Social networks sentiment analysis for cryptocurrencies

Using discriminative aggregation for price prediction

MS&E 448 Final Presentation

Boris Beltinoff and Emile Clastres

June 1st, 2021

Sentiment Analysis - Vanilla

- Define topic space T
- Define sentiment measure s(text) -> [-1, 1]
- Get all tweets for T
- Get sentiment for each tweet using s
- for each time period, average all sentiment in T
- use signal

-> A lot of structural information is lost (users, relations between users etc)

Sentiment Analysis - Discriminative

- Define topic space T, define context C
- Define sentiment measure s(text) -> [-1, 1]
- Get all tweets for T with context
- Get sentiment for each tweet using s
- Cluster users using context
- Define user importance using context
- for each time period, combine all sentiment in T in each cluster using user importance
- use multi-dimensional signal (regression, average...)

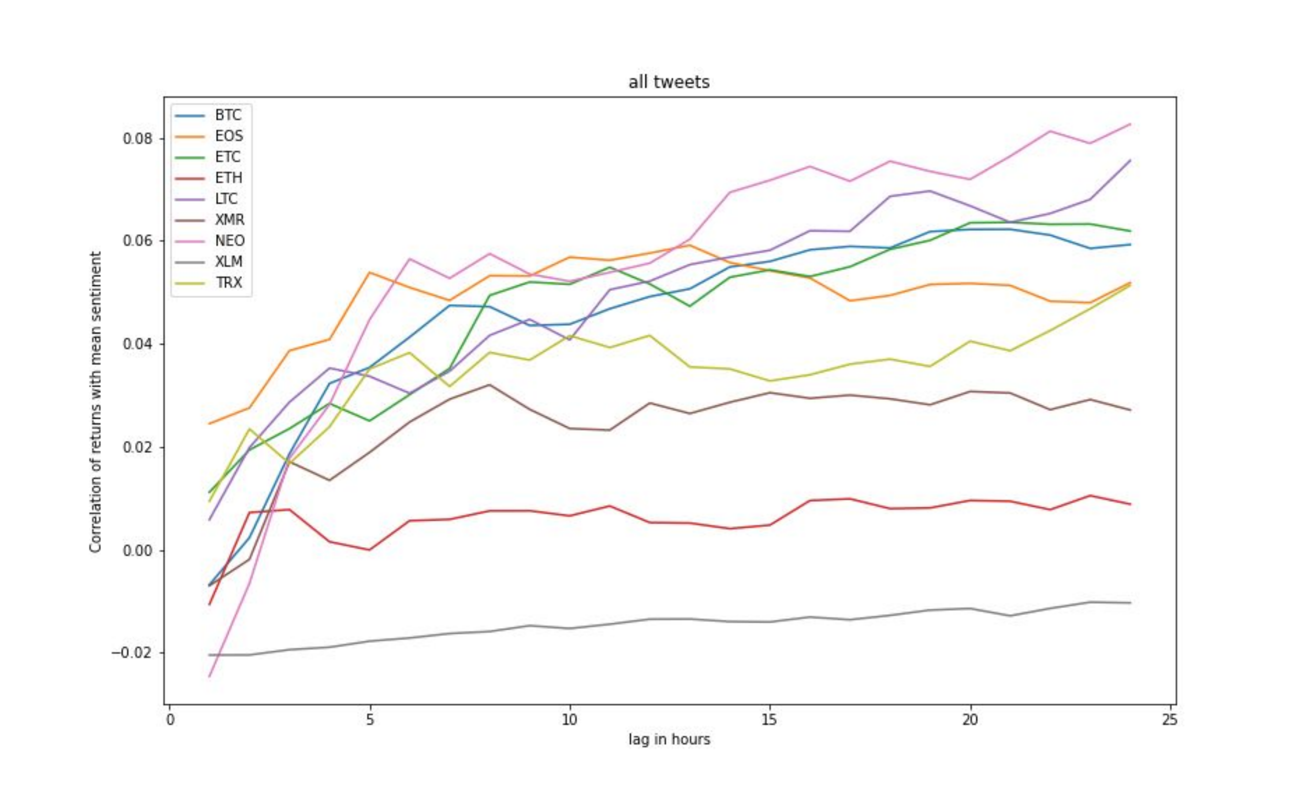
Boris Beltinoff and Emile Clastres

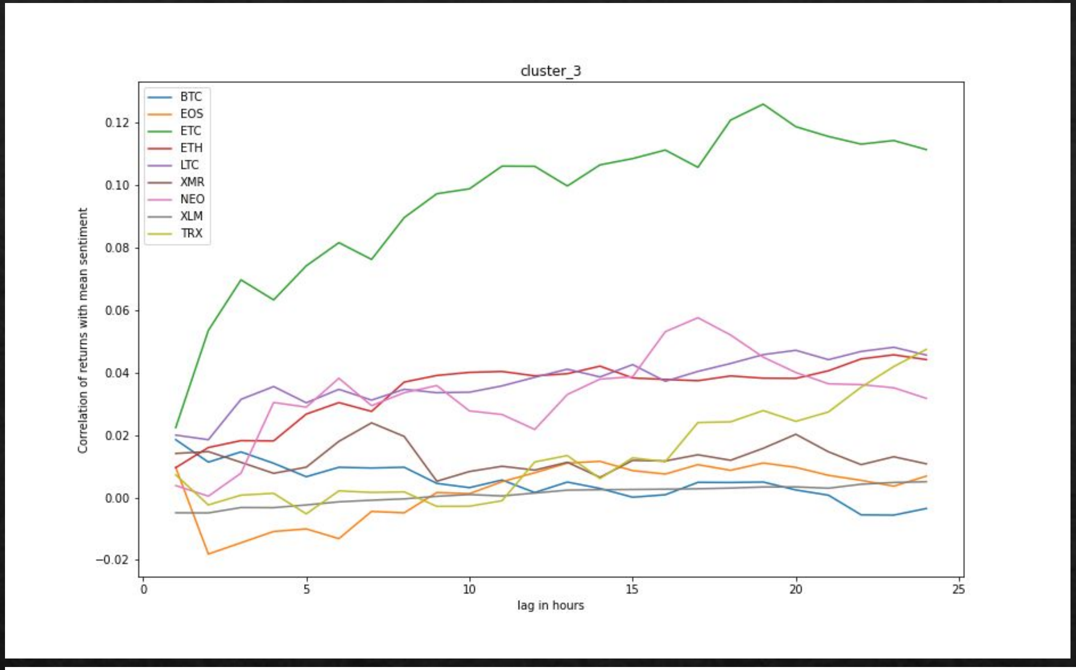
Our simple approach

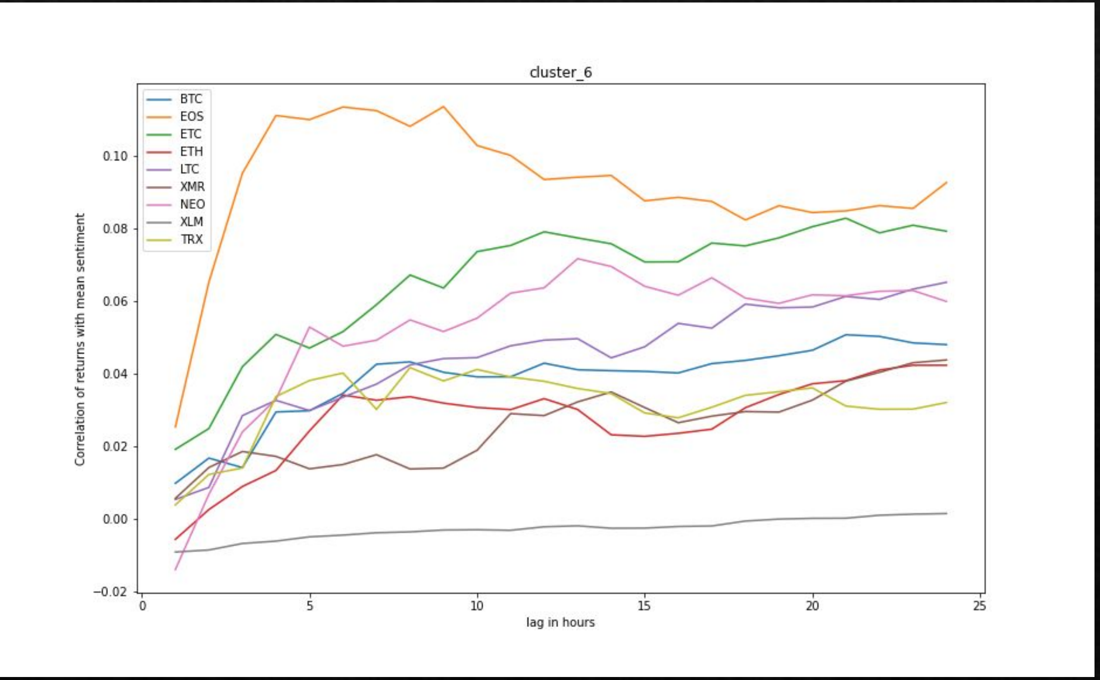
- Social data : tweets on 20 different cryptocurrencies (dirty)
- Price data : hourly prices of 9 cryptocurrency pairs (dirty)
- cluster users based on the proportion of tweets they have for each coin (topic overlap measure)
- give equal importance to each user
- average each cluster's average volume and sentiment (parameter-free)

-> compare to the mean of volume and sentiment across all users (vanilla signal)

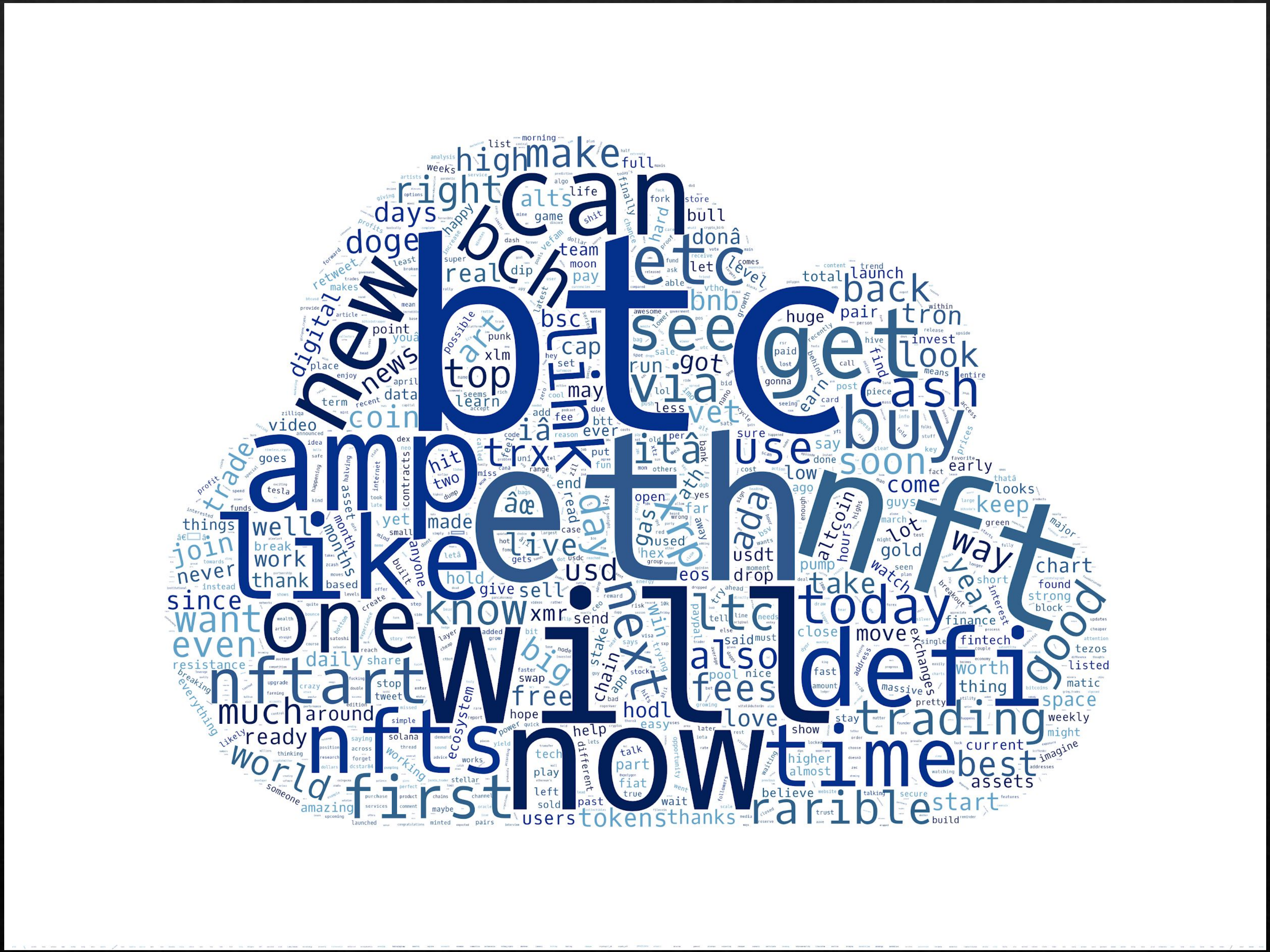
Cluster analysis



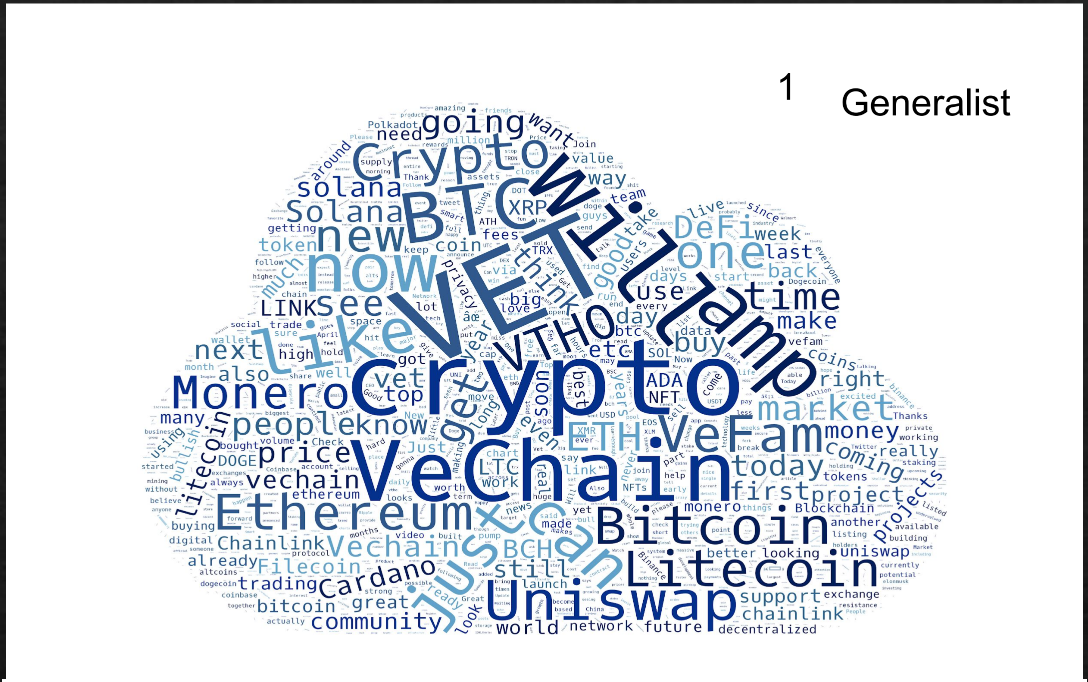


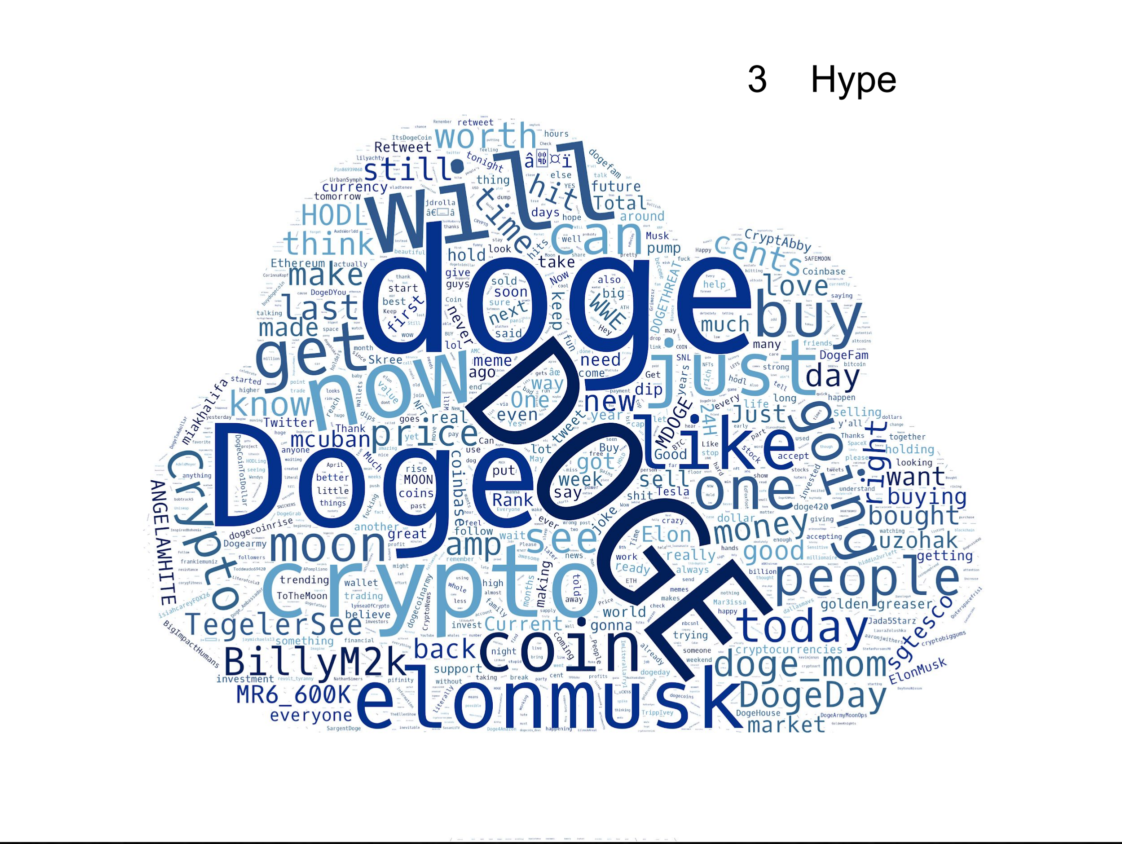


All tweets


btc
eth
will
nft
now
can
get
like
amp
new
one
defi
time
nfts
first
nftart
bch
etc
link
see
buy
cash
via
trading
today
next
rarible
good
make
high
right
doge
days
top
best
fees
much
know
also
want
even
world
tokens
thanks
soon
use
look
back
tron
way
ltc
ada
xrp
trx
usd
coin
news
digital
well
join
since
work
thank
love
hodl
free
hope
around
ready
start
assets
trade
move
come
keep
chart
strong
found
block
finance
fintech
thing
worth
space
weekly
tezos
listed
matic
current
alts
team
moon
pay
bnb
bsc
cap
got
run
vet
amazing
users

Cluster analysis - word clouds





Cluster analysis - 6 (EOS, all positive)

6
actual technical analysis and trading advice

Encouraging results

Table 1: Signal correlation to 18-hour returns

	Discriminative	Vanilla
BTC	**0.0570**	0.0556
EOS	**0.0235**	0.0203
ETC	**0.0262**	0.0249
ETH	**0.0424**	-0.003
LTC	**0.0787**	0.0608
XMR	-0.007	**0.0047**
NEO	**0.0196**	0.0158
XLM	**0.0494**	0.0377
TRX	**0.0170**	0.0105

Technical limitations

- Price data quality (lots missing dates - increases noise when shifting prices relative to signal)
- Sentiment sparsity - most sentiment is 0
- tweet density - we replaced missing sentiment by 0 (a cluster doesn’t mention a coin for an hour)
- Users are clustered rather than tweets : what to do with “new” users ?

This makes the signal ill-suited for prediction, even for linear regression

Room for improvement

- Social media offer a rich contextual structure. Ideally, one could categorize and cluster all users on Twitter prior to restricting on topics
- Temporal clustering/ graph analysis is challenging but important
- Different clusters should be able to evolve at different timescales/horizons
- Data availability (proportion of missing data) should be a feature for the downstream model
- better Price/Twitter data is mandatory for complexification and actual backtests