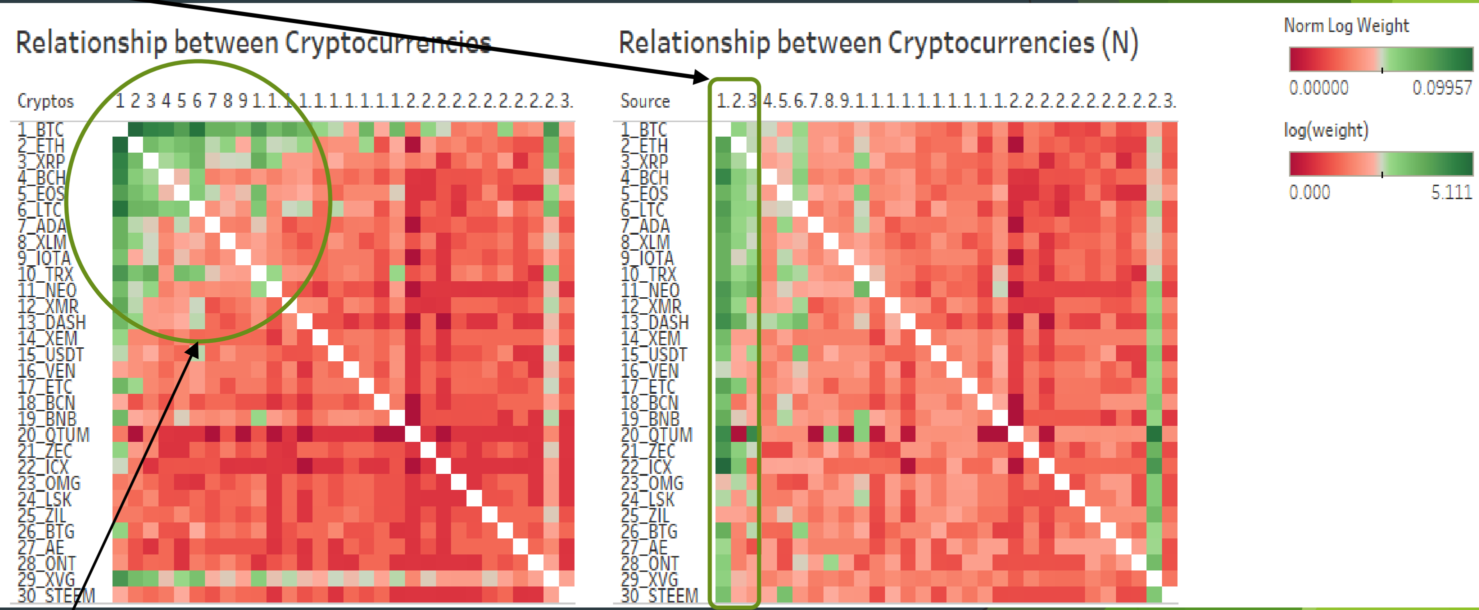




Collect Tweets and Clean data → **Analyse hashtag co-occurrence network** → Collect users and Clean data → Analyse users → Report

- Community detection again – to find group of coins strongly related
- **“How likely that coin X will be mentioned with coin Y?”**
  - “High-ranking cryptocurrencies are more likely to be mentioned with other cryptocurrencies than the low-ranking ones”





“High rank strongly related to each other”
= “People often mention high ranking together”

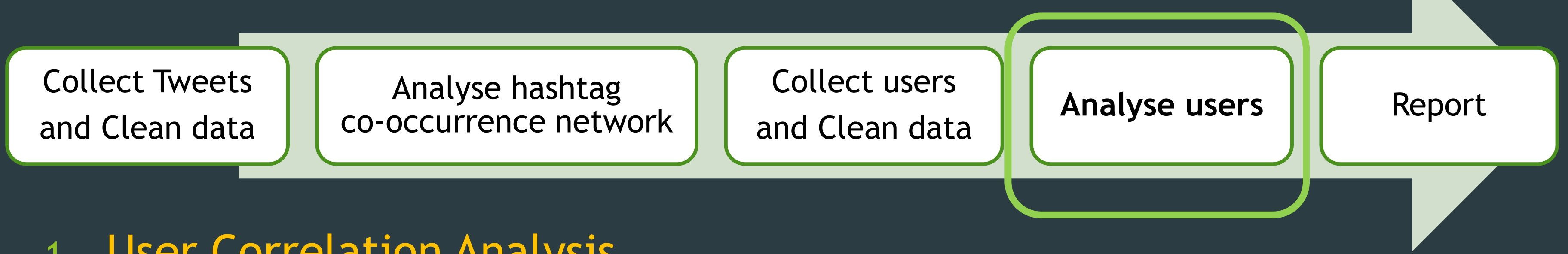
Collect Tweets and Clean data → Analyse hashtag co-occurrence network → **Collect users and Clean data** → Analyse users → Report

1. Take top-1000 users most often tweet about cryptocurrencies as “Based users”
   - Aims to build network of users who are active in cryptocurrencies
2. Collect friends of based users and their connections -> Painful! due to Twitter limitation
   - Collect user data -> friends & followers of base users -> keep only users with >= 10 links to base
   - Collect friendship between them
   - Collect tweets of these users
3. Create user vectors – to measure how much they are interested in each coin and topic.
   - A User Score = P(#coin hashtag) = #coin hashtag / #total hashtags

uid	screen_name	bitcoin	ethere	ripple	bitcoin	eosio	litecoi
_99712395	cloudbreak79	0.165698	0.005814	0.002907	0.002907	0	0.380814
_94561172	vmediahero	0.333333	0.333333	0	0.333333	0	0
_94523500	NeerajNeeraj876	1	0	0	0	0	0
_94519511	danielk15350934	0.333333	0.333333	0	0	0	0
_96705306	WennMoon	0.333333	0.333333	0	0	0	0
_99934346	MikeJam50406782	0.307692	0	0	0	0	0
_99888927	Prene15	0	0	0	0	0	0
_90285150	TheseriousJ1973	0.2	0	0.022222	0.044444	0	0
_95092274	NhaNgoaiCam	0	0	0	0	0	0

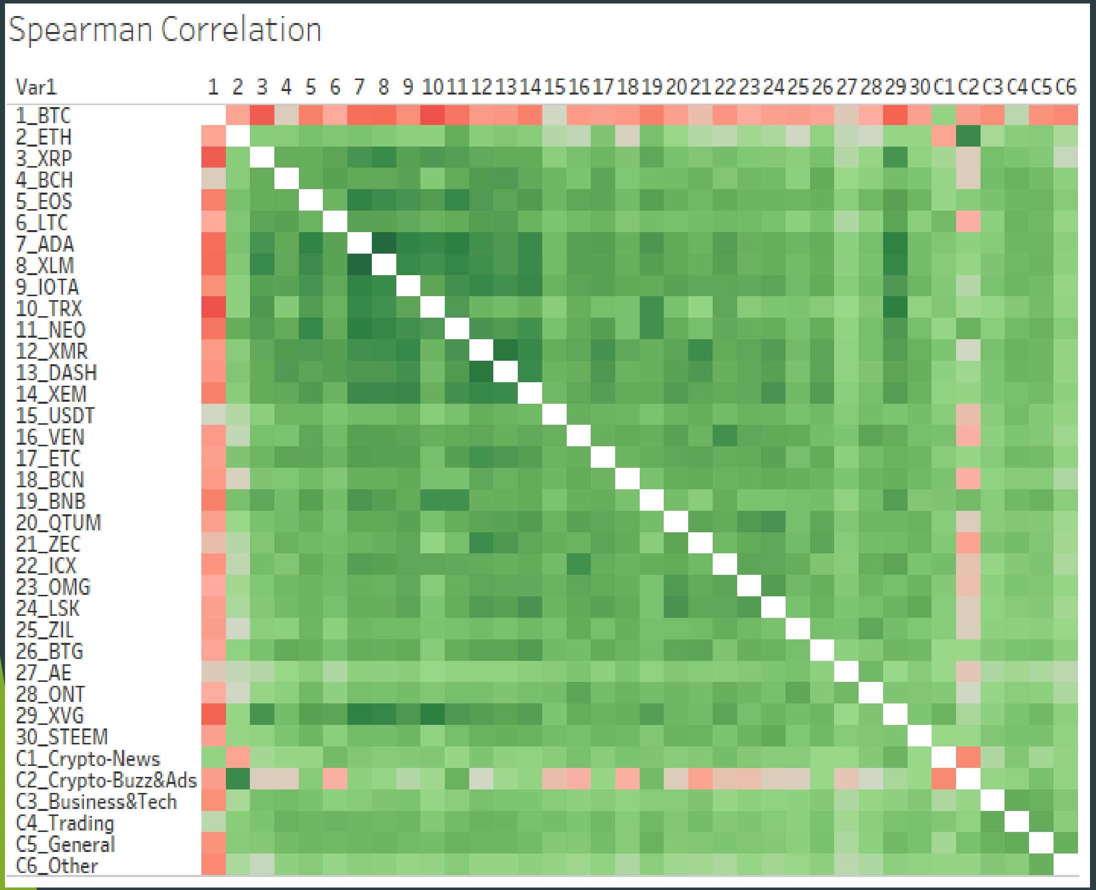
Collect Tweets and Clean data → Analyse hashtag co-occurrence network → Collect users and Clean data → **Analyse users** → Report

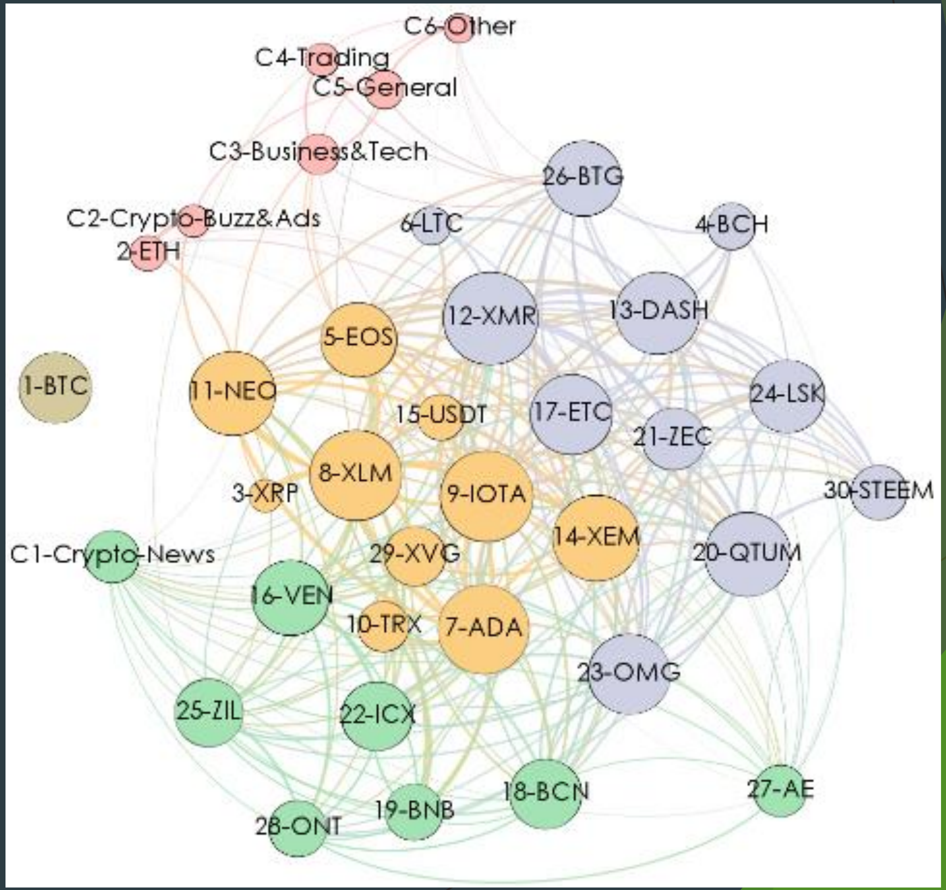
1. User Correlation Analysis – Find correlation between coin scores of each user
   - If users like a specific coin, which other coins they're likely to like?
2. Homophily Analysis – assortative measure
   - If users like a specific coin, how likely their friends will also like it?
   - How likely users who share the same preference will know each other?
3. User Influence factor analysis
   - What makes a user become influential in cryptocurrency social network?



1. User Correlation Analysis

- "Users who like altcoins are likely to like other altcoins around their ranks, while users who actively talked about the top ranks, Bitcoin and Ethereum, show less interest in altcoins but more engagement in general topics".

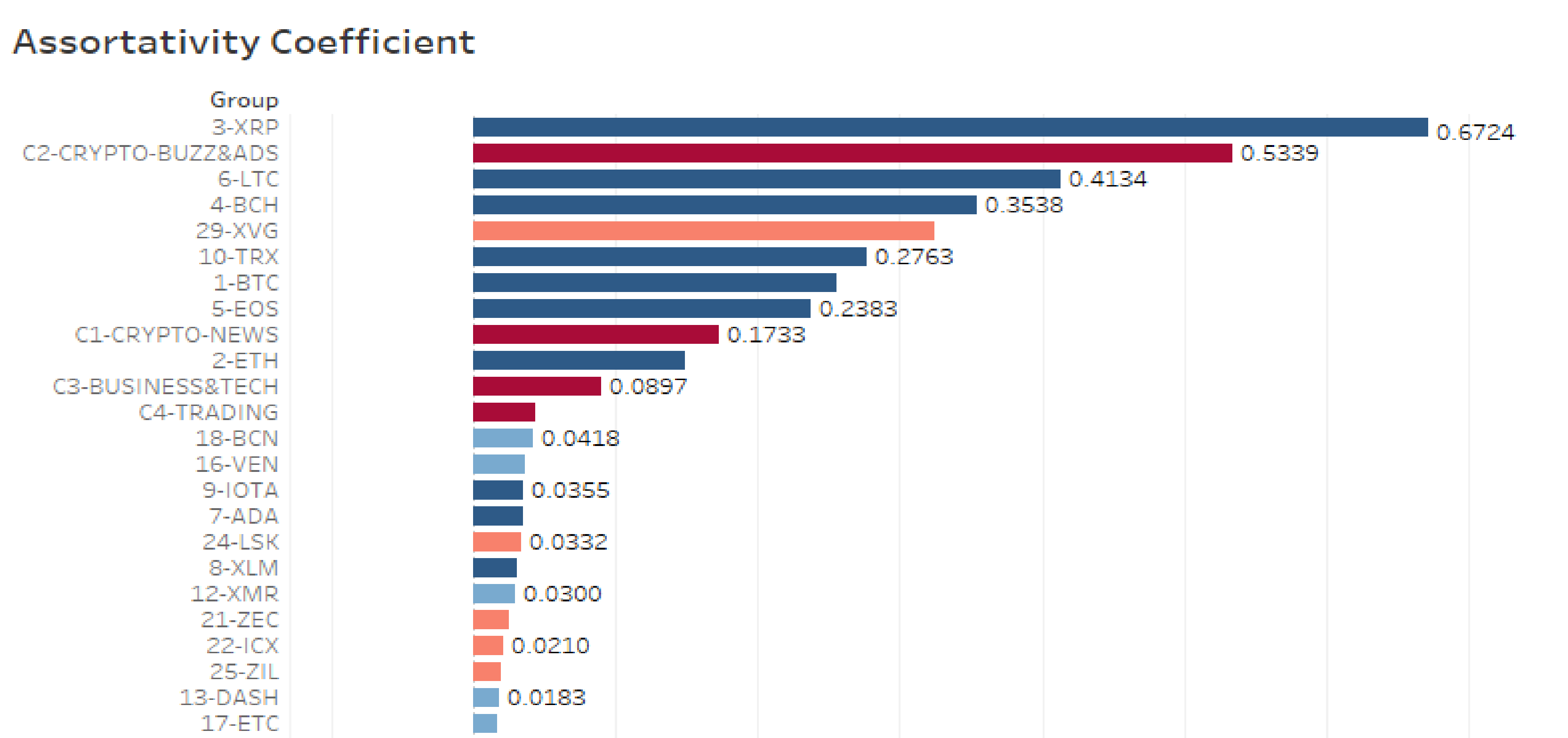




Collect Tweets and Clean data | Analyse hashtag co-occurrence network | Collect users and Clean data | **Analyse users** | Report

2. Homophily Analysis

- If users like a specific coin, how likely their friends are going to like it?
  - Users who like top-10 (Dark Blue) tends to know each other more than the users who like coins in lower ranks



Collect Tweets and Clean data → Analyse hashtag co-occurrence network → Collect users and Clean data → **Analyse users** → Report

3. User Influence factor analysis

- What makes a user become influential in cryptocurrency social network?
  - User who focus in some specific coins or user who talk about many coins in general?
    - Measure by Entropy = How diverse coins that a user talk about ?
- Influence = Be often retweeted
  - Perform linear regression:

```
Call:
lm(formula = rt_count ~ followers_count + friends_count + entropy +
    twt_count + hashtag_count + lifetime_twt_count, data = MyData)

Residuals:
    Min      1Q  Median      3Q     Max
-199071    -428    -221       2  519835

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         9.606e+02  3.143e+02   3.056  0.00225 **
followers_count     4.960e-01  7.887e-03  62.898  < 2e-16 ***
friends_count      -4.654e-01  3.321e-02 -14.015  < 2e-16 ***
entropy            -2.123e+02  1.190e+02  -1.785  0.07438 .
twt_count          -1.962e-01  1.305e-01  -1.504  0.13273
hashtag_count       2.095e-01  4.679e-02   4.477 7.68e-06 ***
lifetime_twt_count -2.293e-02  1.102e-02  -2.082  0.03740 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8390 on 8410 degrees of freedom
Multiple R-squared:  0.3246,    Adjusted R-squared:  0.3241
F-statistic: 673.6 on 6 and 8410 DF,  p-value: < 2.2e-16
```

Questions addressed in this dissertation

<u>**Hashtag co-occurrence network**</u>

- How is coin's ranking related to the social interaction?
  - Top ranks are mentioned in various topics
  - High-High ranks are strongly related.
  - High-Low ranks has stronger relationship than Low-Low = Low attaches to High
- How likely that coin X will be mentioned in topic Y ?
- How likely that coin X will be mentioned with coin Y?

<u>**User network**</u>

- If users like a specific coin, which other coins they're likely to like?
- If users like a specific coin, how likely their friends will also like it?
- How likely users who share the same preference will know each other?
- What makes a user influential in cryptocurrency social network?