Report for Ashley

František Kalvas

2022-03-14

## 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

library(forcats)  
library(sjmisc)

##   
## Attaching package: 'sjmisc'

## The following object is masked from 'package:tidyr':  
##   
## replace\_na

# 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")  
load("shortData1D.RData")  
load("longData1D.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 0.552\*\*\* 0.653\*\*\*   
## (0.002) (0.005)   
##   
## `use\_identity?` 0.105\*\*\* 0.102\*\*\*   
## (0.0004) (0.001)   
##   
## boundary -0.166\*\*\* -0.102\*\*\*   
## (0.003) (0.006)   
##   
## modevaguely-speak -0.142\*\*\* -0.105\*\*\*   
## (0.0003) (0.001)   
##   
## `conformity-level` -0.053\*\*\* -0.022\*\*\*   
## (0.002) (0.004)   
##   
## `p-speaking-level` -0.015\*\*\* -0.001   
## (0.001) (0.003)   
##   
## `tolerance-level` -0.018\*\*\* -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 -0.056\*\*\* -0.193\*\*\*   
## (0.002) (0.005)   
##   
## -------------------------------------------------------------------------------  
## Observations 359,907 89,035   
## R2 0.477 0.492   
## Adjusted R2 0.477 0.492   
## Residual Std. Error 0.098 (df = 359897) 0.082 (df = 89025)   
## F Statistic 36,433.090\*\*\* (df = 9; 359897) 9,580.896\*\*\* (df = 9; 89025)  
## ===============================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ================================================================================  
## Dependent variable:   
## ------------------------------------------------------------  
## normalized\_365 normalized\_3650   
## (1) (2)   
## --------------------------------------------------------------------------------  
## id\_threshold 0.711\*\*\* 0.771\*\*\*   
## (0.002) (0.005)   
##   
## `use\_identity?` 0.101\*\*\* 0.095\*\*\*   
## (0.0004) (0.001)   
##   
## boundary -0.377\*\*\* -0.150\*\*\*   
## (0.003) (0.006)   
##   
## modevaguely-speak -0.171\*\*\* -0.152\*\*\*   
## (0.0003) (0.001)   
##   
## `conformity-level` -0.069\*\*\* -0.009\*\*   
## (0.002) (0.004)   
##   
## `p-speaking-level` -0.022\*\*\* -0.002   
## (0.001) (0.003)   
##   
## `tolerance-level` -0.023\*\*\* -0.022\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.004   
## (0.005) (0.008)   
##   
## `n-neis` -0.0004\*\*\* -0.00002   
## (0.00001) (0.00005)   
##   
## Constant -0.014\*\*\* -0.182\*\*\*   
## (0.003) (0.005)   
##   
## --------------------------------------------------------------------------------  
## Observations 359,907 89,035   
## R2 0.532 0.594   
## Adjusted R2 0.532 0.594   
## Residual Std. Error 0.104 (df = 359897) 0.081 (df = 89025)   
## F Statistic 45,514.780\*\*\* (df = 9; 359897) 14,457.230\*\*\* (df = 9; 89025)  
## ================================================================================  
## 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 ().

### Note on polarization measures:

Please notice, that both polarization measures differ in effect of variable boundary on them. It happened when I introduced just for sure very low level of boundary variable (0.1 besides standard margin 0.22–0.34). From previous analyses I know that low value of this variable has polarization effect, so I was wandering, why Genetic algorithms (GAs) find best fitness (i.e., highest values of polarization for sum of both measures) for boundary == 0.28 and I was a bit unsure whether this relatively wider value is not mistake. So I introduced also value 0.1 for variable boundary and I have been running simulations since Friday.

Narrow boundary (0.1) very probably produces not two, but several groups, probably all relatively dense. Normalized polarization is sensitive to density (i.e. tight aggregation around group’s mean) and distance of groups, so if the groups are evenly distributed in opinion space, it produces relatively high normalized polarization. On the other hand, ESBG is very sensitive to distributions in two groups of equal sizes, their density (i.e. tight aggregation around goup’s mean) and distance. When we receive 4 or 5 dense groups, evenly distributed over opinion space, ESBG assesses it as low polarization.

That is why both measures differ regarding estimation of effect of narrow boundaries and why GAs find wider boundaries as part of most polarizing set of parameters. During GAs I used sum of both polarization measures as fitness parameter. With value of boundary 0.28 and right combination of other parameters we receive high polarization in both measures, with narrow boundary we receive higher polarization through normalized measure, but much weaker through ESBG measure, so when we sum both measures, we receive higher combined polarization with wider boundaries than with narrow ones.

I also introduced high value of probability of speaking, 0.95, just to be sure this variable makes really no difference – but here the previous results are confirmed, now even with such a high values, probability of speaking still makes no difference, so if this assurance will be confirmed with the rest of extra simulations (now are done 40 sets out of 150, i.e. now we are on 25%, but we might be quite sure…), then in further explorations we might set probability of speaking to some constant value – may be to 0.49 which was found by GA, or to 1, but the value 0.49 seems more realistic to me, since not everyone talks always.

## Graphs

### Sampling

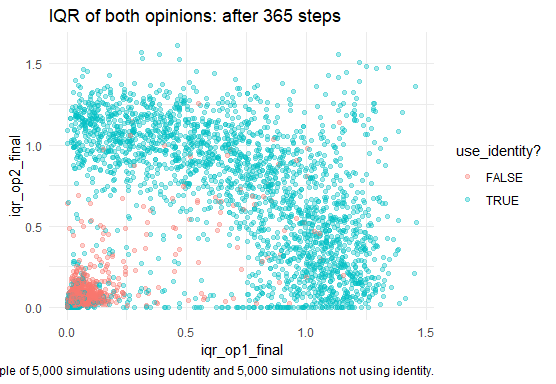
I produced graphs after some random sampling. Both files standard (365 steps) and long (3650 steps) are huge with many thousands of observations. So I created two samples, each of 10,000 observations – 5,000 simulations using identity, 5,000 not using identity.

res\_sample = sample\_n(res[res$`use\_identity?` & res$boundary > 0.1,], 5000) %>%  
 add\_row(sample\_n(res[!res$`use\_identity?` & res$boundary > 0.1,], 5000)) %>%  
 sample\_n(10000)  
  
long\_sample = sample\_n(long[long$`use\_identity?` & long$boundary > 0.1,], 5000) %>%  
 add\_row(sample\_n(long[!long$`use\_identity?` & long$boundary > 0.1,], 5000)) %>%  
 sample\_n(10000)

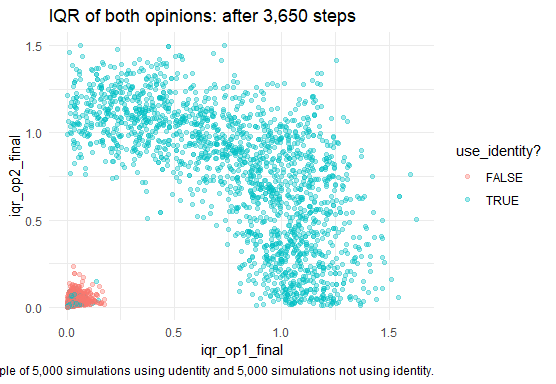
### Inter-quartile range

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()



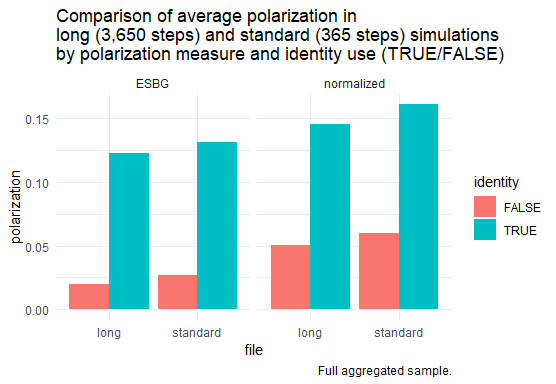
For me the basic logic is same in both graphs: some part of simulations ends up with consensus, mainly its simulations not using identity (red dots). Simulation using identity (turquoise dots) sometimes ends up with consensus as well, but also frequently ends up polarized, which is reflected by turquoise ‘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()



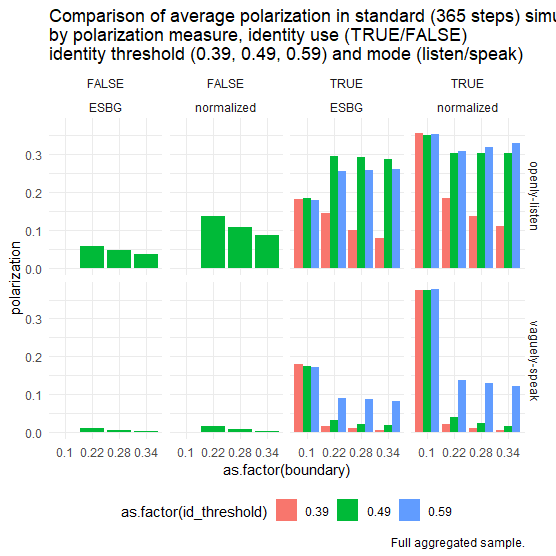
We see that long (3,650 steps) simulation are tiny slightly less polarized than short (365 steps) ones, i.e. on the average, the polarization in further more than 3,000 steps slightly decreases from initial value. We also see that normalized measure shows slightly higher polarization than ESBG. So, we 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")

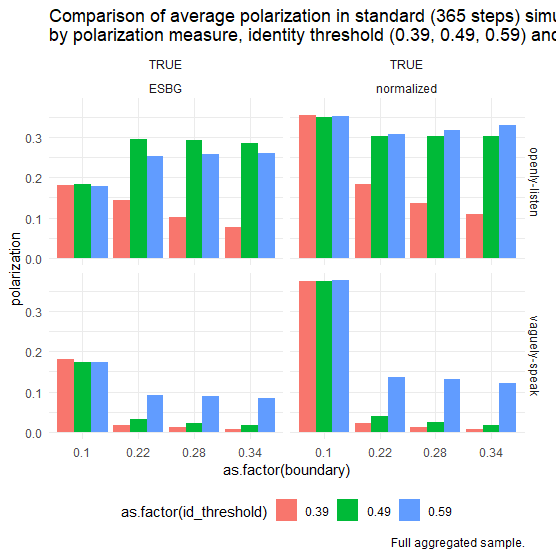
## `summarise()` has grouped output by 'file', 'id\_threshold', 'identity', 'boundary'. You can override using the `.groups` argument.

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 0.552\*\*\* 0.711\*\*\*   
## (0.003) (0.003)   
##   
## `use\_identity?` 0.104\*\*\* 0.101\*\*\*   
## (0.0005) (0.001)   
##   
## Constant -0.244\*\*\* -0.289\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.191 0.191   
## Adjusted R2 0.191 0.191   
## Residual Std. Error (df = 359904) 0.122 0.137   
## F Statistic (df = 2; 359904) 42,486.470\*\*\* 42,572.100\*\*\*   
## ==============================================================  
## 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 359897 3460.0   
## 2 359904 5349.4 -7 -1889.4 28075 < 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 359897 3886.3   
## 2 359904 6719.9 -7 -2833.6 37487 < 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 325120   
## 2 4 246712 -7 156816 < 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 304214   
## 2 4 205668 -7 197091 < 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 -0.190\*\*\* -0.399\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.142\*\*\* -0.171\*\*\*   
## (0.0004) (0.0004)   
##   
## `conformity-level` -0.053\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## Constant 0.254\*\*\* 0.364\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.280 0.337   
## Adjusted R2 0.280 0.337   
## Residual Std. Error (df = 359903) 0.115 0.124   
## F Statistic (df = 3; 359903) 46,691.780\*\*\* 60,941.390\*\*\*   
## ==============================================================  
## 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 359897 3460.0   
## 2 359903 4759.9 -6 -1299.8 22534 < 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 359897 3886.3   
## 2 359903 5510.4 -6 -1624.2 25068 < 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 325120   
## 2 5 267725 -6 114790 < 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 304214   
## 2 5 241376 -6 125676 < 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` -0.017\*\*\* -0.021\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.004 -0.006   
## (0.006) (0.007)   
##   
## `n-neis` -0.0004\*\*\* -0.0004\*\*\*   
## (0.00002) (0.00002)   
##   
## Constant 0.142\*\*\* 0.175\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.006 0.005   
## Adjusted R2 0.006 0.005   
## Residual Std. Error (df = 359903) 0.135 0.152   
## F Statistic (df = 3; 359903) 706.839\*\*\* 637.661\*\*\*   
## ==============================================================  
## 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 359897 3460.0   
## 2 359903 6573.7 -6 -3113.7 53978 < 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 359897 3886.3   
## 2 359903 8265.7 -6 -4379.4 67594 < 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 325120   
## 2 5 209626 -6 230988 < 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 304214   
## 2 5 168410 -6 271608 < 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` 0.008\*\*\* -0.001   
## (0.002) (0.002)   
##   
## Constant 0.102\*\*\* 0.137\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.0001 0.00000   
## Adjusted R2 0.0001 -0.00000   
## Residual Std. Error (df = 359905) 0.136 0.152   
## F Statistic (df = 1; 359905) 20.945\*\*\* 0.152   
## ==============================================================  
## 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 359897 3460   
## 2 359905 6612 -8 -3152 40982 < 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 359897 3886.3   
## 2 359905 8309.6 -8 -4423.4 51204 < 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 325120   
## 2 3 208580 -8 233081 < 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 304214   
## 2 3 167456 -8 273516 < 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 -0.196\*\*\* -0.405\*\*\*   
## (0.004) (0.004)   
##   
## modevaguely-speak -0.142\*\*\* -0.171\*\*\*   
## (0.0004) (0.0004)   
##   
## `conformity-level` -0.053\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## `p-speaking-level` 0.016\*\*\* 0.008\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.017\*\*\* -0.022\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.005   
## (0.005) (0.006)   
##   
## `n-neis` -0.0004\*\*\* -0.0004\*\*\*   
## (0.00001) (0.00001)   
##   
## Constant 0.283\*\*\* 0.401\*\*\*   
## (0.002) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.287 0.343   
## Adjusted R2 0.287 0.343   
## Residual Std. Error (df = 359899) 0.114 0.123   
## F Statistic (df = 7; 359899) 20,654.060\*\*\* 26,822.650\*\*\*   
## ==============================================================  
## 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 359897 3460.0   
## 2 359899 4717.4 -2 -1257.3 65391 < 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 359897 3886.3   
## 2 359899 5460.8 -2 -1574.5 72904 < 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 325120   
## 2 9 269339 -2 111561 < 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 304214   
## 2 9 243005 -2 122417 < 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 0.552\*\*\* 0.711\*\*\*   
## (0.003) (0.003)   
##   
## `use\_identity?` 0.105\*\*\* 0.102\*\*\*   
## (0.0005) (0.001)   
##   
## `p-speaking-level` -0.016\*\*\* -0.024\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.018\*\*\* -0.022\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.006   
## (0.006) (0.006)   
##   
## `n-neis` -0.0004\*\*\* -0.0004\*\*\*   
## (0.00001) (0.00002)   
##   
## Constant -0.198\*\*\* -0.237\*\*\*   
## (0.003) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.198 0.198   
## Adjusted R2 0.198 0.198   
## Residual Std. Error (df = 359900) 0.121 0.136   
## F Statistic (df = 6; 359900) 14,806.760\*\*\* 14,782.620\*\*\*   
## ==============================================================  
## 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 359897 3460.0   
## 2 359900 5303.3 -3 -1843.3 63910 < 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 359897 3886.3   
## 2 359900 6666.7 -3 -2780.4 85828 < 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 325120   
## 2 8 248270 -3 153700 < 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 304214   
## 2 8 207099 -3 194230 < 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 0.552\*\*\* 0.711\*\*\*   
## (0.002) (0.002)   
##   
## `use\_identity?` 0.104\*\*\* 0.100\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.161\*\*\* -0.371\*\*\*   
## (0.003) (0.003)   
##   
## modevaguely-speak -0.142\*\*\* -0.171\*\*\*   
## (0.0003) (0.0003)   
##   
## `conformity-level` -0.053\*\*\* -0.069\*\*\*   
## (0.002) (0.002)   
##   
## Constant -0.104\*\*\* -0.069\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.469 0.525   
## Adjusted R2 0.469 0.525   
## Residual Std. Error (df = 359901) 0.099 0.105   
## F Statistic (df = 5; 359901) 63,692.420\*\*\* 79,669.690\*\*\*   
## ==============================================================  
## 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 359901 3508.2   
## 2 359903 4759.9 -2 -1251.7 64205 < 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 359901 3944.1   
## 2 359903 5510.4 -2 -1566.3 71462 < 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 322633   
## 2 5 267725 -2 109816 < 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 301554   
## 2 5 241376 -2 120357 < 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 359901 3508.2   
## 2 359897 3460.0 4 48.146 1252 < 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 359901 3944.1   
## 2 359897 3886.3 4 57.861 1339.6 < 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 322633   
## 2 11 325120 4 4973.5 < 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 301554   
## 2 11 304214 4 5318.9 < 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 0.552\*\*\* 0.711\*\*\*   
## (0.002) (0.002)   
##   
## `use\_identity?` 0.104\*\*\* 0.100\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.167\*\*\* -0.378\*\*\*   
## (0.003) (0.003)   
##   
## modevaguely-speak -0.142\*\*\* -0.171\*\*\*   
## (0.0003) (0.0003)   
##   
## `conformity-level` -0.053\*\*\* -0.069\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.018\*\*\* -0.023\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.004   
## (0.005) (0.005)   
##   
## `n-neis` -0.0005\*\*\* -0.0004\*\*\*   
## (0.00001) (0.00001)   
##   
## Constant -0.064\*\*\* -0.025\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.477 0.532   
## Adjusted R2 0.477 0.532   
## Residual Std. Error (df = 359898) 0.098 0.104   
## F Statistic (df = 8; 359898) 40,955.900\*\*\* 51,139.860\*\*\*   
## ==============================================================  
## 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 359901 3508.2   
## 2 359898 3461.3 3 46.88 1624.8 < 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 359901 3944.1   
## 2 359898 3888.9 3 55.257 1704.6 < 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 322633   
## 2 10 325054 3 4841.9 < 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 301554   
## 2 10 304093 3 5077.9 < 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 359898 3461.3   
## 2 359897 3460.0 1 1.2656 131.64 < 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` + `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 359898 3888.9   
## 2 359897 3886.3 1 2.6039 241.14 < 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` + `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 325054   
## 2 11 325120 1 131.62 < 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` + `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 304093   
## 2 11 304214 1 241.06 < 2.2e-16 \*\*\*  
## ---  
## 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 observations 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 0.135\*\*\* 0.166\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.005 0.004   
## Adjusted R2 0.005 0.004   
## Residual Std. Error (df = 359905) 0.135 0.152   
## F Statistic (df = 1; 359905) 1,728.564\*\*\* 1,413.788\*\*\*   
## ==============================================================  
## 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 359898 3901.0   
## 2 359897 3886.3 1 14.679 1359.4 < 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 359898 3476.1   
## 2 359897 3460.0 1 16.059 1670.4 < 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 303535   
## 2 11 304214 1 1356.9 < 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` + `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 324287   
## 2 11 325120 1 1666.6 < 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` -0.004 -0.006   
## (0.006) (0.007)   
##   
## Constant 0.108\*\*\* 0.139\*\*\*   
## (0.002) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.00000 0.00000   
## Adjusted R2 -0.00000 -0.00000   
## Residual Std. Error (df = 359905) 0.136 0.152   
## F Statistic (df = 1; 359905) 0.315 0.669   
## ==============================================================  
## 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 359898 3886.3   
## 2 359897 3886.3 1 0.0086149 0.7978 0.3718

## 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 359898 3460   
## 2 359897 3460 1 0.0029234 0.3041 0.5813

## 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 304213   
## 2 11 304214 1 0.7978 0.3717

## 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 325120   
## 2 11 325120 1 0.3041 0.5813

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` -0.028\*\*\* -0.031\*\*\*   
## (0.001) (0.001)   
##   
## Constant 0.127\*\*\* 0.161\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.004 0.004   
## Adjusted R2 0.004 0.004   
## Residual Std. Error (df = 359905) 0.135 0.152   
## F Statistic (df = 1; 359905) 1,373.421\*\*\* 1,400.759\*\*\*   
## ==============================================================  
## 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 359898 3899.8   
## 2 359897 3886.3 1 13.493 1249.6 < 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 359898 3468.4   
## 2 359897 3460.0 1 8.4093 874.7 < 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 303590   
## 2 11 304214 1 1247.4 < 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` + `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 324683   
## 2 11 325120 1 873.66 < 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` 0.008\*\*\* -0.001   
## (0.002) (0.002)   
##   
## Constant 0.102\*\*\* 0.137\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.0001 0.00000   
## Adjusted R2 0.0001 -0.00000   
## Residual Std. Error (df = 359905) 0.136 0.152   
## F Statistic (df = 1; 359905) 20.945\*\*\* 0.152   
## ==============================================================  
## 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 359897 3460.0   
## 2 359898 3461.3 -1 -1.2656 131.64 < 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?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 359897 3886.3   
## 2 359898 3888.9 -1 -2.6039 241.14 < 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?` + boundary + mode +   
## `conformity-level` + `tolerance-level` + `p-random` + `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 325120   
## 2 10 325054 -1 131.62 < 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?` + boundary +   
## mode + `conformity-level` + `tolerance-level` + `p-random` +   
## `n-neis`  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 11 304214   
## 2 10 304093 -1 241.06 < 2.2e-16 \*\*\*  
## ---  
## 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` -0.053\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## Constant 0.131\*\*\* 0.168\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.001 0.001   
## Adjusted R2 0.001 0.001   
## Residual Std. Error (df = 359905) 0.135 0.152   
## F Statistic (df = 1; 359905) 302.681\*\*\* 400.207\*\*\*   
## ==============================================================  
## 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 359898 3895.5   
## 2 359897 3886.3 1 9.2267 854.46 < 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 359898 3465.6   
## 2 359897 3460.0 1 5.5542 577.72 < 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 303787   
## 2 11 304214 1 853.47 < 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 324831   
## 2 11 325120 1 577.27 < 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 -0.142\*\*\* -0.171\*\*\*   
## (0.0004) (0.0004)   
##   
## Constant 0.177\*\*\* 0.222\*\*\*   
## (0.0003) (0.0003)   
##   
## ---------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.274 0.315   
## Adjusted R2 0.274 0.315   
## Residual Std. Error (df = 359905) 0.116 0.126   
## F Statistic (df = 1; 359905) 135,533.400\*\*\* 165,868.900\*\*\*  
## ===============================================================  
## 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 359898 6507.2   
## 2 359897 3886.3 1 2620.9 242712 < 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 359898 5268.5   
## 2 359897 3460.0 1 1808.5 188106 < 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 211457   
## 2 11 304214 1 185514 < 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 249456   
## 2 11 325120 1 151328 < 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 -0.190\*\*\* -0.399\*\*\*   
## (0.004) (0.005)   
##   
## Constant 0.159\*\*\* 0.247\*\*\*   
## (0.001) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.006 0.020   
## Adjusted R2 0.006 0.020   
## Residual Std. Error (df = 359905) 0.135 0.150   
## F Statistic (df = 1; 359905) 2,085.409\*\*\* 7,449.048\*\*\*   
## ==============================================================  
## 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 359898 4036.6   
## 2 359897 3886.3 1 150.27 13916 < 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 359898 3489.3   
## 2 359897 3460.0 1 29.285 3046.1 < 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 297387   
## 2 11 304214 1 13654 < 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 323603   
## 2 11 325120 1 3033.4 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, Boundary explains data better, improves model significantly, but has different magnitude of effect according the polarization measure in use.

NOTE: Following text is repeated (it was presented firstly above), just for sure we don’t miss it in such a long report.

Narrow boundary very probably produces not two, but several groups, probably all relatively dense. Normalized polarization is sensitive to density (i.e. tight aggregation around group’s mean) and distance of groups, so if the groups are evenly distributed in opinion space, it produces relatively high normalized polarization. On the other hand, ESBG is very sensitive to distributions in two groups of equal sizes, their density (i.e. tight aggregation around goup’s mean) and distance. When we receive 4 or 5 dense groups, evenly distributed over opinion space, ESBG assesses it as low polarization.

That is why both measures differ regarding estimation of effect of narrow boundaries and why genetic algorithms (GA) find wider boundaries as part of most polarizing set of parameters. During GA I used sum of both polarization measures as fitness parameter. With value of boundary 0.28 and right combination of other parameters we receive high polarization in both measures, with narrow boundary we receive higher polarization through normalized measure, but much weaker through ESBG measure, so when we sum both measures, we receive higher combined polarization with wider boundaries than with narrow.

### Identity use

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## `use\_identity?` 0.104\*\*\* 0.101\*\*\*   
## (0.001) (0.001)   
##   
## Constant 0.027\*\*\* 0.059\*\*\*   
## (0.0004) (0.001)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.106 0.080   
## Adjusted R2 0.106 0.080   
## Residual Std. Error (df = 359905) 0.128 0.146   
## F Statistic (df = 1; 359905) 42,863.580\*\*\* 31,135.460\*\*\*   
## ==============================================================  
## 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 359898 4533.2   
## 2 359897 3886.3 1 646.89 59907 < 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 359898 4158.5   
## 2 359897 3460.0 1 698.43 72647 < 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 276507   
## 2 11 304214 1 55415 < 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 292033   
## 2 11 325120 1 66175 < 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 0.552\*\*\* 0.711\*\*\*   
## (0.003) (0.003)   
##   
## Constant -0.164\*\*\* -0.211\*\*\*   
## (0.001) (0.002)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.085 0.112   
## Adjusted R2 0.085 0.112   
## Residual Std. Error (df = 359905) 0.130 0.143   
## F Statistic (df = 1; 359905) 33,244.260\*\*\* 45,243.130\*\*\*   
## ==============================================================  
## 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 359898 4814.1   
## 2 359897 3886.3 1 927.77 85918 < 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 359898 4019.1   
## 2 359897 3460.0 1 559.05 58150 < 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 265688   
## 2 11 304214 1 77051 < 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 298167   
## 2 11 325120 1 53906 < 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 0.552\*\*\* 0.552\*\*\*   
## (0.002) (0.002)   
##   
## `use\_identity?` 0.104\*\*\* 0.105\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.166\*\*\*   
## (0.003)   
##   
## modevaguely-speak -0.142\*\*\* -0.142\*\*\*   
## (0.0003) (0.0003)   
##   
## `conformity-level` -0.053\*\*\*   
## (0.002)   
##   
## `p-speaking-level` -0.015\*\*\*   
## (0.001)   
##   
## `tolerance-level` -0.018\*\*\*   
## (0.001)   
##   
## `p-random` -0.003   
## (0.005)   
##   
## `n-neis` -0.0005\*\*\*   
## (0.00001)   
##   
## Constant -0.173\*\*\* -0.056\*\*\*   
## (0.001) (0.002)   
##   
## ----------------------------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.464 0.477   
## Adjusted R2 0.464 0.477   
## Residual Std. Error 0.099 (df = 359903) 0.098 (df = 359897)   
## F Statistic 104,051.700\*\*\* (df = 3; 359903) 36,433.090\*\*\* (df = 9; 359897)  
## ==================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ==================================================================================  
## Dependent variable:   
## --------------------------------------------------------------  
## normalized\_365   
## (1) (2)   
## ----------------------------------------------------------------------------------  
## id\_threshold 0.711\*\*\* 0.711\*\*\*   
## (0.002) (0.002)   
##   
## `use\_identity?` 0.101\*\*\* 0.101\*\*\*   
## (0.0004) (0.0004)   
##   
## boundary -0.377\*\*\*   
## (0.003)   
##   
## modevaguely-speak -0.171\*\*\* -0.171\*\*\*   
## (0.0004) (0.0003)   
##   
## `conformity-level` -0.069\*\*\*   
## (0.002)   
##   
## `p-speaking-level` -0.022\*\*\*   
## (0.001)   
##   
## `tolerance-level` -0.023\*\*\*   
## (0.001)   
##   
## `p-random` -0.004   
## (0.005)   
##   
## `n-neis` -0.0004\*\*\*   
## (0.00001)   
##   
## Constant -0.203\*\*\* -0.014\*\*\*   
## (0.001) (0.003)   
##   
## ----------------------------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.507 0.532   
## Adjusted R2 0.507 0.532   
## Residual Std. Error 0.107 (df = 359903) 0.104 (df = 359897)   
## F Statistic 123,226.000\*\*\* (df = 3; 359903) 45,514.780\*\*\* (df = 9; 359897)  
## ==================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

##   
## ==============================================================  
## Dependent variable:   
## ----------------------------  
## ESBG\_365 normalized\_365  
## (1) (2)   
## --------------------------------------------------------------  
## boundary -0.196\*\*\* -0.405\*\*\*   
## (0.004) (0.005)   
##   
## `conformity-level` -0.053\*\*\* -0.069\*\*\*   
## (0.003) (0.003)   
##   
## `p-speaking-level` 0.015\*\*\* 0.008\*\*\*   
## (0.002) (0.002)   
##   
## `tolerance-level` -0.017\*\*\* -0.022\*\*\*   
## (0.001) (0.001)   
##   
## `p-random` -0.003 -0.005   
## (0.006) (0.007)   
##   
## `n-neis` -0.0004\*\*\* -0.0004\*\*\*   
## (0.00002) (0.00002)   
##   
## Constant 0.213\*\*\* 0.316\*\*\*   
## (0.003) (0.003)   
##   
## --------------------------------------------------------------  
## Observations 359,907 359,907   
## R2 0.013 0.027   
## Adjusted R2 0.013 0.027   
## Residual Std. Error (df = 359900) 0.135 0.150   
## F Statistic (df = 6; 359900) 790.661\*\*\* 1,686.416\*\*\*   
## ==============================================================  
## 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 359903 3541.1   
## 2 359897 3460.0 6 81.079 1405.6 < 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 359903 4099.2   
## 2 359897 3886.3 6 212.87 3285.5 < 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 320952   
## 2 11 325120 6 8336.4 < 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 294618   
## 2 11 304214 6 19192 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

OK, resting 6 another variables improves model slightly, explain 1.2 or 2.5 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 78 or 192 points of RSS only.

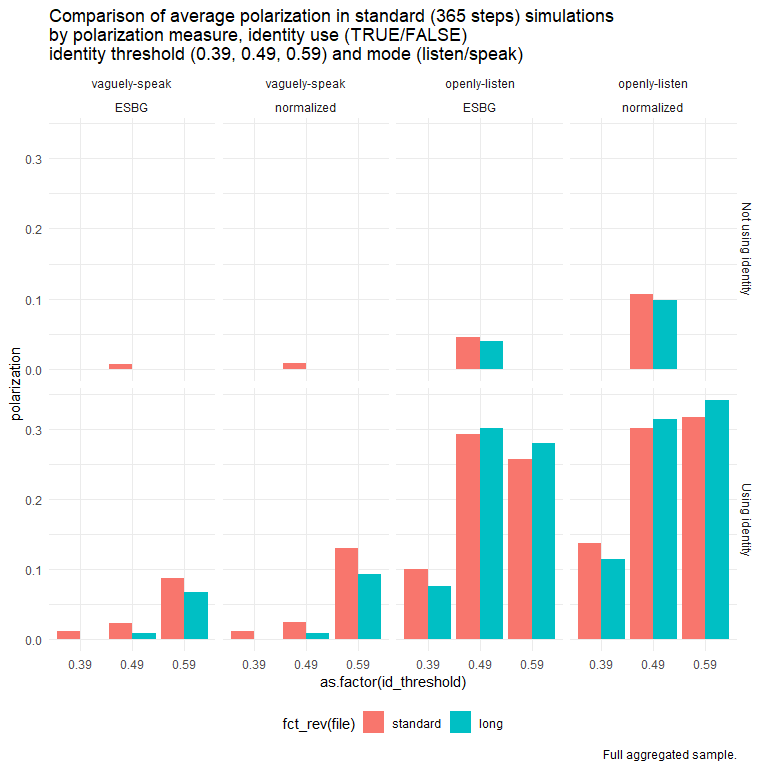
So, we might neglect the 6 resting variables in further experiments and play more smoothly with identity threshold (since the Mode is true/false, as well as Identity use). We also should inspect extreme value of Probability of speaking – if the probability will be 100%, would the results be the same?

## Graph reflecting the best model

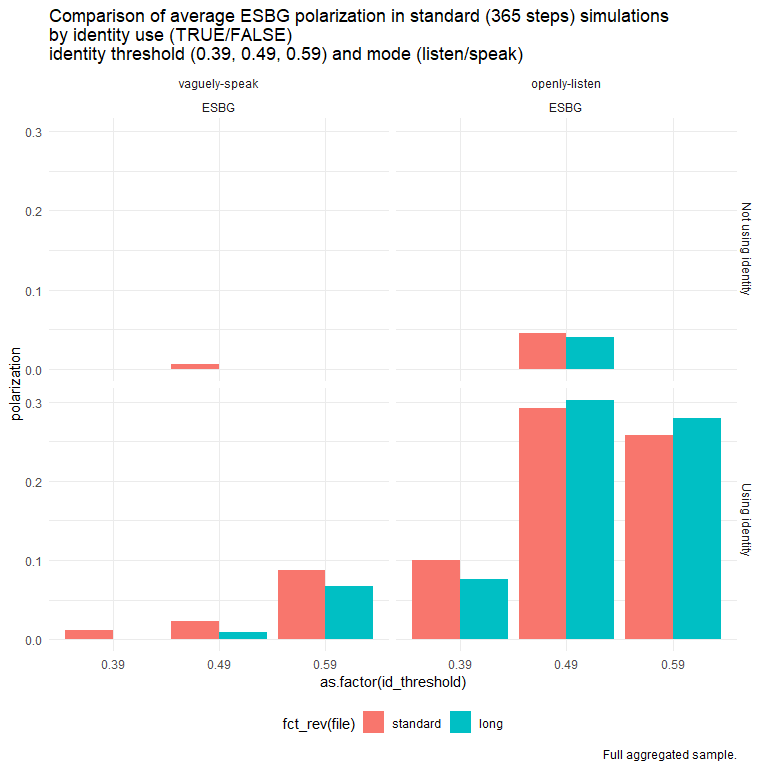
df = res %>% mutate(file = "standard") %>%   
 rename(ESBG = ESBG\_365, normalized = normalized\_365) %>%   
 add\_row(long %>% mutate(file = "long") %>%   
 rename(ESBG = ESBG\_3650, normalized = normalized\_3650)) %>%   
 mutate(identity = if\_else(`use\_identity?`, "Using identity", "Not using identity")) %>%   
 group\_by(file, id\_threshold, identity, mode, boundary) %>%   
 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', 'mode'. You can override using the `.groups` argument.

df %>% filter(boundary == 0.28) %>%   
 ggplot(aes(x = as.factor(id\_threshold), y = polarization, fill = fct\_rev(file))) +  
 facet\_grid(cols = vars(fct\_rev(mode), polarization\_measure), rows = vars(identity)) +  
 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")



df %>% filter(polarization\_measure == "ESBG", boundary == 0.28) %>%   
 ggplot(aes(x = as.factor(id\_threshold), y = polarization, fill = fct\_rev(file))) +  
 facet\_grid(cols = vars(fct\_rev(mode), polarization\_measure), rows = vars(identity)) +  
 geom\_col(position = position\_dodge()) +  
 labs(title = "Comparison of average ESBG polarization in standard (365 steps) simulations\nby 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")



df %>% filter(polarization\_measure == "ESBG", file == "standard") %>%   
 ggplot(aes(fill = as.factor(id\_threshold), y = polarization, x = as.factor(boundary))) +  
 facet\_grid(rows = vars(identity), cols = vars(fct\_rev(mode)), scales = "free\_y", space = "free\_y") +  
 geom\_col(position = position\_dodge()) +  
 labs(title = "Comparison of average ESBG polarization in standard (365 steps) simulations\nby boundary (0.1, 0.22, 0.28, 0.34), 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")

