Extraversion Exploration - Ireland and Pakistan

Danny Rosen

2023-12-04

ROSEN_DANIEL_Pyschometrics_0991

The goal of this project is to assess the validity of a newly developed Extraversion scale composed of six dimensions: activity level, assertiveness, cheerfulness, excitement-seeking, friendliness, and gregariousness. Each dimension is measured through four items, with participants rating each item on a scale from 1 ("very inaccurate") to 5 ("very accurate"). The scale includes reverse-coding for items inversely related to Extraversion, such as "avoid crowds," to ensure positive relationships between all variables. The project aims to use statistical tools to evaluate the effectiveness of this scale in measuring Extraversion across two different countries.

Data: 0991

First let's load our packages and requiste data.

```
library(tidyverse)
library(EGAnet)
```

```
## Warning: package 'EGAnet' was built under R version 4.2.3
```

```
library(ggplot2)

load("./data/0991.RData")
code_book <- read.csv("./data/extraversion_codebook.csv")

survey_data <- extraversion[, -(1:5)]
```

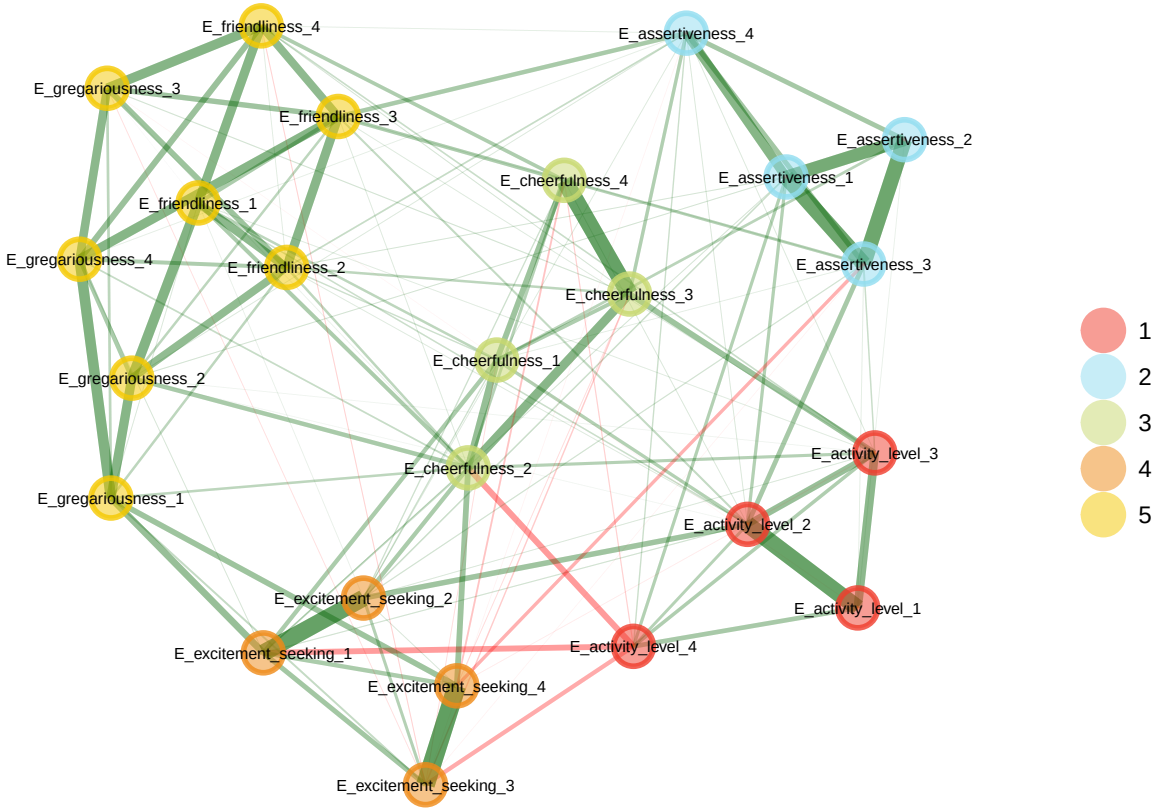
Now, let's help our friend answer some questions

To begin, the questions we need to initially answer focus on if our pilot survey succeeds at measuring "extraversion" the way we expect it to theoretically.

1) The first question we need to answer is if our theoretical dimensions match the empirical dimensions. And further, if the items within those dimensions also match their expected dimensions.

```
original_ega <- EGA(
  survey_data,
  algorithm="louvain",
  node.size = 6,
```

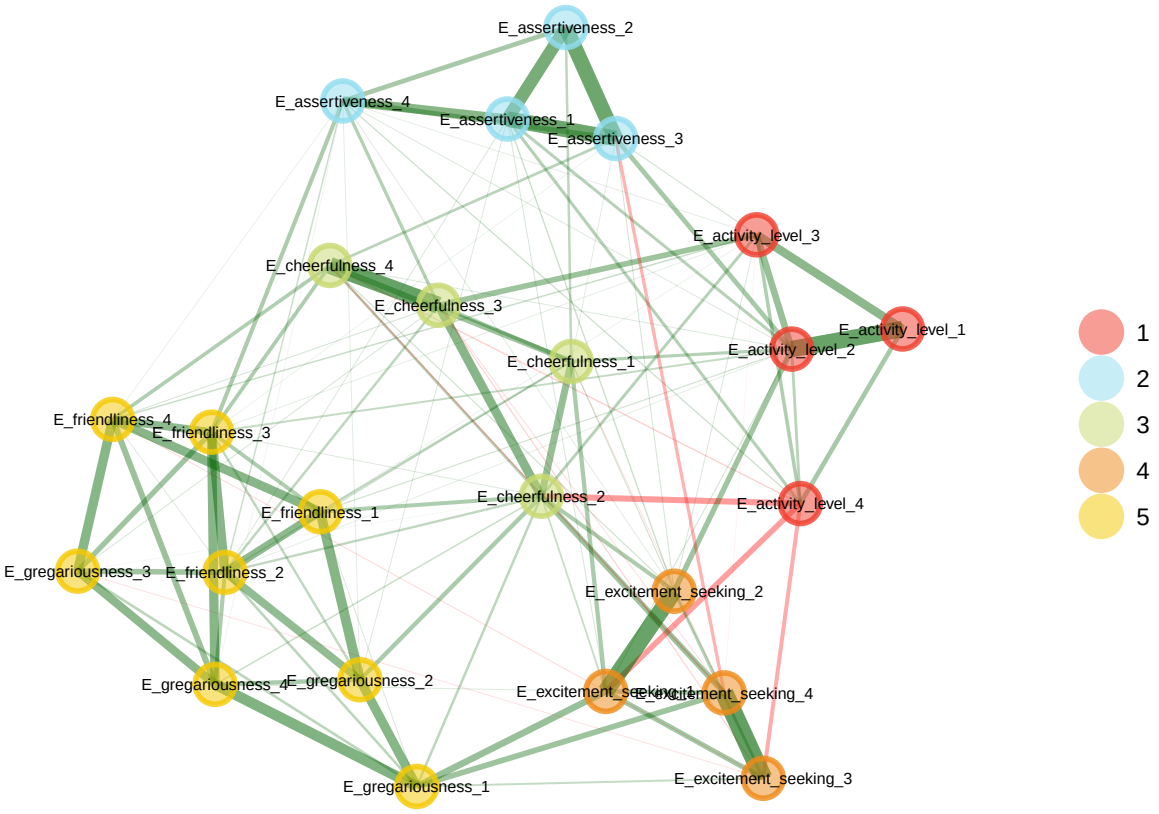
```
  label.size=2,
  plot.EGA = T
)
```



The empirical structure identified 5 dimensions, while theoretically there should be 6 dimensions. Upon examination, the issue lies in gregariousness and friendliness being combined into one empirical dimension. This suggests similarity between the items or an underlying dimensionality problem. Examining the codebook, we can posit as to why these dimensions were grouped together. The trend of the questions the two groups ask are very similar. "Friendliness_3" asks if the participant '[avoids] contacts with others,' while 'gregariousness_3' asks if the participant '[prefers] to be alone.' Although they are not identical questions, as actively avoiding people is not the same as enjoying spending time alone, it is likely that many participants who would avoid people would also enjoy being alone. This idea is present in most of the items within the two dimensions and is likely the reason they are grouped together.

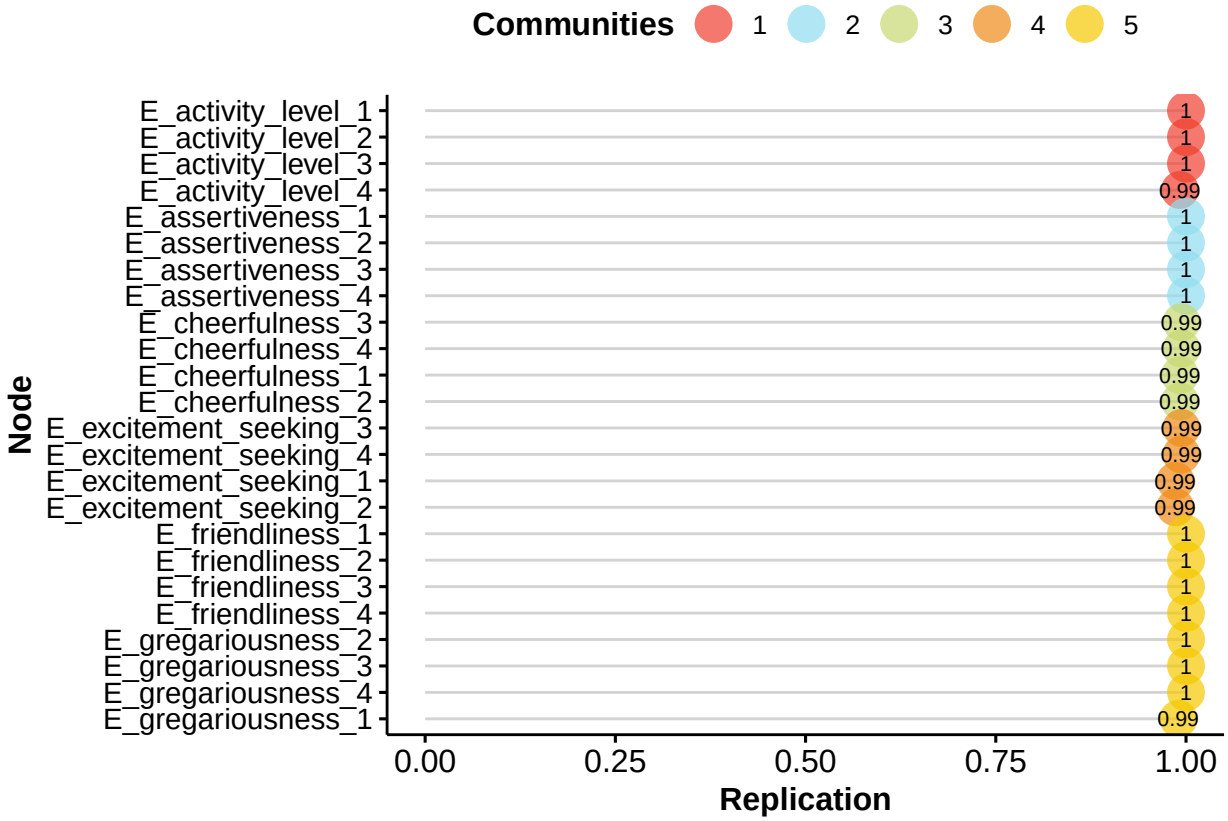
2) Next, we need to establish if this structure of this data is. To do this we can use bootstrapping to create another graph using an estimated distribution of this data.

```
seed <- 123

extraversion_boot_ega <-bootEGA(
  survey_data,
  algorithm="louvain",
  node.size = 6,
  label.size=2,
  seed=seed
)
```



Overall, the estimated structure demonstrates stability, replicating the 5 dimensions nearly 99% of the time under bootstrapping. This indicates it may generalize well. However, given there should theoretically be 6 dimensions, the 5 observed dimensions present an issue for matching the theoretical model. Let's examine this graph further.

```
extraversion_stability <- dimensionStability(extraversion_boot_ega)
```




```
summary(extraversion_stability)
```

```
## EGA Type: EGA
## Bootstrap Samples: 500 (Parametric)
##
## Proportion Replicated in Dimensions:
##
##     E_activity_level_1     E_activity_level_2     E_activity_level_3
##                  1.000                  1.000                  1.000
##     E_activity_level_4      E_assertiveness_1      E_assertiveness_2
##                  0.992                  1.000                  1.000
##      E_assertiveness_3      E_assertiveness_4       E_cheerfulness_1
##                  1.000                  1.000                  0.992
##       E_cheerfulness_2       E_cheerfulness_3       E_cheerfulness_4
##                  0.992                  0.994                  0.994
## E_excitement_seeking_1 E_excitement_seeking_2 E_excitement_seeking_3
##                  0.986                  0.986                  0.994
## E_excitement_seeking_4       E_friendliness_1       E_friendliness_2
##                  0.994                  1.000                  1.000
##       E_friendliness_3       E_friendliness_4     E_gregariousness_1
##                  1.000                  1.000                  0.990
##     E_gregariousness_2     E_gregariousness_3     E_gregariousness_4
##                  1.000                  1.000                  1.000
##
## ----
```

```
##
## Structural Consistency:
##
##     1     2     3     4     5
## 0.992 1.000 0.992 0.986 0.990
```

```
extraversion_stability$item.stability$item.stability$all.dimensions
```

```
##                            1 2     3     4    5
## E_activity_level_1     1.000 0 0.000 0.000 0.00
## E_activity_level_2     1.000 0 0.000 0.000 0.00
## E_activity_level_3     1.000 0 0.000 0.000 0.00
## E_activity_level_4     0.992 0 0.002 0.006 0.00
## E_assertiveness_1      0.000 1 0.000 0.000 0.00
## E_assertiveness_2      0.000 1 0.000 0.000 0.00
## E_assertiveness_3      0.000 1 0.000 0.000 0.00
## E_assertiveness_4      0.000 1 0.000 0.000 0.00
## E_cheerfulness_1       0.006 0 0.992 0.002 0.00
## E_cheerfulness_2       0.006 0 0.992 0.002 0.00
## E_cheerfulness_3       0.006 0 0.994 0.000 0.00
## E_cheerfulness_4       0.006 0 0.994 0.000 0.00
## E_excitement_seeking_1 0.008 0 0.006 0.986 0.00
## E_excitement_seeking_2 0.008 0 0.006 0.986 0.00
## E_excitement_seeking_3 0.000 0 0.006 0.994 0.00
## E_excitement_seeking_4 0.000 0 0.006 0.994 0.00
## E_friendliness_1       0.000 0 0.000 0.000 1.00
## E_friendliness_2       0.000 0 0.000 0.000 1.00
## E_friendliness_3       0.000 0 0.000 0.000 1.00
## E_friendliness_4       0.000 0 0.000 0.000 1.00
## E_gregariousness_1     0.000 0 0.000 0.010 0.99
## E_gregariousness_2     0.000 0 0.000 0.000 1.00
## E_gregariousness_3     0.000 0 0.000 0.000 1.00
## E_gregariousness_4     0.000 0 0.000 0.000 1.00
```

These findings further support our conclusion that the structure and dimensions within are stable.

3) We also need to know if there is local independence between our variables. Our assumption is that, if we account for our goal variable, or in this case "extraversion", our variables will no longer be correlated. We can test for this with Unidimensional Variable Analysis done by using the UVA() function. This function tests for weighted topological overlap which shows to what extent our variables measure the same things.

```
UVA(survey_data)
```

```
## Variable pairs with wTO > 0.30 (large-to-very large redundancy)
##
##                  node_i                 node_j   wto
##  E_excitement_seeking_3 E_excitement_seeking_4 0.305
##      E_activity_level_1     E_activity_level_2 0.305
##       E_assertiveness_2      E_assertiveness_3 0.302
##
## ----
##
## Variable pairs with wTO > 0.25 (moderate-to-large redundancy)
```

```
##
##                  node_i                 node_j   wto
##       E_assertiveness_1      E_assertiveness_2 0.274
##       E_assertiveness_1      E_assertiveness_3 0.269
##  E_excitement_seeking_1 E_excitement_seeking_2 0.268
##        E_cheerfulness_3       E_cheerfulness_4 0.259
##
## ----
##
## Variable pairs with wTO > 0.20 (small-to-moderate redundancy)
```

From the UVA results it is apparent that there are variables within the survey which are not unique and capture the same information as other questions. We know this because the assumption of local independence was violated for some item pairs, specifically between items within the same dimensions. For example, the excitement_seeking and activity_level dimensions had high overlap between items 1 & 2 and 3 & 4 within each respective dimension. To resolve, these items could be revised or dropped to improve uniqueness.

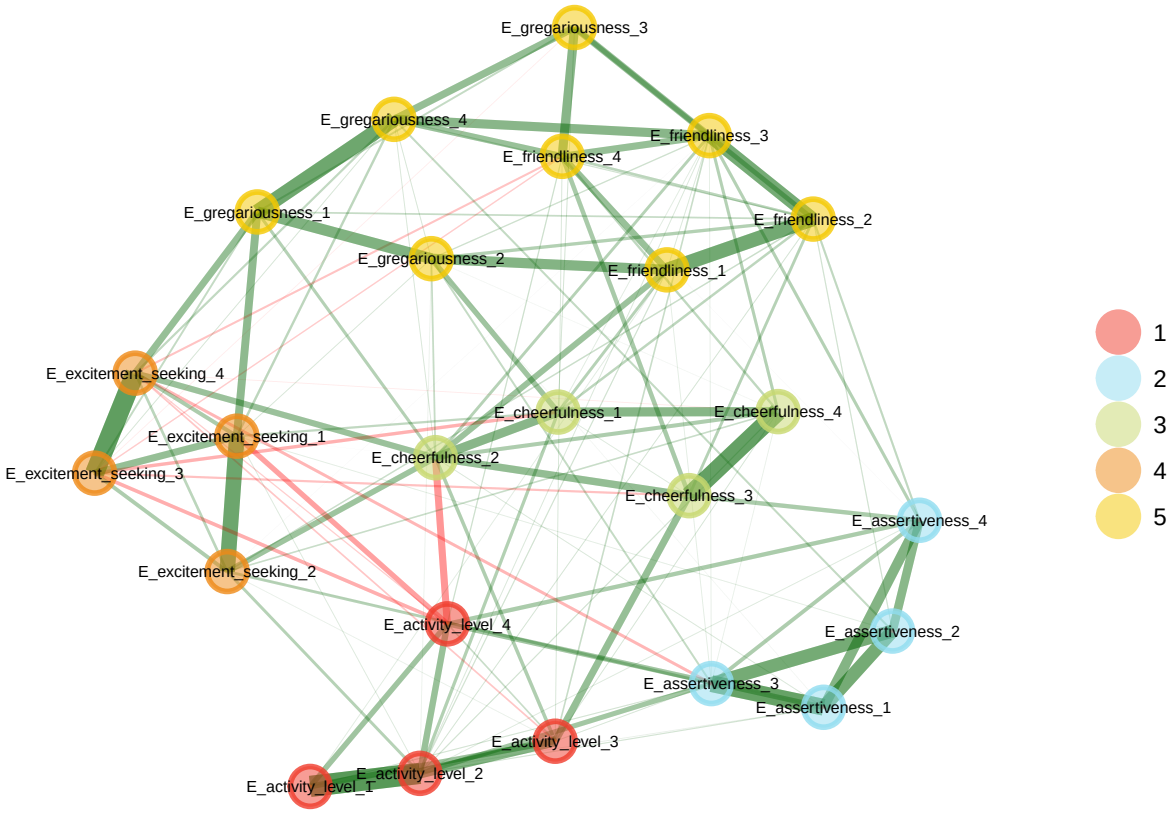
Country Differences

The next questions we need to answer relate to how our survey differs between Ireland and Pakistan. First let's reset our survey dataframe such that the country labels are accessible. Then let's split them into two groups.

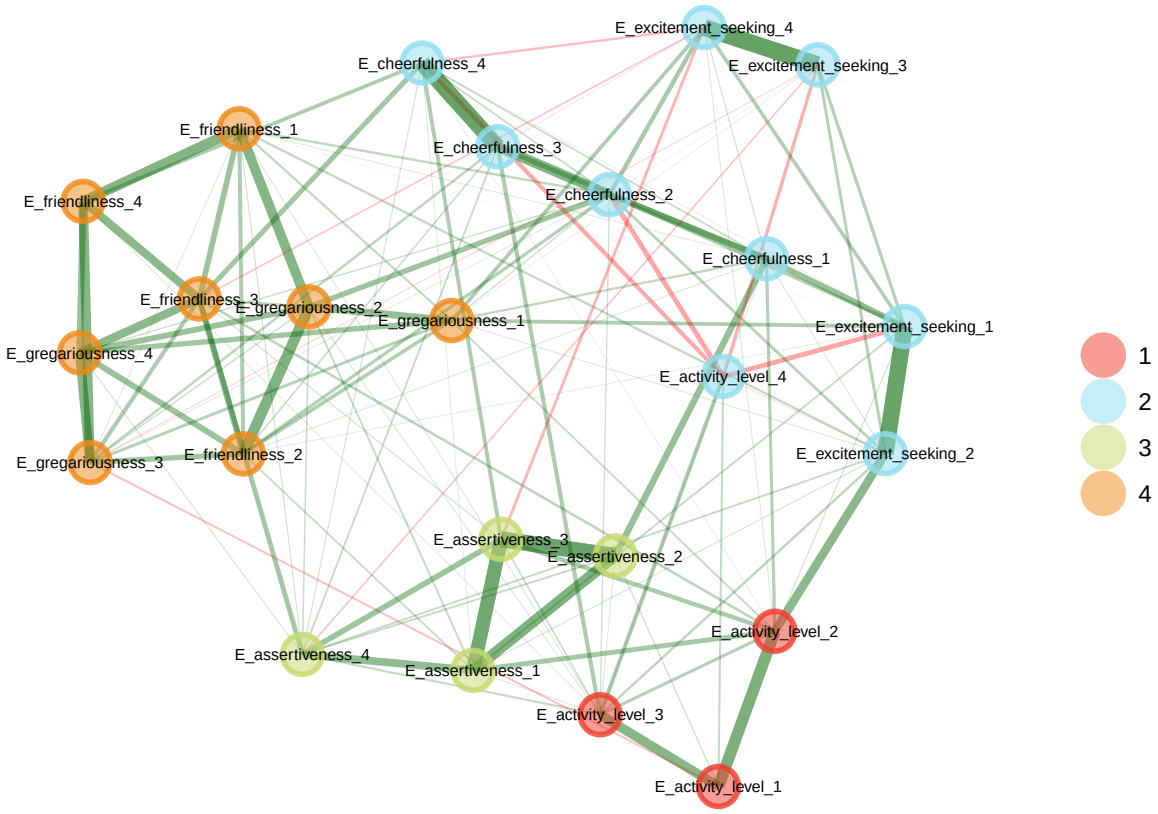
```
survey_data_country <- extraversion[, -(1:4)]

ireland_data <- survey_data_country %>%
  filter(country == 'Ireland') %>%
  select(-country)

pakistan_data <- survey_data_country %>%
  filter(country == 'Pakistan') %>%
  select(-country)
```

4) First, let's see if between countries how our empirical structure compares to the theoretical.

```
ireland_ega <- EGA(
  ireland_data,
  algorithm="louvain",
  node.size = 6,
  label.size=2,
  plot.EGA = T
)
```



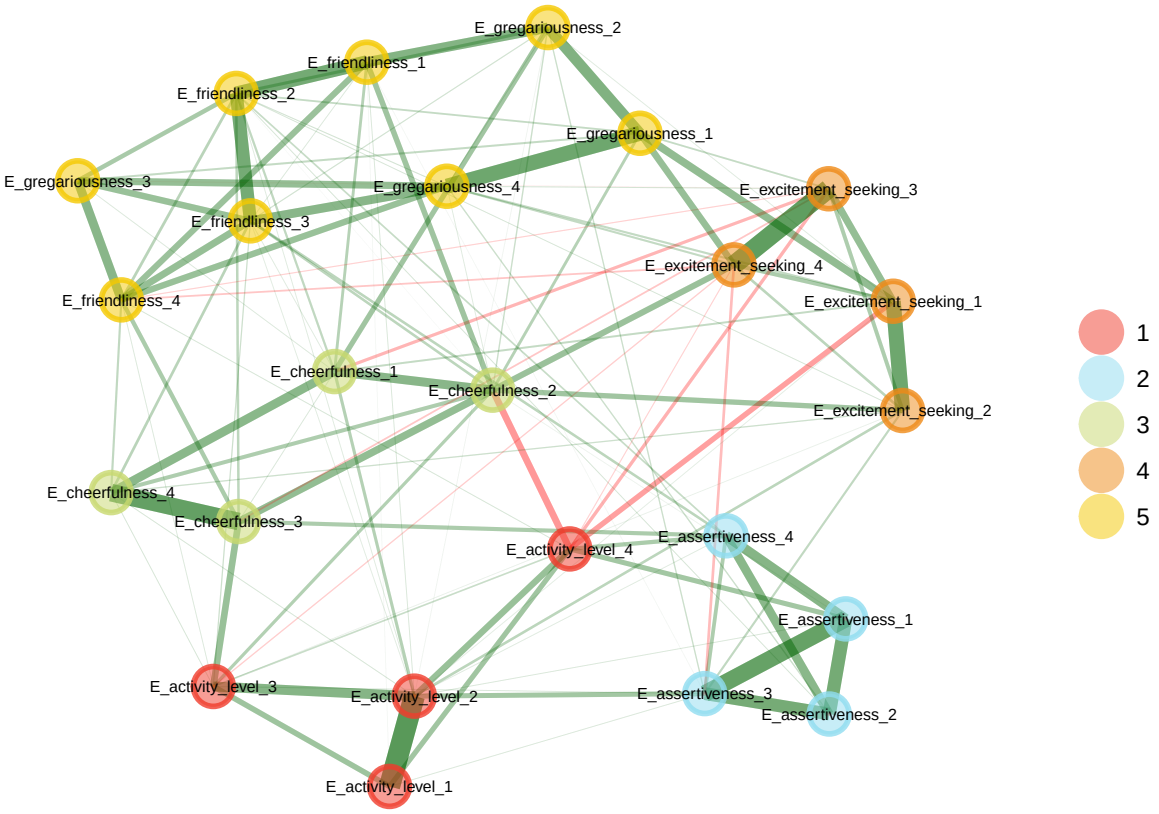

```
pakistan_ega <- EGA(
  pakistan_data,
  algorithm="louvain",
  node.size = 6,
  label.size=2,
  plot.EGA = T
)
```



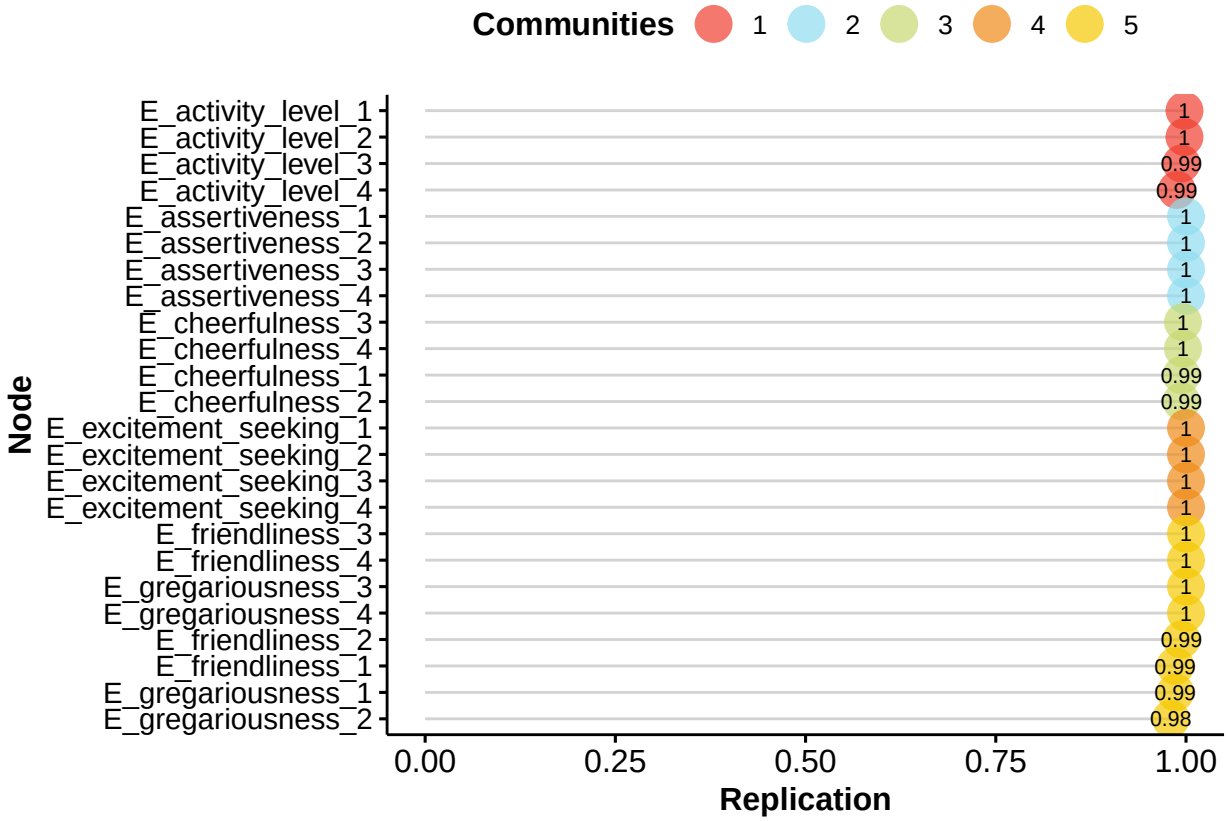
The country analyses revealed notable differences in structure and stability. For Ireland, the structure matched expectations and demonstrated stability under bootstrapping. However for Pakistan, only 4 dimensions were identified and results were highly unstable across iterations. This indicates an issue generalizing the Pakistani survey structure. Differences in specific items were also revealed, suggesting possible cultural influences on the concept of extraversion. The only two dimensions which match between the countries are 'assertiveness' and the grouped-dimension of 'friendliness' and 'gregariousness'.

5) Next, let's look at the instabilities of each of these structures.

```
ireland_boot_ega <-bootEGA(
  ireland_data,
  algorithm="louvain",
  node.size = 6,
  label.size=2,
  seed=seed
)
```




```
ireland_stability <- dimensionStability(ireland_boot_ega)
```




```
summary(ireland_stability)
```

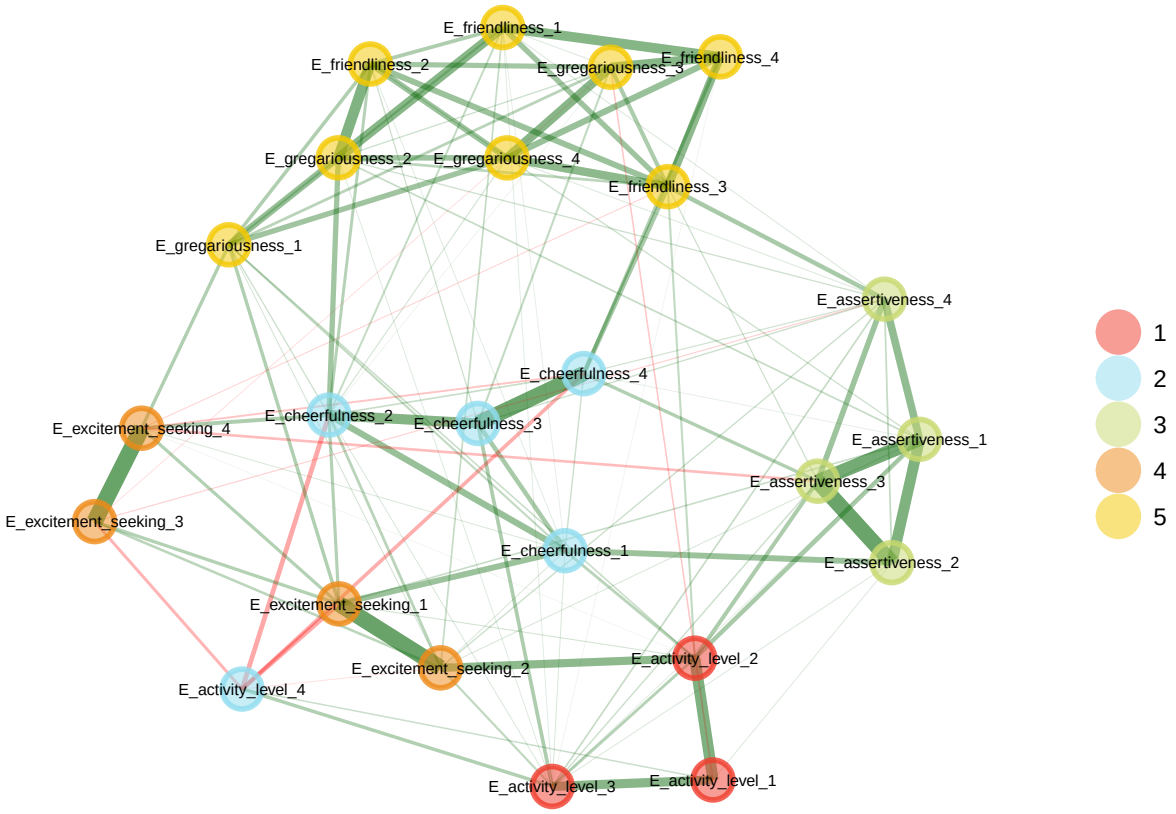
```
## EGA Type: EGA
## Bootstrap Samples: 500 (Parametric)
##
## Proportion Replicated in Dimensions:
##
##     E_activity_level_1     E_activity_level_2     E_activity_level_3
##                  0.998                  0.998                  0.994
##     E_activity_level_4      E_assertiveness_1      E_assertiveness_2
##                  0.988                  1.000                  1.000
##      E_assertiveness_3      E_assertiveness_4       E_cheerfulness_1
##                  1.000                  1.000                  0.994
##       E_cheerfulness_2       E_cheerfulness_3       E_cheerfulness_4
##                  0.994                  0.996                  0.996
## E_excitement_seeking_1 E_excitement_seeking_2 E_excitement_seeking_3
##                  1.000                  1.000                  1.000
## E_excitement_seeking_4       E_friendliness_1       E_friendliness_2
##                  1.000                  0.986                  0.994
##       E_friendliness_3       E_friendliness_4     E_gregariousness_1
##                  1.000                  1.000                  0.986
##     E_gregariousness_2     E_gregariousness_3     E_gregariousness_4
##                  0.980                  1.000                  1.000
##
## ----
```

```
##
## Structural Consistency:
##
##     1     2     3     4     5
## 0.984 1.000 0.994 1.000 0.974
```

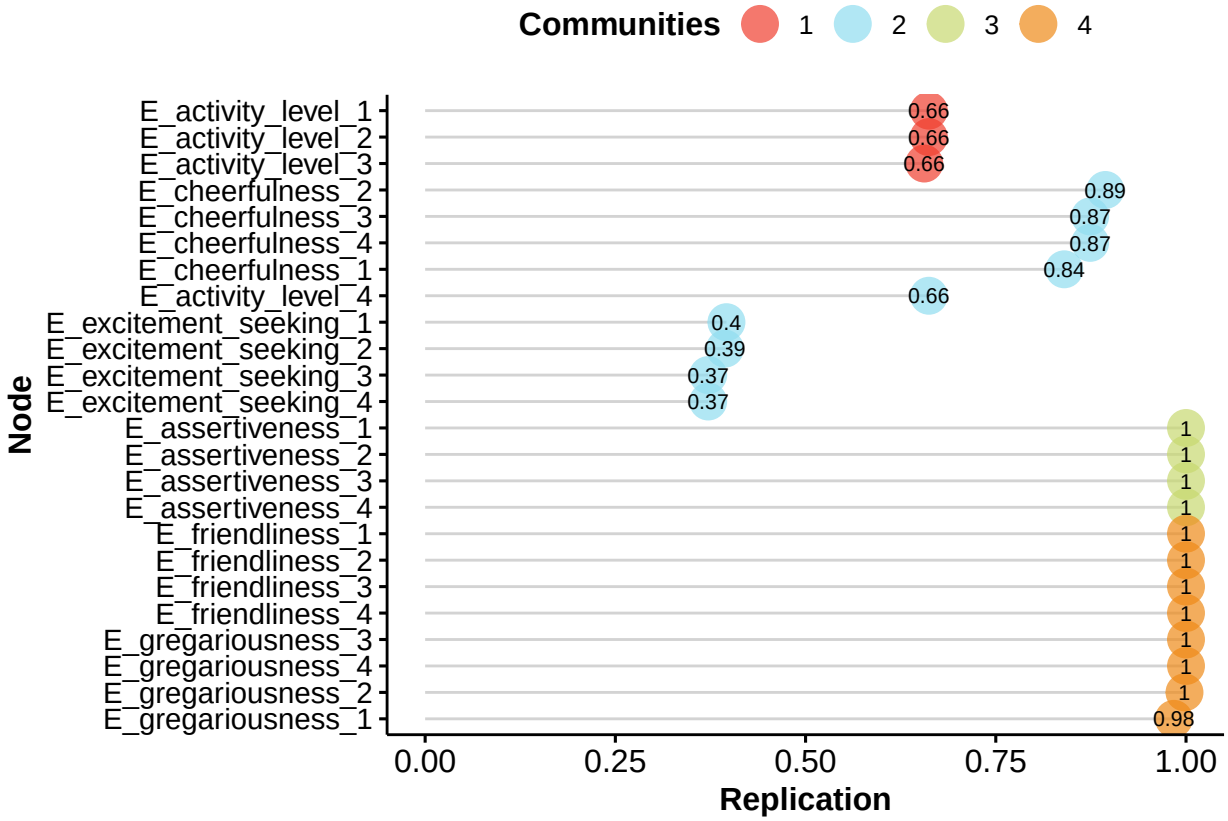
```
ireland_stability$item.stability$item.stability$all.dimensions
```

```
##                            1     2     3     4     5
## E_activity_level_1     0.998 0.002 0.000 0.000 0.000
## E_activity_level_2     0.998 0.002 0.000 0.000 0.000
## E_activity_level_3     0.994 0.000 0.006 0.000 0.000
## E_activity_level_4     0.988 0.012 0.000 0.000 0.000
## E_assertiveness_1      0.000 1.000 0.000 0.000 0.000
## E_assertiveness_2      0.000 1.000 0.000 0.000 0.000
## E_assertiveness_3      0.000 1.000 0.000 0.000 0.000
## E_assertiveness_4      0.000 1.000 0.000 0.000 0.000
## E_cheerfulness_1       0.004 0.000 0.994 0.000 0.002
## E_cheerfulness_2       0.004 0.000 0.994 0.000 0.002
## E_cheerfulness_3       0.004 0.000 0.996 0.000 0.000
## E_cheerfulness_4       0.004 0.000 0.996 0.000 0.000
## E_excitement_seeking_1 0.000 0.000 0.000 1.000 0.000
## E_excitement_seeking_2 0.000 0.000 0.000 1.000 0.000
## E_excitement_seeking_3 0.000 0.000 0.000 1.000 0.000
## E_excitement_seeking_4 0.000 0.000 0.000 1.000 0.000
## E_friendliness_1       0.000 0.000 0.014 0.000 0.986
## E_friendliness_2       0.000 0.000 0.006 0.000 0.994
## E_friendliness_3       0.000 0.000 0.000 0.000 1.000
## E_friendliness_4       0.000 0.000 0.000 0.000 1.000
## E_gregariousness_1     0.000 0.000 0.000 0.014 0.986
## E_gregariousness_2     0.000 0.000 0.012 0.008 0.980
## E_gregariousness_3     0.000 0.000 0.000 0.000 1.000
## E_gregariousness_4     0.000 0.000 0.000 0.000 1.000
```

```
pakistan_boot_ega <-bootEGA(
  pakistan_data,
  algorithm="louvain",
  node.size = 6,
  label.size=2,
  seed=seed
)
```




```
pakistan_stability <- dimensionStability(pakistan_boot_ega)
```




```
summary(pakistan_stability)
```

```
## EGA Type: EGA
## Bootstrap Samples: 500 (Parametric)
##
## Proportion Replicated in Dimensions:
##
##     E_activity_level_1     E_activity_level_2     E_activity_level_3
##                  0.662                  0.662                  0.656
##     E_activity_level_4      E_assertiveness_1      E_assertiveness_2
##                  0.662                  1.000                  1.000
##      E_assertiveness_3      E_assertiveness_4       E_cheerfulness_1
##                  1.000                  1.000                  0.840
##       E_cheerfulness_2       E_cheerfulness_3       E_cheerfulness_4
##                  0.894                  0.874                  0.874
## E_excitement_seeking_1 E_excitement_seeking_2 E_excitement_seeking_3
##                  0.396                  0.394                  0.372
## E_excitement_seeking_4       E_friendliness_1       E_friendliness_2
##                  0.372                  1.000                  1.000
##       E_friendliness_3       E_friendliness_4     E_gregariousness_1
##                  1.000                  1.000                  0.984
##     E_gregariousness_2     E_gregariousness_3     E_gregariousness_4
##                  0.998                  1.000                  1.000
##
## ----
```

```
##
## Structural Consistency:
##
##     1     2     3     4
## 0.656 0.170 1.000 0.984
```

```
pakistan_stability$item.stability$item.stability$all.dimensions
```

```
##                            1     2     3     4     5     6
## E_activity_level_1     0.662 0.052 0.146 0.000 0.140 0.000
## E_activity_level_2     0.662 0.054 0.146 0.000 0.138 0.000
## E_activity_level_3     0.656 0.060 0.146 0.000 0.138 0.000
## E_activity_level_4     0.192 0.662 0.004 0.004 0.138 0.000
## E_assertiveness_1      0.000 0.000 1.000 0.000 0.000 0.000
## E_assertiveness_2      0.000 0.000 1.000 0.000 0.000 0.000
## E_assertiveness_3      0.000 0.000 1.000 0.000 0.000 0.000
## E_assertiveness_4      0.000 0.000 1.000 0.000 0.000 0.000
## E_cheerfulness_1       0.046 0.840 0.004 0.002 0.108 0.000
## E_cheerfulness_2       0.000 0.894 0.000 0.016 0.090 0.000
## E_cheerfulness_3       0.004 0.874 0.000 0.000 0.122 0.000
## E_cheerfulness_4       0.004 0.874 0.000 0.000 0.122 0.000
## E_excitement_seeking_1 0.360 0.396 0.000 0.000 0.244 0.000
## E_excitement_seeking_2 0.362 0.394 0.000 0.000 0.244 0.000
## E_excitement_seeking_3 0.000 0.372 0.000 0.000 0.604 0.024
## E_excitement_seeking_4 0.000 0.372 0.000 0.000 0.604 0.024
## E_friendliness_1       0.000 0.000 0.000 1.000 0.000 0.000
## E_friendliness_2       0.000 0.000 0.000 1.000 0.000 0.000
## E_friendliness_3       0.000 0.000 0.000 1.000 0.000 0.000
## E_friendliness_4       0.000 0.000 0.000 1.000 0.000 0.000
## E_gregariousness_1     0.000 0.010 0.000 0.984 0.006 0.000
## E_gregariousness_2     0.000 0.002 0.000 0.998 0.000 0.000
## E_gregariousness_3     0.000 0.000 0.000 1.000 0.000 0.000
## E_gregariousness_4     0.000 0.000 0.000 1.000 0.000 0.000
```

As we can see, similar to the original structure, the Irish survey is very stable. We can see this because our bootstrapping reproduces the specific dimensions at least at a 97% rate with some being reproduced every time. Conversely, the Pakistani survey has a wildly unstable structure. Looking at the 'excitement-seeking', 'cheerfulness', and 'activity-level' dimensions, these are highly inconsistent across bootstrapping iterations. Given that the empirical structure has four dimensions, we know it is inconsistent because dimension two is only reproduced in that way 16.4% of the time and dimension one is only reproduced 65% of the time. In conclusion, the Irish survey is generalizable while the Pakistani survey is not.

Let's also look at the weighted topological overlap of the countries' surveys.

```
UVA(ireland_data)
```

```
## Variable pairs with wTO > 0.30 (large-to-very large redundancy)
##
##              node_i              node_j wto
##  E_activity_level_1 E_activity_level_2 0.4
##
## ----
##
```

```
## Variable pairs with wTO > 0.25 (moderate-to-large redundancy)
##
##                  node_i                 node_j   wto
##  E_excitement_seeking_3 E_excitement_seeking_4 0.291
##       E_assertiveness_1      E_assertiveness_3 0.284
##       E_assertiveness_1      E_assertiveness_2 0.282
##        E_cheerfulness_3       E_cheerfulness_4 0.277
##       E_assertiveness_2      E_assertiveness_3 0.262
##
## ----
##
## Variable pairs with wTO > 0.20 (small-to-moderate redundancy)
##
##                  node_i                 node_j   wto
##  E_excitement_seeking_1 E_excitement_seeking_2 0.244
##        E_friendliness_1       E_friendliness_2 0.213
##      E_gregariousness_1     E_gregariousness_4 0.208
```

For the Irish survey, the only items with what is considered 'large-to-very large redundancy' is activity_level 1 and 2. This has a wTO of .4 which is extremely large and needs to be managed. There are also several item-pairs with moderate to large wTO including excitement_seeking 3 and 4, assertiveness 1 with both assertiveness 2 and 3, cheerfulness 3 and 4, and assertiveness 2 and 3.

```
UVA(pakistan_data)
```

```
## Variable pairs with wTO > 0.30 (large-to-very large redundancy)
##
## ----
##
## Variable pairs with wTO > 0.25 (moderate-to-large redundancy)
##
##                  node_i                 node_j   wto
##       E_assertiveness_2      E_assertiveness_3 0.295
##  E_excitement_seeking_3 E_excitement_seeking_4 0.294
##       E_assertiveness_1      E_assertiveness_3 0.251
##
## ----
##
## Variable pairs with wTO > 0.20 (small-to-moderate redundancy)
##
##                  node_i                 node_j   wto
##  E_excitement_seeking_1 E_excitement_seeking_2 0.249
##        E_cheerfulness_3       E_cheerfulness_4 0.236
##       E_assertiveness_1      E_assertiveness_2 0.225
##      E_activity_level_1     E_activity_level_2 0.202
```

Interestingly, despite the instability of the Pakistani survey, there are no item pair with large to very large wTOs. The 3 pairs with noticeable wTOs are assertiveness 3 with both assertiveness 1 and 2 and excitement_seeking 3 and 4.

Invariance

6) Now that we have established that the survey has shown different results in Ireland and Pakistan, the question we need to answer is which of the items are the invariant across country lines and which are not.

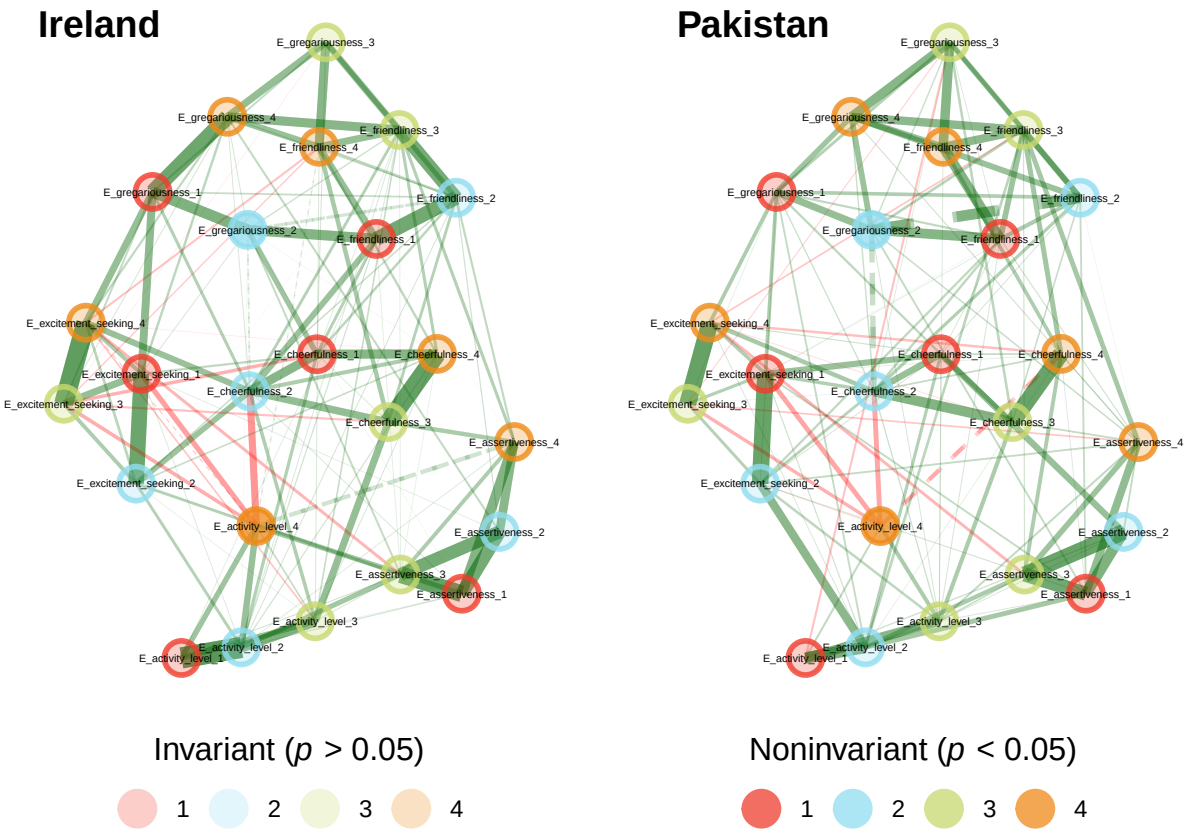
First, we need to set up our theoretical structure.

```
theoretical <- rep(1:4, 6)
```

```
country_invariance <- invariance(
  data = survey_data,
  groups = survey_data_country$country,
  structure = theoretical,
  seed = 123
)
```

```
## Testing metric invariance...
```

```
plot(country_invariance, node.size = 5, label.size = 1.5)
```




```
country_invariance$results %>%
  filter(!(sig %in% c("", '.')))
```

```
##                     Membership Difference     p  p_BH sig          Direction
## E_gregariousness_2           2     -0.100 0.010 0.120  ** Ireland < Pakistan
## E_activity_level_4           4      0.105 0.002 0.048  ** Ireland > Pakistan
```

These results show us that there are two items which are not being measured the same way across the countries. The first of these is 'gregariousness_2' which asks the participants if they "Talk to a lot of

different people at parties," and 'activity_level_4' which ask if the participants if they "Like to take it easy." Interestingly, it would appear that talking to alot of different people at parties is a better measure of extraversion for Pakistani participants than for Irish participants, while enjoying taking it easy is a better measure extraversion for Irish participants than the Pakistani participants. We know this by looking at the significance and direction measures provided by the invariance() function.

With these differences in mind, we now need to come up with a hypothesis about why these might be occurring and an experiment which could be used evaluate them.

Hypothesis We believe the differences in measurement for 'gregariousness_2' and 'activity_level_4' may be due to respective cultural influence on perceptions of extraversion:

Pakistani Participants: In Pakistan, the concept of extraversion might be more closely linked to social engagement in large groups, such as talking to various people at parties. This reflects a cultural emphasis on community and social interactions.

Irish Participants: For the Irish group, extraversion may be more associated with a relaxed, easy-going attitude rather than overt social engagement. This could indicate a cultural preference for subtler forms of social interaction.

Experiment: To test this hypothesis, I would perform a qualitative experiment which would involve interviewing a diverse group of people from both Pakistan and Ireland and exploring participants' personal interpretations of extraversion, focusing on their views on socializing at parties and relaxation as indicators of extraversion. Because we already have plenty of quantitative data about extraversion in the current survey, I believe collecting additional qualitative data would help us answer our questions.

Once the interview data was collected, we could perform thematic analysis to identify common themes and cultural nuances in the perception of extraversion. For example, topic modeling could be done on the responses to highlight key cultural tropes or large language models could be used to summarize the responses and to make the large amount of data digestible for researchers. The questions could then be adjusted to make our survey more generalization and informative.

7) Finally, we need to help our friend in determining if the responses given in their survey can be used to predict the age of the respondent. And further, if it can be predicted, is it different between the countries and which variables are most correlated. To start, let's break down our original 'extraversion' dataframe so we can proceed.

```
voi <- !(colnames(extraversion) %in% c("id", "sex", "year"))
age_data <- extraversion[, voi]
row.names(age_data) <- NULL

age_data_ireland <- age_data %>%
  filter(country == 'Ireland') %>%
  select(-country)

age_data_pakistan <- age_data %>%
  filter(country == 'Pakistan') %>%
  select(-country)
```

Next, let's break these down into test and training sets. We will use 80% of rows for training and 20% for testing.

```
set.seed(123)
#Ireland
i_size = nrow(age_data_ireland)
ireland_train_idx <- sample(1:i_size, as.integer(.8*i_size), replace = F)
```

```r
ireland_test_idx <- (1:i_size)[!(1:i_size %in% ireland_train_idx)]
ireland_train = age_data_ireland[ireland_train_idx, ]
ireland_test = age_data_ireland[ireland_test_idx, ]

#Pakistan
p_size = nrow(age_data_pakistan)
pakistan_train_idx <- sample(1:p_size, as.integer(.8*i_size), replace = F)
pakistan_test_idx <- (1:i_size)[!(1:i_size %in% pakistan_train_idx)]
pakistan_train = age_data_pakistan[pakistan_train_idx, ]
pakistan_test = age_data_pakistan[pakistan_test_idx, ]
```

Now, let's create a linear model between the dimensions and age for both countries' data and see the results. First, Ireland.

```r
ireland_lm <- lm(
  age ~ .,
  ireland_train
)

summary(ireland_lm)
```

```
##
## Call:
## lm(formula = age ~ ., data = ireland_train)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.165  -5.442  -1.285   4.068  26.831
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               29.3554     2.8365  10.349  < 2e-16 ***
## E_activity_level_1         1.1591     0.4729   2.451 0.014666 *
## E_activity_level_2         0.3163     0.5155   0.613 0.539893
## E_activity_level_3        -0.3548     0.3806  -0.932 0.351748
## E_activity_level_4         0.5482     0.4075   1.345 0.179240
## E_assertiveness_1         -0.2789     0.5331  -0.523 0.601144
## E_assertiveness_2         -0.5601     0.4557  -1.229 0.219713
## E_assertiveness_3          1.4452     0.5586   2.587 0.010030 *
## E_assertiveness_4          0.5660     0.4223   1.340 0.180850
## E_cheerfulness_1           0.2915     0.4325   0.674 0.500775
## E_cheerfulness_2          -1.1735     0.5087  -2.307 0.021577 *
## E_cheerfulness_3          -0.3092     0.4639  -0.667 0.505411
## E_cheerfulness_4           0.9417     0.4606   2.044 0.041571 *
## E_excitement_seeking_1    -0.8242     0.5563  -1.482 0.139251
## E_excitement_seeking_2    -1.1915     0.4635  -2.571 0.010509 *
## E_excitement_seeking_3     0.5140     0.4058   1.267 0.206020
## E_excitement_seeking_4    -1.5899     0.4208  -3.778 0.000182 ***
## E_friendliness_1          -0.1632     0.4926  -0.331 0.740644
## E_friendliness_2           0.8171     0.5402   1.513 0.131192
## E_friendliness_3           0.8709     0.5056   1.722 0.085754 .
## E_friendliness_4           0.5559     0.4393   1.266 0.206412
## E_gregariousness_1        -1.0305     0.4360  -2.363 0.018586 *
```

```
## E_gregariousness_2      -0.2266     0.4007  -0.566 0.572035
## E_gregariousness_3      -0.5877     0.4285  -1.372 0.170939
## E_gregariousness_4      -0.6016     0.4641  -1.296 0.195590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.056 on 402 degrees of freedom
## Multiple R-squared:  0.308,  Adjusted R-squared:  0.2667
## F-statistic: 7.455 on 24 and 402 DF,  p-value: < 2.2e-16
```

```
rmse <- function(preds, target){
  error <- preds - target
  se <- error ^ 2
  mse <- mean(se)
  rmse <- sqrt(mse)
  return(rmse)
}

mae <- function(preds, target){
  me <- abs(preds - target)
  mae <- mean(me)
  return(mae)
}
```

```
ireland_targets <- ireland_test$age

ireland_predictions <- predict.lm(
  ireland_lm,
  ireland_test
)

paste("Mean Absolute Error for Ireland LM:", mae(ireland_predictions, ireland_targets))
```

```
## [1] "Mean Absolute Error for Ireland LM: 6.66542636877239"
```

```
paste("Root Mean Squared Error for Ireland LM:", rmse(ireland_predictions, ireland_targets))
```

```
## [1] "Root Mean Squared Error for Ireland LM: 9.12458755456151"
```

Next, Pakistan.

```
pakistan_lm <- lm(
  age ~ .,
  pakistan_train
)

summary(pakistan_lm)
```

```
##
## Call:
## lm(formula = age ~ ., data = pakistan_train)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -15.767  -4.279  -1.152   2.671  61.874
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             29.00463    2.99838   9.673  < 2e-16 ***
## E_activity_level_1       0.84698    0.35465   2.388  0.01739 *
## E_activity_level_2      -0.34233    0.39774  -0.861  0.38993
## E_activity_level_3       0.24936    0.34449   0.724  0.46958
## E_activity_level_4       0.11981    0.35013   0.342  0.73239
## E_assertiveness_1        0.07973    0.48535   0.164  0.86960
## E_assertiveness_2       -0.77635    0.39019  -1.990  0.04730 *
## E_assertiveness_3        0.69192    0.50179   1.379  0.16869
## E_assertiveness_4       -0.18298    0.35636  -0.513  0.60791
## E_cheerfulness_1         1.26052    0.40094   3.144  0.00179 **
## E_cheerfulness_2        -1.17062    0.42220  -2.773  0.00582 **
## E_cheerfulness_3        -0.71191    0.40891  -1.741  0.08245 .
## E_cheerfulness_4         0.52139    0.33974   1.535  0.12565
## E_excitement_seeking_1 -1.47785    0.47705  -3.098  0.00209 **
## E_excitement_seeking_2 -0.31377    0.39811  -0.788  0.43108
## E_excitement_seeking_3 -0.29569    0.33761  -0.876  0.38164
## E_excitement_seeking_4 -0.32440    0.31247  -1.038  0.29980
## E_friendliness_1        -0.96566    0.39562  -2.441  0.01508 *
## E_friendliness_2         0.50947    0.40227   1.266  0.20608
## E_friendliness_3         0.23388    0.41331   0.566  0.57180
## E_friendliness_4         0.10126    0.39065   0.259  0.79560
## E_gregariousness_1       0.56575    0.33949   1.666  0.09640 .
## E_gregariousness_2       0.49701    0.38647   1.286  0.19917
## E_gregariousness_3      -0.02065    0.37601  -0.055  0.95622
## E_gregariousness_4       0.33465    0.37951   0.882  0.37842
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.612 on 402 degrees of freedom
## Multiple R-squared:  0.1528, Adjusted R-squared:  0.1023
## F-statistic: 3.022 on 24 and 402 DF,  p-value: 3.721e-06
```

```
pakistan_targets <- pakistan_test$age

pakistan_predictions <- predict.lm(
  pakistan_lm,
  pakistan_test
)

paste("Mean Absolute Error for Pakistan LM:", mae(pakistan_predictions, pakistan_targets))
```

```
## [1] "Mean Absolute Error for Pakistan LM: 5.88692307428531"
```

```
paste("Root Mean Squared Error for Pakistan LM:", rmse(pakistan_predictions, pakistan_targets))
```

```
## [1] "Root Mean Squared Error for Pakistan LM: 7.4256659735178"
```

In analyzing the survey data from Ireland and Pakistan for age prediction in ad campaigns, the Irish model shows a Mean Absolute Error (MAE) of 6.67 years and a Root Mean Square Error (RMSE) of 9.12 years. This indicates that while the model generally predicts within about 6.5 years of the actual age, it occasionally makes larger errors, as evidenced by the RMSE being 137% larger than the MAE. A significant predictor in this model is 'excitement_seeking_4', which seems to correlate with younger age groups, particularly teenagers and college students, who might agree more with 'acting wild and crazy'. In contrast, the Pakistani model, despite its less stable graph analysis structure, is more accurate in predicting ages, with a MAE of 5.89 and a RMSE of 7.43. This model's consistency and fewer large-scale errors are highlighted by the RMSE being only 126% of the MAE. Unlike the Irish model, the Pakistani model relies on multiple moderate-to-strong predictors such as 'cheerfulness_1', 'cheerfulness_2', and 'excitement_seeking_1', which may capture a broader range of age-related behaviors. This difference in significant predictors between the models could be attributed to inherent cultural variations influencing responses in each country, a crucial factor in cross-cultural data analysis. Overall, these findings provide valuable insights for tailoring effective ad campaigns in each respective country.