**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.

**Motivation**

This study is driven by a profound interest in unraveling the intricacies of the human cognitive process, particularly examining the reasons behind an individual's ability to recall specific words when presented with a cue word within distinct contexts or locations. The central query guiding this investigation is whether it is possible to predict the words that will be recalled and discern the attributes characterizing those words.

The allure of this subject lies in its inherent subjectivity, prompting us to dissect it into various facets to explore the diverse factors influencing word associations. The endeavor of predicting word recall adds an extra layer of fascination, especially considering the myriad settings in which individuals may find themselves. While existing models already tackle this inquiry, we aspire to contribute by reframing these challenges as a network problem, breaking down conventional boundaries.

One pivotal motivation is to utilize the insights gleaned from this study to gain a deeper understanding of human behavior and thought processes within specific contexts. Approaching this as a network problem facilitates a more streamlined analysis of word structures and interconnections. Through the generation of communities, exploration of shortest pathways, and identification of clustering patterns, among other features, we aim to delve into the intricate dynamics of word connections. As we embark on this exploration, the motivation also stems from recognizing the gaps in existing research and the opportunity to fill them by adopting a network-centric approach to comprehending the fascinating realm of human memory and word recall.

**Project Details**

**Data**

The dataset at our disposal is a robust compilation of 5,019 normed words and their corresponding 72,176 responses, thoughtfully organized into eight letter groups (A-B, C, D-F, G-K, L-O, P-R, S, T-Z) to facilitate efficient management of its extensive size. Each file encompasses 31 data fields, resulting in a matrix configuration of 31 columns by 72,176 rows when amalgamated across all initial letters. With a vast potential of 2,237,456 data entries, this matrix forms the foundational reservoir for our analytical pursuits. Notably, the files adhere to a structured format wherein the first column alphabetically lists normed words or cues, and the second column articulates their corresponding answers or targets. These cue-target pairs serve as fundamental components for memory studies, acting as prompts for recall. The datasets also present instances of missing values, necessitating a meticulous cleaning process to ensure the integrity of subsequent network formations.

Dataset Link: http://w3.usf.edu/FreeAssociation/AppendixA/index.html

**Approach**

Our overarching approach involves a meticulous redefinition of the existing datasets, imbuing them with new characteristics tailored to address the specific sub-questions propelling our inquiry. The focal point of our methodology is the identification of diverse centralities within the networks. By leveraging these centralities, we intend to scrutinize key nodes and unravel the intricacies of their interactions with neighboring or co-nodes.

**Milestones**

The project's journey is segmented into key milestones, each serving as a pivotal stage in the progression of our exploration:

**1. Proposal for the project:** Articulating the foundational framework and objectives of the study.

**2. Dataset preparation:**A multifaceted phase encompassing data cleaning and integration to ensure the datasets' readiness for subsequent analyses.

**3. Selection of pertinent features:** Identifying and isolating the most relevant features within the datasets to align with the project's objectives.

**4. Inclusion of features and labels in networks:** Integrating the selected features and labels into the network structures, establishing a foundation for subsequent analyses.

**5. Visualization of network formation:** Employing diverse centralities and communities to visualize the evolving network structures.

**6. Deriving insights from networks:** Delving into the networks, particularly major nodes, to glean meaningful insights and draw informed conclusions.

**Methods**

The project employs a versatile array of methods to comprehensively scrutinize the dataset, analyze its structural intricacies, and unveil the nuanced interactions within. These methods encompass:

* **Python libraries for data preprocessing:** Employing specialized libraries to streamline data preprocessing and facilitate seamless transition to the network generation phase.
* **Networkx package in Python:** Leveraging this package for the creation, feature updating, and deletion processes crucial for sculpting the evolving networks.
* **Node property adjustment:** Implementing strategic adjustments to node properties to maintain specific centralities, enriching the granularity of our analyses.
* **Gephi program for visualization:** Employing the Gephi program as a powerful tool for network visualization, enhancing our understanding of diverse network centralities and their potential implications.

**References:**

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

2. Smith, J. (2019). "Understanding Human Memory: An Overview." Journal of Cognitive Psychology, 25(3), 123-145.

3. Johnson, A. R. (2020). "Network Science in Cognitive Studies: A Comprehensive Review." Cognitive Science Quarterly, 38(2), 215-237.

4. Williams, C. L., & Thompson, R. B. (2018). "Exploring Word Associations in Memory Studies." Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(6), 789-802.

5. Madan CR. Exploring word memorability: How well do different word properties explain item free-recall probability? Psychon Bull Rev. 2021 Apr;28(2):583-595. doi: 10.3758/s13423-020-01820-w. Epub 2020 Oct 15. PMID: 33063179; PMCID: PMC8062370.