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.

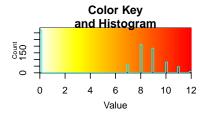
- The further things are from the origin, the more discriminating they are.
- Look at the length of the line connecting the row label to the origin. Longer lines indicate that the row label is highly associated with some of the column labels (i.e., it has at least one high residual).
- Look at the length 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.

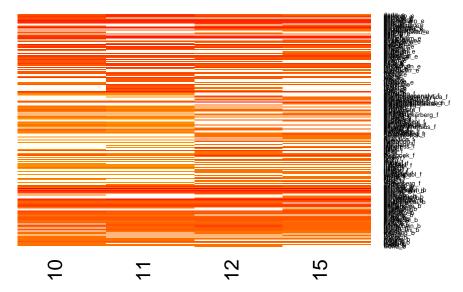
Dataset

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

Analysis

Heatmap



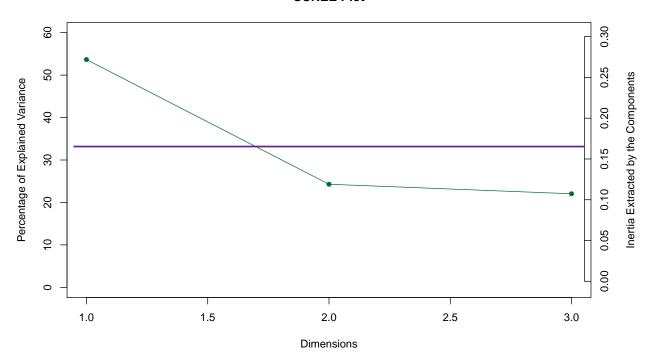


Scree Plot

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

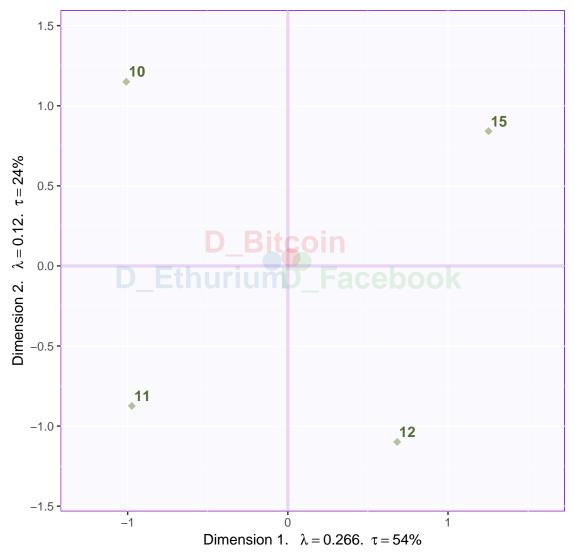


Factor Scores

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>
```

```
money#uttokensation pripplb #crowdfund tege_e
                          ocapit#ether_b #token_e #neo_b hour_b #freecoin_e
                             #xrp_e #trade#blecoin/drypto_econgress/f/ceo_f
        day
                                 #bounti e btc e #eth/b#coinbas_senat#maga f
    1.0 -
Dimension 2. \lambda = 0.12. \tau = 24\%
    0.5
    0.0
                                                facebook f
                                                 ha#cyberseenaly# @@fofacebook_f
        busi
                                                       ad #deletefattatyottmew_e
                              etetiene_e
                                              world b
           con_#a#lgraktaetokemmeandiff#bts b
   -1.5 -
                              Dimension 1. \lambda = 0.266. \tau = 54\%
```

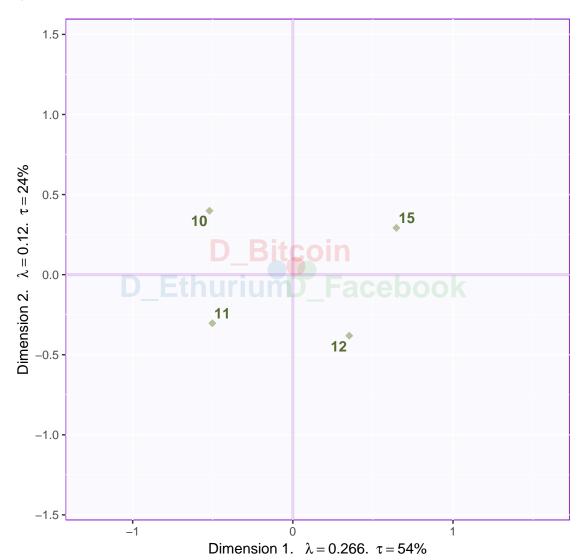
```
#### Asymmetric Plot

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>
```

```
#newstart /#xrb#enoneyorbject#byitter#ether b /#ico e #freec#bydata_e
       #c#emergytokepaea_f btc#bour#tir/pptoc#fintech b #crowdfundiguess_f
               #bch/#pinterest_f | #music_f
            eauth a tyourntts yptonew b ##eth/b/
                                                en_f hour_b
                                  hbinanabeunjicebookdatalaak_f
                                *xr#bleckchaimyeto_e question_#geoloc_b
              f
#média f
                          #cryptocu#crypto/ #itc/b
                                                    #tge_#opirium#reaga_f
Dimension 2. \lambda = 0.12. \tau = 24\%
                      #neo_e #freetokence#digitizecoin_ceyptocurr_e ceo_f
       #new_f
       0.5
                                             #invest time_f
       #amazon #bitcoin_b#litecoin e
       radictapit befrista orface fe
                                                           000fadelegicolocfe
       #diustalntacke#lilm_e#dash_en tracebtackbrookalytice #cryptonew_e
                                              bersecur b
cambridg\f account_f
                   #dash_#itc #x/p btrx b #erc20_eel_#deleteface#polvaci_f
                      #trxteke#a/ecoin#wade b #xkg b
                                                             share f fb f
       #linkedin_#grabcoin_b/join_price_blockchain_the_bplattiofm_e le_f
       video_fldhanc_eairdrop_e/
                                                                   #data f
                  eroject_eeth/elink_e_nogl_f
                                                     ether e peopl_f #socialmedia f
       #ada_b/reach_e #eth_eegistrypto_b
       mani_effer_e#xvg_e time_e
                                      dato f scandal compani_fmedia f
  _1.5 - goffice state eplanks have teum profrid_bdelet_f
                                                  persondiffuttobiabafidgean
```

Dimension 1. $\lambda = 0.266$. $\tau = 54\%$

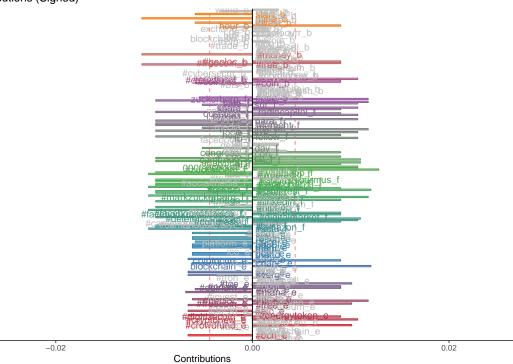
Symmetric Plot

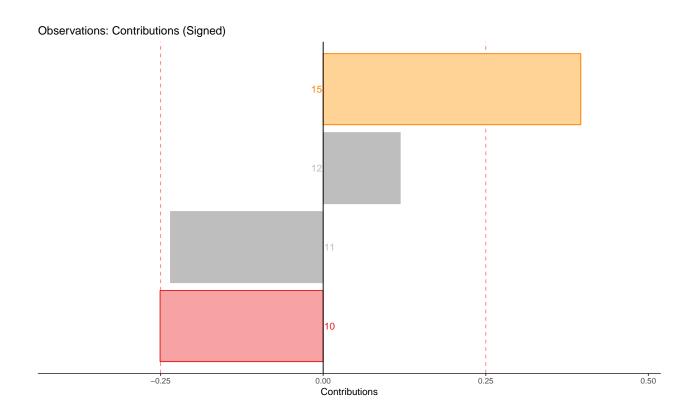


```
1.5 price_e#coim_test_b#etc e
                                                                                                      ##eagebookdatalgedre/iptocurr_e
                  start #new_e _#bch_e _ #xrb_e _#freetok#ether@unip#skighizecoin_e
                                   #free e earn e#youtu#new b #ico b
                                                                                                                                              hour #pdata e
                                   #energytokercreyptocktimyest by #eth b #ltc_bsenat_kfreecoin_b
                n#akki#boapitiolgasitoldoplemyptonemicb/e##actebrookdatabookuffuncada f
                  sur#webradiorojektokens#trædelockc##winceiiteeoirydto hear ____ceo f
                 #faitsdiv#free_beth_epage#boundthere.u#bbc#tekenestion_fcongress_f
                         #actu_f /#btc#tokensatohetcryptocumee #ripplatfofmedopiriuanen f
= 0.12. \tau = 24\%
        0.5 #smn#niediabtc #ewitterbounti #xrp#erypta#teoexeh#freebothgroedfundle
                 liksnáp#hotnévygbogbihandottco#binan#me/estoevdsaice e#altcoin eser f
                 btushatsaminé_b #eo_e#neo #en ech b et#ciybersecutix no #geologart f
        #new_f #litec@bloekchaftm.by #sticialmackbusckenkenkengkerberg_f
0.0 - #hip#pubmesty.ottdensis#tippl_ee social_f ad #privaci_fpeotpt_owdsal_b
                 radiocapitathuridelojoinf e#ada capto b el_f privaci_f 000facebook_f
Dimension 2.
                 news#inkedin_f ce_f free_processoryptonedia_share flaccount_f
                        follow f para_f#xlm_airdrop_eth_b
                                                                                                              la f cambritayptonew_e
                                                             #dash_e#music f de b
                                                                                                                                  post f\time f#data f
                 ##MAN Harman Har
                                                              #ltc e #altcolblockchai#xbg#bacebooblockchain e
                                     peopleocury #eth_e join_b#trade #ether_e de_fanalytica f
                             goal_e
                                                             #trx e market b#trx West 20: egeants lighter fallyt 5
      -1.0 -
                 #ada_b/refer_feee//b
                                                                #xvg registinestagramed b
                 startfreend_e #info_f
                                                                     share e link e
                 chanc_etrade_project_e time e #facebookg compani_f world b
      _1.5 offer_enani_leitcoin#dirdrop_e person_f
                                                                                                                           #bts b\#trump f
                                                              Dimension 1. \lambda = 0.266. \tau = 54\%
```

Most Contributing Variables

Observations: Contributions (Signed)





Inference CA

Bootstrap ratio 1 -5000 -150

