Psychological & Cognitive Networks: Introvert vs Extrovert

Submitted By:
Waseem Salami
Nadia Sirhan

Intro: what

Informative Background

- The MBTI personality test distinguishes and classifies personalities based on four dichotomous dimensions:
  - Introversion (I) / Extroversion (E)
  1. Intuition (N) / Sensing (S)
  2. Thinking (T) / Feeling (F)
  3. Judging (J) / Perceiving (P)
- The two personality types we chose to focus on are:
  - **ENFJ** – **Extroverted** Intuitive Feeling Judging
  - ❑ **INFJ** – **Introverted** Intuitive Feeling Judging

- It can be seen that these two personality types differ only in the first dimension, **Introversion (I) / Extroversion (E).** The reason we chose to compare these two types is that we are interested in examining the difference between semantic networks of introverted individuals compared to extroverted individuals. We are also interested in understanding the impact of introversion/extroversion on our personalities. We thought that if we take two individuals whose personalities are "similar" in every dimension except for the dimension of introversion/extroversion, we can more precisely separate and compare them, and also compare our results with the research findings based on this personality test.

Intro: what

- **Objective and Research Questions:**

Investigating the semantic structure of these two different personality subreddits raises several interesting research questions:

- Primary Research Question:

How do the semantic networks of individuals with introverted personalities differ from the networks of individuals with extroverted personalities?

- Secondary Research Questions:
- How do the semantic networks of individuals with INFJ personalities differ from the networks of individuals with ENFJ personalities?
- Does personality type influence the organization of the semantic network?

In all questions, it is also possible to incorporate a check for correlation with the findings of the MBTI research.

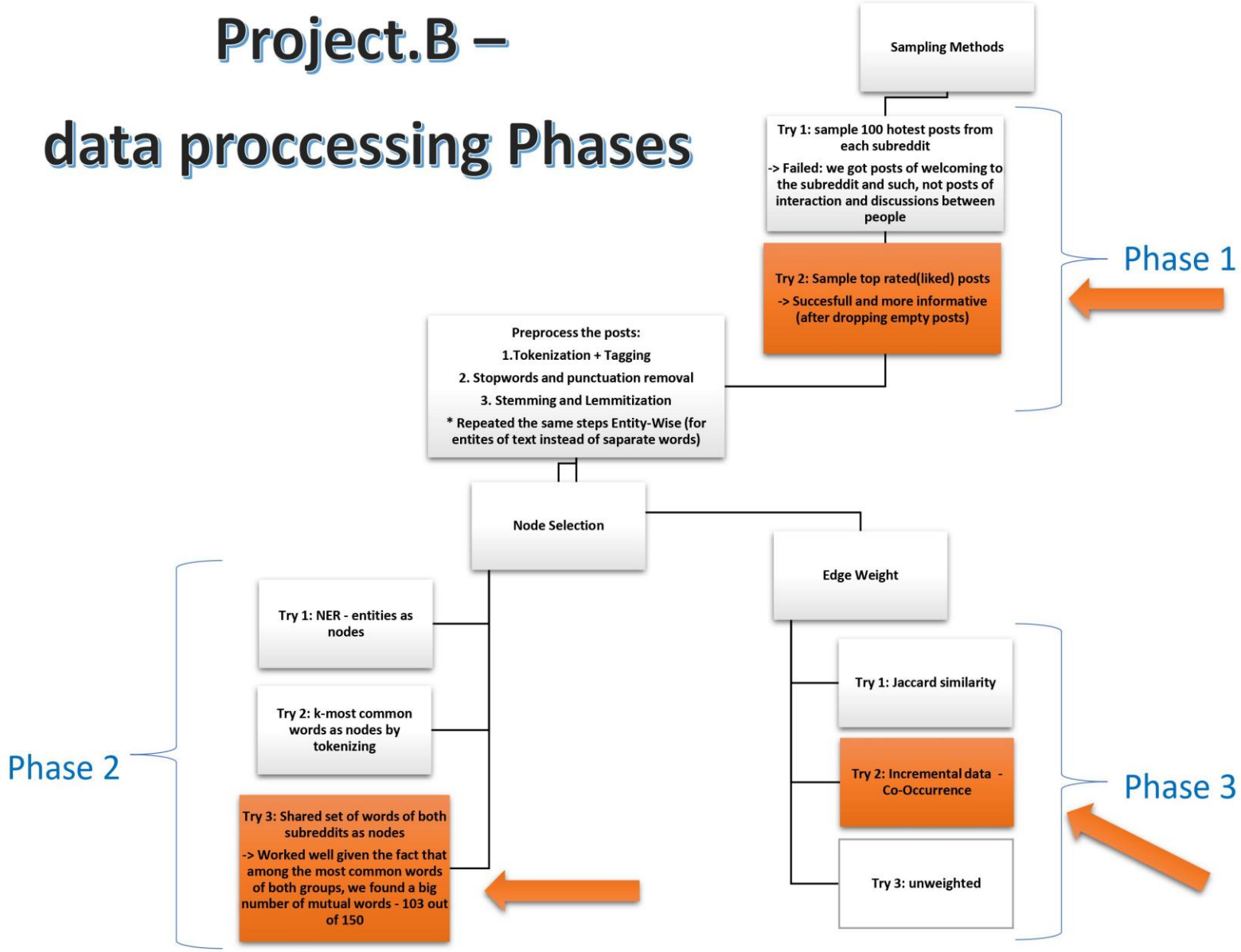
Intro: why?

- **<u>Why analyze this phenomenon in networks?</u>**

1. Visualization: Network analysis provides a visual representation of the structure and relationships within semantic networks.
2. Quantitative measures: It offers quantitative metrics such as centrality or network density to understand the importance and connections of concepts.
3. Community detection: Network analysis can identify clusters of related terms within the networks.
4. Path analysis: It allows exploring paths between terms, illustrating how ideas are interconnected.

- In summary, network analysis offers a unique combination of visualization, quantitative measures, and dynamic analysis, making it a powerful tool for investigating and comparing semantic networks.

Intro: how?


Project.B –
data proccessing Phases
Sampling Methods
Try 1: sample 100 hotest posts from each subreddit
-> Failed: we got posts of welcoming to the subreddit and such, not posts of interaction and discussions between people
Try 2: Sample top rated(liked) posts
-> Succesfull and more informative (after dropping empty posts)
Phase 1
Preprocess the posts:
1.Tokenization + Tagging
2. Stopwords and punctuation removal
3. Stemming and Lemmitization
* Repeated the same steps Entity-Wise (for entites of text instead of saparate words)
Node Selection
Edge Weight
Try 1: NER - entities as nodes
Try 2: k-most common words as nodes by tokenizing
Try 3: Shared set of words of both subreddits as nodes
-> Worked well given the fact that among the most common words of both groups, we found a big number of mutual words - 103 out of 150
Phase 2
Try 1: Jaccard similarity
Try 2: Incremental data - Co-Occurrence
Try 3: unweighted
Phase 3

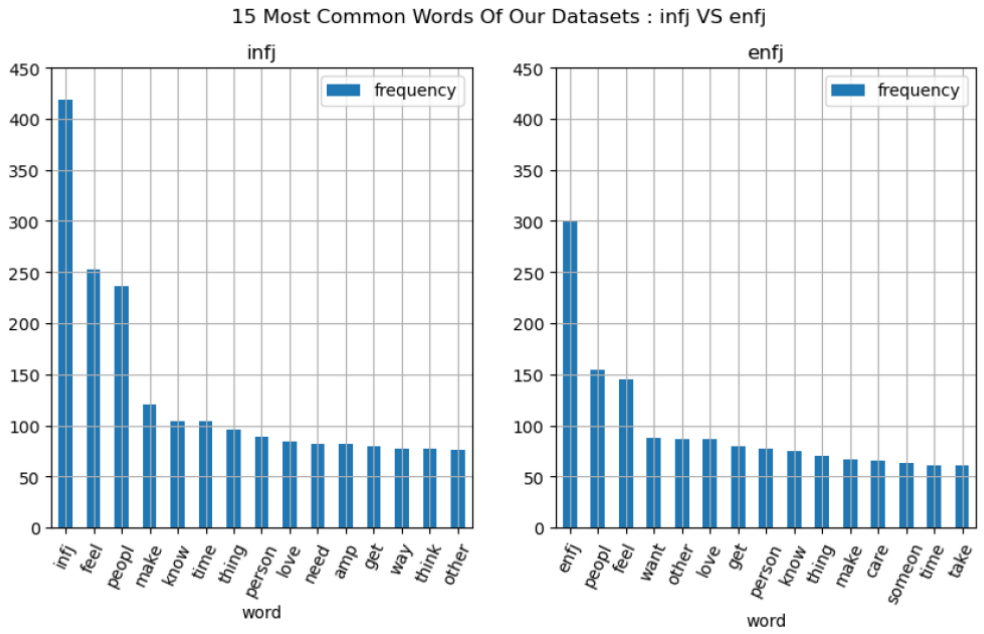
Creating the semantic networks:

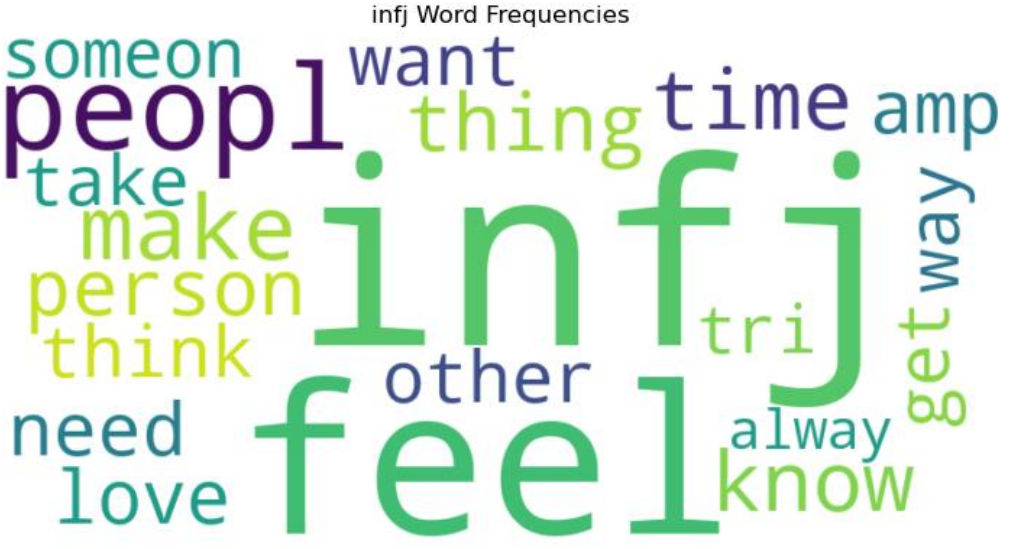
- We created our semantic network as follows:

Nodes: **77 mutual words**
after exploring the retrieved Reddit text from both subreddits, Interestingly enough - we got that out of 100 most common words of each subreddit, 77 were mutual! so we saw this as an opportunity to choose this set of words to be our nodes in order to examine the differences in the structure of the semantic networks which we will make based on these nodes.

- Edges: edges in the graph are in the form of :
  (word1, word2, Co-Occourence of word1 and word2)

meaning that we defined weight of two words to be the Co-Occourence(number of times both words appeared in the same entity)

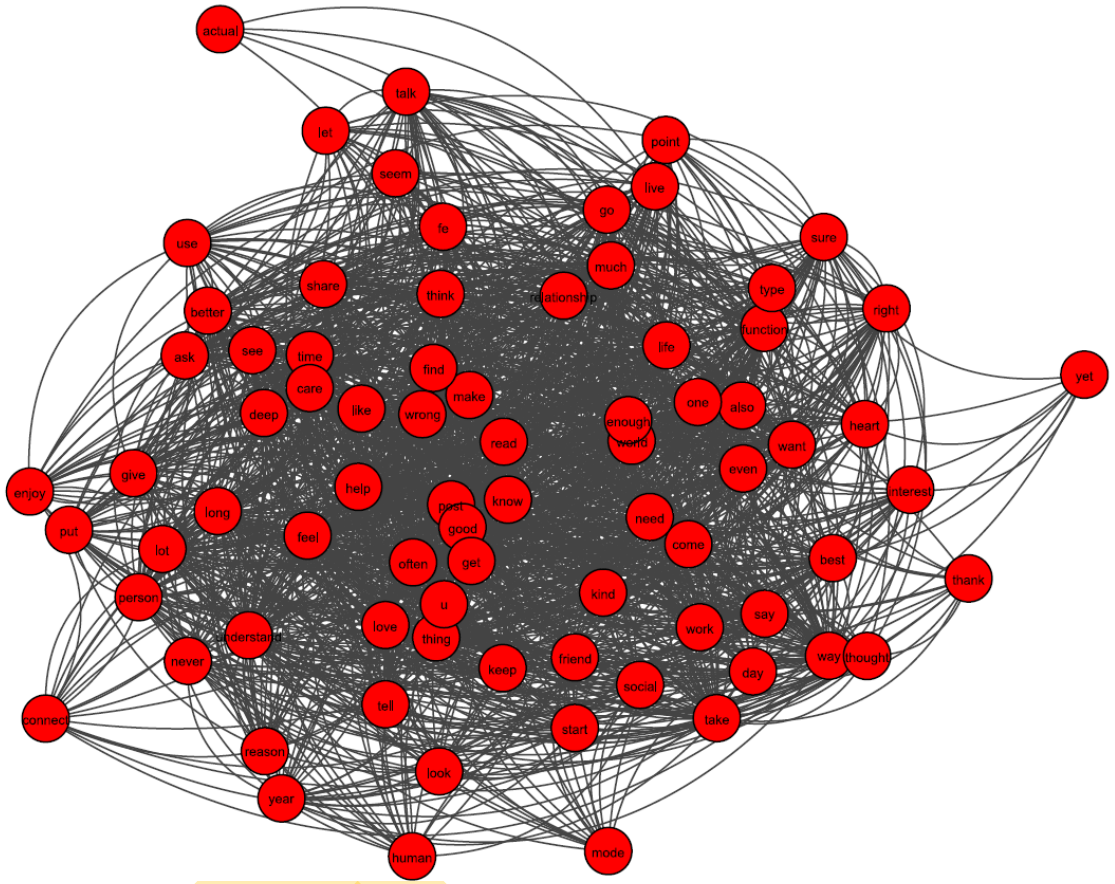

15 Most Common Words Of Our Datasets : infj VS enfj
infj
enfj
frequency
word
infj feel peopl make know time thing person love need amp get way think other
enfj peopl feel want other love get person know thing make care someon time take


infj Word Frequencies
someon
want
peopl
thing
time
amp
take
infj
make
person
tri
get
way
think
other
need
feel
alway
know
love


enfj Word Frequencies
everyon
get
feel
other
make
life
enfj
love
need
tri
know
care
friend
person
thing
peopl
want
take
someon
time

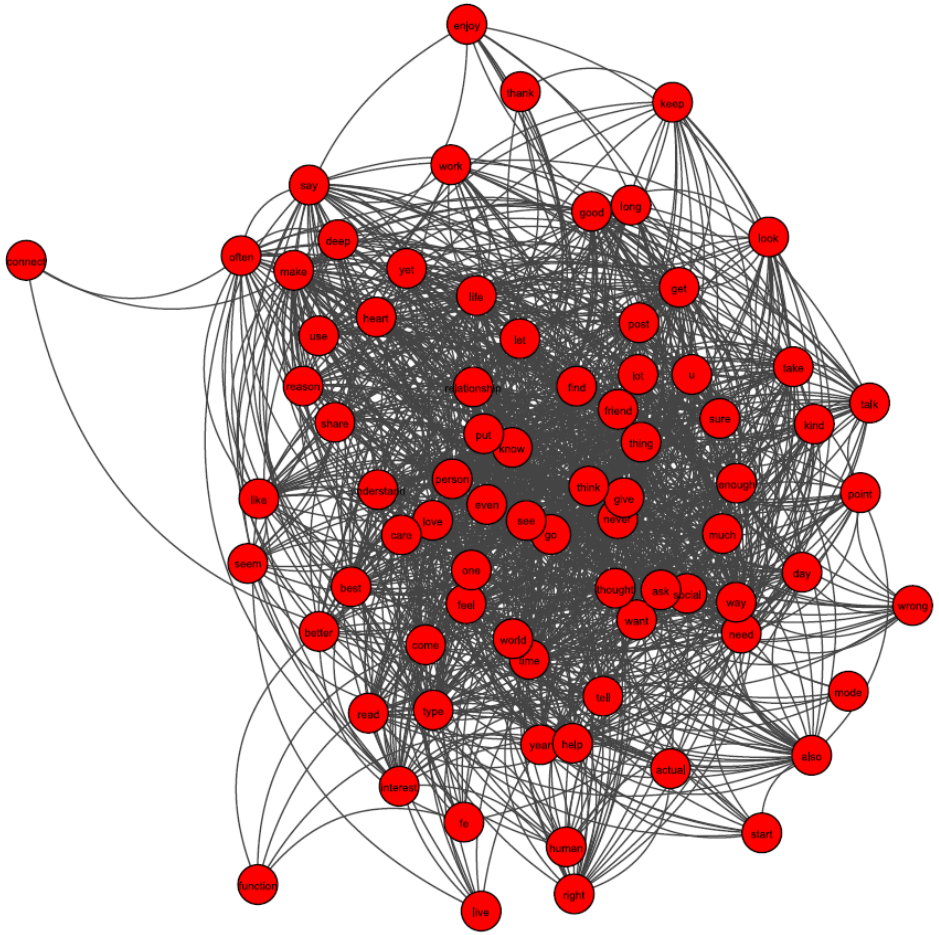
infj semantic network

```
number of nodes in the infj graph: 77
number of edges in the infj graph: 1504
```



enfj semantic network

```
number of nodes in the enfj graph: 77
number of edges in the enfj graph: 1207
```



Statistics on our semantic networks

```
biggest words pair Co_Occourence: infj vs enfj:   26    23
median infj words pair Co_Occourence:             2.0   1.0
smallest infj words pair Co_Occourence:           1     1
```

```
infj_Graph.edges[('feel', 'love')]
```

```
{'weight': 7}
```

```
enfj_Graph.edges[('feel', 'love')]
```

```
{'weight': 3}
```

```
number of word pairs formed in the infj network:   1535
```

```
an example of the first 10 pairs of infj Co-Occourance pairs:
```

```
[(('long', 'time'), 15),
 (('long', 'work'), 6),
 (('long', 'post'), 2),
 (('time', 'work'), 10),
 (('post', 'time'), 10),
 (('post', 'work'), 1),
 (('know', 'life'), 9),
 (('know', 'world'), 5),
 (('life', 'world'), 11),
 (('good', 'know'), 2)]
```

```
number of word pairs formed in the enfj network:   1218
```

```
an example of the first 10 pairs of enfj Co-Occourance pairs:
```

```
[(('go', 'thing'), 3),
 (('thing', 'time'), 10),
 (('never', 'thing'), 5),
 (('let', 'thing'), 2),
 (('thing', 'want'), 9),
 (('go', 'time'), 5),
 (('go', 'never'), 3),
 (('go', 'let'), 1),
 (('go', 'want'), 1),
 (('never', 'time'), 6)]
```

t-statistic of the betweness centrality:

infj vs enfj: -1.292254

p-value of the betweness centrality:

infj vs enfj: 0.198230

Correlation of the betweness centrality:

infj vs enfj: 0.806170

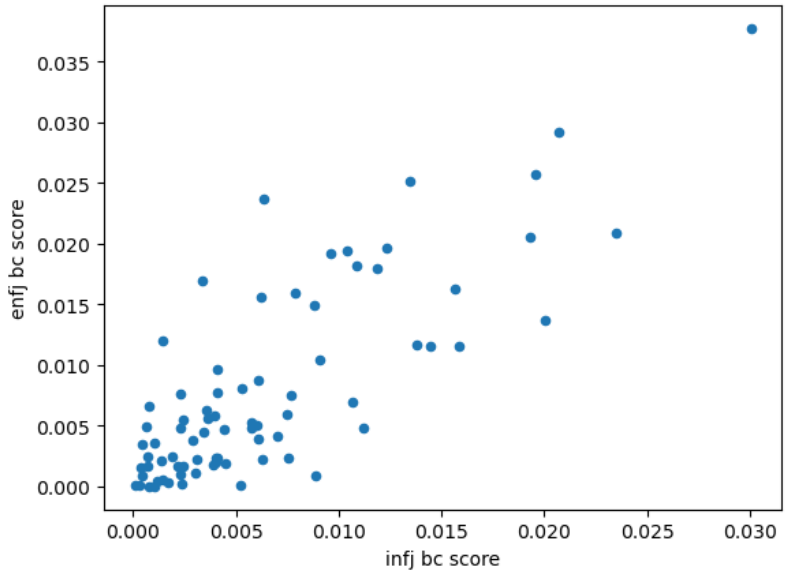
```
top 10 enfj nodes sorted by highest betweenness centrality

Key: feel, Value: 0.037753630725573466
Key: time, Value: 0.029177418892781234
Key: want, Value: 0.025763568516201308
Key: come, Value: 0.02516832130148998
Key: person, Value: 0.023689873662393453
Key: thing, Value: 0.020896489787817323
Key: make, Value: 0.020607754976916647
Key: much, Value: 0.019692141799315777
Key: love, Value: 0.019376109483794368
Key: need, Value: 0.019172764527140392
```

```
top 10 infj nodes sorted by highest betweenness centrality

Key: feel, Value: 0.030040340754029407
Key: thing, Value: 0.023457997177415396
Key: time, Value: 0.020705607661215767
Key: take, Value: 0.02000524435602027
Key: want, Value: 0.019590787883291313
Key: make, Value: 0.019305966195317246
Key: way, Value: 0.015830127581873236
Key: know, Value: 0.0156545742450279
Key: think, Value: 0.014470061894170185
Key: often, Value: 0.013816682768711943
```

	node	infj bc score	enfj bc score
0	feel	0.030040	0.037754
1	thing	0.023458	0.020896
2	time	0.020706	0.029177
3	take	0.020005	0.013654
4	want	0.019591	0.025764
...	...	...	...
72	human	0.000471	0.000906
73	reason	0.000421	0.003473
74	yet	0.000362	0.001561
75	thank	0.000340	0.000095
76	actual	0.000090	0.000093




```
infj's most central edge: ('thing', 'actual')
enfj's most central edge: ('connect', 'come')
```

Analyzing the semantic networks

Network Measurements:

betweenness centrality:

t-statistic of the eigenvector centrality:
infj vs enfj: 0.452380

p-value of the eigenvector centrality:
infj vs enfj: 0.651639

Correlation of the eigenvector centrality:
infj vs enfj: 0.716519

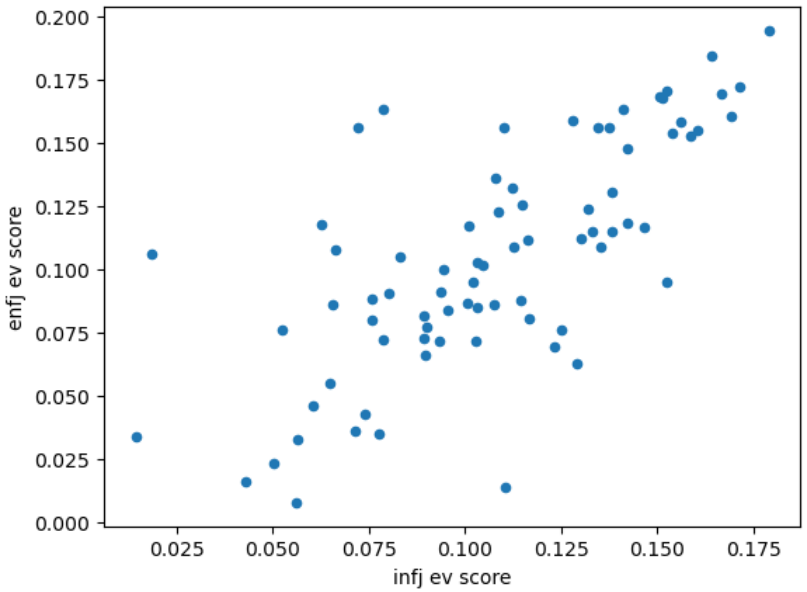
top 10 enfj nodes sorted by highest betweenness centrality

Key: feel, Value: 0.037753630725573466
Key: time, Value: 0.029177418892781234
Key: want, Value: 0.025763568516201308
Key: come, Value: 0.02516832130148998
Key: person, Value: 0.023689873662393453
Key: thing, Value: 0.020896489787817323
Key: make, Value: 0.020607754976916647
Key: much, Value: 0.019692141799315777
Key: love, Value: 0.019376109483794368
Key: need, Value: 0.019172764527140392

top 10 infj nodes sorted by highest betweenness centrality

Key: feel, Value: 0.030040340754029407
Key: thing, Value: 0.023457997177415396
Key: time, Value: 0.020705607661215767
Key: take, Value: 0.02000524435602027
Key: want, Value: 0.019590787883291313
Key: make, Value: 0.019305966195317246
Key: way, Value: 0.015830127581873236
Key: know, Value: 0.0156545742450279
Key: think, Value: 0.014470061894170185
Key: often, Value: 0.013816682768711943

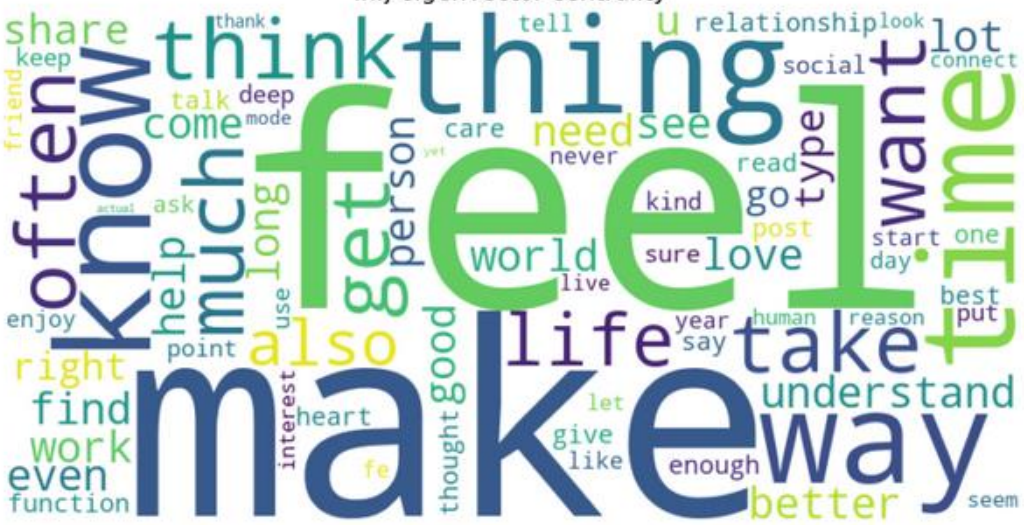
	node	infj ev score	enfj ev score
0	feel	0.178928	0.194393
1	thing	0.169297	0.160576
2	time	0.164011	0.184239
3	take	0.153789	0.153595
4	want	0.152318	0.170720
...	...	...	...
72	human	0.065060	0.054853
73	reason	0.066405	0.107505
74	yet	0.018559	0.106294
75	thank	0.043029	0.016396
76	actual	0.014439	0.033893



Analyzing the semantic networks
Network Measurements:

eigenvector centrality:

infj eigenvector centrality



infj betweenness centrality

enfj eigenvector centrality

enfj betweenness centrality

Community detection

- Tried several algorithms, mainly the **Girvan Newman Algorithm** and the **Louvain Algorithm,** the algorithm that obtained best Modularity score was the Louvain Algorithm, so we chose it for community detection.

```
Louvain's chosen 5 enfj communities:


community number 1:

{'start', 'enjoy', 'need', 'think', 'give', 'thought', 'find', 'long', 'also', 'human', 'help', 'time', 'love', 'thing', 'l
ook'}


community number 2:

{'one', 'best', 'often', 'kind', 'read', 'say', 'day', 'take', 'friend', 'talk', 'go', 'let', 'mode', 'like', 'heart'}


community number 3:

{'point', 'put', 'interest', 'yet', 'sure', 'even', 'share', 'reason', 'u', 'feel', 'type', 'never', 'understand', 'tell',
'lot', 'much', 'enough', 'year', 'make'}


community number 4:

{'fe', 'connect', 'live', 'better', 'deep', 'world', 'care', 'come', 'function', 'way', 'life', 'work'}


community number 5:

{'ask', 'post', 'get', 'good', 'relationship', 'person', 'actual', 'right', 'thank', 'seem', 'see', 'know', 'keep', 'want',
'social', 'use'}
```

```
infj Community Modularity: Louvian vs GirvanNewman 0.09112579110237094 0.012273022442903481

enfj Community Modularity: Louvian vs GirvanNewman  0.13504201738077876 0.005917112891342587
```

Louvain's chosen 5 infj communities:

community number 1:

{'point', 'put', 'enjoy', 'kind', 'think', 'interest', 'yet', 'share', 'live', 'thought', 'feel', 'type', 'way', 'life', 'connect', 'world', 'come', 'like', 'post', 'heart'}

community number 2:

{'good', 'person', 'tell', 'actual', 'enough', 'right', 'thing'}

community number 3:

{'start', 'understand', 'fe', 'sure', 'relationship', 'human', 'better', 'deep', 'social', 'help', 'function', 'reason', 'friend', 'mode', 'never', 'year', 'make', 'use'}

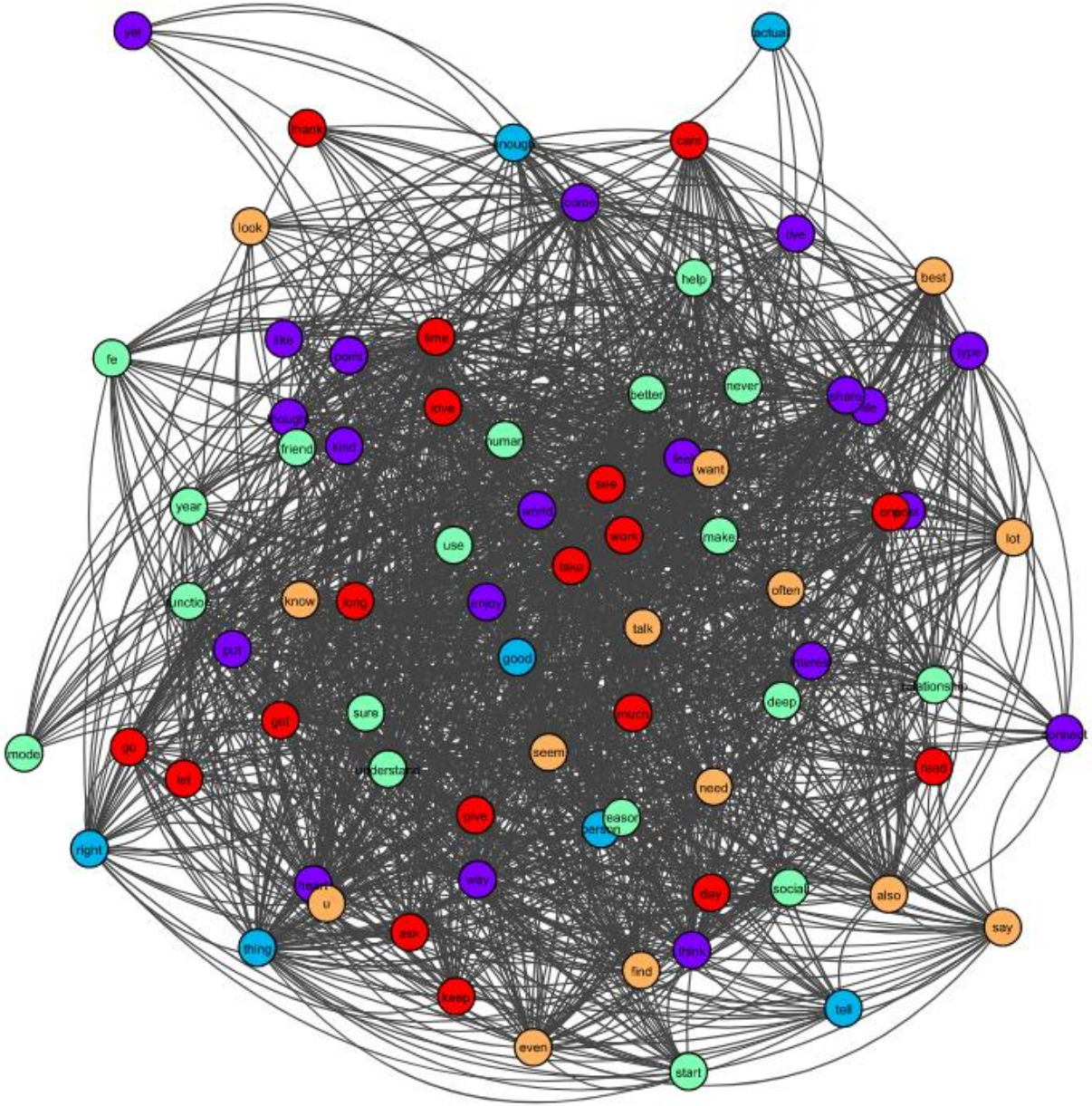
community number 4:

{'best', 'often', 'need', 'also', 'know', 'even', 'want', 'lot', 'find', 'u', 'talk', 'say', 'look', 'seem'}

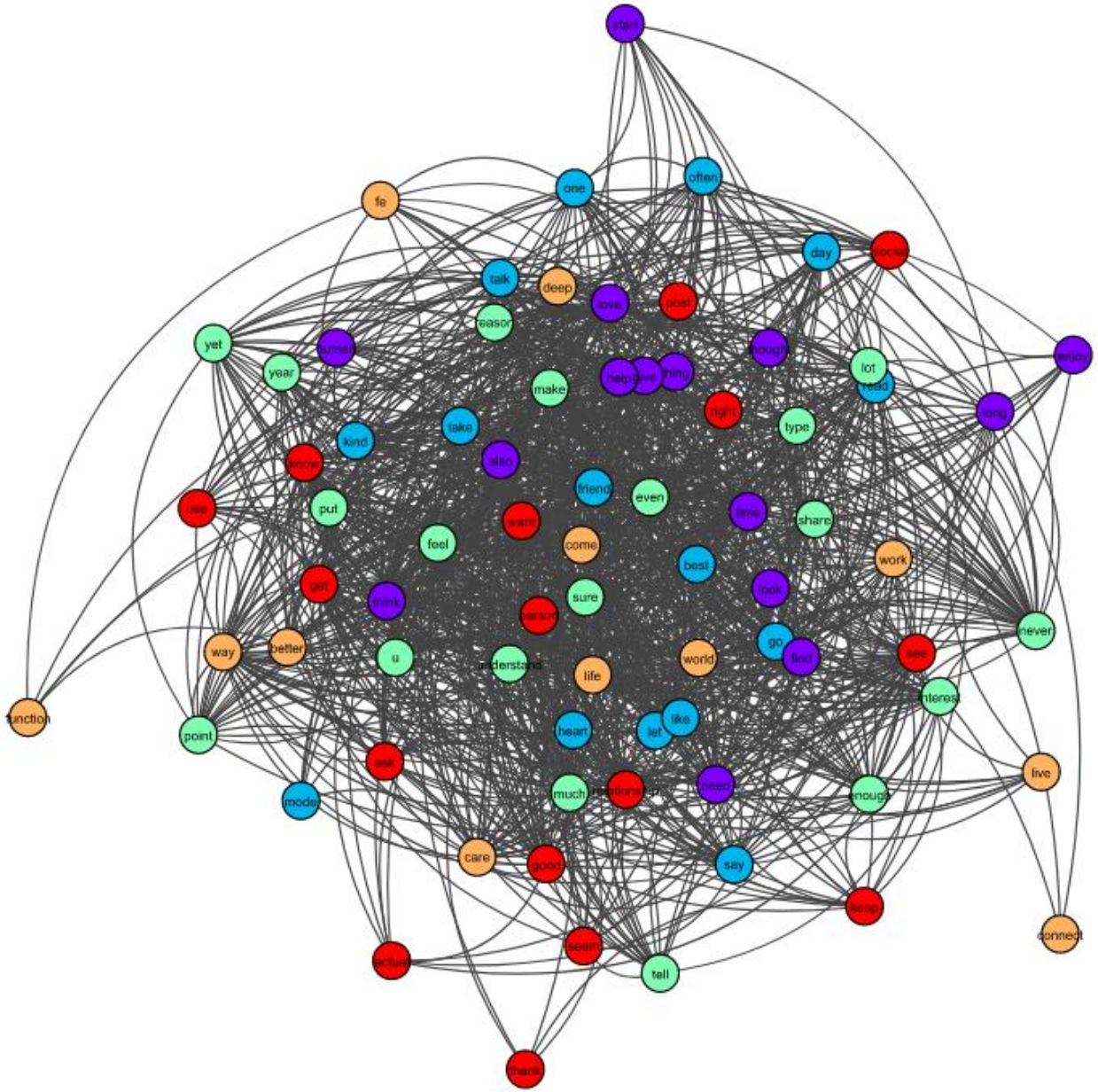
community number 5:

{'one', 'ask', 'get', 'read', 'give', 'day', 'care', 'take', 'long', 'thank', 'see', 'keep', 'much', 'time', 'love', 'go', 'let', 'work'}

infj Communities



enfj Communities



Conclusions

For both of our groups: infj which represents the introverts and enfj which represents the extroverts, we tried several approaches to compare their semantic networks, we compared their most common words, we compared their metrics such as the betweenness and eigenvector centrality both in word-to-word score comparison and in t-tests and statistics, we also tried to detect communities within each group and compare them

-> from all these comparisons we inferred that **we reject our null hypothesis**: introverts and extroverts have relatively-**similar** semantic networks, which in our opinion is also an interesting direction to test.

Future Steps

- To gain more insights, we can conduct the following steps:
- **Time-based Analysis**: we can conduct a temporal analysis. Where we look at how the semantic networks of introverts and extroverts evolve over time. This could reveal interesting dynamic features in the data.
- **Node-level Analysis**: we can do a more detailed analysis at the node level. Which nodes (words) are more central or peripheral in each group's network? How do these 'key' nodes differ between the groups? And other questions my pop-up.
- **Incorporate External Data**: we thought about bringing in external data related to the words in our network. This might include psychological scales or measures that relate to the words in our network, which could provide additional context or information, e.g. sentiment analysis.

Thank you for listening!

And thank you for an awesome course.