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.

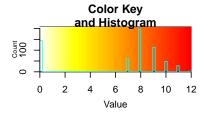
Dataset

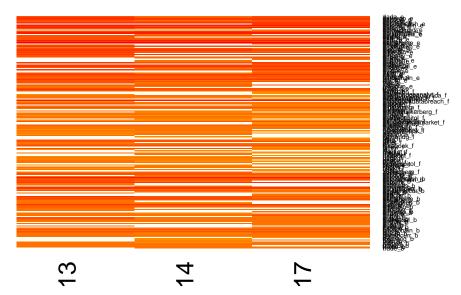
```
## #ada_e 9 8 9
## #airdrop_e 11 10 10
## #altcoin_e 10 10 10
## #bch_e 9 8 9
## #binanc_e 9 9 9
```

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

Analysis

Heatmap



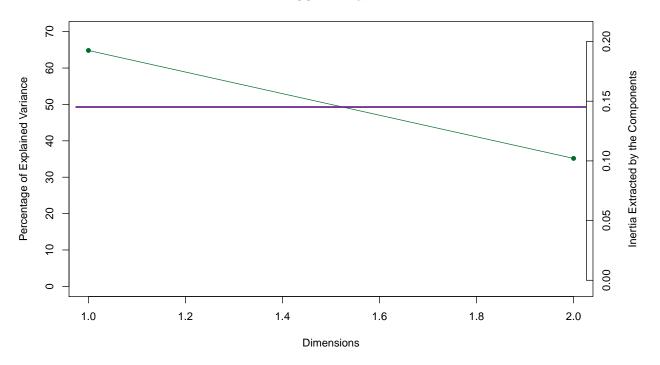


Scree Plot

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

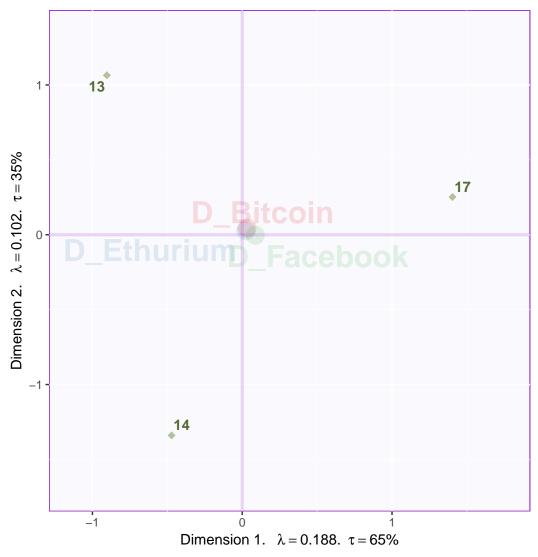


Factor Scores

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

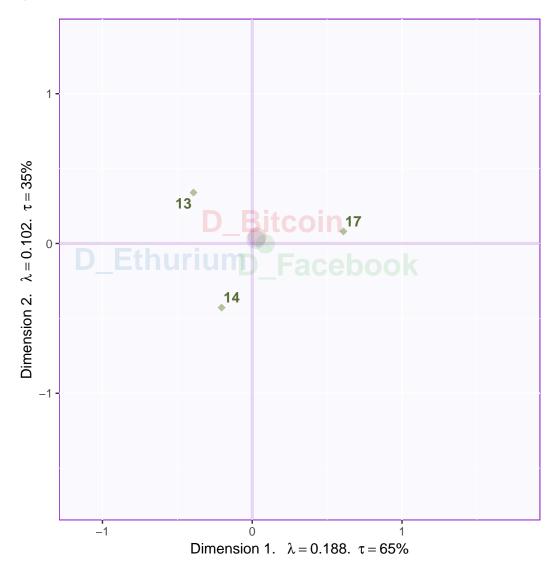
map.IJ.asym <- asymMap$baseMap + asymMap$I_labels +
   asymMap$I_points + labels4CA + legend$zeMap_dots + legend$zeMap_text
print(map.IJ.asym)</pre>
```

```
#moneWousi/#learnbbidge#rijap/#lefch_ele_foken#laccial/mediamarket_f
      #free e #fasebookdatabreach_f join b
                                                           #média f
                                                                    #escort b
      #earn#eame_e #free_b #ada_e #bitcoincas#cbypto b
                             #airdrop_e de b#instagram f
      call f
                                    #eo_e/#market f radiocapitol_f/#np_e
                     #market_b
          delet
                                                    #music_#hiphop_/#tge_e
                 price e token_e
      time f #token_#coin_#litecoin_e
                                          ///market_b #eo_#blinke@li#mfp_b
           #eth_bpost_#bounti_ebuy/bpr/ce_b#justimome#aniazon#busi_f
\lambda = 0.102. \tau = 35\%
      #bbiletefacebeippl#freetoken je meshaptilaaliofcapitodeffilitechcen f
      free_e#freetoker#googi_f#xlm/e#webradienfew_#twitter_f #travel_b
      #geoletoobn#moneycrypto a epopler #new withdogitatson ketcloud f
          peopl i projettiata_f projettialta_f projettialta_f projettialta_f projettialta_f projettialta_f projettialta_f
      privaci_f*saandadigennavitaticoin_de_project_ethcrewdfuadee_f
platformate/icmediaryprocetterrance_bet_f*sunidtiaftsdfvideerreexico_b
Dimension 2.
      newschempnigati la bragintechara page f#bt#betcoincasblew_f
      aconniace #acktopowdsatte #ico bairchooke #faine bottbitco #xeen e
      zucke#beygton#ether#er##wad#rteo#bch #xvg #x#eth#beorkchain e
      user_f data_f#xy#etheretlitedoir#neo_teadtcryptocutidaeh_e#erc20_e
      ad f#inves#teken #btc#bintec#cbypto_e #new e#iso_e#et#binanc e
      #cyberseblockchain_bos_f #invest_b btc_e #tron_e #ltc_e
         #ge#lolockcha#snbm#coinbas #trade#eltcoin_e free_b
      #ethe#tokensal b hour b #tokensal_e #dash_b
                             million f #pdata e
         cambrida f
                      ico_e
      exchang b
                                                 share f
          #markzucke#logyotonew #epirium_e
           -1
                         Dimension 1. \lambda = 0.188. \tau = 65\%
```

```
#### 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>
```

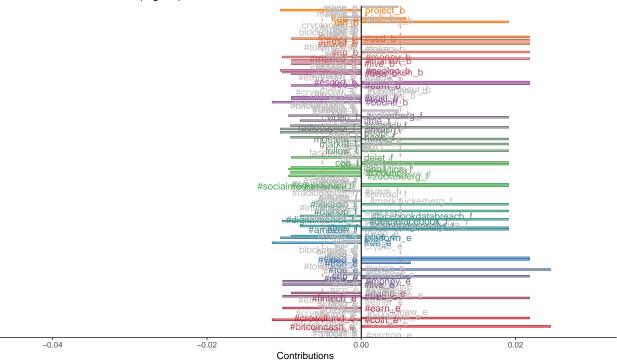
Symmetric Plot

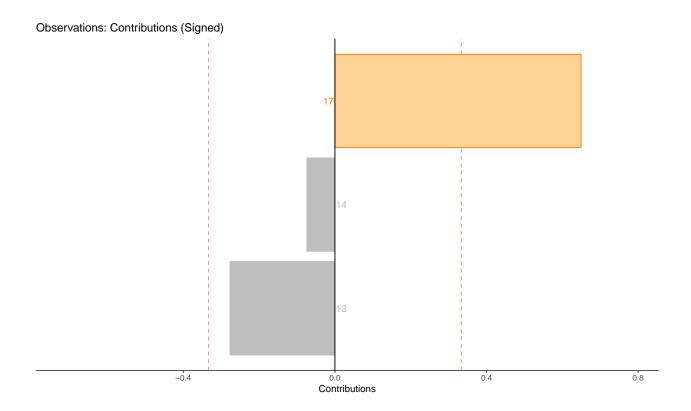


```
#busi<u>##eam#Neick|eamladettaf#eeacjoin_#trx#e</u>##Ndetcofiedesdejbm#ttyet_e
      price e #game e#free b
                                xlm_e//le_f /price b
                                                        #linkedin f
                           oin_b#bch_e#cry/pto_b/#justic_f
                                                            #amazon f
     #live_b #free
                     #usd_#airdrop_ede /#instag#ahiqifalmarkeltofw f
       #live_e #earn_e#bounti_e buy / market rhoment#bus#hiphopce_f
      call_#money_crypto_#litecoin_facel#oatliocapitotlefactiocapitotlefactiocapitotlefactiocapitotlefactiocapitotlefaction
     free_e_#token_b#freetoken#enapchat#fnusic_#bitcoin#ashnecloud_f
\lambda = 0.102. \tau = 35\%
     #deleterace#etbkutosttóken#eippl eocia#socialmedia_feth/b
     #bounti_money_#googl #fada teke#twitte#inflowebradio_f
     platione belt btc e join travel #bsa_b
Dimension 2.
     nearsalf/ticapfiva#ethe#teadth manganeprojedockchain#etc_e #np_b
     account_fid_f data ###lockchamabf #xvg_e#faitsdiairfdrop_enp_e
      zucker#eybers#coowdsallecon#lico #ico e #xrp #bitcoin#eideo_b
        user f la blockchatfintechirtyest #neo_ethew_efacebook_f #eo_b
     media#étherbitcottetbereun#ltb_bbtc_trade_ttrate_tbe#blockchain_e
      #cryptone #btc_#tokensal_#bokensal_e#trade_e#ltc_e#eth#leinanc_e
     #airdrop_b/cryptocurcabbridg_f #altcoin_e#bch_b free_b#erc20_e
                                  #crypto_e \#dash_e \#cryptocurr_e
      #invest_#smm_f__ico_e
                              millio#markzuckerberg_#dash_b
      #token #coinbas b
                                             #opirium e
               #cryptonew #pdata_e
                                     hour_b
           exchang b
                         share f
          -1
                       Dimension 1. \lambda = 0.188. \tau = 65\%
```

Most Contributing Variables

Observations: Contributions (Signed)





Inference CA

