Report for Ashley

František Kalvas

2022-03-11

## Packages etc.

library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(ggplot2)  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

# My own functon for renaming in Tidyverse  
prejmenuj = function(data, positions, new.names) {  
 names(data)[positions] = new.names  
 data  
}

## Loading data

Data are at <http://github.com/frantisek901/Spirals/Experiment>. Experiment is still running and I, Francesco, from time to time actualize the \*.csv files at GitHub, then I run script experiment.R which loads the data. Later version probably finds better names for variables, but now, I use default names from NetLogo experiment.

Who is not interested in working with megabytes of \*.csv files, might use compiled \*.RData, there are two files: shortData.RData, which is main data file from experiments running only 365 steps, these data are extended by extra simulations with low size of small-world network neighborhood; and longData.RData, which is additional data file from experiments running 3650 steps – thanks to it we might test the effect of simulation length.

Now we load these data:

load("shortData.RData")  
load("longData.RData")

## Regressions

On the two following pages, there are 4 regressions in 2 tables (I’m starting with stargazer, later I will produce better output, but for now…). The first table uses ESBG polarization measure, after 365 and 3650 steps, the second uses my normalized polarization measure after same number of steps.

##   
## ===============================================================================  
## Dependent variable:   
## -----------------------------------------------------------  
## ESBG\_365 ESBG\_3650   
## (1) (2)   
## -------------------------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.648\*\*\*   
## (0.002) (0.005)   
##   
## `use\_identity?` 0.104\*\*\* 0.102\*\*\*   
## (0.0004) (0.001)   
##   
## boundary -0.142\*\*\* -0.106\*\*\*   
## (0.004) (0.006)   
##   
## modevaguely-speak -0.139\*\*\* -0.103\*\*\*   
## (0.0003) (0.001)   
##   
## `conformity-level` -0.054\*\*\* -0.023\*\*\*   
## (0.002) (0.004)   
##   
## `p-speaking-level` -0.013\*\*\* -0.001   
## (0.002) (0.003)   
##   
## `tolerance-level` -0.019\*\*\* -0.012\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.001   
## (0.005) (0.008)   
##   
## `n-neis` -0.0005\*\*\* -0.0001   
## (0.00001) (0.00005)   
##   
## Constant -0.069\*\*\* -0.190\*\*\*   
## (0.003) (0.005)   
##   
## -------------------------------------------------------------------------------  
## Observations 313,883 88,101   
## R2 0.481 0.491   
## Adjusted R2 0.481 0.491   
## Residual Std. Error 0.097 (df = 313873) 0.081 (df = 88091)   
## F Statistic 32,352.440\*\*\* (df = 9; 313873) 9,436.357\*\*\* (df = 9; 88091)  
## ===============================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ================================================================================  
## Dependent variable:   
## ------------------------------------------------------------  
## normalized\_365 normalized\_3650   
## (1) (2)   
## --------------------------------------------------------------------------------  
## id\_threshold 0.725\*\*\* 0.770\*\*\*   
## (0.003) (0.005)   
##   
## `use\_identity?` 0.097\*\*\* 0.094\*\*\*   
## (0.0004) (0.001)   
##   
## boundary -0.191\*\*\* -0.154\*\*\*   
## (0.004) (0.006)   
##   
## modevaguely-speak -0.170\*\*\* -0.150\*\*\*   
## (0.0004) (0.001)   
##   
## `conformity-level` -0.069\*\*\* -0.009\*\*   
## (0.002) (0.004)   
##   
## `p-speaking-level` -0.015\*\*\* -0.002   
## (0.002) (0.003)   
##   
## `tolerance-level` -0.026\*\*\* -0.022\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.004   
## (0.005) (0.008)   
##   
## `n-neis` -0.0005\*\*\* -0.00002   
## (0.00001) (0.00005)   
##   
## Constant -0.073\*\*\* -0.182\*\*\*   
## (0.003) (0.005)   
##   
## --------------------------------------------------------------------------------  
## Observations 313,883 88,101   
## R2 0.543 0.593   
## Adjusted R2 0.543 0.593   
## Residual Std. Error 0.100 (df = 313873) 0.081 (df = 88091)   
## F Statistic 41,374.930\*\*\* (df = 9; 313873) 14,288.110\*\*\* (df = 9; 88091)  
## ================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Note:

1. Variables mode:vaguely-speak and use\_identity? are binary, n-neis is measured on scale 1–64, and all other variables (id\_threshold, boundary etc.) are measured on scale 0–1.
2. I check the problem of use\_identity? – I estimated same regression model on sub-sample of simulation with use\_identity?==TRUE, naturally, effect of mere use\_identity? is not estimable, but good news is that effect of id\_threshold is completely same (OK, up to 5th decimal place).
3. Just for curiosity I estimated the model for subsample use\_identity?==FALSE, I was surprised that all effects were roughly by one order lower ().

## Graphs

### Sampling

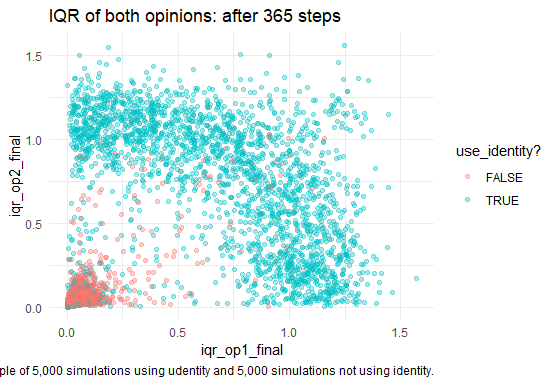
I produced graphs after some random sampling. Both files standard (365 steps) and long (3650 steps) are huge with many thousands of observations. So I created two samples, each of 10,000 observations – 5,000 simulations using identity, 5,000 not using identity.

res\_sample = sample\_n(res[res$`use\_identity?`,], 5000) %>%  
 add\_row(sample\_n(res[!res$`use\_identity?`,], 5000)) %>%  
 sample\_n(10000)  
  
long\_sample = sample\_n(long[long$`use\_identity?`,], 5000) %>%  
 add\_row(sample\_n(long[!long$`use\_identity?`,], 5000)) %>%  
 sample\_n(10000)

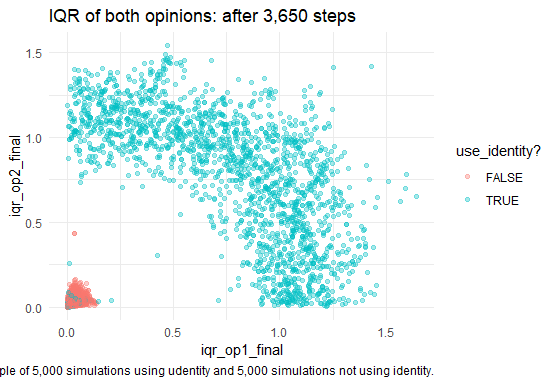
### Inter-quartile range

Here we look at depiction of distribution of interquartile range of both opinions. The first graph is made from standard (365 steps) data, the second from long (3,650 steps) data.

res\_sample %>%   
 ggplot(aes(x = iqr\_op1\_final, y = iqr\_op2\_final, col = `use\_identity?`)) +  
 geom\_point(alpha = 0.35) +  
 labs(title = "IQR of both opinions: after 365 steps",   
 caption = "Sample of 5,000 simulations using udentity and 5,000 simulations not using identity.") +  
 theme\_minimal()



long\_sample %>%   
 ggplot(aes(x = iqr\_op1\_final, y = iqr\_op2\_final, col = `use\_identity?`)) +  
 geom\_point(alpha = 0.35) +  
 labs(title = "IQR of both opinions: after 3,650 steps",   
 caption = "Sample of 5,000 simulations using udentity and 5,000 simulations not using identity.") +  
 theme\_minimal()



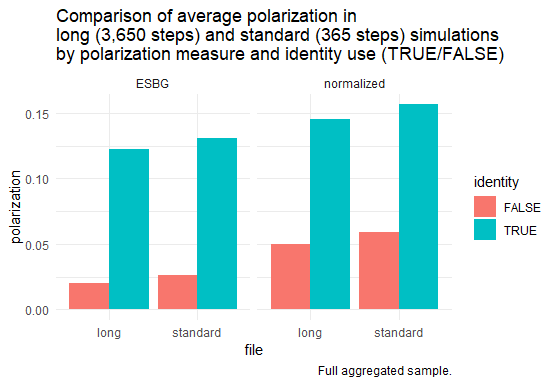
For me the basic logic is same in both graphs: some part of simulations ends up with consensus, mainly its simulations not using identity (red dots). Simulation using identity (turquoise dots) sometimes ends up with consensus as well, but also frequently ends up polarized, which is reflected by turquoise ‘perimeter’. It seems to me that this basic logic – identity use = perimeter of discord – is same regardless the length of simulation.

But different is cleanness of this pattern. In long data (3,650 steps) it is very clear and there are almost no observations between ‘red consensus dot’ in left down corner and ‘turquoise discord perimeter’. In standard data (365 steps) there are some observations and the perimeter seems fatter. The result is obvious: some standard simulations (365 steps) ended too early, because their ‘longer twins’ moved from ‘discord perimeter’ or space in between to ‘concensus dot’. So, let’s check the differences in polarization between standard and long data:

df = res %>% mutate(file = "standard") %>%   
 rename(ESBG = ESBG\_365, normalized = normalized\_365, identity = `use\_identity?`) %>%   
 add\_row(long %>% mutate(file = "long") %>%   
 rename(ESBG = ESBG\_3650, normalized = normalized\_3650, identity = `use\_identity?`)) %>%   
 group\_by(file, identity) %>%   
 summarise(ESBG = mean(ESBG), normalized = mean(normalized)) %>%   
 pivot\_longer(cols = c(ESBG, normalized), names\_to = "polarization\_measure", values\_to = "polarization")

## `summarise()` has grouped output by 'file'. You can override using the `.groups` argument.

ggplot(df, aes(x = file, y = polarization, fill = identity)) +  
 facet\_wrap(vars(polarization\_measure)) +  
 geom\_col(position = position\_dodge()) +  
 labs(title = "Comparison of average polarization in \nlong (3,650 steps) and standard (365 steps) simulations\nby polarization measure and identity use (TRUE/FALSE)", caption = "Full aggregated sample.") +  
 theme\_minimal()



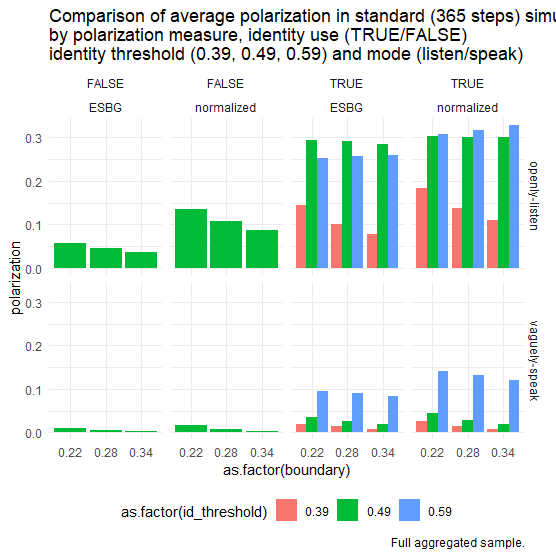
We see that long (3,650 steps) simulation are tiny slightly less polarized than short (365 steps) ones, i.e. on the average, the polarization in further more than 3,000 steps slightly decreases from initial value. We also see that normalized measure shows slightly higher polarization than ESBG. So, we might be quite confident that the length of simulation doesn’t spoil the results that much – since there is some tiny differences in aggregate results, it makes sense to do further analyses on individual level, i.e. level of individual simulation, and compute and plot how many times polarization increases from 365th to 3,650th step and how much, but for now we see that after 365 steps we received almost same picture as after 3,650 steps.

But the main difference is obviously whether we use identity process or not – regardless the level of identity threshold (but note that we simulate it only for values 0.39, 0.49, 0.59, since it is so important parameter, we now could look at it in more detail). So, let’s look now graphically in same way on data, as we did in regression tables:

df = res %>% mutate(file = "standard") %>%   
 rename(ESBG = ESBG\_365, normalized = normalized\_365, identity = `use\_identity?`) %>%   
 add\_row(long %>% mutate(file = "long") %>%   
 rename(ESBG = ESBG\_3650, normalized = normalized\_3650, identity = `use\_identity?`)) %>%   
 group\_by(file, id\_threshold, identity, boundary, mode) %>%   
 summarise(ESBG = mean(ESBG), normalized = mean(normalized)) %>%   
 pivot\_longer(cols = c(ESBG, normalized), names\_to = "polarization\_measure", values\_to = "polarization")

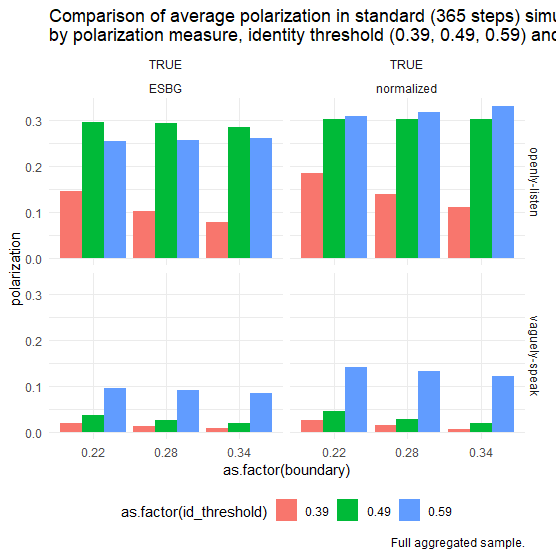
## `summarise()` has grouped output by 'file', 'id\_threshold', 'identity', 'boundary'. You can override using the `.groups` argument.

# `tolerance-level` -0.024\*\*\* -0.025\*\*\*   
# (0.001) (0.001)   
#   
# `p-speaking-level` -0.018\*\*\* -0.011\*\*\*   
# (0.002) (0.003)   
#   
# `conformity-level` -0.056\*\*\* 0.011\*\*\*   
 # (0.003) (0.003)   
  
df %>% filter(file == "standard") %>%   
 ggplot(aes(fill = as.factor(id\_threshold), y = polarization, x = as.factor(boundary))) +  
 facet\_grid(cols = vars(identity, polarization\_measure), rows = vars(mode)) +  
 geom\_col(position = position\_dodge()) +  
 labs(title = "Comparison of average polarization in standard (365 steps) simulations\nby polarization measure, identity use (TRUE/FALSE)\nidentity threshold (0.39, 0.49, 0.59) and mode (listen/speak)", caption = "Full aggregated sample.") +  
 theme\_minimal() +  
 theme(legend.position = "bottom")



Again same graph, just for better view only simulations using identity.

df %>% filter(file == "standard", identity) %>%   
 ggplot(aes(fill = as.factor(id\_threshold), y = polarization, x = as.factor(boundary))) +  
 facet\_grid(cols = vars(identity, polarization\_measure), rows = vars(mode)) +  
 geom\_col(position = position\_dodge()) +  
 labs(title = "Comparison of average polarization in standard (365 steps) simulations\nby polarization measure, identity threshold (0.39, 0.49, 0.59) and mode (listen/speak)", caption = "Full aggregated sample.") +  
 theme\_minimal() +  
 theme(legend.position = "bottom")



In previous graph we saw that while some polarisation might happen even without using identity (especially with narrower boundaries), more polarized simulations on average are that using identity. Effect of identity threshold is non-linear: in simulations with ‘openly listen’ mode the main polarization increase is between 0.39 and 0.49 values, in mode ‘vaguely speak’ between values 0.49 and 0.59 (but generally, the later mode is less polarized). It is also interesting, that in mode ‘vaguely speak’ with boundary widening the polarization always decreases, but in mode ‘openly listen’ this happens only for the lowest identity threshold value (0.39), for other threshold values (0.49, 0.59) the polarization stays same with widening of boundary or even very slightly increases!

The last result is very surprising – Hegselmann-Krause model usually finds overall concensus and avoids polarization with wider boundary, it’s one of basic results. But when we introduce identity, then this old true changes or is contingent on simulation mode (speaking/listening) and identity threshold. The classical HK findings still hold true, but only for ‘vaguely speak’ mode and low identity threshold values.

## More regressions

Here we compare full models le, ln, ne and nn with single varibles models, with full models omiting these single variables and with variables blocks. Let’s start with blocks!

### Identity block

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.725\*\*\*   
## (0.003) (0.003)   
##   
## `use\_identity?` 0.104\*\*\* 0.097\*\*\*   
## (0.0005) (0.001)   
##   
## Constant -0.249\*\*\* -0.296\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.203 0.201   
## Adjusted R2 0.203 0.201   
## Residual Std. Error (df = 313880) 0.120 0.133   
## F Statistic (df = 2; 313880) 40,034.190\*\*\* 39,543.350\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313880 4515.8 -7 -1575.6 24028 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313880 5516.4 -7 -2357.6 33466 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 4 220278 -7 134689 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 4 188867 -7 175001 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Hegselmann-Krause block

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## boundary -0.143\*\*\* -0.193\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.139\*\*\* -0.170\*\*\*   
## (0.0004) (0.0004)   
##   
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## Constant 0.236\*\*\* 0.300\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.270 0.332   
## Adjusted R2 0.270 0.332   
## Residual Std. Error (df = 313879) 0.115 0.121   
## F Statistic (df = 3; 313879) 38,633.840\*\*\* 52,001.240\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313879 4139.3 -6 -1199.1 21334 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313879 4613.4 -6 -1454.6 24089 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 5 233941 -6 107363 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 5 216922 -6 118891 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### Network block

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `tolerance-level` -0.020\*\*\* -0.027\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.004   
## (0.007) (0.008)   
##   
## `n-neis` -0.001\*\*\* -0.001\*\*\*   
## (0.00002) (0.00002)   
##   
## Constant 0.145\*\*\* 0.179\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.010 0.011   
## Adjusted R2 0.010 0.011   
## Residual Std. Error (df = 313879) 0.134 0.148   
## F Statistic (df = 3; 313879) 1,022.445\*\*\* 1,114.979\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313879 5612.9 -6 -2672.7 47553 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313879 6833.5 -6 -3674.7 60856 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 5 186145 -6 202953 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 5 155263 -6 242208 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

### SoS block

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `p-speaking-level` -0.013\*\*\* -0.015\*\*\*   
## (0.003) (0.003)   
##   
## Constant 0.110\*\*\* 0.139\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.0001 0.0001   
## Adjusted R2 0.0001 0.0001   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 22.057\*\*\* 26.784\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ `p-speaking-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313881 5667.3 -8 -2727.1 36391 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ `p-speaking-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313881 6905.7 -8 -3746.9 46539 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ `p-speaking-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 3 184630 -8 205984 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ `p-speaking-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 3 153613 -8 245508 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, we see, that we might omit SoS and Network blocks, we may focus only on Identity and HK blocks.

## Without identity

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## boundary -0.143\*\*\* -0.193\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.139\*\*\* -0.170\*\*\*   
## (0.0004) (0.0004)   
##   
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## `p-speaking-level` -0.013\*\*\* -0.015\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.020\*\*\* -0.027\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.004   
## (0.006) (0.006)   
##   
## `n-neis` -0.001\*\*\* -0.001\*\*\*   
## (0.00001) (0.00001)   
##   
## Constant 0.285\*\*\* 0.357\*\*\*   
## (0.003) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.279 0.343   
## Adjusted R2 0.279 0.343   
## Residual Std. Error (df = 313875) 0.114 0.120   
## F Statistic (df = 7; 313875) 17,390.130\*\*\* 23,374.960\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313875 4083.9 -2 -1143.7 61046 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313875 4539.7 -2 -1381 68609 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 9 236056 -2 103133 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 9 219447 -2 113839 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Without HK block

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.725\*\*\*   
## (0.003) (0.003)   
##   
## `use\_identity?` 0.104\*\*\* 0.097\*\*\*   
## (0.0005) (0.001)   
##   
## `p-speaking-level` -0.013\*\*\* -0.015\*\*\*   
## (0.002) (0.003)   
##   
## `tolerance-level` -0.019\*\*\* -0.026\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.003   
## (0.006) (0.007)   
##   
## `n-neis` -0.0005\*\*\* -0.0005\*\*\*   
## (0.00001) (0.00002)   
##   
## Constant -0.203\*\*\* -0.242\*\*\*   
## (0.003) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.212 0.211   
## Adjusted R2 0.212 0.211   
## Residual Std. Error (df = 313876) 0.119 0.132   
## F Statistic (df = 6; 313876) 14,042.820\*\*\* 13,962.360\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313876 4468.3 -3 -1528.1 54375 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313876 5451.3 -3 -2292.6 75934 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 8 221938 -3 131368 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + `p-speaking-level` +   
## `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 8 190728 -3 171278 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.725\*\*\*   
## (0.002) (0.003)   
##   
## `use\_identity?` 0.104\*\*\* 0.097\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.141\*\*\* -0.191\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.139\*\*\* -0.170\*\*\*   
## (0.0003) (0.0004)   
##   
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.002) (0.002)   
##   
## Constant -0.116\*\*\* -0.127\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.473 0.533   
## Adjusted R2 0.473 0.533   
## Residual Std. Error (df = 313877) 0.098 0.101   
## F Statistic (df = 5; 313877) 56,304.560\*\*\* 71,697.290\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 2987.9   
## 2 313879 4139.3 -2 -1151.4 60479 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 3224.1   
## 2 313879 4613.4 -2 -1389.3 67629 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ boundary + mode + `conformity-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 285098   
## 2 5 233941 -2 102315 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ boundary + mode + `conformity-level`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 273157   
## 2 5 216922 -2 112471 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 2987.8   
## 2 313873 2940.2 4 47.668 1272.2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 3224.1   
## 2 313873 3158.8 4 65.268 1621.3 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 285098   
## 2 11 287622 4 5048 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 273157   
## 2 11 276367 4 6419.5 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, then! Evidently the weaker block (Identity) improves model with stronger block (Hegselmann-Krause) a lot. Other blocks improve model also (statistically), but it is weak improvement. It is evident that present data these blocks (Identity, HK) explain satisfactorily. But just let’s look at Network block – would it improve the present model (Identity + HK)?

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.725\*\*\*   
## (0.002) (0.003)   
##   
## `use\_identity?` 0.104\*\*\* 0.097\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.142\*\*\* -0.191\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.139\*\*\* -0.170\*\*\*   
## (0.0003) (0.0004)   
##   
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.019\*\*\* -0.026\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.004   
## (0.005) (0.005)   
##   
## `n-neis` -0.0005\*\*\* -0.0005\*\*\*   
## (0.00001) (0.00001)   
##   
## Constant -0.076\*\*\* -0.081\*\*\*   
## (0.003) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.481 0.543   
## Adjusted R2 0.481 0.543   
## Residual Std. Error (df = 313874) 0.097 0.100   
## F Statistic (df = 8; 313874) 36,386.390\*\*\* 46,530.980\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 2987.8   
## 2 313874 2940.6 3 47.27 1681.9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313877 3224.1   
## 2 313874 3159.4 3 64.68 2141.9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 285098   
## 2 10 287601 3 5005.6 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 7 273157   
## 2 10 276338 3 6361 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 2940.6   
## 2 313873 2940.2 1 0.39739 42.422 7.366e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3159.4   
## 2 313873 3158.8 1 0.58796 58.423 2.12e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 287601   
## 2 11 287622 1 42.421 7.361e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 276338   
## 2 11 276367 1 58.419 2.118e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Network and SoS improve model, but it is very weak. Thanks to huge number of obsefvations we receive statistically significant results, even in case of SoS, but substantively it’s not big improvement (instead of thousands, only tens of point of $^2 $). Only just for sure, we continue with testing effect of individual variables.

## Individual variables

This testing is not very useful, I think, because we just tested effect of SoS – block of one variable – and even this weak variable is statistically significant. But, OK, for sure, we do it! :-) And we will start from the tail, from weakest variables. We will estimate model with just one variable, and also just without this one variable.

### Number of neighbors

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `n-neis` -0.001\*\*\* -0.001\*\*\*   
## (0.00001) (0.00002)   
##   
## Constant 0.137\*\*\* 0.169\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.008 0.008   
## Adjusted R2 0.008 0.008   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 2,533.751\*\*\* 2,552.522\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3176.9   
## 2 313873 3158.8 1 18.11 1799.5 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 2957.1   
## 2 313873 2940.2 1 16.874 1801.3 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 275470   
## 2 11 276367 1 1794.4 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 286724   
## 2 11 287622 1 1796.2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Size of neighborhood explains data better, improves model significantly, but still it is weak improvement.

### Probability of random links

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `p-random` -0.003 -0.004   
## (0.007) (0.008)   
##   
## Constant 0.104\*\*\* 0.132\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.00000 0.00000   
## Adjusted R2 -0.00000 -0.00000   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 0.251 0.263   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 313874 3158.8   
## 2 313873 3158.8 1 0.0054733 0.5439 0.4608

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 313874 2940.2   
## 2 313873 2940.2 1 0.0040896 0.4366 0.5088

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 10 276367   
## 2 11 276367 1 0.5439 0.4608

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 10 287622   
## 2 11 287622 1 0.4366 0.5088

Whoa! Probability of random links doesn’t improve model significantly! This probability is useless.

### Tolerance to opponents in neighborhood

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `tolerance-level` -0.034\*\*\* -0.042\*\*\*   
## (0.001) (0.001)   
##   
## Constant 0.128\*\*\* 0.161\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.006 0.008   
## Adjusted R2 0.006 0.008   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 1,968.405\*\*\* 2,388.906\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3174.2   
## 2 313873 3158.8 1 15.424 1532.6 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 2948.3   
## 2 313873 2940.2 1 8.1227 867.12 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 275603   
## 2 11 276367 1 1528.9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 287189   
## 2 11 287622 1 865.95 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Tolerance to opponents in neighborhood explains data better, improves model significantly, but still it is weak improvement.

### Probability of speaking

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `p-speaking-level` -0.013\*\*\* -0.015\*\*\*   
## (0.003) (0.003)   
##   
## Constant 0.110\*\*\* 0.139\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.0001 0.0001   
## Adjusted R2 0.0001 0.0001   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 22.057\*\*\* 26.784\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 2940.2   
## 2 313874 2940.6 -1 -0.39739 42.422 7.366e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313873 3158.8   
## 2 313874 3159.4 -1 -0.58796 58.423 2.12e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 287622   
## 2 10 287601 -1 42.421 7.361e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 276367   
## 2 10 276338 -1 58.419 2.118e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Probability of speaking explains data better, improves model significantly, but still it is very weak improvement.

### Conformity level

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.003) (0.004)   
##   
## Constant 0.127\*\*\* 0.162\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.001 0.001   
## Adjusted R2 0.001 0.001   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 275.707\*\*\* 368.772\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `tolerance-level` + `p-speaking-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3166.9   
## 2 313873 3158.8 1 8.1028 805.13 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 2945.2   
## 2 313873 2940.2 1 4.9727 530.85 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `tolerance-level` + `p-speaking-level` + `p-random` +   
## `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 275965   
## 2 11 276367 1 804.13 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 287357   
## 2 11 287622 1 530.41 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Conformity level in neighborhood explains data better, improves model significantly, but still it is weak improvement.

### Mode

##   
## ===============================================================  
## Dependent variable:   
## -----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## ---------------------------------------------------------------  
## modevaguely-speak -0.139\*\*\* -0.170\*\*\*   
## (0.0004) (0.0004)   
##   
## Constant 0.172\*\*\* 0.215\*\*\*   
## (0.0003) (0.0003)   
##   
## ---------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.266 0.327   
## Adjusted R2 0.266 0.327   
## Residual Std. Error (df = 313881) 0.115 0.122   
## F Statistic (df = 1; 313881) 113,798.000\*\*\* 152,363.200\*\*\*  
## ===============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## `conformity-level` + `tolerance-level` + `p-speaking-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 5415.7   
## 2 313873 3158.8 1 2256.9 224258 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 4448.2   
## 2 313873 2940.2 1 1508 160986 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## `conformity-level` + `tolerance-level` + `p-speaking-level` +   
## `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 191758   
## 2 11 276367 1 169219 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 222644   
## 2 11 287622 1 129957 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WHOOOOAAA! Mode is hugely significant, explains data so much better, improves model very significantly, it is strong improvement.

### Boundary

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## boundary -0.143\*\*\* -0.193\*\*\*   
## (0.005) (0.005)   
##   
## Constant 0.143\*\*\* 0.185\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.003 0.004   
## Adjusted R2 0.003 0.004   
## Residual Std. Error (df = 313881) 0.134 0.148   
## F Statistic (df = 1; 313881) 854.218\*\*\* 1,273.656\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3186.4   
## 2 313873 3158.8 1 27.565 2739 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 2955.3   
## 2 313873 2940.2 1 15.083 1610.2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 275003   
## 2 11 276367 1 2727.2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 286819   
## 2 11 287622 1 1606.1 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Boundary explains data better, improves model significantly, but still it is weak improvement.

### Identity use

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `use\_identity?` 0.104\*\*\* 0.097\*\*\*   
## (0.001) (0.001)   
##   
## Constant 0.027\*\*\* 0.059\*\*\*   
## (0.0004) (0.0005)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.118 0.085   
## Adjusted R2 0.118 0.085   
## Residual Std. Error (df = 313881) 0.126 0.142   
## F Statistic (df = 1; 313881) 41,996.720\*\*\* 29,038.510\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3734.6   
## 2 313873 3158.8 1 575.79 57213 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3600.7   
## 2 313873 2940.2 1 660.56 70517 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 250088   
## 2 11 276367 1 52558 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 255815   
## 2 11 287622 1 63614 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WHOOOOAAA! Identity use is very significant, explains data much better, improves model significantly, it is strong improvement.

### Identity level

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.725\*\*\*   
## (0.003) (0.004)   
##   
## Constant -0.172\*\*\* -0.225\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.085 0.117   
## Adjusted R2 0.085 0.117   
## Residual Std. Error (df = 313881) 0.129 0.139   
## F Statistic (df = 1; 313881) 29,243.480\*\*\* 41,415.220\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ `use\_identity?` + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3964.0   
## 2 313873 3158.8 1 805.23 80012 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ `use\_identity?` + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313874 3423.4   
## 2 313873 2940.2 1 483.18 51581 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ `use\_identity?` + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 240730   
## 2 11 276367 1 71273 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ `use\_identity?` + boundary + mode + `conformity-level` +   
## `tolerance-level` + `p-speaking-level` + `p-random` + `n-neis`  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 10 263743   
## 2 11 287622 1 47758 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

WHOOOOAAA! Identity level is very significant, explains data much better, improves model significantly, it is strong improvement.

OK, then! So, it seems that there are only 3 variables really significant, all explaining 10+% of variability, they are: Identity use, Identity level and Mode. So for the final regression test, let’s compare model with just these three vars and full model, let’s see how different these models are.

### Best model?

##   
## =================================================================================  
## Dependent variable:   
## -------------------------------------------------------------  
## ESBG\_365   
## (1) (2)   
## ---------------------------------------------------------------------------------  
## id\_threshold 0.562\*\*\* 0.562\*\*\*   
## (0.003) (0.002)   
##   
## `use\_identity?` 0.104\*\*\* 0.104\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.142\*\*\*   
## (0.004)   
##   
## modevaguely-speak -0.139\*\*\* -0.139\*\*\*   
## (0.0003) (0.0003)   
##   
## `conformity-level` -0.054\*\*\*   
## (0.002)   
##   
## `p-speaking-level` -0.013\*\*\*   
## (0.002)   
##   
## `tolerance-level` -0.019\*\*\*   
## (0.001)   
##   
## `p-random` -0.003   
## (0.005)   
##   
## `n-neis` -0.0005\*\*\*   
## (0.00001)   
##   
## Constant -0.179\*\*\* -0.069\*\*\*   
## (0.001) (0.003)   
##   
## ---------------------------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.469 0.481   
## Adjusted R2 0.469 0.481   
## Residual Std. Error 0.098 (df = 313879) 0.097 (df = 313873)   
## F Statistic 92,524.260\*\*\* (df = 3; 313879) 32,352.440\*\*\* (df = 9; 313873)  
## =================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ==================================================================================  
## Dependent variable:   
## --------------------------------------------------------------  
## normalized\_365   
## (1) (2)   
## ----------------------------------------------------------------------------------  
## id\_threshold 0.725\*\*\* 0.725\*\*\*   
## (0.003) (0.003)   
##   
## `use\_identity?` 0.097\*\*\* 0.097\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.191\*\*\*   
## (0.004)   
##   
## modevaguely-speak -0.170\*\*\* -0.170\*\*\*   
## (0.0004) (0.0004)   
##   
## `conformity-level` -0.069\*\*\*   
## (0.002)   
##   
## `p-speaking-level` -0.015\*\*\*   
## (0.002)   
##   
## `tolerance-level` -0.026\*\*\*   
## (0.001)   
##   
## `p-random` -0.004   
## (0.005)   
##   
## `n-neis` -0.0005\*\*\*   
## (0.00001)   
##   
## Constant -0.211\*\*\* -0.073\*\*\*   
## (0.001) (0.003)   
##   
## ----------------------------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.528 0.543   
## Adjusted R2 0.528 0.543   
## Residual Std. Error 0.102 (df = 313879) 0.100 (df = 313873)   
## F Statistic 117,053.800\*\*\* (df = 3; 313879) 41,374.930\*\*\* (df = 9; 313873)  
## ==================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## boundary -0.143\*\*\* -0.193\*\*\*   
## (0.005) (0.005)   
##   
## `conformity-level` -0.054\*\*\* -0.069\*\*\*   
## (0.003) (0.004)   
##   
## `p-speaking-level` -0.013\*\*\* -0.015\*\*\*   
## (0.003) (0.003)   
##   
## `tolerance-level` -0.020\*\*\* -0.027\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.004   
## (0.007) (0.008)   
##   
## `n-neis` -0.001\*\*\* -0.001\*\*\*   
## (0.00002) (0.00002)   
##   
## Constant 0.216\*\*\* 0.272\*\*\*   
## (0.003) (0.004)   
##   
## --------------------------------------------------------------  
## Observations 313,883 313,883   
## R2 0.013 0.016   
## Adjusted R2 0.013 0.016   
## Residual Std. Error (df = 313876) 0.133 0.147   
## F Statistic (df = 6; 313876) 708.204\*\*\* 843.406\*\*\*   
## ==============================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Analysis of Variance Table  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + mode  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313879 3007.8   
## 2 313873 2940.2 6 67.631 1203.3 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + mode  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 313879 3259.6   
## 2 313873 3158.8 6 100.79 1669.2 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: ESBG\_365 ~ id\_threshold + `use\_identity?` + mode  
## Model 2: ESBG\_365 ~ id\_threshold + `use\_identity?` + boundary + mode +   
## `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 5 284053   
## 2 11 287622 6 7138.3 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Likelihood ratio test  
##   
## Model 1: normalized\_365 ~ id\_threshold + `use\_identity?` + mode  
## Model 2: normalized\_365 ~ id\_threshold + `use\_identity?` + boundary +   
## mode + `conformity-level` + `p-speaking-level` + `tolerance-level` +   
## `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 5 271438   
## 2 11 276367 6 9858.9 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, resting 6 another variables improves model slightly, explain 1.5 percent points of more. So, yes, the triumvirate variables, 3 the most important, explain almost all variability, the resp explains something, but it’s so weak. ANOVA and Likelihood ratio tests tell us that these 6 variables improve, but 67 or 100 points of RSS only.

So, we might neglect the 6 resting variables in further experiments and play more smoothly with identity threshold (since the Mode is true/false, as well as Identity use). We also should inspect extreme value of Probability of speaking – if the probability will be 100%, would the results be the same?