Election Topic Network - Czech Presidential Election 2018

Matous Pilnacek 24 ÄŤervence 2018

Contents

R Markdown . . . R Markdown library(foreign) library(readxl) library(plyr) library(igraph) ## Attaching package: 'igraph' ## The following objects are masked from 'package:stats': ## decompose, spectrum ## The following object is masked from 'package:base': ## union library(knitr) NS1802 = as.data.frame(read.spss("./../data/NS_1802_DataFile-FINAL.sav", use.value.labels = F, reencode ## Warning in read.spss("./../data/NS_1802_DataFile-FINAL.sav", ## use.value.labels = F, : ./../data/NS_1802_DataFile-FINAL.sav: Very long ## string record(s) found (record type 7, subtype 14), each will be imported ## in consecutive separate variables ## Warning in read.spss("./../data/NS_1802_DataFile-FINAL.sav", ## use.value.labels = F, : ./../data/NS_1802_DataFile-FINAL.sav: Long string ## value labels record found (record type 7, subtype 21), but ignored ## Warning in read.spss("./../data/NS_1802_DataFile-FINAL.sav", ## use.value.labels = F, : ./../data/NS_1802_DataFile-FINAL.sav: Long string ## missing values record found (record type 7, subtype 22), but ignored ## re-encoding from utf8 topics = read_excel("./../data/temata_2kolo.xlsx", sheet = "temata") NS1802\$topic1 = NS1802\$PS.23aa NS1802[NS1802\$topic1 %in% c(98,99,0,96,97), "topic1"] = NANS1802\$topic2 = NS1802\$PS.23ab NS1802[NS1802\$topic2 %in% c(98,99,0,96,97), "topic2"] = NA

```
\# NS1802$vote = mapualues(NS1802$PV.149B, c(96,99,0,9,1),c(NA,NA,NA,"Zeman", "Drahoš"))
NS1802$topic1 = mapvalues(NS1802$topic1, topics$val, topics$label)
## The following `from` values were not present in `x`: 98, 99
NS1802$topic2 = mapvalues(NS1802$topic2, topics$val, topics$label)
## The following `from` values were not present in `x`: 98, 99, 21
ties = data.frame ( table ( NS1802$topic1, NS1802$topic2) )
ties = ties[ties$Freq > 0,]
names(ties)[3] = "weight"
head(ties)
##
                            Var1
                                                   Var2 weight
     Důchody, důchodová reforma Bezpečnost, terorismus
## 4
## 5
                       Ekonomika Bezpečnost, terorismus
                                                              3
## 6
                        EU, euro Bezpečnost, terorismus
                                                              1
## 8
          Kritika současné vlády Bezpečnost, terorismus
                                                              1
                         Migrace Bezpečnost, terorismus
## 9
                                                              4
## 11
              Osobnost kandidáta Bezpečnost, terorismus
                                                              2
nodes = data.frame ( table ( c(NS1802$topic1, NS1802$topic2 )) )
kable(head(nodes))
```

Var1	Freq
Bezpečnost, terorismus	33
Církevní restituce	7
Daně	7
Důchody, důchodová reforma	24
Ekonomika	38
EU, euro	159

```
g = graph_from_data_frame(ties, directed = FALSE, vertices = nodes)
g = simplify(g, edge.attr.comb = list("sum"))
plot(g, vertex.size = log(V(g)$Freq)*4,edge.width = E(g)$weight)
```

Pravomoc prezidenta Pravomoc prezidenta Minulost kandidátu Útoky a linkandidátu Útoky a linkandidátu Proce voleb Kdo zimena politické situace Proce voleb Kdo zimena politické situace Osobne tropy vya mika Spoluprzow kinkalitati kandinakanictví Bezpechost duchodová reforma Done Církevn restituce

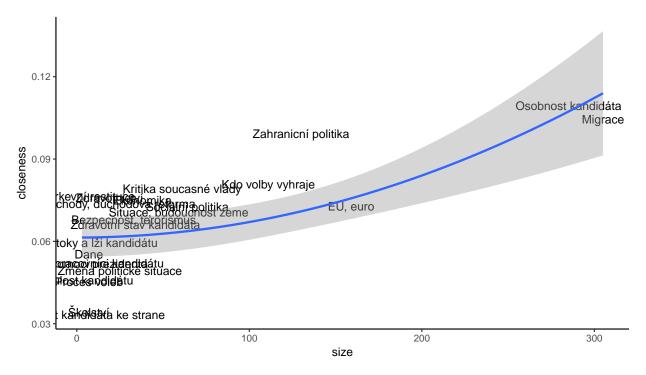
```
closeness = data.frame(closeness(g, weights = 1/E(g)$weight))
closeness = cbind(rownames(closeness), closeness)

closeness = merge(nodes, closeness, by.x = "Var1", by.y = "rownames(closeness)")
names(closeness) = c("label", "size", "closeness")
```

Vztah mezi dvěma tématy vzniká díky tomu, že jsou ** 1) pro danou osobu obě významná ** 2) je mezi nimi v paměti asociační vazba

```
library(ggplot2)

ggplot(closeness, aes(x=size, y=closeness)) +geom_text(label=closeness$label)+
   stat_smooth(method = "lm", formula = y ~ I(x^2), size = 1)+ theme_bw() +
   theme(panel.border = element_blank(), panel.grid.major = element_blank(),
```



Takto agregovaná síť je problematická, protože agreguje dohromady všechny respondenty, kteří mohou mít ale síť postojů výrazně odlišnou. Je proto vhodné síť rozdělit mezi více skupin. V našem případě na voliče Zemana a Drahoše

Jak ověřit důležitost zapojení v síti? Nárh: pomocí regrese

Závislá proměnná by bylo zapojení do diskuze pod internetovým zpravodajstvím na iDNES.cz během druhého kola prezidentské volby pro podpůrce Zemana a Drahoše (počet lajků/dislajků u komentářů pro Zemana/Drahoše). Články by byly tagovány k jednotlivým tématům ze sítě.

Nezávislé proměnné by byla velikost uzlu, clossenes a míra zájmu politiku lidí, kteří dané téma zmiňovali. Otázka je, zda closeness hraje roli a jak velkou.