Evolution on Psychology Research

Computational Content Analysis Final Project

Tianyue Niu & Yinjiang Xiong

Introduction

Question of Interest:

'How has the focus of psychological research changed in the last 50 years?'

- Psychology has had a long history of development
- Cognitive psychology & neuroscience in particular have made huge breakthroughs in recent decades
- Past research patterns inform us of past changes in cultural trends, significant sociological implication.
- Maybe even allow us to predict the future.

Introduction

Question of Interest:

'How has the focus of psychological research changed in the last 50 years?'

- Psychology has had a long history of development
- Cognitive psychology & neuroscience in particular have made huge breakthroughs in recent decades
- Past research patterns inform us of past changes in cultural trends, significant sociological implication.
- Maybe even allow us to predict the future.

Data & Method

Data:

- PubMed database (a total of 30 million citations), accessed and parsed each 1 out of 8 files
- Filter by keywords ('psyc', 'neur', etc) to include only psychology related articles

Techniques used:

- Large Scale Computing Techniques: AWS EC2, S3 bucket, EMR Notebook with PySpark
- Topic Modeling: macro changes in topics
- Word Embedding: changes in our understanding for specific topics
- Network Analysis: focus on changes in 'Alzheimer' in the past 5 years
- Text Generation: an exploration

Data & Method

Data:

- PubMed database (a total of 30 million citations), accessed and parsed each 1 out of 8 files
- Filter by keywords ('psyc', 'neur', etc) to include only psychology related articles

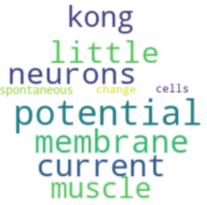
Techniques used:

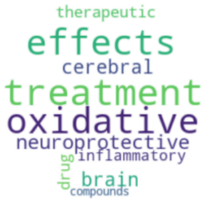
- Large Scale Computing Techniques: AWS EC2, S3 bucket, EMR Notebook with PySpark
- Topic Modeling: macro changes in topics
- Word Embedding: changes in our understanding for specific topics
- Network Analysis: focus on changes in 'Alzheimer' in the past 5 years
- Text Generation: an exploration

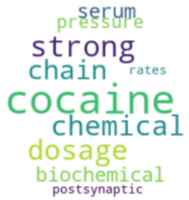
Topic Modeling: macro-changes in research focus

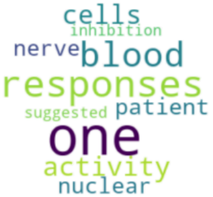
Six periods: before 1970, 70-80, 80-90, 90-00, 00-10, after 2010

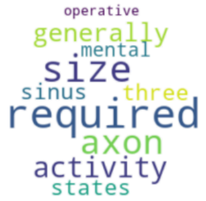
Before 1970



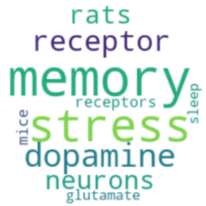








After 2010

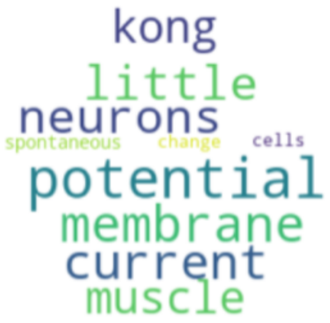


Topic Modeling: macro-changes in research focus

Advancement in specific topics

Action Potential

Before 1970



1970 - 1980

1980-1990

Neurotransmitter

1970 - 1980

1980-1990

1990-2000

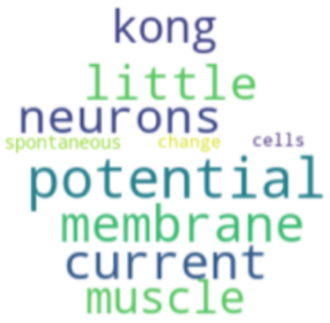
2000-2010

Topic Modeling: macro-changes in research focus

Advancement in specific topics

Action Potential

Before 1970



1970 - 1980

1980-1990

Neurotransmitter

1970 - 1980

1980-1990

1990-2000

2000-2010

Word Embedding Analysis

Compare synonyms extracted for ‘mental’, ‘abnormal’ and ‘alzheimer’

	mental	alzheimer	abnormal
1970 - 1980	applicable 0.992386	alkaline 0.997977	last 0.994673
	care 0.990944	desirable 0.991022	preferred 0.989865
	resolve 0.990482	added 0.988666	myopia 0.986158
	states 0.990316	glyoxylic 0.986688	divided 0.984945
	asked 0.990215	elimination 0.984513	fit 0.983737
	fluoroscopy 0.988398	efficiently 0.983935	regressed 0.983211
	needed 0.987659	oxidation 0.983633	unknown 0.982618
	setting 0.987610	suspension 0.982340	orthopaedic 0.982572
	relapse 0.986823	periodate 0.981462	hypothyroidism 0.982553
	today 0.986600	equilibrium 0.981461	words 0.982152
After 2010	stave 0.994489	neurometabolic 0.993835	quinolones 0.998634
	health 0.994171	microbiologically 0.993593	acylated 0.998501
	psychosocial 0.993910	fragile 0.993417	pentapeptide 0.998122
	lifestyles 0.992894	foodborne 0.987663	congophilic 0.997481
	lebanon 0.992530	rett 0.984513	ascs 0.996644
	narrated 0.991415	disease 0.982809	myoinhibitory 0.992948
	psrfs 0.991413	wolfram 0.981952	glycosylated 0.992174
	rita 0.991092	affecting 0.980769	heptapeptide 0.992013
	bereavement 0.989977	fibrinolytic 0.980083	dodecyl 0.990582
	affordability 0.988072	encephalopathies 0.978776	vasoconstrictive 0.990035

Before 1970:

- ‘alzheimer’ and ‘abnormal’ not in vocabulary
- ‘mental’ top synonym was ‘psychotic'

Word Embedding Analysis

Compare synonyms extracted for ‘mental’, ‘abnormal’ and ‘alzheimer’

	mental	alzheimer	abnormal
1970 - 1980	applicable 0.992386	alkaline 0.997977	last 0.994673
	care 0.990944	desirable 0.991022	preferred 0.989865
	resolve 0.990482	added 0.988666	myopia 0.986158
	states 0.990316	glyoxylic 0.986688	divided 0.984945
	asked 0.990215	elimination 0.984513	fit 0.983737
	fluoroscopy 0.988398	efficiently 0.983935	regressed 0.983211
	needed 0.987659	oxidation 0.983633	unknown 0.982618
	setting 0.987610	suspension 0.982340	orthopaedic 0.982572
	relapse 0.986823	periodate 0.981462	hypothyroidism 0.982553
	today 0.986600	equilibrium 0.981461	words 0.982152
After 2010	stave 0.994489	neurometabolic 0.993835	quinolones 0.998634
	health 0.994171	microbiologically 0.993593	acylated 0.998501
	psychosocial 0.993910	fragile 0.993417	pentapeptide 0.998122
	lifestyles 0.992894	foodborne 0.987663	congophilic 0.997481
	lebanon 0.992530	rett 0.984513	ascs 0.996644
	narrated 0.991415	disease 0.982809	myoinhibitory 0.992948
	psrfs 0.991413	wolfram 0.981952	glycosylated 0.992174
	rita 0.991092	affecting 0.980769	heptapeptide 0.992013
	bereavement 0.989977	fibrinolytic 0.980083	dodecyl 0.990582
	affordability 0.988072	encephalopathies 0.978776	vasoconstrictive 0.990035

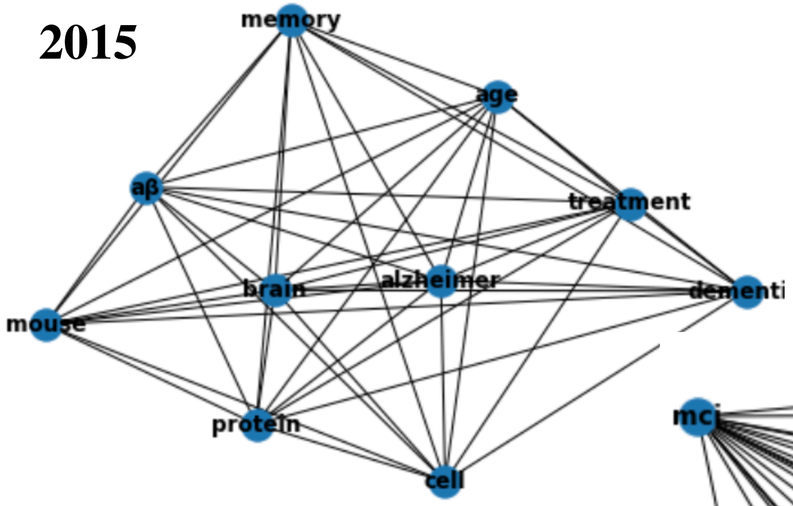
Before 1970:

- ‘alzheimer’ and ‘abnormal’ not in vocabulary
- ‘mental’ top synonym was ‘psychotic'

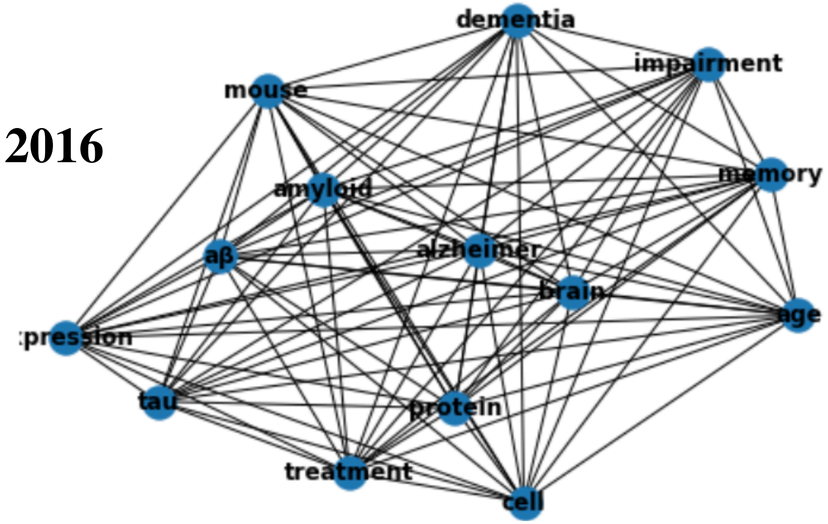
Network Analysis

Changes in Alzheimer research

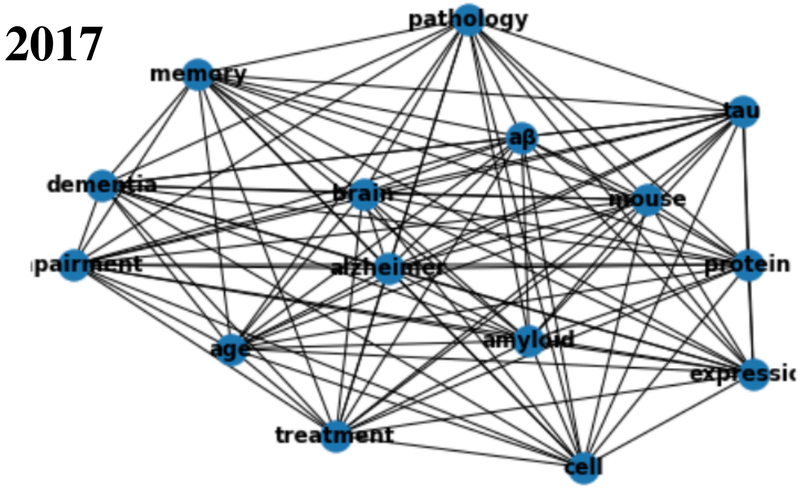
2015



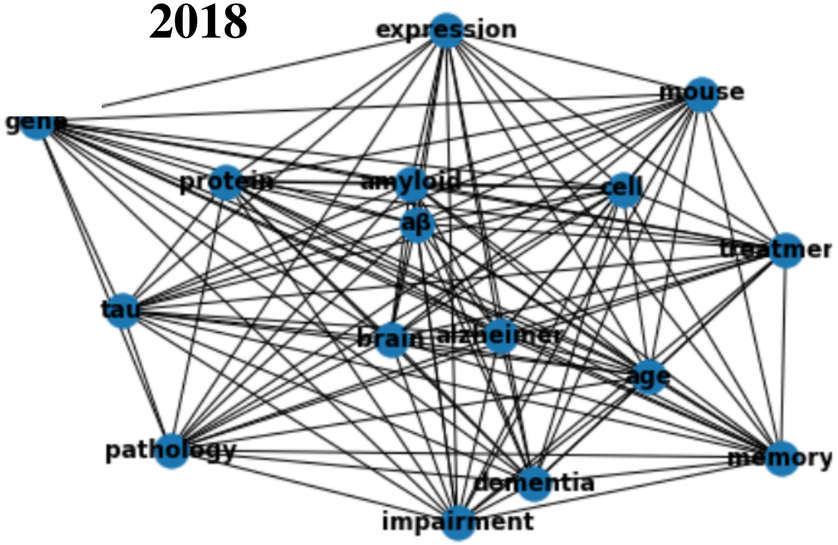
2016



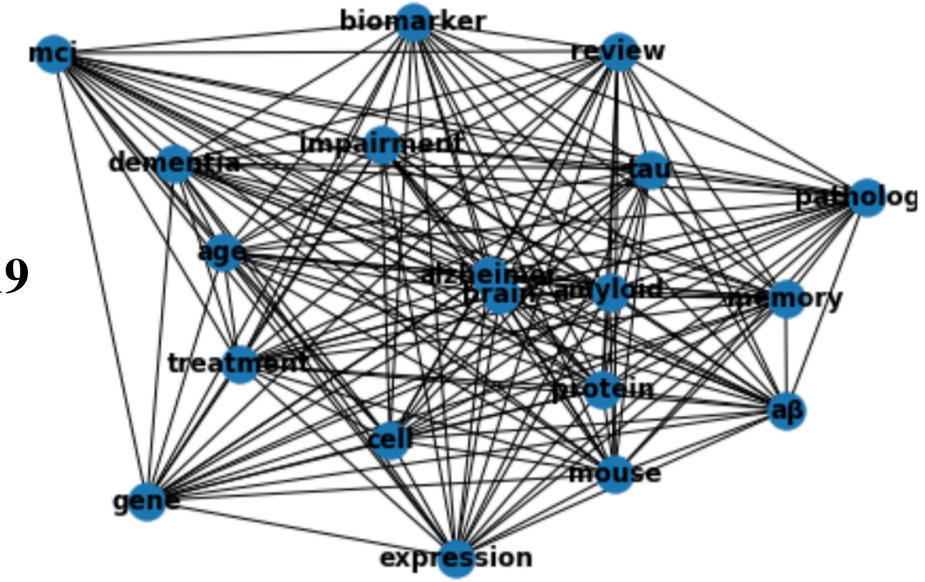
2017



2018



2019



Network Analysis

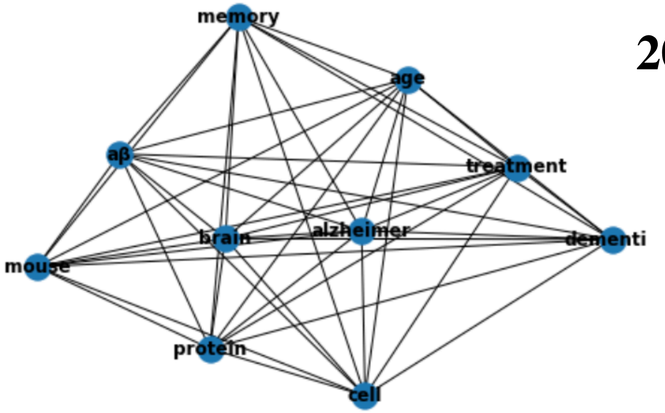
Changes in Alzheimer research

1. Growing number of terms included

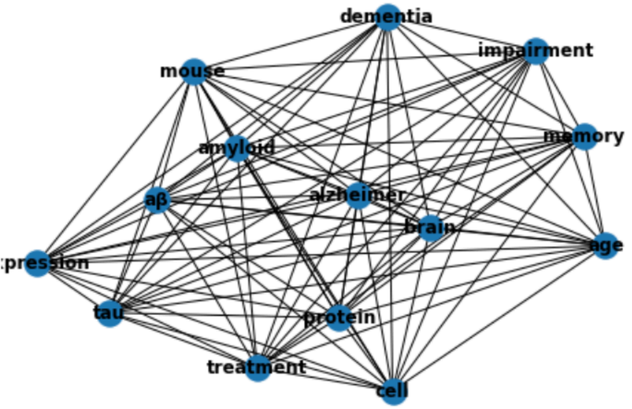
2. Appearance of “tau” in 2016 and move closer to “alzheimer” in 2019

3. Appearance of “biomarker” and “mci” in 2019

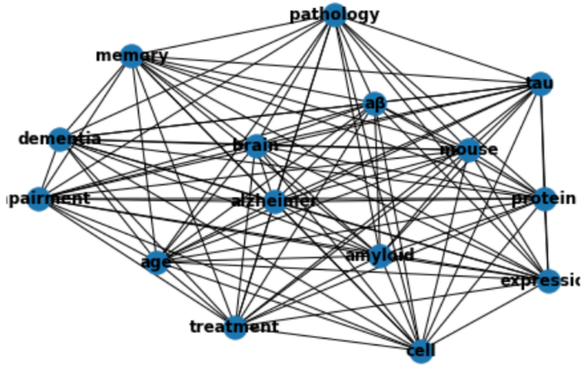
2015



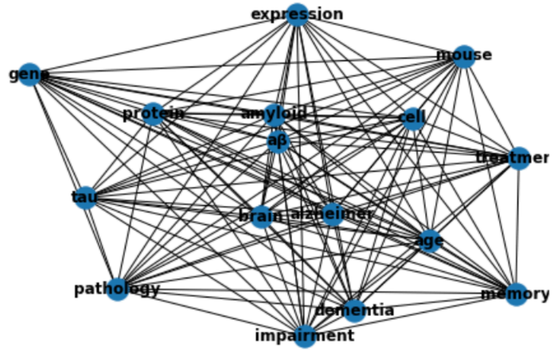
2016



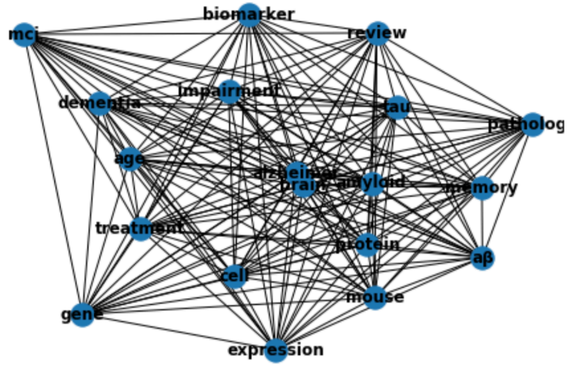
2017



2018



2019



Network Analysis

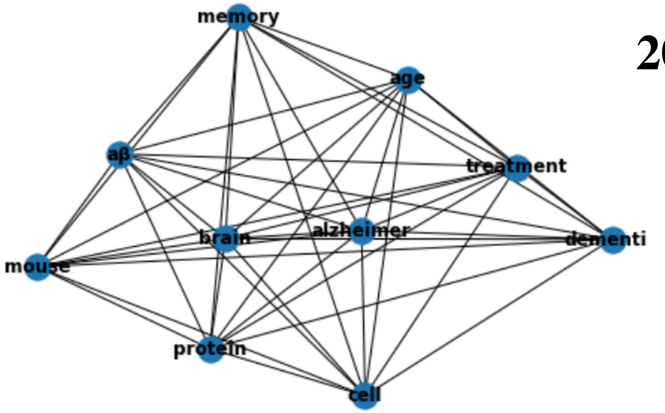
Changes in Alzheimer research

1. Growing number of terms included

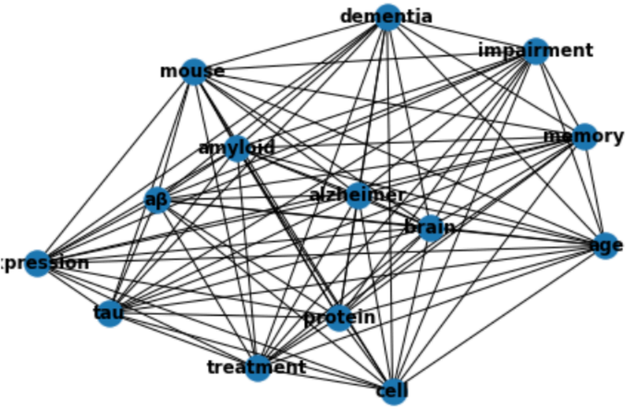
2. Appearance of “tau” in 2016 and move closer to “alzheimer” in 2019

3. Appearance of “biomarker” and “mci” in 2019

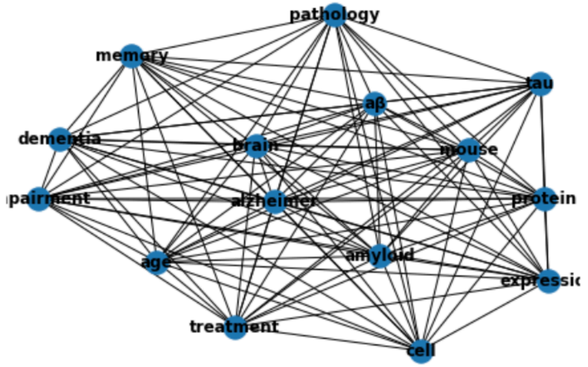
2015



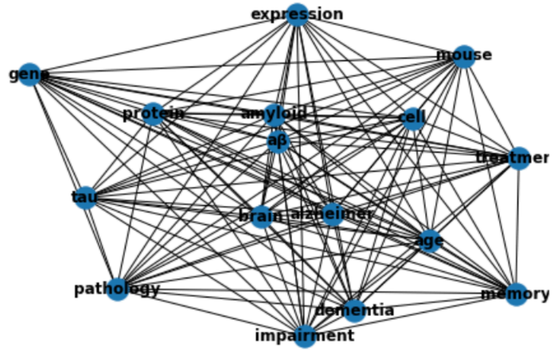
2016



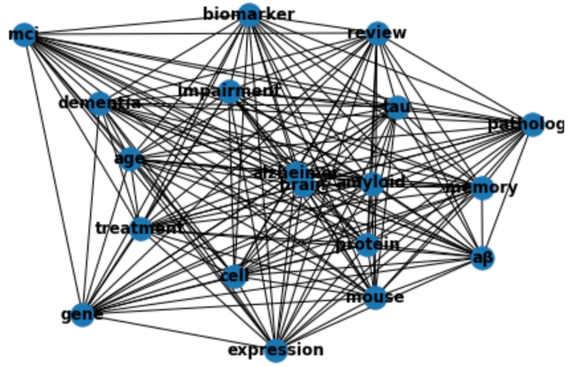
2017



2018



2019



Network Analysis

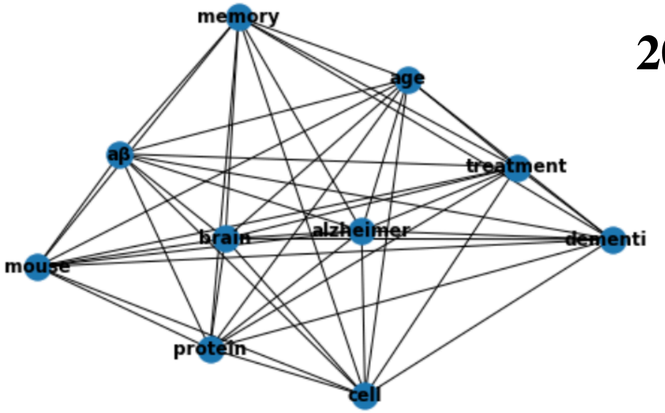
Changes in Alzheimer research

1. Growing number of terms included

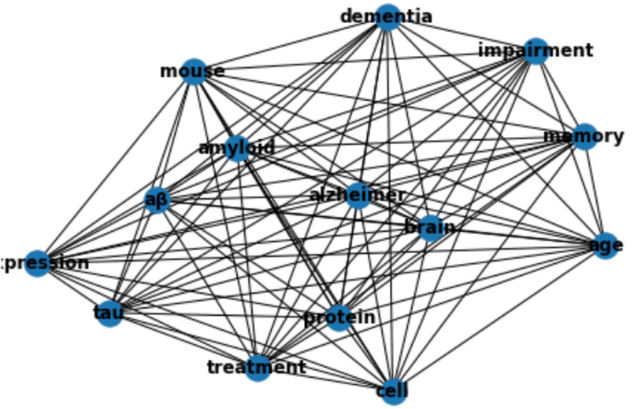
2. Appearance of “tau” in 2016 and move closer to “alzheimer” in 2019

3. Appearance of “biomarker” and “mci” in 2019

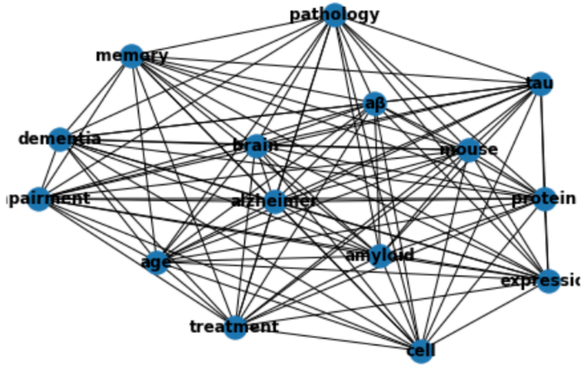
2015



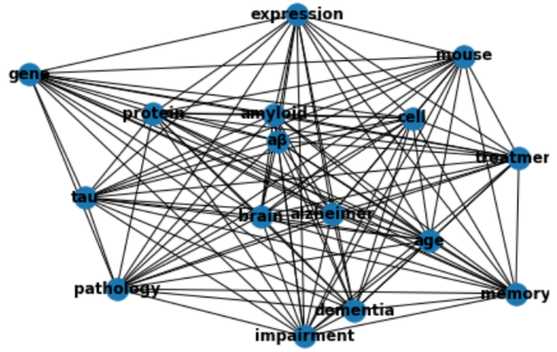
2016



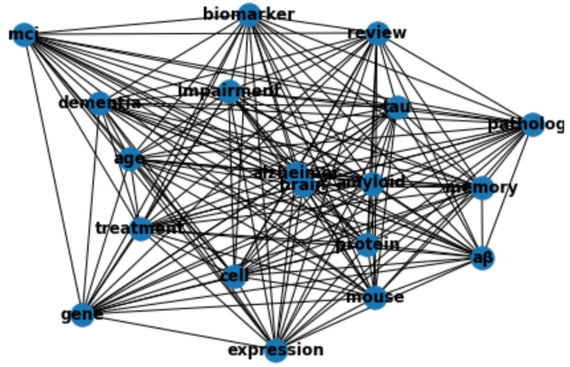
2017



2018



2019



Text Generation

Fine-tuned Models using 1989, 1999, 2009 and 2019 data

1. GPT2 pre-trained model fine-tuned with context
2. Sequence prompt “Alzheimer is a disease”
3. Sequence prompt “Rats are used”

Original Model

Alzheimer is a disease that affects about 1 in 10 people. It is caused by a genetic mutation that causes the brain to become more active.

Text Generation

Fine-tuned Models using 1989, 1999, 2009 and 2019 data

1. GPT2 pre-trained model fine-tuned with context
2. Sequence prompt “Alzheimer is a disease”
3. Sequence prompt “Rats are used”

Original Model

Alzheimer is a disease that affects about 1 in 10 people. It is caused by a genetic mutation that causes the brain to become more active.

Text Generation

Fine-tuned Models using 1989, 1999, 2009 and 2019 data

1. GPT2 pre-trained model fine-tuned with context
2. Sequence prompt “Alzheimer is a disease”
3. Sequence prompt “Rats are used”

1989

Alzheimer is a disease of the brain that affects the brain's ability to process information. It is a neurodegenerative disease that affects the brain's ability to process information.

1999

Alzheimer is a disease of the brain that affects the central nervous system. It is a neurodegenerative disease of the brain that affects the central nervous system.

2009

Alzheimer is a disease of the brain that affects the brain's ability to process information. The disease is associated with a wide range of neurological and psychiatric disorders, including Alzheimer's disease, Parkinson's disease, Huntington's disease.

2019

Alzheimer is a disease of the central nervous system that affects the brain and spinal cord.

Text Generation

Fine-tuned Models using 1989, 1999, 2009 and 2019 data

1. GPT2 pre-trained model fine-tuned with context
2. Sequence prompt “Alzheimer is a disease”
3. Sequence prompt “Rats are used”

Original Model

Rats are used to kill the most common types of birds, including the black-footed, the black-tailed, and the black-tailed eagle.

Text Generation

Fine-tuned Models using 1989, 1999, 2009 and 2019 data

1. GPT2 pre-trained model fine-tuned with context
2. Sequence prompt “Alzheimer is a disease”
3. Sequence prompt “Rats are used”

1989

Rats are used in the study of the effects of the various drugs on the nervous system.

1999

Rats are used to study the effects of a variety of drugs on the immune system. The purpose of this study was to determine the effects of a variety of drugs on the immune system of rats.

2009

Rats are used to study the effects of a novel antipsychotic on the development of schizophrenia. The aim of this study was to determine the effects of a novel antipsychotic on the development of schizophrenia in rats.

2019

Rats are used to study the effects of a variety of drugs on the brain. The aim of this study was to investigate the effects of a novel drug, the selective serotonin reuptake inhibitor (SSRI), on the brain of rats.

Summary & Future Direction

- Techniques used accurately captured changes in research trends, have the potential of complimenting, or even replacing meta literature review.
- Huge advancements in neuroscience and neurodegenerative diseases in the past 5 decades.
- Network analysis and text generation can be used to analyze trends of specific topics
- Techniques can be easily extended to study other subjects, or the entire PubMed data.