

Computational Analysis of Religious and Ideological Linguistic Behavior

University of Virginia, MS in Data Science Capstone

Samantha Garofalo, Seth Green, Leigh Harton, Megan Stiles



Advisor: Don Brown

Client: Peter Ochs

Sponsor: ARL

Problem: How to work with Value-Based Groups?

What is a Value-Based Group?

- *Any group motivated by a set of values*

Who needs to work with Value-Based groups?

- *US Government*
- *Humanitarian Groups*
- *Business Leaders*

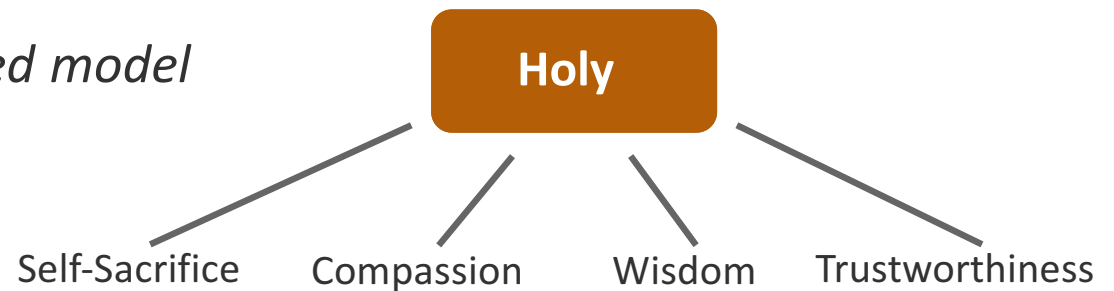
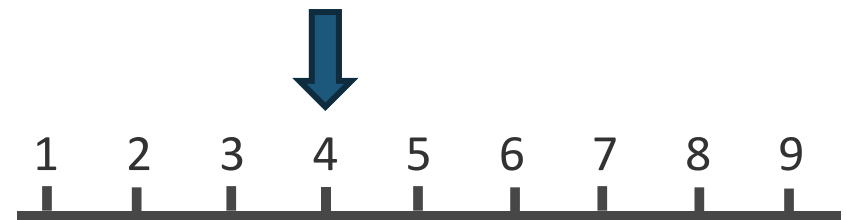
Large-Scale Goals:

- *Real-time recommendations to organizations on the ground*

UVa Center for Religion, Politics, and Conflict

1-9 Linguistic Rigidity Scale

- *Number of definitions not the definition itself is what matters*
- *Venuti et al. automated model*



About Our Data

Input

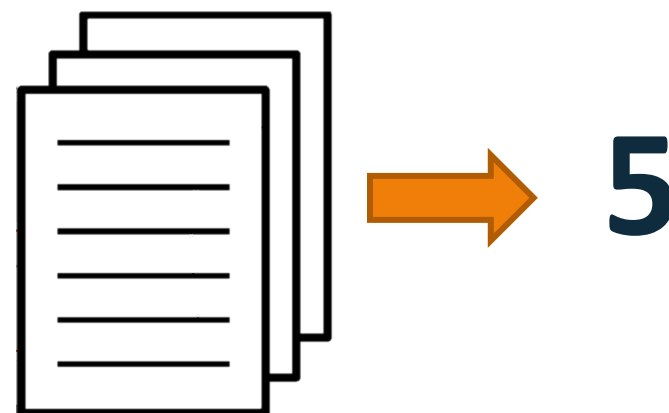
- *Raw ideological text*

Output

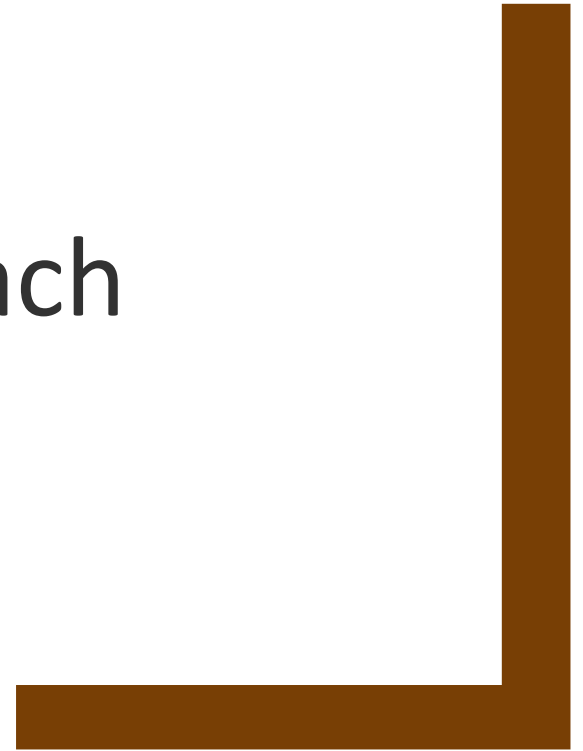
- *Number on Linguistic Rigidity Scale*

Sources:

- *Last year's corpus*
- *New Groups*
 - Web Scraping
 - OCR for PDFs



Technical Approach



Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders demonstrated no mercy at all. I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

Raw Text

Context Vectors

sinner should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders demonstrated no **mercy** at all. I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

Identify Keywords

Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders **demonstrated no mercy at all.** I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, "I desire mercy, not sacrifice." Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: justice, mercy and faithfulness. You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

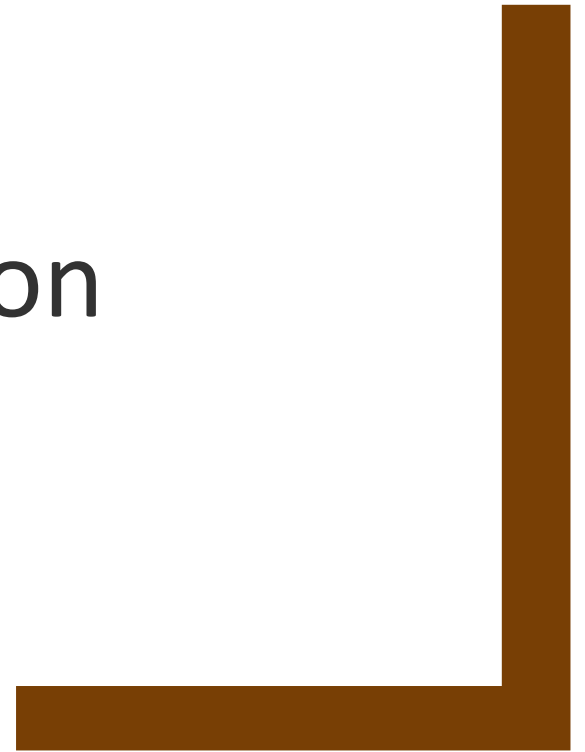
Collect words in context vector

Context Vectors

sinners should be welcome. However, some churches are quick to boot people out if they don't live up to their standards. I heard about a preacher whose wife was in the hospital and dying of cancer. The elders of his church came to the hospital to see them. The preacher thought they were coming to pray for his wife, but instead, they came to fire him. I have no idea what that preacher did that was so bad, but those elders **demonstrated no mercy at all.** I have heard of other horror stories where people have been driven out of a church for one thing or another. No mercy was demonstrated, only rugged legalism. But Jesus said, **"I desire mercy, not sacrifice."** Matt. 23:23 "Woe to you, teachers of the law and Pharisees, you hypocrites! You give a tenth of your spices: mint, dill and cummin. But you have neglected the more important matters of the law: **justice, mercy and faithfulness.** You should have practiced the latter, without neglecting the former." The Pharisees of old were scrupulous or careful

- Repeat for all occurrences of keyword
- Singular Value Decomposition to vector of length 100
- Calculate Average Semantic Density and Distributional Co-occurrence Score

Keyword Selection

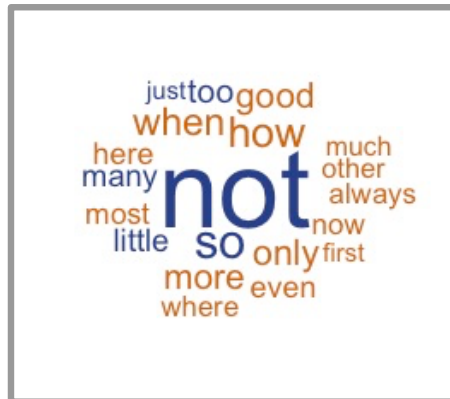


TFIDF Matrix

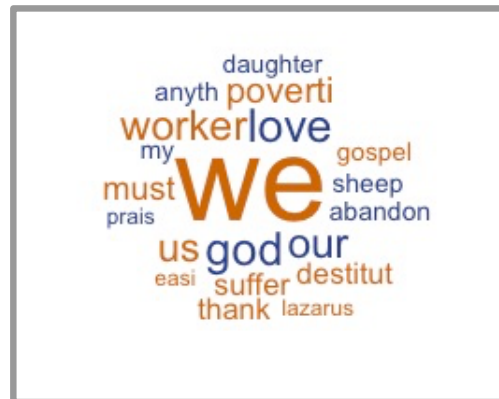
	freq	idf	logidf	tfidf	logtfidf
thee	18	139.000000	4.934474	2502.000000	88.820531
prophet	12	139.000000	4.934474	1668.000000	59.213687
me	12	139.000000	4.934474	1668.000000	59.213687
we	21	75.818182	4.328338	1592.181818	90.895101
uniti	9	139.000000	4.934474	1251.000000	44.410265
fate	8	139.000000	4.934474	1112.000000	39.475791
muhammad	8	139.000000	4.934474	1112.000000	39.475791
decre	8	139.000000	4.934474	1112.000000	39.475791
christ	10	104.250000	4.646792	1042.500000	46.467919

Value Words

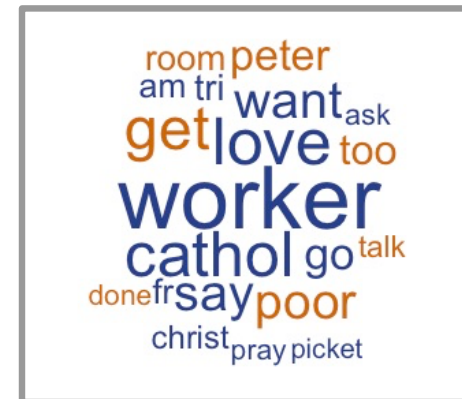
Previous Method



TFIDF Method w/Pronouns



TFIDF Method w/out Pronouns



Judgment Classification

Previous Method:

- *Judgment = (noun + “to be” verb + adj/adv)*

New Method:

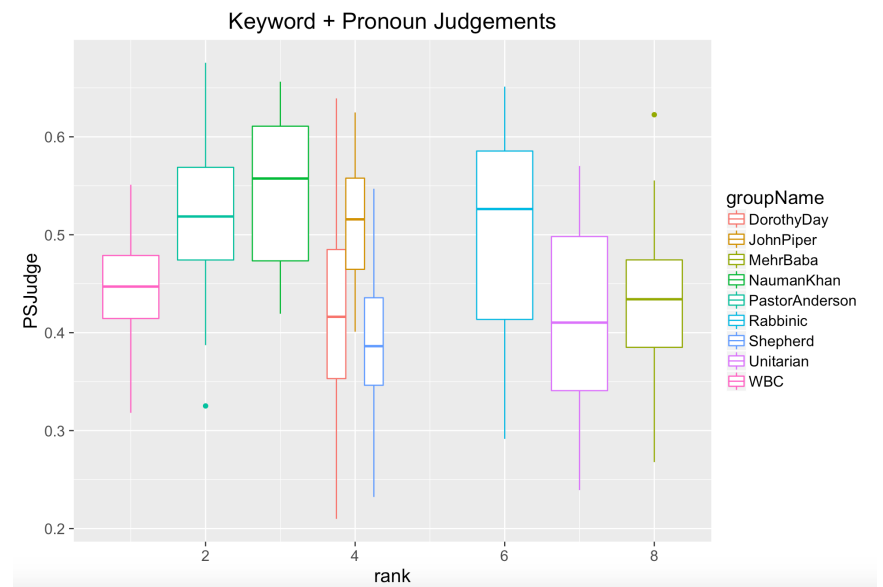
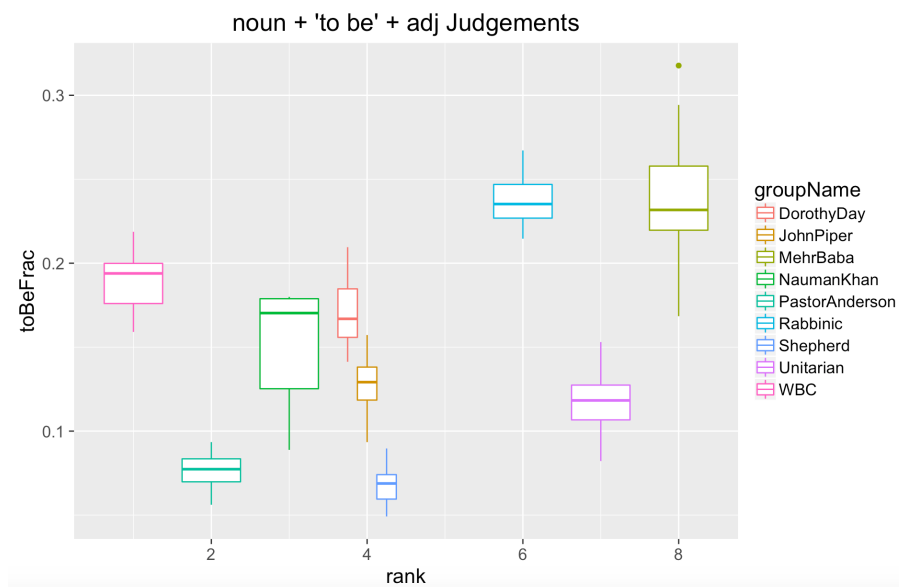
- *Judgment = pronoun + key word*

The first priority is to perform hijrah from wherever you are to the Islamic State, from dārul-kufr to dārul-Islām. Rush to perform it as Mūsā (‘alayhis-salām) rushed to his Lord, saying, {And I hastened to You, my Lord, that You be pleased} [Tāhā: 84]. Rush to the shade of the Islamic State with your parents, siblings, spouses, and children. There are homes here for you and your families. You can be a major contributor towards the liberation of Makkah, Madīnah, and al-Quds. Would you not like to reach Judgment Day with these grand deeds in your scales.

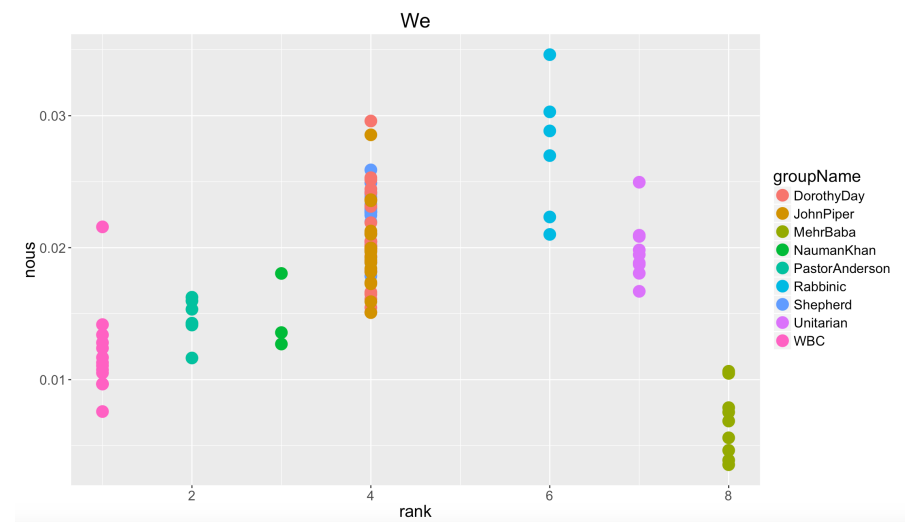
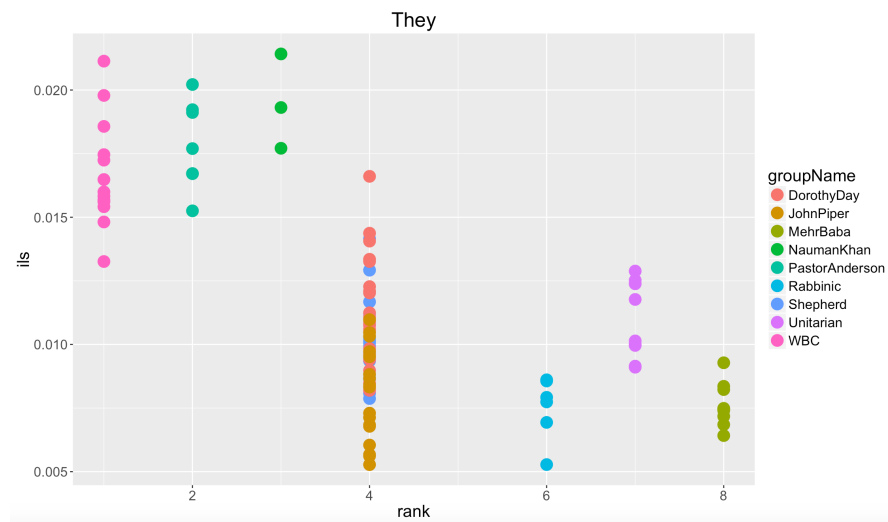
Second, if you cannot perform hijrah for whatever extraordinary reason, then try in your location to organize bay’āt (pledges of allegiance) to the Khalīfah Ibrāhīm. Publicize them as much as possible. Gather people in the masājid, Islamic centers, and Islamic organizations, for example, and make public announcements of bay’ah. Try to record these bay’āt and then distribute them through all forms of media including the Internet. It is necessary that bay’ah becomes so common to the average Muslim that he considers those holding back as grossly abnormal. This effort, inshā’allah, will encourage Islamic groups to abandon their partisanship and also announce their bay’ah to the Khalīfah Ibrāhīm.

If you live in a police state that will arrest you over such bay’āt, then use means of anonymity to con-

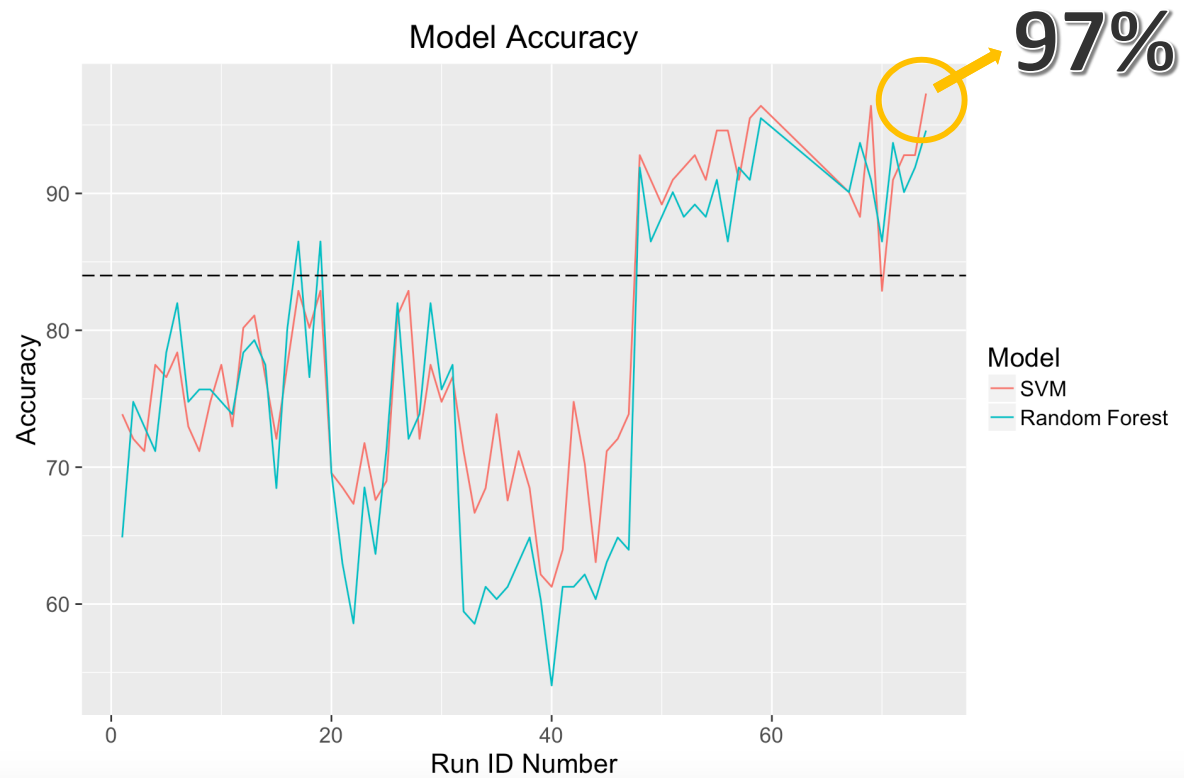
Judgment Classification



Pronoun Analysis



Results – Group Scores



Results – Group Scores

*Compared to Previous Methods using
Previous Corpus*

Model	Accuracy	Previous Work
Random Forest	96.5	86.0
SVM	97.2	84.0
Boosted Tree	96.2	N/A
Artificial Neural Net	N/A	80.0
<i>Average</i>	<i>96.6</i>	<i>83.3</i>

Using Full Data Set

Model	Accuracy
Random Forest	96.1
SVM	96.3
Boosted Tree	95.4
<i>Average</i>	<i>96.0</i>

Single Document Analysis

Why Single Documents?

- *Identifying potential spikes or deviations from a groups baseline*
- *Moving towards real-time, on-the-ground analysis*

of Documents

- 342

Results

- *~77% accuracy*



Word Embeddings

What are Word Embeddings?

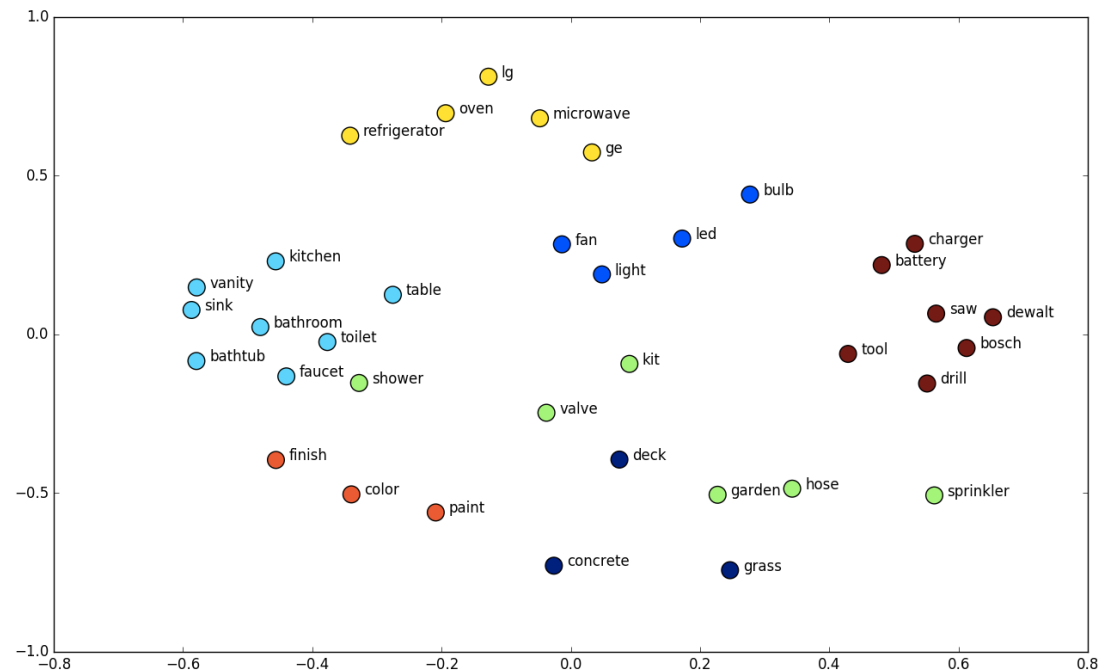
- *Vocabulary mapped to vectors*

When trained on Google News articles, similar words to “book”:

- *memoir*
- *novels*
- *paperback*
- *hardback*
- *autobiography*

When trained on Christian News articles, similar words to “book”:

- *church*
- *self*
- *son*
- *priesthood*
- *tradition*



Questions?

