Locus of Control

Li Ying

12 July 2015

### Executive Summary

* This report performed exploratory analysis on language pattern produced by two groups of people: External locus control and Internal locus control (refer to as External Group (EG) and Internal Group (IG) in the following text)
* Word Cloud, Minimum spanning tree, hierarchical clustering graph, and association network were plotted.
* In the end, statistical test were performed on parameters of association network using *partial network analysis / bootstrape with 1000 iteration*. Detailed steps can be found in the appendix.
* Statistical testing method refers to paper written by [Kenett, Gold and Faust( 2015)](http://scottbarrykaufman.com/wp-content/uploads/2015/06/Kenett-Gold-and-Faust-2015.pdf). Parameters of association network includes: **Diameter(D)**, [Clustering Coefficient(CC)](http://mathinsight.org/definition/clustering_coefficient), [Average Path Length(APL)](http://mathinsight.org/definition/network_mean_path_length). The detailed definition of these concepts can be found in the appendix.

#### We found that:

* t-test shows that D and CC are significantally different between EG and IG, while APL shows marginal significance.

### Data Source

* Number of participants: 33
* Locus of Control: [Rotter Internal-External Locus of Control Scale, (Rotter, 1966)](http://www.wilderdom.com/psychology/loc/RotterLOC29.html)
* Free Association task: each participants were given 73 **cue words**. For each of cue words, they were instructed to write down 3 words/responses that they think is most relevant to the **cue words**. The result will be mentioned as **word data** in the following text.
* Other information on *locus of control* can be found here:
  + [Rotter's Scoring Instruction](http://www.wilderdom.com/psychology/loc/RotterLOC29Scoring.html)
  + [What is locus of control?](http://www.wilderdom.com/psychology/loc/LocusOfControlWhatIs.html#InternalExternal)
  + [28 measures of locus of control](http://teachinternalcontrol.com/uploads/LOC_Measures__1_.pdf)

### Data Process

* Control score were averaged for each participants. Participants were categorized into External Group (Low in locus of control) and Internal Group (High in locus of control) using group mean as cut-off line.
* Steps 2, 3, 4 are identical to Free Will and Language: Network Analysis.

### Exploratory Data Analysis

#### Normal Distribution of locus of control score

#### Distribution of number of unique response of each cue words

#### Word Cloud

[Tutorial](https://trinkerrstuff.wordpress.com/2012/11/13/gradient-word-clouds/)

#### Clustering Analysis (single linkage)

#### Minimum Spanning Trees (MST)

## Initial stress : 0.05315  
## stress after 10 iters: 0.03827, magic = 0.092  
## stress after 14 iters: 0.03102

## Initial stress : 0.07392  
## stress after 10 iters: 0.05610, magic = 0.018  
## stress after 20 iters: 0.04089, magic = 0.213  
## stress after 30 iters: 0.03750, magic = 0.020  
## stress after 40 iters: 0.03617, magic = 0.045  
## stress after 50 iters: 0.03133, magic = 0.500  
## stress after 60 iters: 0.03064, magic = 0.500  
## stress after 70 iters: 0.03054, magic = 0.500

#### Association Network Graph

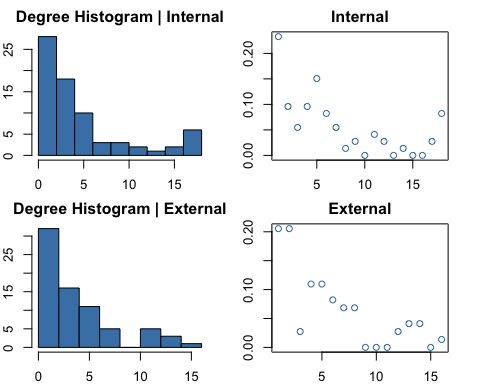
* Association Network were only plotted on those cue words that have correlation > 0.2
* The edge were colored according to correctional strength between nodes, followed sequence of red, blue and grey representing correlation >0.5, >0.4, and else.
* The length of edge is weighted by correlational strength between nodes.

Method: [Tutorial](http://stackoverflow.com/questions/19965600/igraph-r-how-to-create-correlation-network-with-only-strong-r-values)

##### Network parameters from Internal and External groups were calculated:

## High Low  
## APL 2.8644578 3.1527464  
## Diameter 2.0543328 2.1380380  
## CC 0.5488902 0.5275754  
## Degree 4.8767123 3.8630137

##### Degree distribution

Most of nodes do not have a lot edges with other nodes. But there are some nodes having lots of connections. Those nodes could be hubs in the network that connects between different clusters.  


#### Statistical text on social network graph

**Diameter** and **clustering coefficient** show statistical significance between two groups (**p-value < 0.05**). Average path length shows marginally significance (p<0.06)

##### Summary:

## High.Mean Low.Mean p.value   
## APL "2.4286479" "2.3798213" "0.0563"   
## D "1.6906234" "1.7590928" "0.00206"  
## CC "0.4984709" "0.4853743" "0.026"

Detailed information on t-test

# t-test on average path length  
t.test(comAPL[,1],comAPL[,2])

##   
## Welch Two Sample t-test  
##   
## data: comAPL[, 1] and comAPL[, 2]  
## t = 1.9101, df = 1997.8, p-value = 0.05626  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.001303983 0.098957048  
## sample estimates:  
## mean of x mean of y   
## 2.428648 2.379821

# t-test on diameter   
t.test(comD[,1],comD[,2])

##   
## Welch Two Sample t-test  
##   
## data: comD[, 1] and comD[, 2]  
## t = -3.0852, df = 1989.8, p-value = 0.002062  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.11199285 -0.02494607  
## sample estimates:  
## mean of x mean of y   
## 1.690623 1.759093

# t-test on cluster coefficient   
t.test(comCC[,1],comCC[,2])

##   
## Welch Two Sample t-test  
##   
## data: comCC[, 1] and comCC[, 2]  
## t = 2.2284, df = 1965.8, p-value = 0.02596  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.001570774 0.024622361  
## sample estimates:  
## mean of x mean of y   
## 0.4984709 0.4853743

### Appendix

##### Statistical testing on association network:

* "Second, we examined whether differences between the parameters of two association networks were statistically significant by applying the bootstrap method (Efron, 1979) to simulate partial random External and Internal networks and compared these networks (Kenett et al., 2013; Kenett, Anaki, et al., 2014). This procedure had a twofold rationale:
  + if the two networks truly differ from each other, then any sub-network consisting of the same nodes in both networks should also be different; and
  + the bootstrap method enables the generation of many simulated partial External and External networks, allowing for statistical examination of the difference between the two networks.
* In order to conduct the bootstrapping procedure, half of the cue words (nodes) were randomly chosen. Then partial external and internal networks were constructed separately using these random nodes, based on the entire sample. This method is known as the without replacement bootstrap (Bertail, 1997; Politis & Romano, 1994; Shao, 2003). Finally, for each partial AS and MC network, CC, ASPL,and D measures were computed. This procedure was simulated with 1000 realizations."

###### Terms:

[Degree](http://mathinsight.org/degree_distribution)  
[Unweighted Adjacency Matrix](http://mathinsight.org/definition/adjacency_matrix)  
[Small Word Network](http://mathinsight.org/small_world_network): Featured by large clustering coefficient and small average path length  
[Transitivity](http://mathinsight.org/evidence_additional_structure_real_networks#transitivity)  
[Transitivity math](http://mathinsight.org/definition/transitivity_graph)  
[Clustering Coefficient](http://mathinsight.org/definition/clustering_coefficient)  
[Free Will Scale Scoring Document](http://agencyandresponsibility.typepad.com/files/competing-free-will-and-responsibility-scales.pdf)  
[Minimum Spanning Tree Tutorial](http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf)  
[Average Path Length](http://mathinsight.org/definition/network_mean_path_length)