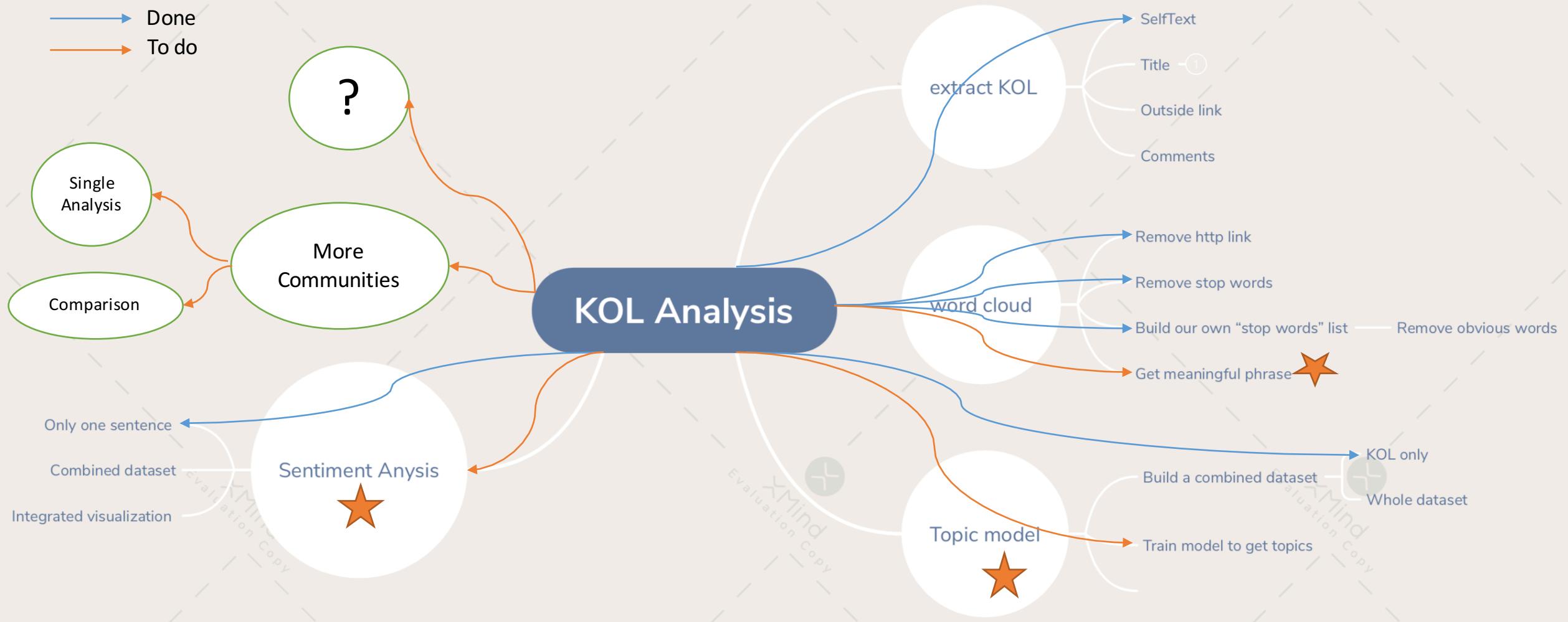


→ Done
→ To do

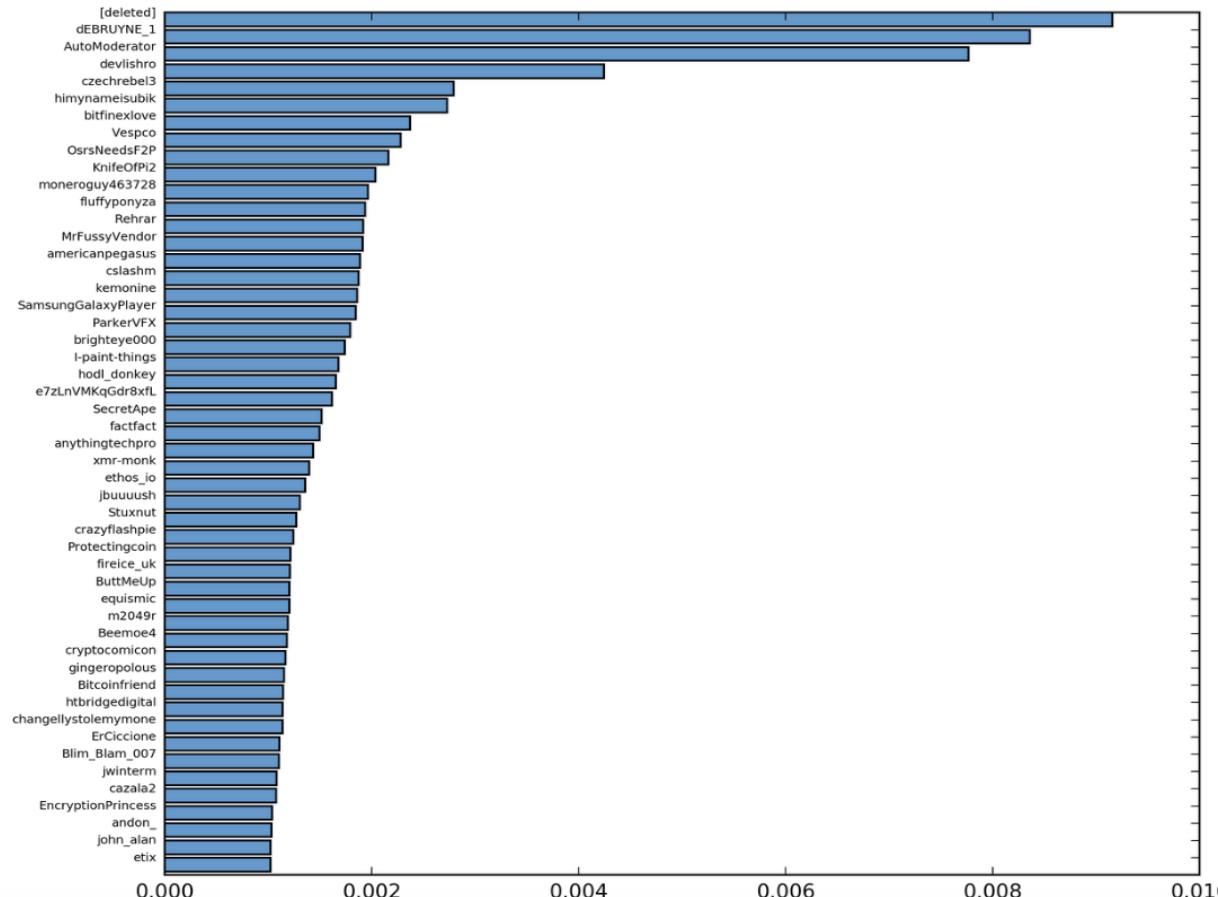


Extract KOL

- **Algorithm: pageRank**
- The PageRank algorithm outputs a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called “iterations”, through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value. It's simple though proved to be effective and already has been implemented in the networkx library.
-
- Tools: networkx, jupyter notebook
-
- Data: all the data of Monero Subreddit in table: *neo_crawl_data*, *reddit_comment_all* and *reddit_link_all*

PageRank Relationship

```
['[deleted]', 'dEBRUYNE_1', 'AutoModerator', 'devlishro', 'czechrebel3', 'himynameisubik', 'bitfinexlove', 'Vespo', 'OsrsNeedsF2P', 'KnifeOfPi2', 'moneroguy463728', 'fluffyponza', 'Rehrar', 'MrFussyVendor', 'americanpegasus', 'cslashm', 'kemonine', 'SamsungGalaxyPlayer', 'ParkerVFX', 'brighteye000', 'I-paint-things', 'hodl_donkey', 'e7zLnVMKqGdr8xfl', 'SecretApe', 'factfact', 'anythingtechpro', 'xmr-monk', 'ethos_io', 'jbuuuush', 'Stuxnut', 'crazyflashpie', 'Protectingcoin', 'fireice_uk', 'ButtMeUp', 'equismic', 'm2049r', 'Beemoe4', 'cryptocomicon', 'gingeropolous', 'Bitcoinfriend', 'htbridgedigital', 'changellystolemymone', 'ErCiccione', 'Blim_Blam_007', 'jwinterm', 'cazala2', 'EncryptionPrincess', 'andon_', 'john_alan', 'etix']  
[0.009155669761507078, 0.008359672974546142, 0.0077665719127893015, 0.004244532094394332, 0.002790  
061607616219, 0.0027313376269972977, 0.0023722869679683256, 0.0022785490549155846, 0.0021616193643  
97211, 0.002034931815845783, 0.0019629880020479206, 0.0019378958437326994, 0.001915951427370004, 0  
.001910699099229476, 0.00188677193022558, 0.001872758940523257, 0.0018606568616968148, 0.001844568  
5909814476, 0.0017917040940030554, 0.0017399140706085905, 0.0016776799204557755, 0.001653836895086  
2927, 0.0016159461812756304, 0.0015151743675082031, 0.0014964000076377255, 0.0014362886582267214,  
0.0013934789598807468, 0.0013593346867983175, 0.001302718249041582, 0.0012696093911748663, 0.00124  
16996856732293, 0.0012118981588578976, 0.001210546072145787, 0.0012060325009185387, 0.001204494514  
442522, 0.0011886387203558465, 0.001179262489832522, 0.0011673738052312209, 0.001150406430992682,  
0.0011413456501873245, 0.0011384546255665967, 0.001136566222358644, 0.001111083576986854, 0.001105  
9390859540403, 0.0010798461384961516, 0.0010772954248441928, 0.0010394364227588788, 0.001032977275  
0822023, 0.001021547887167431, 0.001021407267271823]
```



Result Analysis

- Compare result of Leveraging Conversation Structure on Social Media to Identify Potentially Influential Users and result of pageRank

	A	B	C	D	E	F	G	H	I	J	K	L
1	data.author	post_count	conversation	atk_in	atk_out	def_in	def_out	weighted(post25%+atk_in pageRank)				pageRank.author
2	dEBRUYNE_1	12943	0	10543	12419	2727	1703	1			2 [deleted]	
3	[deleted]	8592	0	4995	5051	2218	2469	0.6529918			1 dEBRUYNE_1	
4	gingeropolou	5586	0	4772	5301	1565	1751	0.48316567			AutoModerator	
5	fluffyponza	3193	0	3546	3100	1059	1152	0.31909635			12 devlishro	
6	AutoModera	396	0	1789	0	2111	0	0.30903646			3 czechrebel3	
7	uy88	4681	0	2644	4629	671	2192	0.29101006			himynameisubik	
8	KnifeOfPi2	1938	0	2037	1906	1384	672	0.27653079			10 bitfinexlove	
9	OsrsNeedsF2	1785	0	1665	1658	1159	416	0.23509915			9 Vespc	
10	snirpie	2958	0	2187	2816	678	974	0.22844101			OsrsNeedsF2P	
11	cryptochangi	1989	0	1544	1978	1048	693	0.22460313			KnifeOfPi2	
12	americanpeg	1543	0	2367	1235	947	340	0.22381811			15 moneroguy463728	
13	SamsungGalax	2176	0	2079	2037	790	747	0.21558993			18 fluffyponza	
14	mWo12	2105	0	2012	1778	691	370	0.20006397			Rehrar	
15	smooth_xmr	2945	0	1828	2918	452	1530	0.19134151			MrFussyVendor	
16	Vespc	1751	0	1893	1452	665	327	0.18328778			8 americanpegasus	
17	Ant-n	2404	0	1515	2325	392	895	0.15953364			cslashm	
18	Rehrar	724	0	1237	657	725	224	0.14133279			13 kemonine	
19	rrunner7	1441	0	1069	1423	564	784	0.13857068			SamsungGalaxyPlayer	
20	Flenst	1159	0	824	1140	614	325	0.12906321			ParkerVFX	
21	hyc_symas	1480	0	1374	1393	379	608	0.12705705			brighteye000	
22	ferretinjapan	1529	0	1221	1506	399	534	0.12635925				
23	moneroguy4	479	0	1158	256	632	27	0.1210967				
24	pinkphloid	769	0	747	701	629	163	0.11722975				
25	binaryFate	1388	0	1087	1335	344	539	0.1119672				

As we can see, these two algorithms have 50% of top_20 KOL in common.

The original top_11 KOL's posts without any data cleaning:



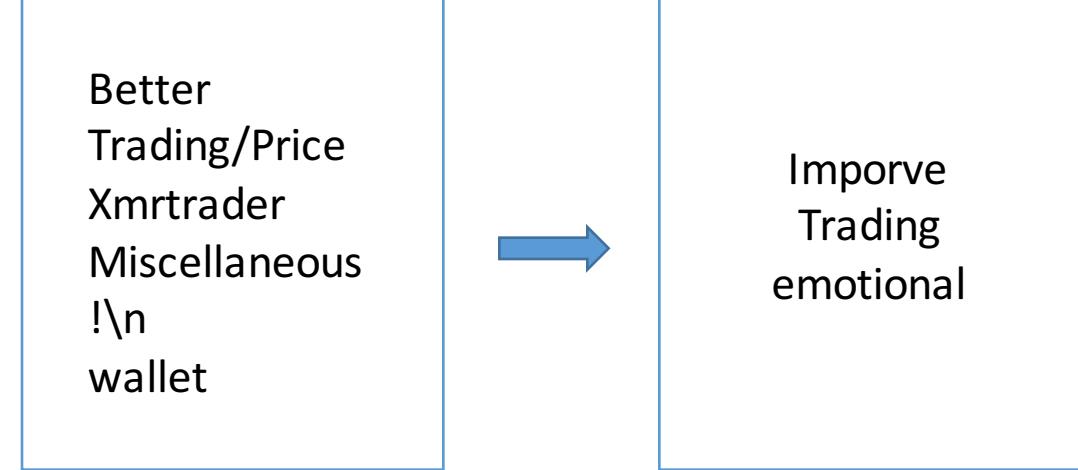
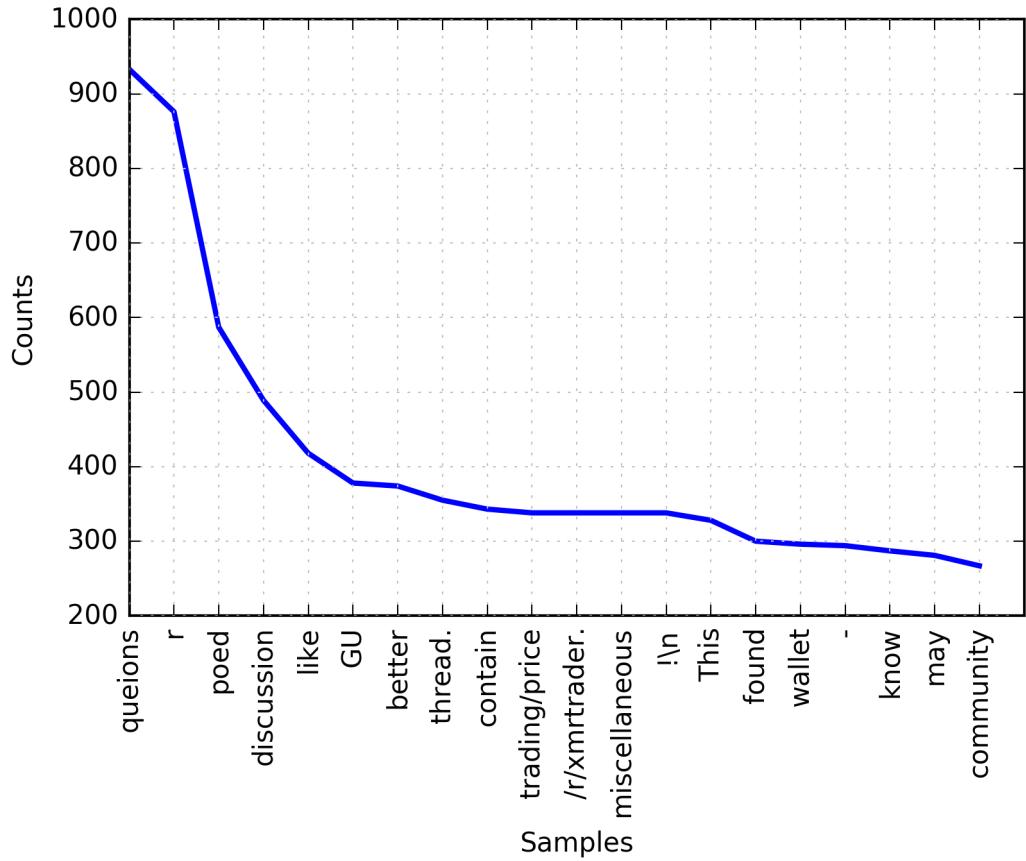
Data Cleaning

- English stop words
Tools: nltk.corpus
- Monero Subreddit stop words
Version 0.1.0:
['monero', 'Monero', 'MONERO', 'get', 'org', 'https', 'reddit', 'MAAM', 'comments', 'org', 'v0', 'tar', 'x64', 'will', '/n', 'The', '/nThe']
- Reddit Special:
http links, line breaks, slash, back slash, link number, etc..
-

NLP Insights

- Word Cloud
- Topic Modeling
- Sentimental Analysis

The top_11 KOL's posts word frequency after preliminary cleaning



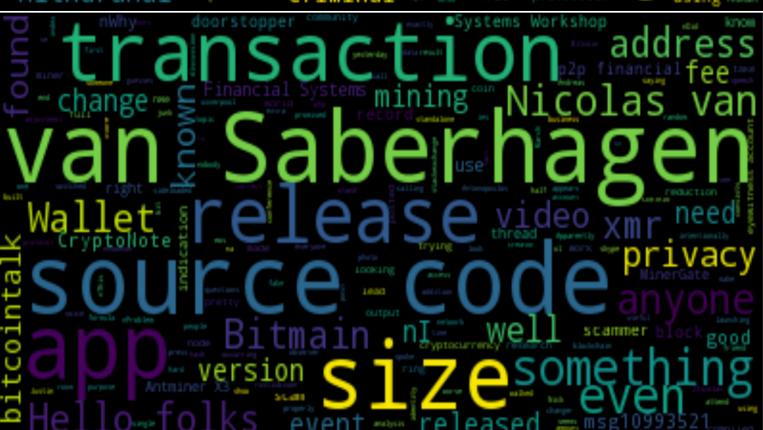
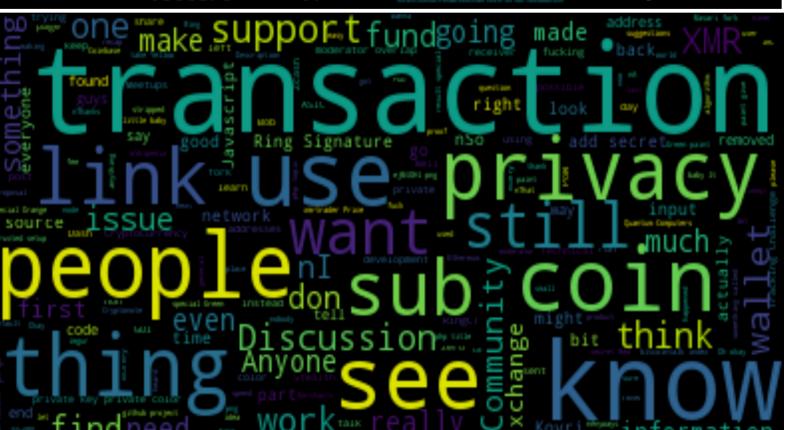
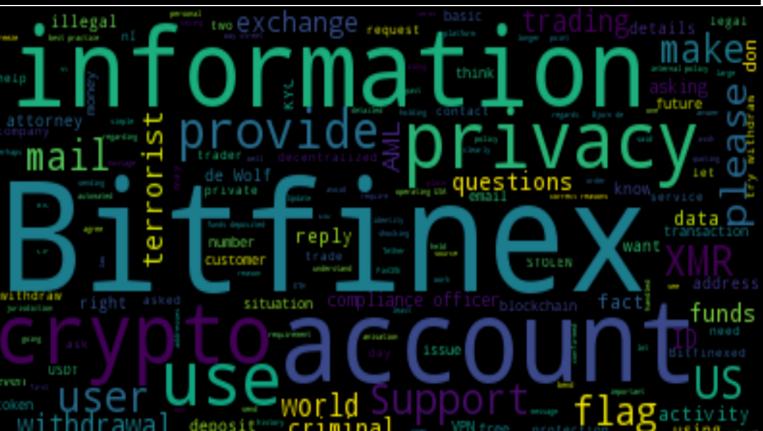
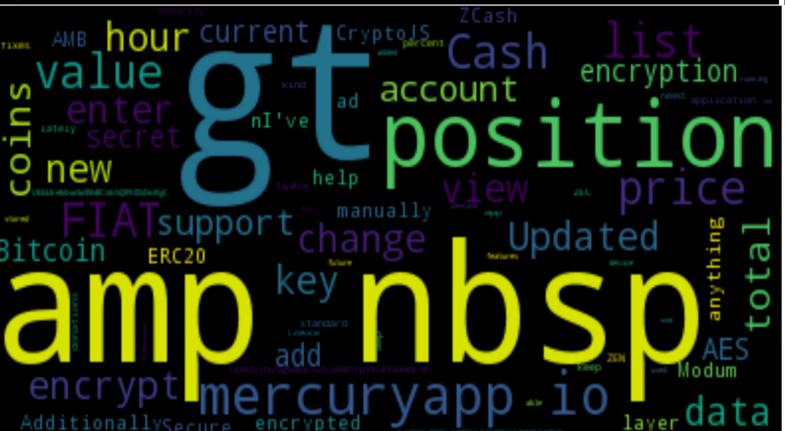
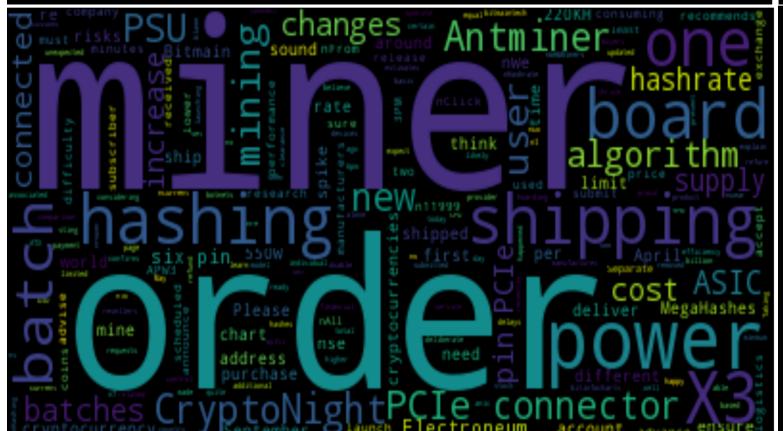
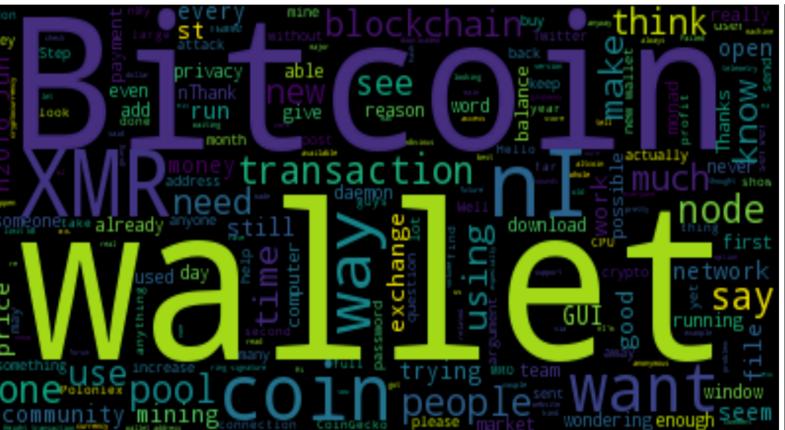
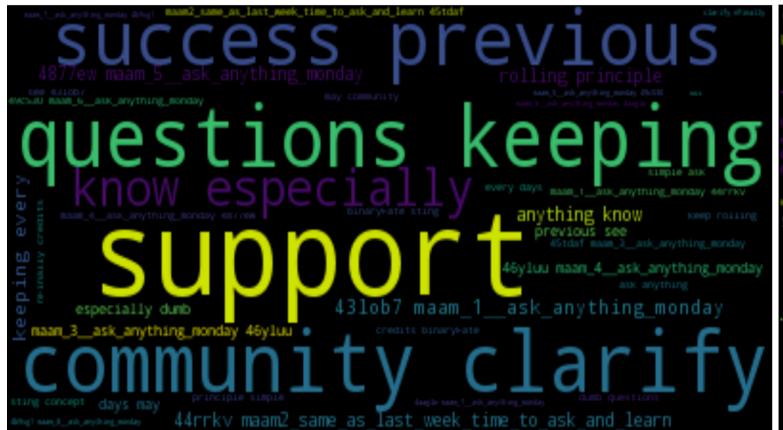
Positive?

The top_11 KOL's posts word frequency (count > 230):

```
for key,val in freq.items():
    if val > 230:
        print (str(key) + ':' + str(val))
```

```

suited:231
know:287
buying:239
directly:249
community:267
would:267
preferably:234
\n\n-:233
trading/price:338
reading:239
Mining:237
SE:231
better:374
/r/xmrtrader.:338
discussion:489
wealth:235
miscellaneous:338
thread.:355
queions:933
poed:587
fiat:235
!\n:338
found:300
available:262
make:239
options:239
li:258
like:418
GU:378
This:328
contain:343
-:294
may:281
contains:253
wallet:296
(and:267
new:234
related:253
r:876
currency:261
```



The top_9 KOL's posts data after preliminary cleaning

From left to right, each cube belongs to No.1~ No.9 KOL .

Only a few stop_words have been filtered in the result above, further data cleaning shall be down after discussion with our domain experts. Although the words above seems very general, we can differentiate them through intuition. Some basic conclusions we can get:

- People are discussing Monero together with Bitcoin, maybe some techniques comparison, price comparison...
- The No.1 KOL likes to talk about the community, and support, maybe he/she is one of the organizer of Monero Community.
- No. 3 ~ No.6 KOL seems to care a lot about the order, power, encryption, privacy. Maybe we can regard them as the “true believers” of the Monero Community.
- No. 7~ No.9 KOL talk about bitcoin, transaction a lot. Maybe they are investing in Monero recently.
-

Need your help

- What kind of insights do you get from the word cloud above?
- Please list the words you regard as useless, and grade them from 5 to 1 (5 is the most useless word).
- Did this help you get a general understanding of the community?
- Any suggestions ?

Topic Modeling and Latent Dirichlet Allocation(LDA)

- Topic modeling is a type of statistical modeling for discovering the abstract “topics” that occur in a collection of documents. Latent Dirichlet Allocation (LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions. In order to monitor a community, we apply LDA to cache the trending topics.

5 Topics

- Tools: python sklearn package
- Data: top_11 KOL posts (preliminary cleaned)
- Topics: 5
- Findings:

```
tf_feature_names = tf_vectorizer.get_feature_names()
print_top_words(lda, tf_feature_names, n_top_words)
```

Topic #0:

com www ackexchange currency discussion information contains se po r_newcomers_read_everything_need 5zgail 12 fir github work fluffypony simple
i_01200_lithium_luna_megathread_download

Topic #1:

poed wallet like available directly 10 contain know community use transaction 1201 ju think mo project bz2 release downloads way

Topic #2:

options fiat price xmrtrader new bitcoin nthank wealth newcomers amp transactions time privacy nf especially sure going previous really network

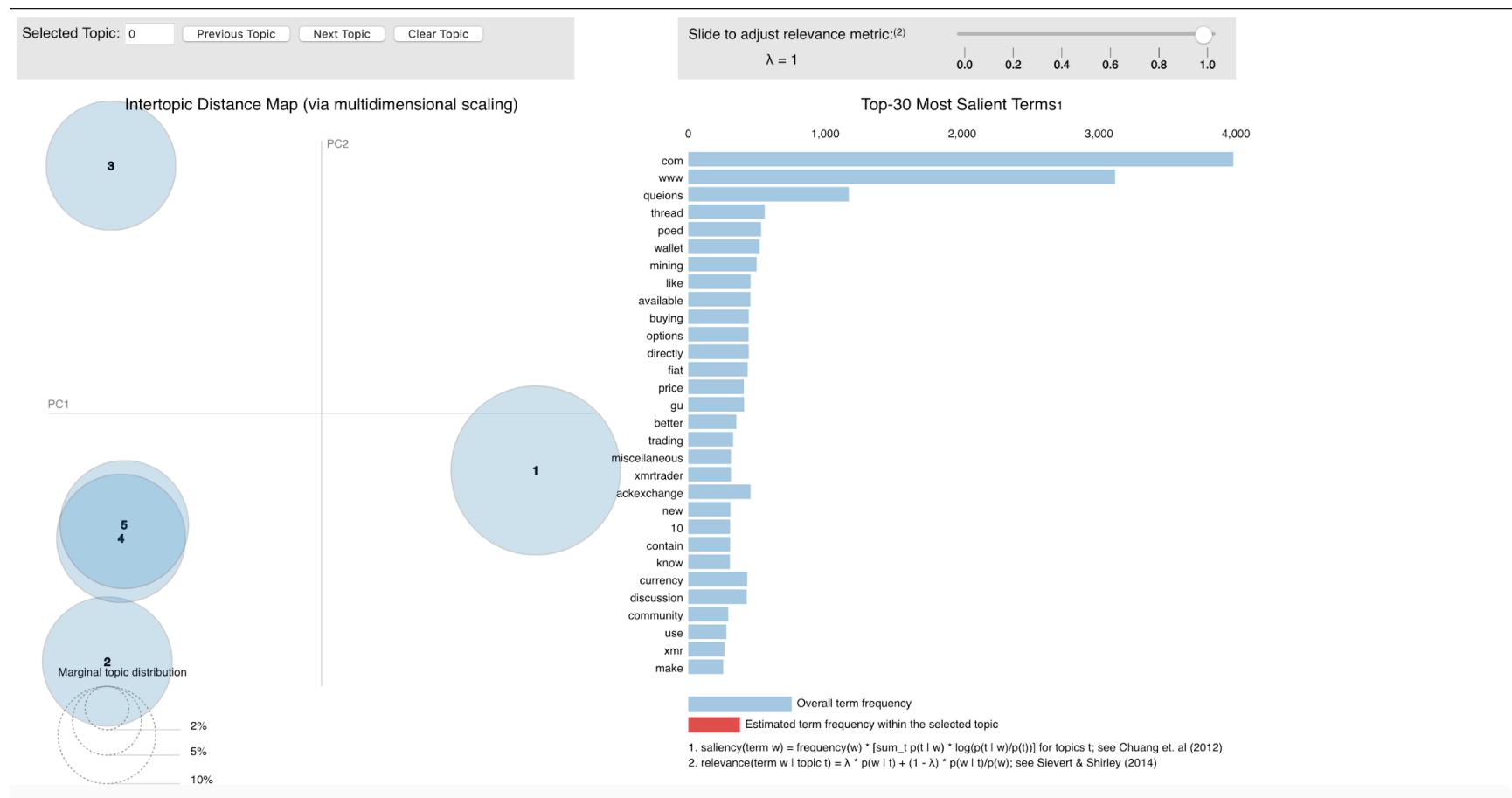
Topic #3:

thread mining gu better make people li related bit reading preferably pointers suited support removed private ask money ing gui

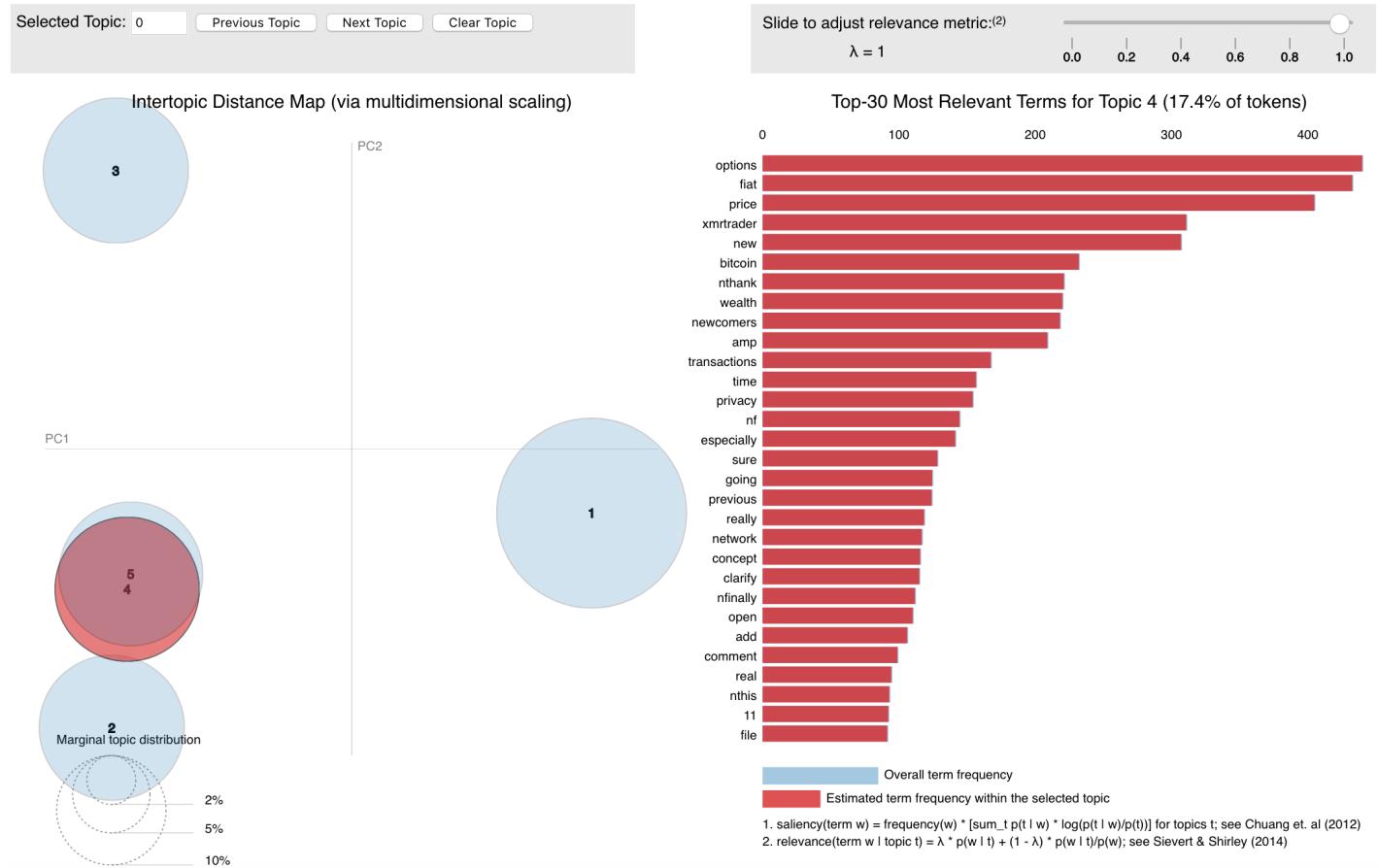
Topic #4:

queions buying trading miscellaneous xmr linux selling typicall http address need blockchain days issues code help using success dumb binaries

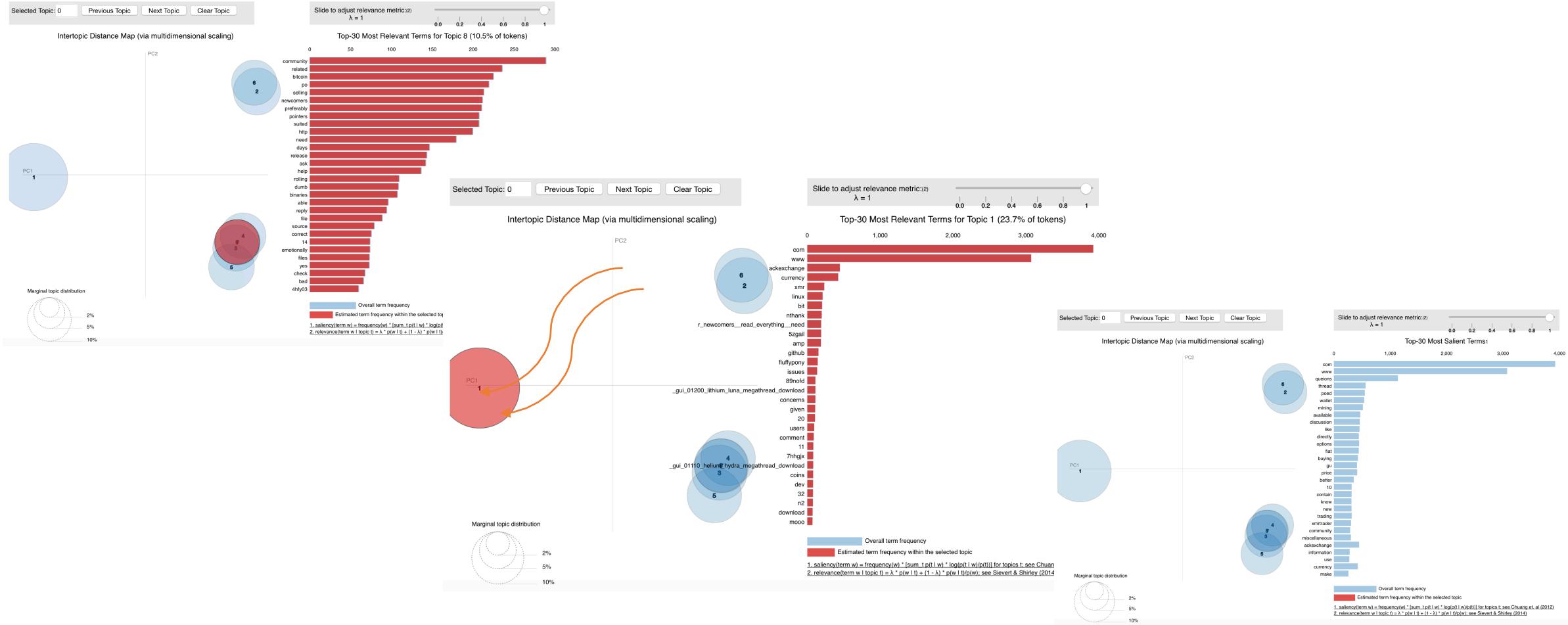
5 Topics



5 Topics



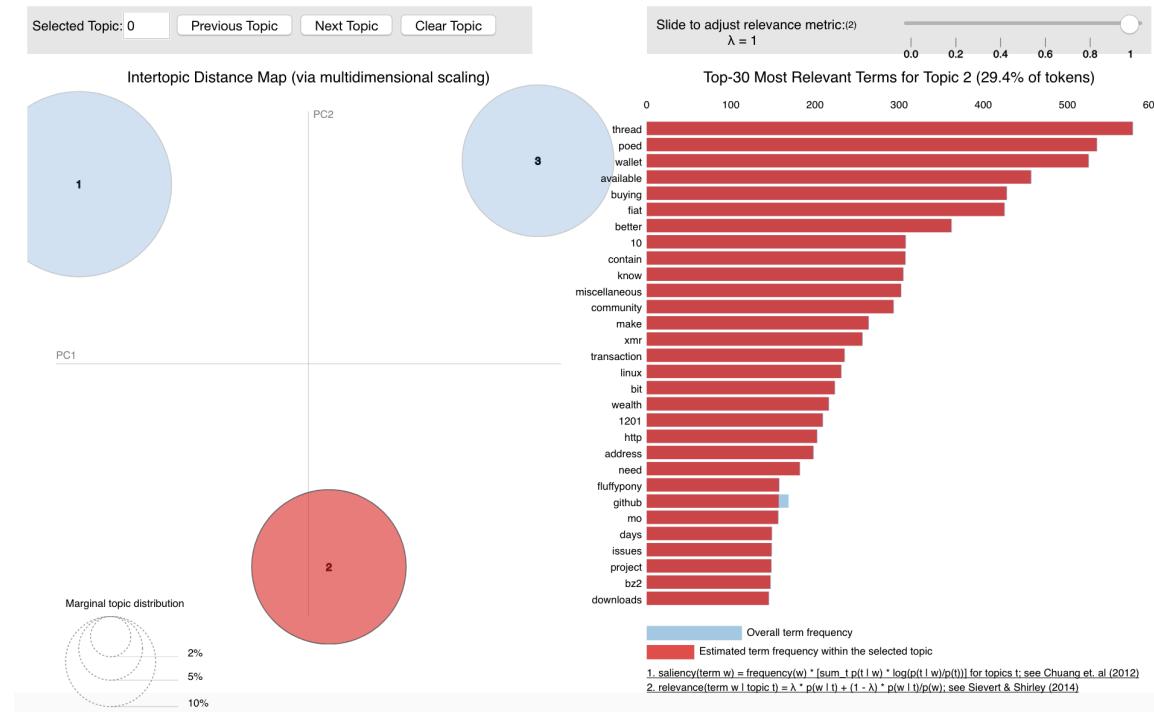
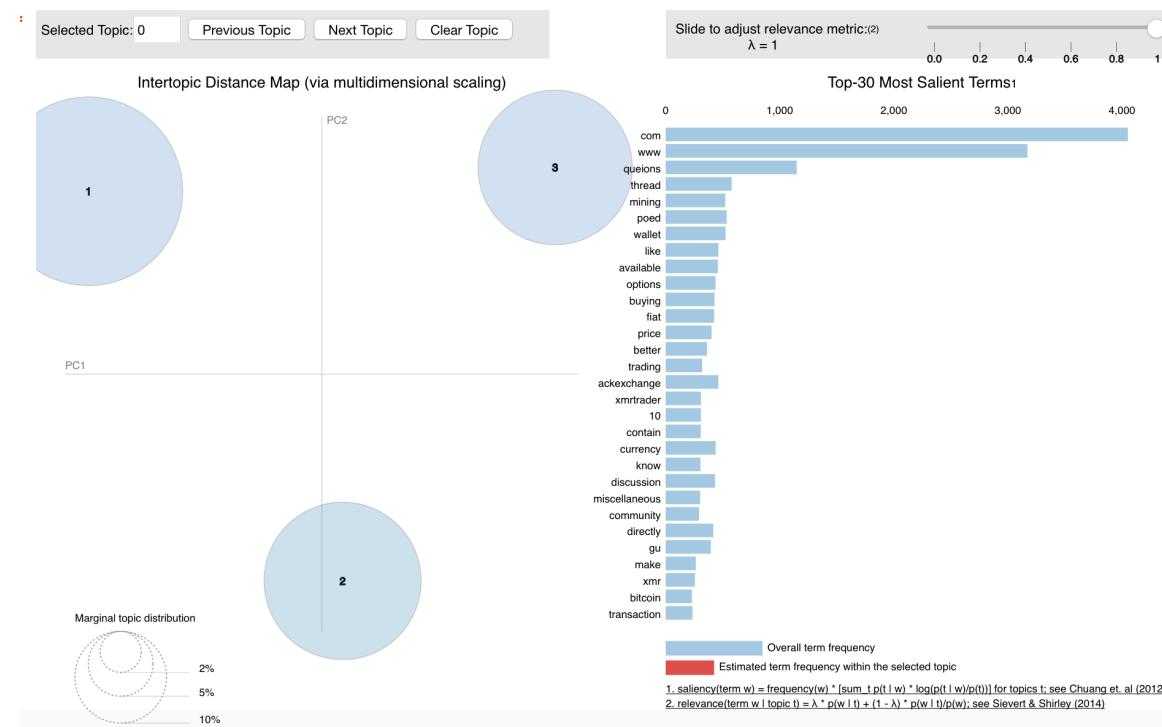
8 Topics



8 Topics

- Topic #0: information use fir blockchain work privacy simple good pool keeping binaryfate ill principle nfinally don nt idea aspects user fork keys 24 unto things following 13 4gakjp maam_13__askAnythingMonday easily ideas extremely cryptonote send great 19 thing maam_19__askAnythingMonday 4mq0ys expect maam_24__askAnythingMonday
- Topic #1: poed wallet available 10 contain know transaction wealth mo project bz2 downloads want windows possible thank cli mac node cryptocurrency topic xa0 22 crypto link 4d8kdb maam_10__askAnythingMonday zip teing 4q1891 maam_22__askAnythingMonday simplewallet php ve index reore bitcointalk x86 lea little
- Topic #2: options fiat price new 1201 typcall time way especially money sure going previous really concept clarify credits open nthis forum version 23 n1 upgrade major bug talk change bring funding seed ringct believe nto 4rqxn6 maam_22__askAnythingThursdayJuThree provide 26 27 maam_26__askAnythingMonday
- Topic #3: thread gu better make people li reading address support removed private nf ing key added ring got dont daemon 64 ns 15 n3 ll coin win point maybe bitfinex 21 hard process 4igtts maam_15__askAnythingMonday important technical running miners 4ozr3w maam_21__askAnythingMonday
- Topic #4: queions buying trading miscellaneous contains se 12 transactions gui network success real right week used funds 16 error command account hash works core value maam_12__askAnythingMonday 4f939i non world maam_16__askAnythingMonday 4igtts development developers problems set software mnemonic number solutions 42 future
- Topic #5: com www ackexchange currency xmr linux bit nthank r_newcomers__readEverything__need 5zgail amp github fluffypony issues 89nofd __gui_01200_lithium_luna_megathread_download concerns given 20 users comment 11 __gui_01110_helium_hydra_megathread_download 7hhgjx coins dev 32 n2 download mooo issue calm 25 syem feel maam_3__askAnythingMonday 44rrkv45tdaf46yluu maam_1__askAnythingMonday
- Topic #6: community related bitcoin po selling newcomers preferably pointers suited http need days release ask help rolling dumb binaries able reply file source correct 14 emotionally files yes check bad maam_14__askAnythingMonday 4hf03 multisig la payment addition te nice gt io marketplace
- Topic #7: mining discussion like directly xmrtrader ju think code using add block cl build main tru size 18 note currently ledger sort working 30 data 28 maam_18__askAnythingMonday 4lmn1y means allow wanted potential default maer allows directory 4z0azk maam_28__askAnythingMonday try small unconfrrmed

3 Topics



3 Topics

- Topic #0: com www ackexchange **currency** discussion directly gu new **information** use **people** li related contains se po r_**newcomers**_read_everything_need 5zgail amp fir blockchain work **privacy** **release** help simple good pool sure gui 01200 lithium_luna_megathread_download 89nofd **concerns** 4igtts 20 keeping binaryfate ill **principle** **rolling** **credits**
- Topic #1: **thread** poed **wallet** available **buying** fiat better 10 **contain** know **miscellaneous** **community** make **xmr** **transaction** linux bit **wealth** 1201 http address need fluffypony **github** mo days issues project bz2 downloads ask want windows gui possible previous thank cli success 12
- Topic #2: queions **mining** like options price trading xmrtrader bitcoin reading nthank selling newcomers preferably pointers suited typicall ju support removed think transactions private time nf way especially code using money ing going given really network key concept clarify added nfinally open

Need your help

- What kind of insights do you get from the topics above?
- How many topics you expect to get from a community like Monero sub-reddit?
- Did this help you get a general understanding of the community?
- Any suggestions ?

→ Done
→ To do

