Project Title: Mnemonic Mesh: Exploring the intricacies of word connections

Akriti Kumari: akkumari@iu.edu

Abstract

This project delves into the intriguing phenomenon of word recall, exploring the challenges individuals face in remembering specific terms as contextual factors evolve. Our objective is to comprehend the mechanisms underlying word recollection, examining how cues to other words influence this process. Through this exploration, we seek to uncover patterns in human memory, shedding light on how individuals memorize and interconnect words.

Introduction

Within the realm of human cognition, how do we naturally establish connections between words, whether or not they share identical meanings but maintain some form of association? Often, this cognitive mapping occurs subconsciously, with individuals deducing or recalling words through indirect connections with others. The term "mnemonics mesh" suggests the presence of subtle prompts that aid in memorizing specific words, presenting a compelling avenue for delving into the intricacies of how people think and interrelate words.

The Mnemonic Mesh project extends beyond a mere exploration of word associations; it encompasses a comprehensive dataset that includes not only word associations, but also additional properties associated with these words. The project's scope transcends understanding association patterns alone; it integrates elements of network science, such as centralities and community detection, to unravel how different cues cluster together and identify which words garner the most usage among cues or recalled words.

This endeavor serves as an applied study with implications spanning psychology, linguistics, neuroscience, and network science. By addressing key questions, the project aims to enhance our comprehension of word transitions within the Mnemonic Mesh:

1. Can we discern the starting letters or length of the remembered word based on the initial or final letters or the sound of the cue word?
2. Is it feasible to identify additional properties of the recalled word based on the length of the first word?

By systematically investigating these questions, the project seeks to unravel the complexities of word connections, offering valuable insights into human cognition and the intricate web of associations that shape our understanding of language.

Proposal

In the "Mnemonic Mesh" project, I plan to deeply analyze the relationships between cued and recalled words, employing network science concepts like centrality measures and community detection to uncover and understand the underlying connections. By categorizing these words into distinct topics, sub-topics, or parts of speech and examining their interconnections, we aim to elucidate the broader network of associations that facilitate word recall and cognitive mapping. This comprehensive approach will enhance our understanding of how words are interconnected within the human memory system.

Reference

Nelson, McEvoy & Schreiber: **http://w3.usf.edu/FreeAssociation/AppendixA/index.html**

Acknowledgment

I extend our heartfelt thanks to Prof. Yong Yeol Ahn and all the associate teachers for their invaluable guidance and timely feedback throughout our project.

Final Project content

Introduction

Words represent one of the most crucial elements of human interaction, and it's intriguing to observe which words individuals recall in response to specific cues. Typically, words that evoke personal memories or have contextual relevance are more readily remembered. In this study, I aim to explore how people recall words when prompted by other words and to identify patterns in how words are memorized and connected.

The primary objective of this research is to enhance our understanding of the human cognitive process, particularly why certain words are remembered in response to specific cue words within a given context. The question driving this investigation is: Can I analyze which words will be remembered in response to a cue word, and can I determine the characteristics of those words?

Although subjective by nature, this topic can be segmented into various sections to examine different factors influencing word associations. It has always been captivating to discover what words a person recalls in diverse scenarios, and numerous models currently explore this phenomenon. My goal is to contribute to breaking down these complexities by framing these challenges within the context of network science.

In the "Mnemonic Mesh" project, the dataset not only includes word associations but also other properties of the cues and target words. Details such as the number of people responding to a cue, and the forward and backward strength between cue and target words offer valuable insights into word relationships. I have employed specific filters to focus on a more manageable subset of the data: excluding records with no backward strength (indicating the target word does not impact the cue word), requiring at least three responses per cue word, and limiting the parts of speech to nouns, adjectives, and verbs. Only 'normed' words, as defined by the dataset, are considered valid for analysis. Additionally, I have utilized parts of speech tags of both cue and target words to further explore their connections.

Due to the extensive array of cue-target terms, I have linked the words with their hypernyms to derive their topics and subtopics, which provides a broader perspective on their associations. This approach helps in understanding the intricate mesh of word connections within a cognitive framework.

Methods

I utilize various techniques to explore the connections between cue words and the corresponding words they evoke, as detailed below:

1. **Part-of-Speech Tag**
2. **Vowels Association**
3. **Subtopics Association**
4. **Topics Association**

Part-of-Speech Tag

In the dataset, both cue and target words are labeled with their respective Parts of Speech. I use these tags to construct a cue-target network, which I then visualize using Gephi software. During preprocessing, I observed that the most frequent part-of-speech tags were Noun (N), Verb (V), and Adjective (ADJ), while other tags were minimal and thus excluded from the dataset. Furthermore, the average Forward Strength between these tags is utilized as the weight for the network's edges, illustrating the influence of one tag type on another. I load the network as a directed graph in Gephi, employing attributes such as Node Size, and Edge Color.

Node Size: Ranked by Degree
Edge Color: Weight(FSG)

Vowel Association

I use the dataset and filter the cue words to include only those starting with vowels. This approach is adopted because of the dataset's extensive size, allowing me to focus specifically on certain word associations. After preprocessing, I not only analyze the cue-target words but also use the Forward Strength observed between them as the weights for the edges in the network, illustrating the influence of one word on another.

The network is loaded as a directed graph in Gephi and the following attributes are used: **Node Size:** Ranked by Betweenness centrality
Node Color: Community formation using Modularity
Edge Color: Weight(FSG)

Subtopics Association

Given the large volume of cue-target words, I identify the hypernyms of these words to serve as subtopics, offering a generalized view of the words. During the initial data preprocessing, I discovered that many records lacked associated subtopics, leading to their removal from the dataset. Moreover, I use the mean Forward Strength observed between these subtopics as the weight for the network edges, indicating the influence of subtopics on each other.

I load the network as a directed graph in Gephi, employing the following attributes:

Node Size: Ranked by Betweenness centrality
Node Color: Community formation using Modularity
Edge Color: Weight(FSG)

Topics Association

Due to the substantial count of cue-target words, I also retrieve and use the hypernyms of the subtopics, which represent the broader topics of the given words. Similar to the subtopics, during preprocessing, I found that many records did not have an associated topic, resulting in their exclusion from the dataset. The mean of the Forward Strength observed between these topics is taken as the weight for the edges in the network, illustrating the influence of topics on one another.

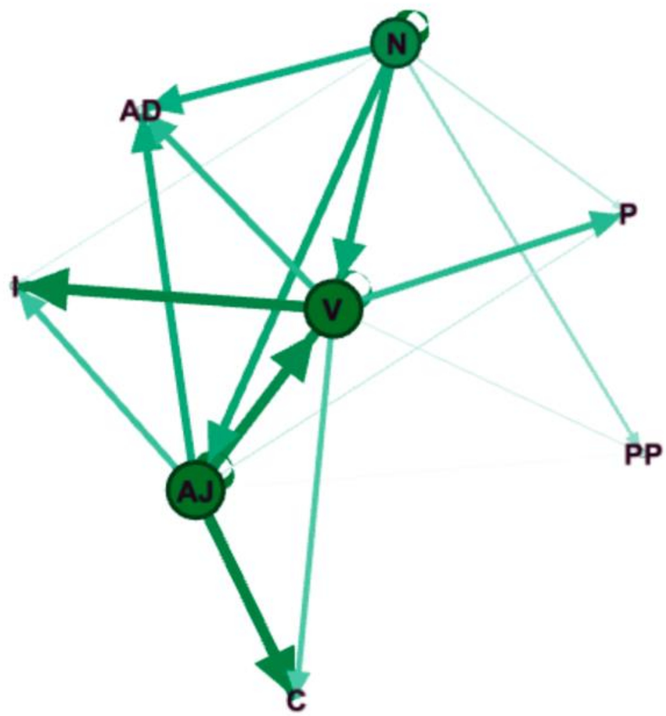
The network is set up as a directed graph in Gephi with the same visualization attributes: **Node Size:** Ranked by Betweenness centrality
Node Color: Community formation using Modularity
Edge Color: Weight(FSG)

Results

The four methodologies outlined above have been applied to gain deeper insights into the characteristics and selection patterns of the words, as detailed below:

Part-of-Speech Tag

The network visualized for the Part-of-Speech is shown below:



In the network analyzed, the following characteristics are noted:

- Nodes: 8; Edges: 20
- Average Degree: 2.5
- Average Weighted Degree: 0.124

From this network, we can derive the following insights:

- The most common Parts-of-Speech present are Nouns (N), Verbs (V), and Adjectives (AJ).
- Adverbs (AD) are the most frequently recalled words in response to any of the prominent Parts-of-Speech.
- Verbs tend to be recalled in response to Adjectives or Nouns, but not the reverse.
- Given that the dataset predominantly contains these prominent Parts-of-Speech, the insights might be somewhat skewed, suggesting that more balanced data could yield more accurate observations.

Vowels Association

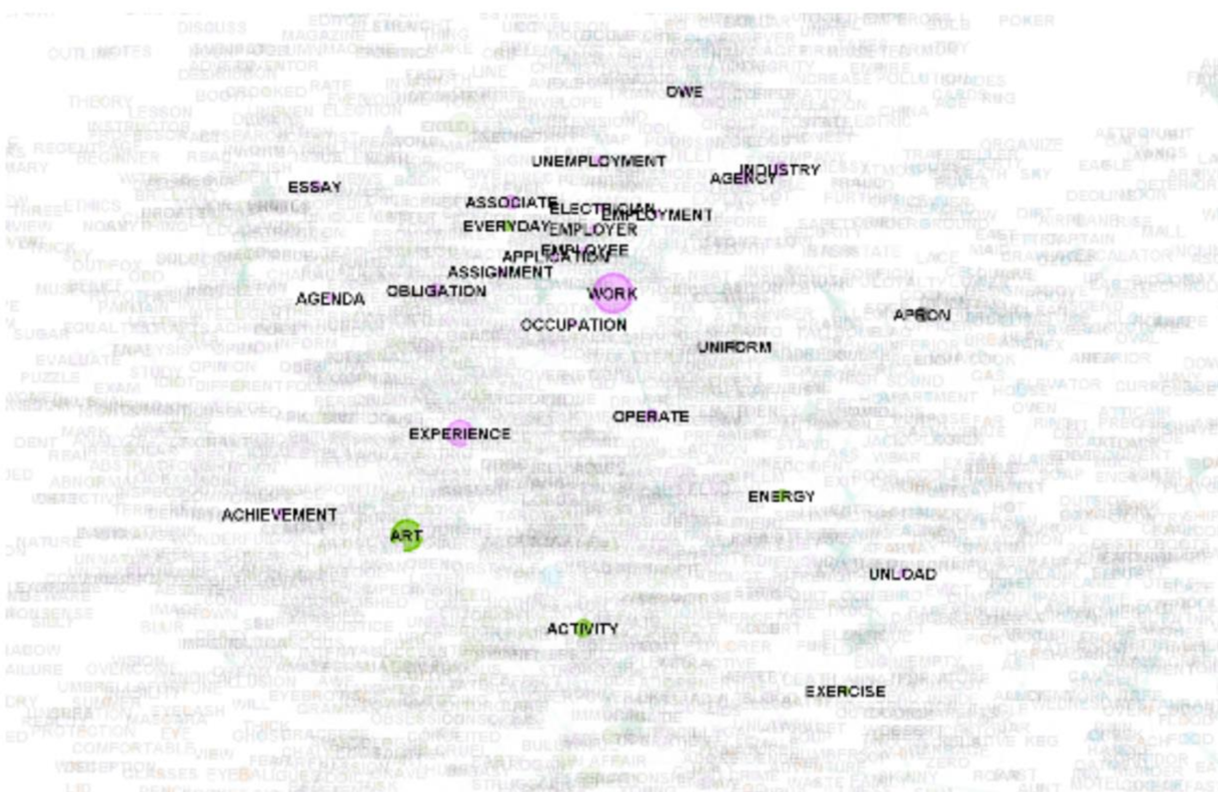
The network visualized for Vowels Association is shown below:

In the network under analysis, we observe the following characteristics:

- Nodes: 2126; Edges: 4208
- Average Degree: 1.979
- Average Weighted Degree: 0.122
- Average Clustering Coefficient: 0.014
- Average Path Length: 4.031
- Communities: 8 (Resolution=5.0)

This network is relatively sparse, indicating that few nodes have strong connections with their neighboring nodes. The larger nodes represent the betweenness centers, and through specific examples, we will delve deeper to uncover additional insights in this network.

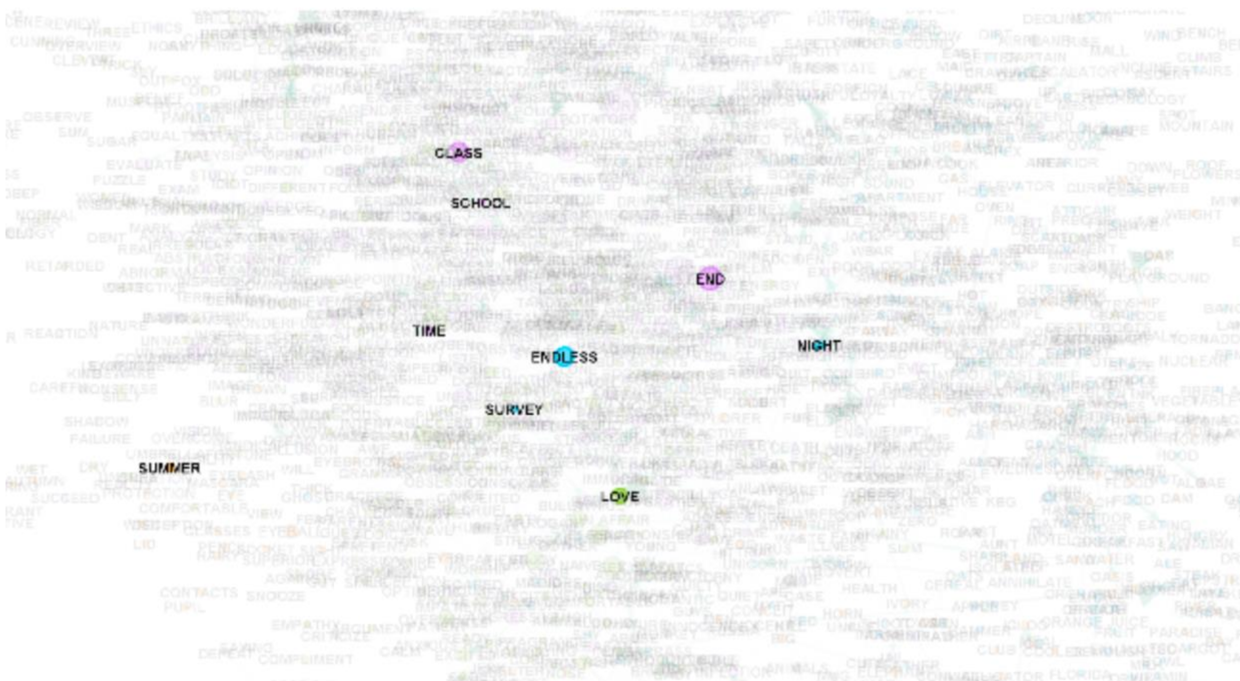
Vowel Association: Work



In the network, the node labeled "Work" has been examined with the following observations:

- It is a key betweenness center, showing strong connections across various communities.
- This node has a notably high number of neighbors.
- Interestingly, although the initial filter was set to include only cue words starting with vowels, the node "Work," which begins with a consonant, emerges as one of the major betweenness centers.
- This suggests that consonant-starting words might play a significant role in linking other words within the network, a hypothesis that could be further validated with additional examples.

Vowel Association: Endless

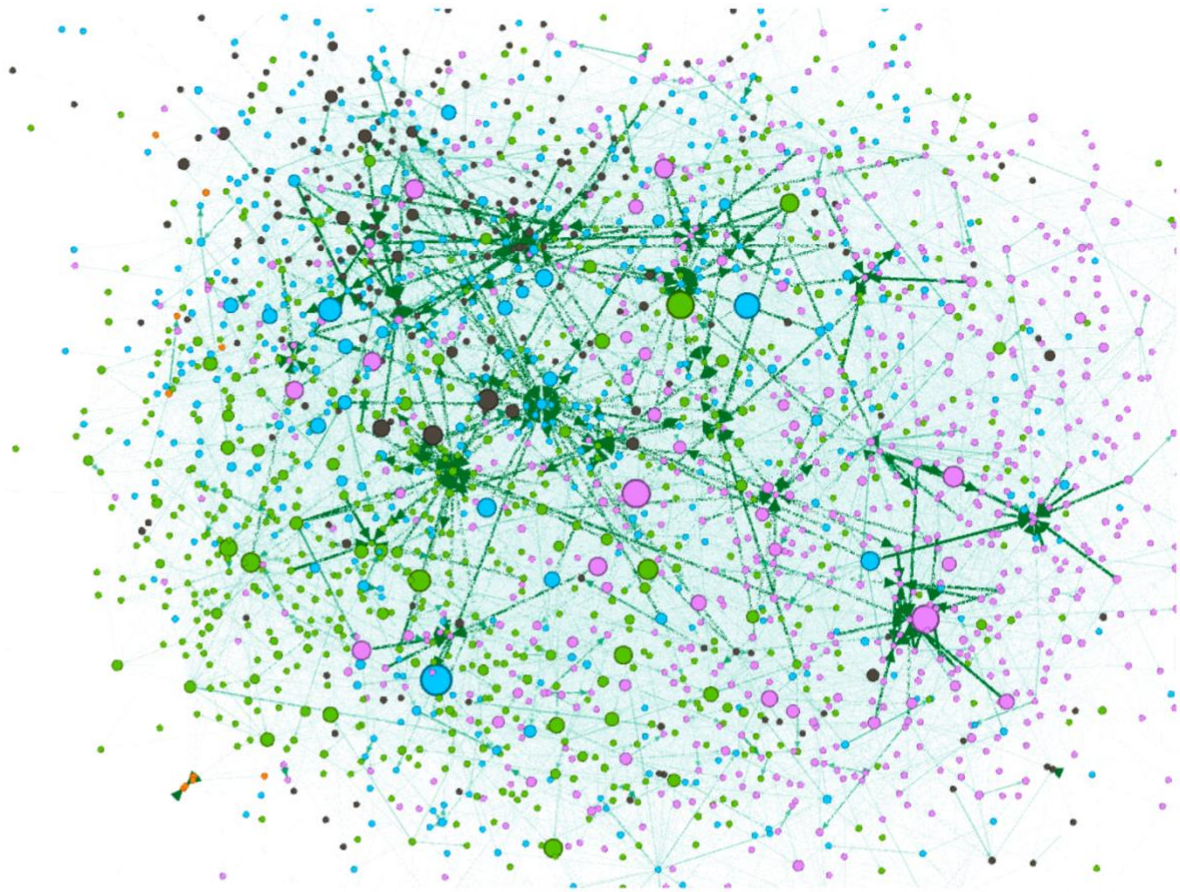


In the network, the node labeled "Endless" has been analyzed with the following observations:

- It is one of the smaller betweenness centers and exhibits limited connectivity across different communities.
- The number of neighbors this node has is relatively low.
- These observations might suggest that words beginning with vowels may not play a crucial role in connecting multiple words within the network.

Subtopics Association

The network visualized for the Subtopics Association is shown below:

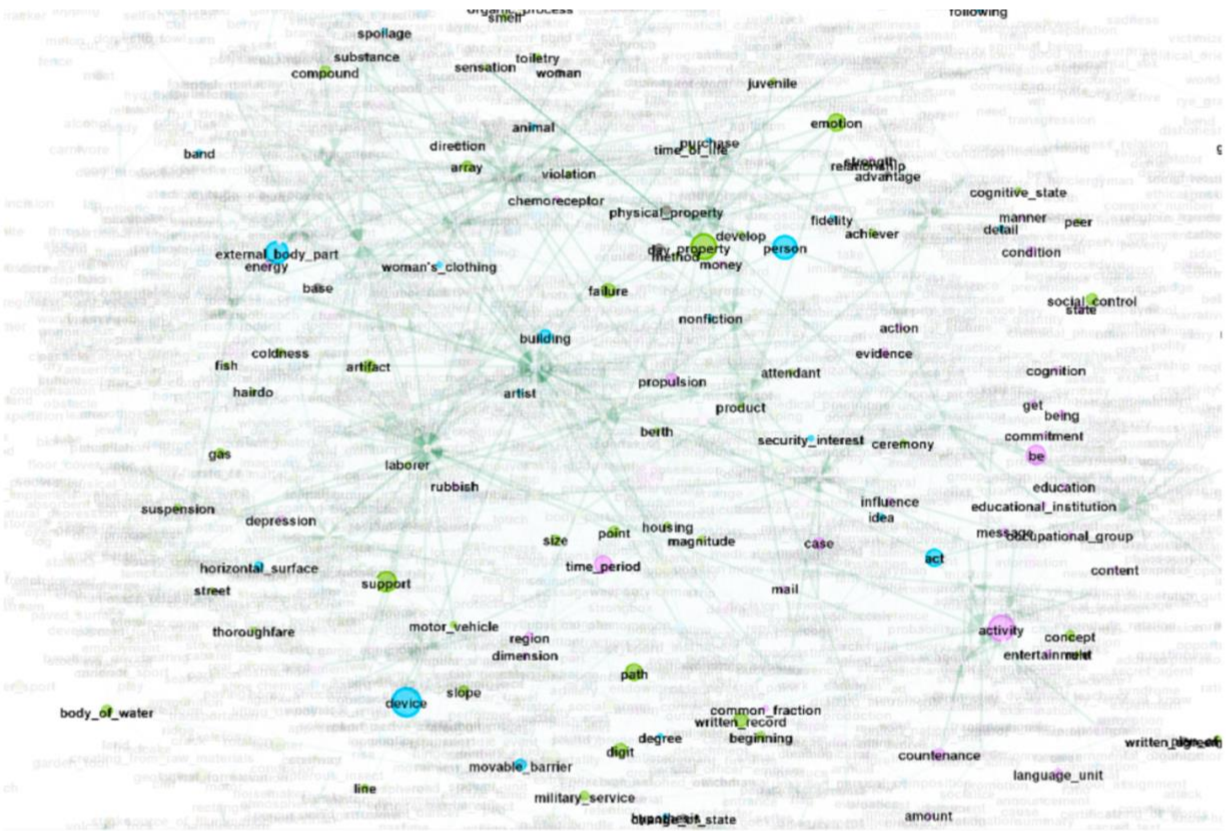


In the network, we observe the following characteristics:

- Nodes: 2154
- Edges: 17933
- Average Degree: 8.325
- Average Weighted Degree: 0.651
- Average Clustering Coefficient: 0.082
- Average Path Length: 3.505
- Communities: 12 (Resolution=2.0)

This network is notably dense, featuring many nodes with strong connections to their neighbors. Although the large nodes represent betweenness centers, it is particularly noteworthy that the nodes with strong associations are not necessarily the betweenness centers. We will examine specific examples to derive more insights from this network.

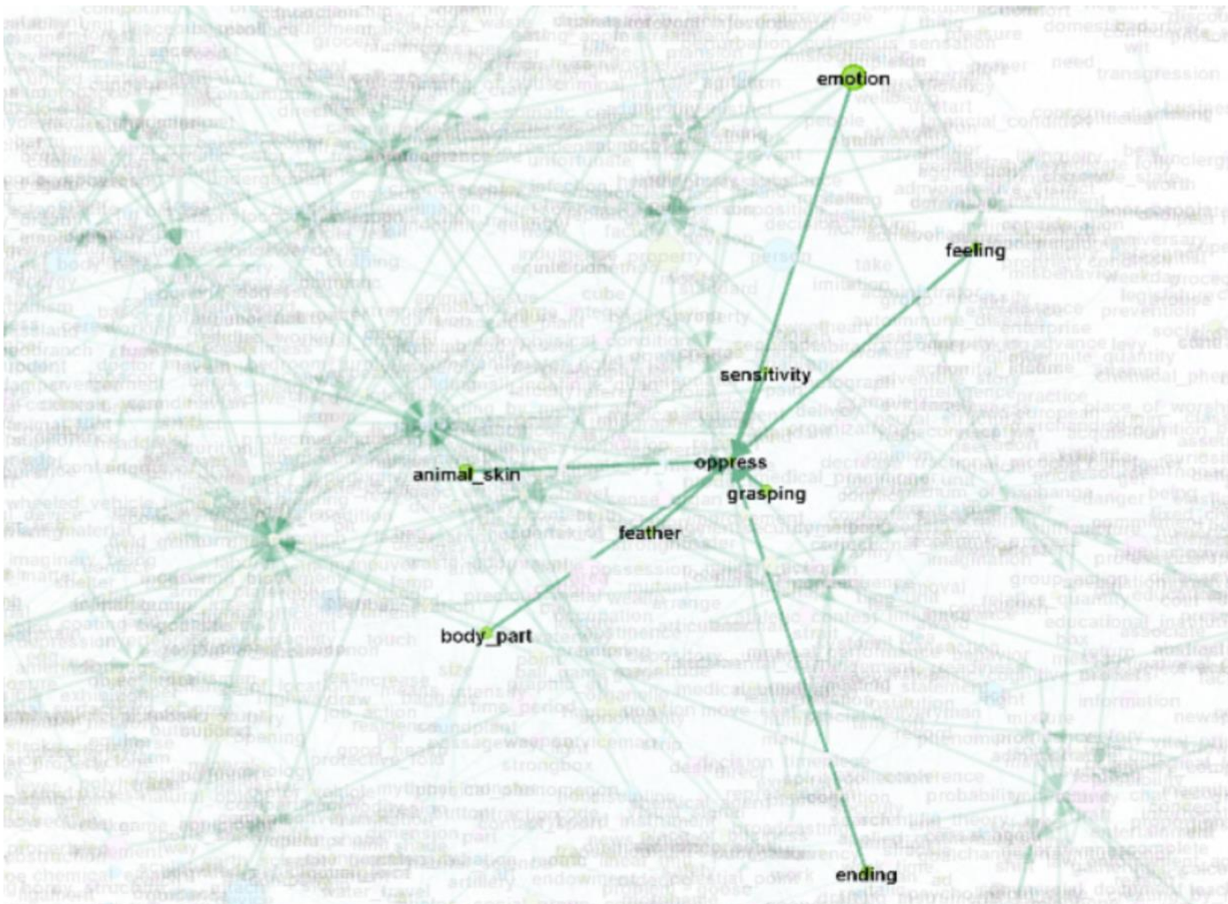
Sub-Topic: Property



In the network, the node labeled "Property" has been analyzed with the following observations:

- It is identified as one of the betweenness centers with robust connections spanning various communities.
- This node boasts a high number of neighbors.
- Notably, despite its extensive connectivity, the associations with its neighbors are relatively weak. This suggests that there may be underlying factors or hidden properties that could provide deeper insights with further investigation.

Sub-Topic: Oppress

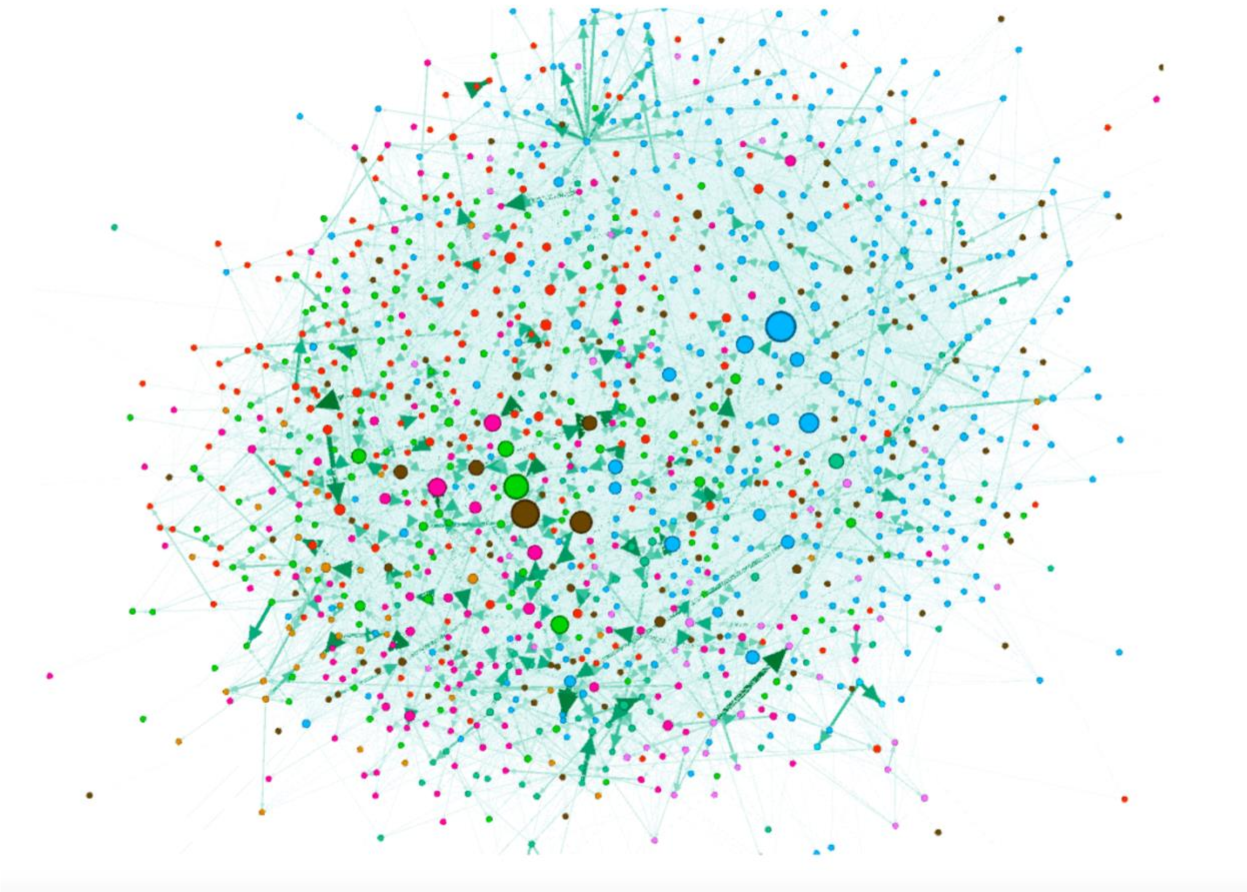


In the network, the node labeled "Oppress" has been examined with the following observations:

- It has strong associations with all of its neighbors within the network.
- The number of neighbors it has is relatively low.
- Interestingly, despite its limited number of neighbors, the relationships with them are robust. This suggests that there may be underlying attributes or specific connections that could reveal more about its associations upon deeper analysis.

Topics Association

The network visualized for the Topics Association is shown below:

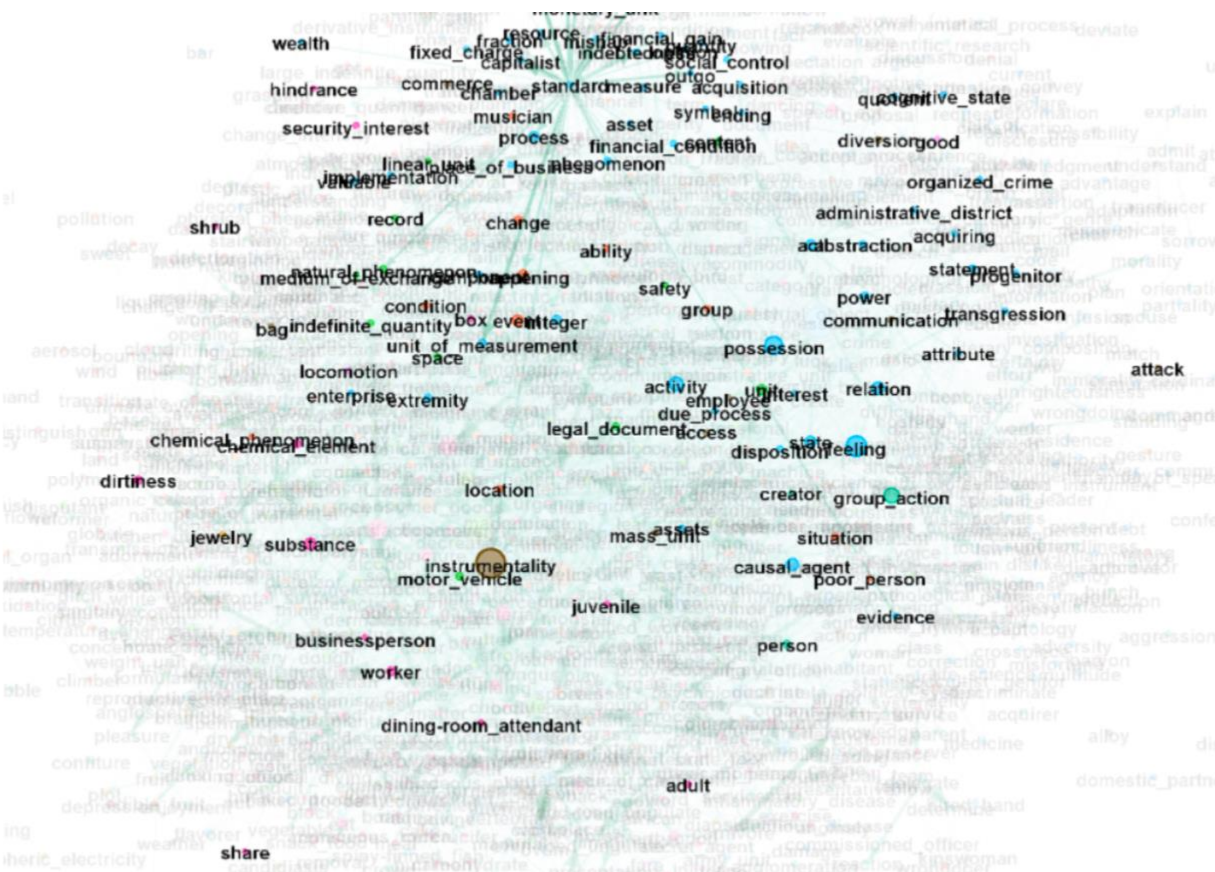


In the network, we can identify the following characteristics:

- Nodes: 1185
- Edges: 14951
- Average Degree: 12.617
- Average Weighted Degree: 0.795
- Average Clustering Coefficient: 0.155
- Average Path Length: 2.76
- Communities: 10 (Resolution=1.25)

This network is notably dense and features a variety of nodes, some with strong and others with weak associations with their neighbors. The larger nodes represent betweenness centers and are distributed across various communities. This arrangement suggests the possibility of a scale-free network structure, potentially exhibiting small-world or even ultra-small-world properties. We will delve into specific examples to glean further insights from this network.

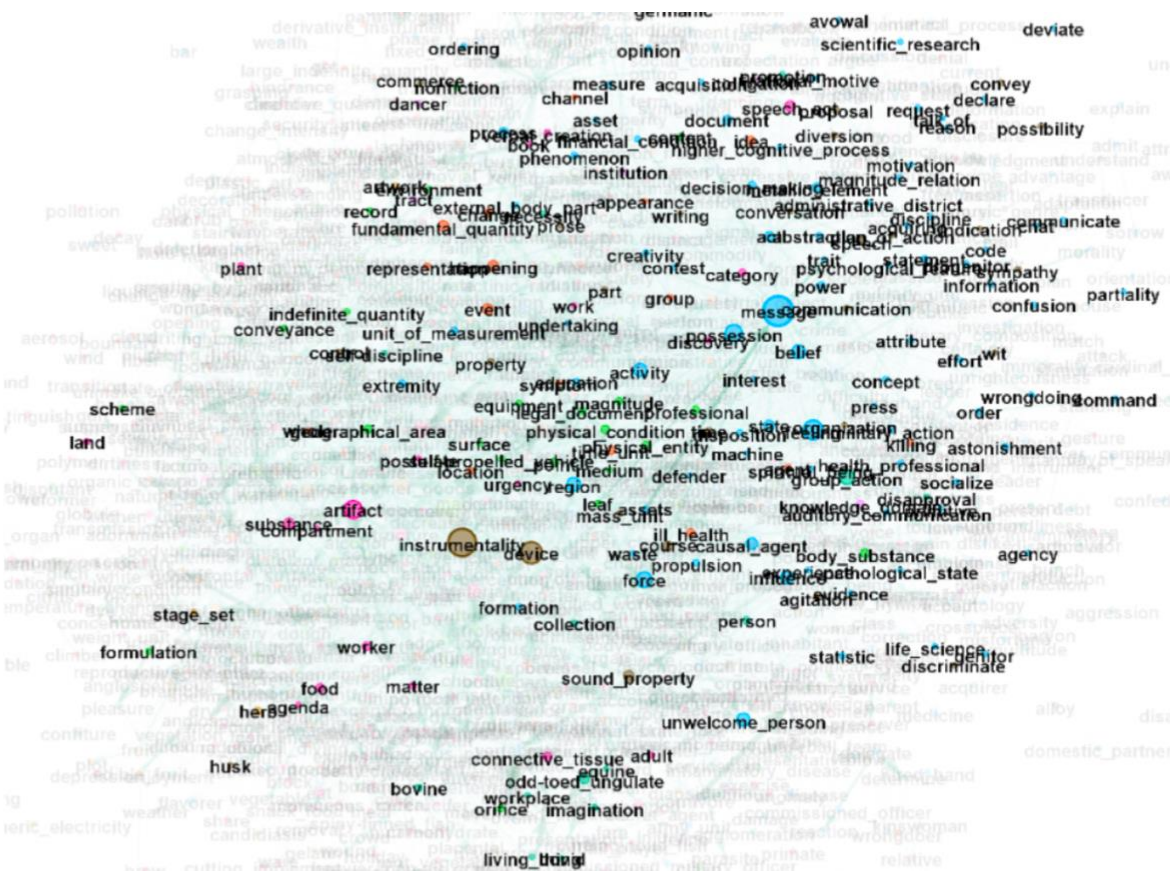
Topic: Standard



In the network, the node labeled "Standard" has been evaluated with these observations:

- Although it is not one of the betweenness centers, this node has extensive connections across various communities.
- It has a significant number of neighbors.
- The associations with its neighbors vary, including both strong and weak connections.
- These characteristics strongly suggest that the network might exhibit small-world or even ultra-small-world properties, which warrants further investigation to confirm.

Topic: Message



In the network, the node labeled "Message" has been analyzed with the following observations:

- It serves as one of the betweenness centers, with extensive connections spanning various communities.
- This node is characterized by a high number of neighbors.
- The associations with these neighbors display a mix of both strong and weak connections.
- These features further suggest that the network may exhibit small-world or even ultra-small-world properties.

Discussion and Conclusion

- From the Part-of-Speech tagging, I discerned a robust relationship among Nouns, Verbs, and Adjectives, although this might be biased due to the scarcity of data on other parts of speech.
- The vowels association network exhibits a heterogeneous structure.
- In the Subtopics association, some subtopics demonstrate high connectivity, while others are less well-connected.
- The Topics association reveals that the network behaves like a real-world, scale-free graph, exhibiting small-world properties.

Limitations

Despite implementing various methodologies and techniques, I encountered some limitations:

- The visualization and analysis of a large number of nodes and edges in Gephi presented computational challenges, necessitating the selection of a dataset subset to generate results.
- Another limitation was the difficulty in acquiring highly relevant data sources that could be effectively mapped to the current dataset.

Future Scope

- Future research could explore the use of machine learning models to predict target words based on cue words.
- Expanding the study to include sub-features like word sounds, topic modeling, and the application of community detection concepts on a larger dataset with more diverse features could enhance the richness and relevance of the study.

Citations

- Adaptive Factorization Network: Learning Adaptive-Order Feature Interactions Weiyu Cheng, Yanyan Shen,∗Linpeng Huang, Shanghai Jiao Tong University{weiyu cheng, shenyy, lphuang}@sjtu.edu.cn
- Nelson, McEvoy & Schreiber: http://w3.usf.edu/FreeAssociation/AppendixA/index.html
- https://unsplash.com/photos/C5SUkYZT7nU
- https://pixabay.com/photos/abc-alphabet-alphabet-letter-blank-3523454/