

# Word Embeddings Quantify Exogenous Cultural Shocks

## Introduction

How does culture evolve and change over time? Scholars have postulated that it is through a combination of **endogenous** (fads & fashions, incremental technological innovation, linguistic drift) and **exogenous** (natural disasters, war, economic crisis). While endogenous cultural change has been well observed through the temporal evolution of word embedding models, scant research has sought to identify the impact of exogenous shocks. This project seeks to do so by examining the 2007/2008 financial crisis on the business community.

## Data

We scraped over 100,000 quarterly earning calls of publicly traded firms from the financial website seekingAlpha. We analyze the Q&A section of the calls, where analysts ask questions and firm executives respond. All sentences are converted to lowercase, stripped of punctuation, and tokenized.

Data	Count
Calls	170,470
Firms	5,985
Tokens	190 million
Start Date	2006
End Date	2016

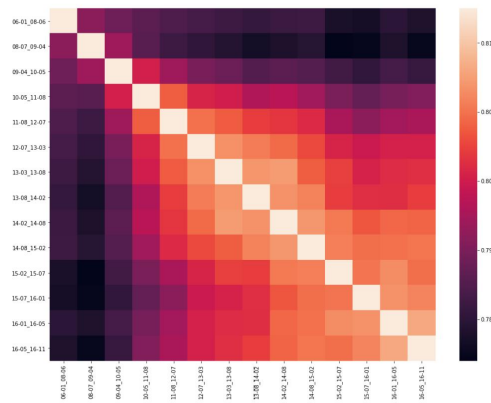
## Analysis

We segment the data into equally sized bins, grouped by year, to ensure data quantity does not impact the results. We then train a 100-dimensional GloVe embedding model on each bin of the data, initialized with the 2014 Wikipedia gigawords pretrained to ensure model alignment. Pairwise cosine similarity scores quantify embedding similarity across periods, capturing cultural change

$$J = \sum_{i,j=1}^V f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2$$

## Results

Pairwise model cosine similarity for top 4,000 words



Paul Vicinanza  
pvicinan@stanford.edu

## Evolving Words

Word	Early NNs	Late NNs
Housing	economy, home, residential, economic, issue, recession	senior, starts, home, urban, funding, economy
Bankruptcy	filed, filing, retirement, severe, lenders, pool	court, filing, patent, filed, insurance, recession
Border	zone, forces, river, strip, territory	cross, territory, zone, men, forces, agents, federal
Allocation	capital, management, power, wealth, allocated	capital, buybacks, priorities, dividend, strategy
Borrowing	revolver, debt, funding, bank, refinancing, credit	base, installed, facility, banks, bank
Learning	curve, learn, centers, experience, schools	learn, machine, technology, training, software, creative
Cloud	atlantic, 2013, tower, 2012, facebook, tablets	services, platform, enterprise, software, customers, service
Analytics	presentation, map, deflation, reinvestment, web	tools, data, technology, solutions, capabilities, cloud
Social	networking, political, reform, education, physical	media, political, internet, digital, facebook, web
Storage	server, systems, application, servers, space	system, capacity, equipment, gathering, processing, power

## Discussion / Future

We find strong evidence that the financial crisis disrupted the existing linguistic space. We replicate the results exactly using alternative model specifications such as smaller/larger vocabulary size, smaller/larger window, larger embedding dimension, and analyzing with n-nearest neighbors instead of cosine similarity.

For future research we are working to obtain additional data for the years prior to 2006 in order to establish a pre-crisis baseline. We are also expanding to alternative datasets, including the collection of house and senate committee transcripts before and after 9/11.