

# Thesis Plans

Tobin South

January 9, 2020

## 1 Title Ideas

## 2 Sections

### 2.1 Introduction

Introduce the problem we are trying to solve.  
What actually is the goal

### 2.2 Background

#### 2.2.1 Entropy Basics

Basics and History, essentially a literature review on entropy as a topic.

#### 2.2.2 A introduction to Natural Language Processing

Tokenizing, Def: BOW, Token Sequence, Time of Speaking A small literature review on the work of tokenizing and different techniques.

#### 2.2.3 Process Entropy and Natural Language

An introduction to the idea of language as a non-markovian process. A description of the core maths that underlies process entropy in stationary ergodic processes. Def: Stationary, Ergodic, Markvoian.

### 2.3 Information Measures and Distances

Predictability and Self-information  
KL Divergence

JS Distance  
Mutual information?

## **2.4 Chapter 1**

### **2.4.1 Data**

Twitter?  
Describe Text and Provide Analysis  
Allsides explained and evaluated with reliability.

### **2.4.2 Basic Works**

Initial clustering of data and TFIDF  
LDA?  
Exploration of why this doesn't work.

### **2.4.3 Entropy on this text**

Full entropy (not time based)  
Plots of distributions, groupings by bias  
signal to noise ratio

## **2.5 Chapter 2**

### **2.5.1 A background on networks**

A background on networks  
directed graphs  
undirected graphs  
networks vs graphs  
dense graphs centrality measures

### **2.5.2 Directed Networks**

Create a directed network, look for clustering and report metrics  
more to come