Introduction

Narrative Economic is the study of the spread and dynamics of the collective verbal output of a society in order to understand economic fluctuations. Many reasons motivate individuals to focus on and propagate specific narratives from the many they come across. Each narrative originates in the mind of an individual before starting its evolutionary course through time. The Kermack-McKendrick (1927) SIR Model lends itself well to representing the spread of a narrative over time. An individual at a given point in time either has or has not been exposed to specific narrative. If this person has not come across a narrative, she is susceptible to coming across it for the first time. Otherwise, she is either infected (she will pass pass the narrative along to another susceptible individual) or recovered (she will not pass the narrative along to another susceptible individual). Given that more infected individuals implies a more powerful social epidemic, we hypothesize that narratives of certain valences should evoke aggregate market responses that manifest themselves as measurable economic fluctuations.

We hope to use Natural Language Processing on a carefully chosen set of corpora in order to track the appearance and subsequent growth and evolution of narratives. Once we build the necessary tools to automatically identify narratives and their dynamics through time, we will explore the predictive power of the valence and infectivity of specific narratives on economic fluctuations during the course of their evolution.

Methods

In our previous work on the project we found preliminary evidence that the most promising way to automatically identify the development and temporal evolution of narratives in corpora was through
meme tracking. The idea for this came from the paper “Meme-Tracking and the Dynamics of the News Cycle.” In this paper, the authors construct a directed acyclic graph to track news cycle dynamics. We will develop and fine-tune an algorithm based on the approach in this paper to track narratives. Although our algorithm resembles the one in the paper, it differs in two fundamental ways. First, instead of forcing our graph to be directed by restricting by increasing meme length, we will restrict chronologically. This approach is more consistent with the definition of economic narratives as idea epidemics propagating among individuals through time. Secondly, we will focus a great deal of effort on developing an edit distance metric between memes that accurately identifies which memes are most closely part of the same narrative. Our edit distance metric must correctly identify memes that are part of the same narrative even if naive minimum edit distances between the memes is high. This algorithm will generate a directed acyclic graph for each narrative in the corpora. Analysis of the properties of the graphs produced by the algorithm will hopefully provide insight into economic fluctuations.

A more precise explanation of how this specific meme tracking approach could be implemented to track and quantify narratives is as follows. We begin with a set of vertices—which are the “memes” extracted from our corpora. There will be a set of criteria that we will use to determine what to extract from the text as a meme. The most naive approach—which is used in the “Meme-Tracking and the Dynamics of the News Cycle” and that we have experimented with—is to simply extract any text that is inside of quotation marks. Each of these extracted quotations is a “meme,” and thus a node in the directed acyclic graph we will construct. So, each node in the graph is a string which is taken from the corpora (most naively any string found inside of quotation marks). How do we draw edges between nodes? For each node \( v \), we compute a string edit distance between \( v \) and every node which chronologically preceded \( v \) (within a certain time frame limit). The purpose of this is to find a meme from which \( v \) evolved. The assumption is that unless \( v \) was the start of a specific narrative, the node with the closest edit distance to \( v \) which chronologically preceded it will be part of the same narrative from which \( v \) evolved. If there is no chronologically preceding node to \( v \) with an edit distance below a defined threshold, \( v \) is assumed to be the start of a specific narrative (so no edge will be drawn going into it). In the article from which the idea for this algorithm was drawn, DAG Partitioning is defined as: “given a directed acyclic graph with edge
weights, delete a set of edges of minimum total weight so that each of the resulting components is single-rooted.” Each resulting component in our case would be the narrative which we have mined from the text (the number of nodes in a component on a given day is a measure of how prevalent that narrative was on that day). Unfortunately the paper also goes on to prove that DAG Partitioning is NP-hard.

So, one challenge in implementing this approach includes developing heuristics to overcome the fact that DAG Partitioning is NP-hard. We must also develop a method to compute edit distances which accurately identifies which quotes are most closely related in a computationally feasible manner. Likewise, our method of “meme” identification within the text must extract strings from the text that are relevant to the spread and dynamics of narratives as they appear and evolve throughout the corpora. Although this algorithm is only one of many possible approaching to mining the corpora for narratives and their dynamics, it is immediately obvious that the independent computations of similarity between the pairs of nodes lends itself well to parallelization. This fact makes meme tracking an appealing start to quantifying the prevalence of certain narratives in corpora related to the financial markets and corresponding macroeconomic fluctuations.

We have started preliminary work on the topic by inspecting an historic Wall Street Journal corpus of articles for the validity of this approach at identifying narratives. In future work we hope to mine investor surveys which have been conducted continuously for a long time. We hope to capture specific narratives within the institutional investor community from these surveys and observe their coincidence with changes in economic variables like unemployment and overall asset price indices. If, for example, we compare two periods of similarly high unemployment, perhaps we can observe that peaks in negative valence narratives coincide with slower recoveries. A systematic trace of memes associated with negative affect through the tools I hope to develop would provide the necessary data for a behavioral economist to draw such conclusions.
Timeline

9/06-9/20
- Study existing literature and understand goals (i.e. identify which macroeconomic variables we hope to gain predictive power of through the algorithm).

9/20-10/04
- Preprocess corpora and plan exact algorithm/approach to be coded in collaboration with Professors Slade, Goetzmann, and Shiller.

10/04-10/25
- Modify algorithm/approach as we will likely find many areas of improvement that would make the narrative tracker more useful.

10/25-11/16
- Code modifications which were determined to be useful in the previous weeks.
- Buffer to insure that all previous goals have been met. Seek out direction from professors to identify areas of potential improvement to project before beginning final work.

11/16-11/30
- Code and organize everything required to write the final paper.

12/01-…
- Write final paper and submit database containing code from throughout the semester.

Deliverables

1. A final paper presenting any evidence of correspondence between the peaks of specific narratives and relevant macroeconomic fluctuations (e.g. changes in asset price or other economic indices, or economic metrics such as unemployment).
2. Code for the meme tracking algorithm. This algorithm will automatically identify narratives within corpora as directed acyclic graphs where the nodes are extracted quotations from the text and the edges are edit distances.

Bibliography


Lescovek, Jure. “Meme-Tracking and the Dynamics of the News Cycle.”