



Parkinson's Disease

Alexis Jones

(1) ENABLE Program, (2) Carolina Health Informatics Program, UNC Chapel Hill, Chapel Hill, NC

Abstract

In this project I will analyze the triggers which activates the cascade of neurodegeneration using text mining. Through clustering and Network analysis I will analyze the prominent gender difference in PD and how age, environment and lifestyle affect these differences. The goal is to see what hyper-parameters need to be set through text mining to get the maximum amount of information to answer my three research questions.

What is Parkinson's Disease?

Parkinson's disease is a neurodegenerative disorder that affects predominantly dopamine-producing neurons in a specific area of the brain called substantia nigra.

Statistics

- More than 10 million people worldwide are living with PD
- Men are 1.5 times more likely to have Parkinson's disease than women
- Approximately 60,000 people are diagnosed with PD every year
- Nearly one million will be living with Parkinson's disease in the U.S. by 2020

What is Text Mining?

Text mining is the process of deriving high quality information from text using algorithms to examine relationship, patterns and trends through various means of text.

Objective

- Gender Difference
- Environment
- Genetic Mechanisms

Research Questions:

- What is the trigger which activates the cascade of neurodegeneration?
- Why are there such prominent gender differences in Parkinson's disease?
- Two Part Question:
 - Part A: What relationship, if any, exist between environment/air quality, lung disease and Parkinson's disease?
 - Part B: If a relationship is discovered, why does such little research exist in this area?

Methods

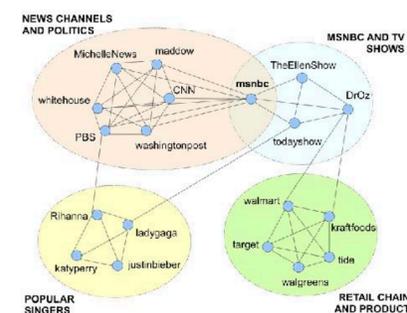
Clustering:

- Clustering is grouping things together, but not all members of that group are equal
- Clusters give you the idea of where they are, but does not give you the idea of how they are connected

Network Analysis:

- Networks are ideas or concepts that have similarities
 - Nodes represent an object
 - Edges represent relationships between objects
- Cliques are a collection of nodes that have mutual similarity
 - Maximum degree centrality are nodes that have the most edges or things connected to it
 - Transitive property
 - Cosine calculation measures the similarities between documents

Example:



Results

Cluster: Exploratory Analysis

- Oxidative stress
- Mitochondrial dysfunction
- Higher urate levels in plasma or cerebrospinal fluid (CSF) have been linked to both a lower risk of developing PD
- Non-gene domination
- SLC18A2, TH or DRD2
- CCNH, DLK1, PCDH8, SLIT1, DLD, PBX1, INSM1, and BMI1 were found to be significantly associated to biological process affected in PD

Networks: Exploratory Analysis

- There may be a link between Parkinson's Disease and Lung Cancer
- Exposure in certain pesticides can lead to PD
- Damage in BER genes
- Oxidative Stress and Depression
- 104 GSTP1 polymorphism is associated with males who have PD and not females
- Detected a protective effect of wild type genotype of XRCC1 in women
- Drugs and environment
- MTH1, OGG1, MUTYH all 3 enzymes upregulated in the SN of PD patients

Future Implications

- Use data mining to determine what region of the country is Parkinson's disease most prevalent in:
 - Examine patients work history and Mental health history
 - Depression and lung cancer statistics in this region
- How much money do these patients spend on healthcare a year in these regions?

References

Division of Neurofunctional Genomics, Department of Immunobiology and Neuroscience, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan
 Department of Neurology, Pittsburgh Institute for Neurodegenerative Diseases, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.
 Center for Environmental Health Sciences, Department of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, MT

Acknowledgements:

Special thanks to my primary advisor on this project Michael Ortiz.
 Thank you to Heejun Kim and Eric Cui for mentorship.
 Thank you to the ENABLE program at UNC Chapel Hill.
 Thank you to Shikha Yadav the ENABLE Program Coordinator