Latent Topic Modeling on Twitter Data

Method: CA

Correspondence Analysis (CA) is a multivariate graphical technique designed to explore relationships among categorical variables. The outcome from correspondence analysis is a graphical display of the rows and columns of a contingency table that is designed to permit visualization of the salient relationships among the variable responses in a low-dimensional space. Such a representation reveals a more global picture of the relationships among row-column pairs which would otherwise not be detected through a pairwise analysis.

Calculate CA:

- Step 1: Compute row and column averages
- Step 2: Compute the expected values
- Step 3: Compute the residuals
- Step 4: Plotting labels with similar residuals close together
- Step 5: Interpreting the relationship between row and column labels

How to Interpret Correspondence Analysis Plots

Correspondence analysis does not show us which rows have the highest numbers, nor which columns have the highest numbers. It instead shows us the relativities.

- The further things are from the origin, the more discriminating they are.
- Look at the length of the line connecting the row label to the origin. Longer lines indicate that the row label is highly associated with some of the column labels (i.e., it has at least one high residual).
- Look at the length of the label connecting the column label to the origin. Longer lines again indicate a high association between the column label and one or more row labels.
- Look at the angle formed between these two lines. Really small angles indicate association. 90 degree angles indicate no relationship. Angles near 180 degrees indicate negative associations.

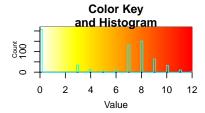
Dataset

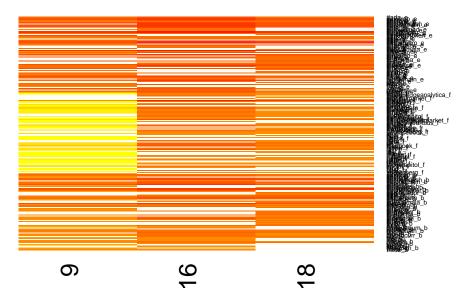
```
## # 9 16 18
## #ada_e 8 9 8
## #airdrop_e 9 10 9
## #altcoin_e 9 10 9
## #bch_e 7 8 8
## #binanc_e 8 8 8
## #bitcoin_e 10 11 10
```

- Research Question
 - Do we see new words appearing on a particular week

Analysis

Heatmap



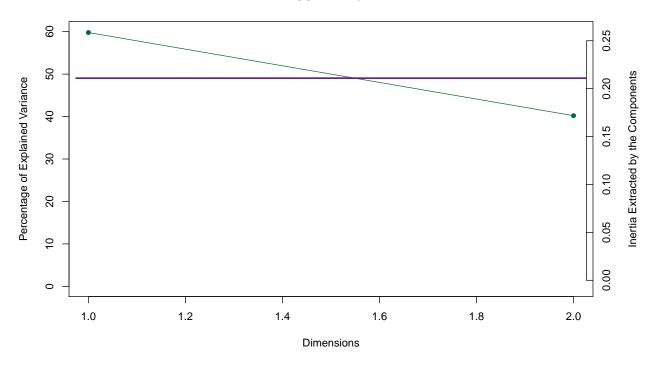


Scree Plot

Gives amount of information explained by corresponding component. Gives an intuition to decide which components best represent data in order to answer the research question.

P.S. The most contribution component may not always be most useful for a given research question.

SCREE Plot

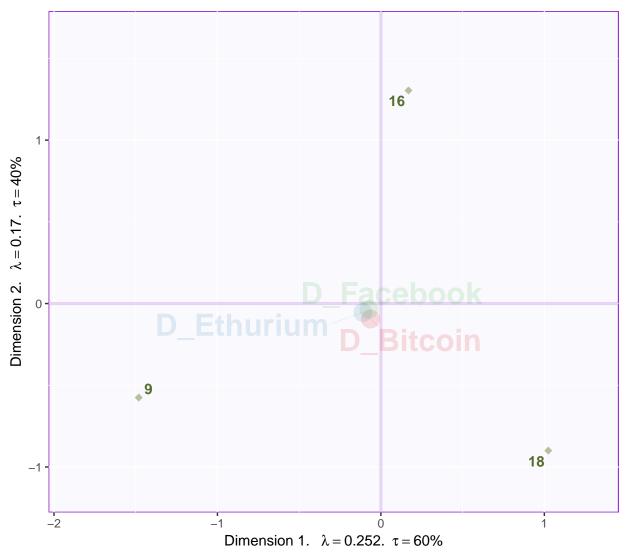


Factor Scores

Asymmetric Plot

```
#### Asymmetric Plot

map.IJ.asym <- asymMap$baseMap + asymMap$J_labels +
   asymMap$J_points + labels4CA + legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)</pre>
```



```
#### Asymmetric Plot

map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
   asymMap$I_points + labels4CA + legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)</pre>
```

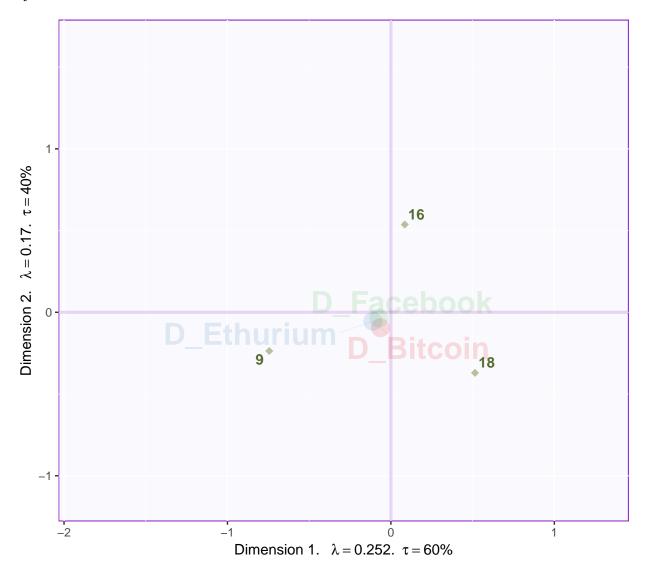
```
000guarium_b /#geoloc_e /#smm_f
                               crypto_e ico_b start_b blockchain_e ico_e #data_f
                       #crowdsal_e sell_b
               project e #trade_b
                                       al_f #gdpr_f su_f
        #bounti_e token_e#ethereu#aoendcloud_#bts_e #digitalmarket_f #bittrex_b
            #invest e market b moment f en f #music_f #webradio#actu_f
                 #xvg b #fintech b#linkedi#instagram#justic_f la_f /#média_f
       free b
                                 #boun#tdkensal_#market_fp#gænfbridgeanalytidata_f
Dimension 2. \lambda = 0.17. \tau = 40\%
                                            facebook#googl/#faitsdiv_fuser_fbuy_b
             token b#cybersexyptbcurr b
                  trade prowdfund caypterfacebook f #radiocapite/privaci_f de_b
      telegran#airdrop_e#trade_e bilcoin_b\btc_e #in#socialn#ediake##singoinqushaei_f
                                                    social/f/el fzuckerberdescort e
           #ltc_e #altcoin_e eth_e #binanc
      #bch b #btc b #ico b #altcoin b #invest#byitter#wrigu for thew f#tge #new e
      secur b#blockchain_e#xrp_e #ada_e#tokensal_e
                   #crypto_e #btc_e #ico
       #coin\b
                                                         para f #escort b #girl b
               busi f #bitcoin_@tc_b #xvg_e
                                                             media f ce_f #video_b
                                                 hew ad f
          #ada b #neo_e join_e #eth_e#ta** efollowde #busi_fost#finder_f
      #energytoken_#cryptocurr##Im_e #rippl_e
                                                   #token beth #snapd/catsfarica b
      #free_eardrop_market_#tokeblockchainy_btockether_price_lderc20#goutub #rt_b
                 timethiphousasomityou#unksonoteeth_thoch_etxr#bucoincas#f@news_f
      #coinbats/eng##plravliocapitoideo#litecoin bt#litecoirettee#rlopldateffatur f
      #liv#soojæ#netdianiareket#fusa#airdro/ockchain_bles_f #eo_b app_f #whatsapp_f
      #etc_eree e tip t peoplaifdrop#deash #eron e#vip b #guater#elaabador b de_e
      #neo_b #gto e por fcheck #edeehereun#usa#mexico#et e #usa_b##polwplayipbe
     friendearn_#xrb_econ#binan#biec#guatemal#ne #trav#pana#pan#batbatborice_b
                             Dimension 1. \lambda = 0.252. \tau = 60\%
```

```
#### Asymmetric Plot

map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
   asymMap$I_points + asymMap$J_labels +
   asymMap$J_points + labels4CA + legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)</pre>
```

6

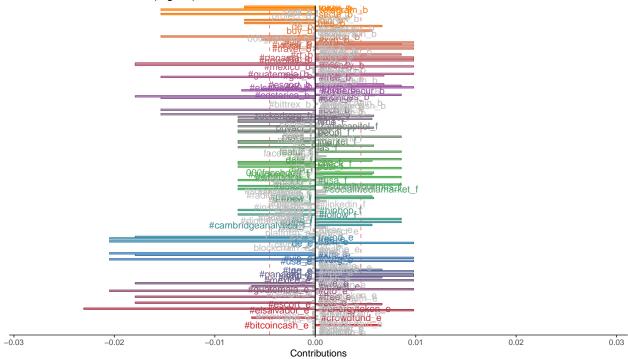
Symmetric Plot

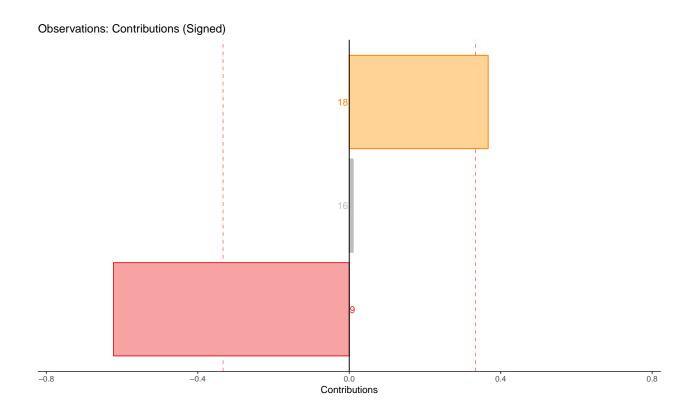


```
000quarium b #geoloc_e #data_f /crypto_e
                                                   b #gdpr_f
                                                     #info f
                  #bounti_#invest#soundcloud_fen_f
                                                         #justic#fuamdwidgeanalytica f
          trade project #altcoin_bnoment_f#tokensal_market_f el_f, #média_#privaci_f
      #free b joi#cybersecultrade #txad#libkedinacebooksbcial#faitsdi#bitcoincash_e
\lambda = 0.17. \tau = 40\%
                                             #facebockmtisitkwtebradio_f/data_f / de_b
            #btc b #altcomerewdforge weether eting to gram the Lucker be specific kerberg f
      secur b #bitcoin_e#ico b #dash nbarket bbtc_e page f /buy_
              #btc_e #ico_e #xrp_e #binanc_btwitte#racliocapito #new_f#escort_e
       #bch b#crypto#blorclechaitheth #adaceyp#tto
                     join e #ltc b#crypto e #bounti b
Dimension 2.
                                                      follow f ce f
       #verg\e
                   radiocapitol#trx e #token e
                                                           ad #youttahapttchettarica_b
                                                     #new b eth b de f #rt_beatur_f
                  spendhoediammadtef_f#trx ##thereum
                                                        #token b#xrp_b date_fnews_f
      #liventergy oken e con_f #litecoin#wy#freetoken e rc2fbiecoincash#wideo_b
         #free_e #follow#hipblepckich#einy betocurr_b#litecoin_e #ether_b #eo_p#tinder_f
      #xmr_e free_eme#binantaiedrop#dash_e #eth_b #kippl_b price #whatsap#f8_f
      #etcombassbentityourtinustripsbiscoin_ttron_e btc_b
   ___ friend_#xrb_erideo_busi #eo_e#blockchain_les_f
                                                                 #travel b
      #neo#gto_e#usa_f
                                                                        #panama_b
                                   #mexico #usa b #mexico#et_e
      #quatem#lap el#v#pamauraemallaphelsalvado#rlowplay_b#elsalvado#vep_b
                                                                                  de_e
                               Dimension 1. \lambda = 0.252. \tau = 60\%
```

Most Contributing Variables







Inference CA

