

THE PROBLEM OF SEMANTIC SHIFT IN LONGITUDINAL MONITORING OF SOCIAL MEDIA

A CASE STUDY ON MENTAL HEALTH DURING THE COVID-19 PANDEMIC

KEITH HARRIGIAN & MARK DREDZE

JOHNS HOPKINS UNIVERSITY

Education

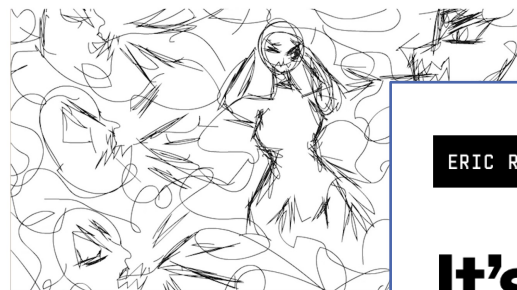
Partly hidden by isolation, many of the nation's schoolchildren struggle with mental health

Perspective Mental Health and the Covid-19 Pandemic

Betty Pfefferbaum, M.D., J.D., and Carol S. North, M.D., M.P.E.

'I'm literally breaking inside': As COVID-19 leaves millions jobless and struggling, the mental health toll rises

Charisse Jones USA TODAY
Published 12:01 a.m. ET Feb. 15, 2021 | Updated 3:27 p.m. ET Feb. 15, 2021



Schizophrenia: Amid COVID-19, mental health center clients m
Mental health treatment through telemedicine helps, but patients with schizophrenia at D.C. to face contact and routine. Urban Health Media Project

ERIC RAVENSCRAFT SCIENCE 01.18.2021 09:00 AM

It's Not Just You: Everyone's Mental Health Is Suffering

If you're thinking, "Oh, I just need to suck it up," stop. What you're feeling is real. Here's how to cope.



This Issue Views **196,192** | Citations **260** | Altmetric **859** | Comments **4**



Viewpoint

FREE



April 10, 2020

More ▾

The Mental Health Consequences of COVID-19 and Physical Distancing
The Need for Prevention and Early Intervention

Sandro Galea, MD¹; Raina M. Merchant, MD²; Nicole Lurie, MD³

[Author Affiliations](#) | [Article Information](#)

JAMA Intern Med. 2020;180(6):817-818. doi:10.1001/jamainternmed.2020.1562

Related Articles

MENTAL HEALTH DURING THE COVID-19 PANDEMIC

A Mental Health Crisis

- Healthcare experts hypothesized that COVID-19 would have broad negative effects on mental health

Call to Action

- Provide data that allows inspection of this hypothesis
- Guide policy makers (e.g., public funding decisions, creation of new social programs)





Social Media Monitoring Has Advantages Over Traditional Mechanisms



Real-time updates at scale



Multiple perspectives (e.g., municipalities, outcome measures)



Retrospective analysis



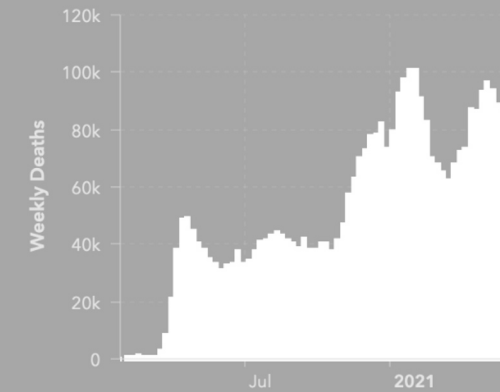
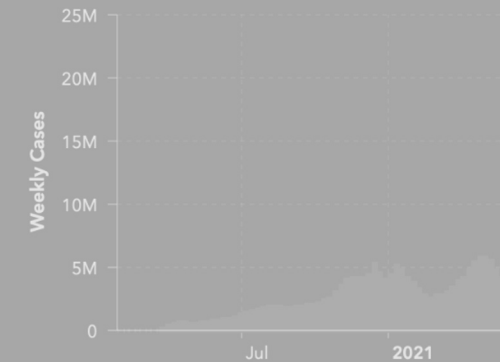
Mitigate sample biases

Stigmatized topics

Survey “opt-in” bias

... Doses Administered
4,609,836

... Doses Administered
6,608,730



Las
2/

US
28-Da
Totals:

Franc
28-Da
Totals:

India
28-Da
Totals:

Italy
28-Da
Totals:

Brazil
28-Da
Totals:

Germ
28-Da
Totals:

Unite
28-Da
Totals:

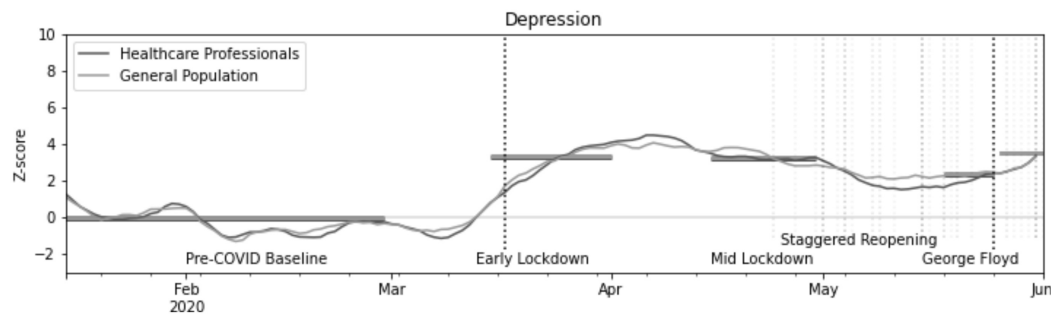
Spain
28-Da
Totals:

Argentina
28-Day: 2,279,035 | 5,192
Totals: 1,000,000 | 1,234,567

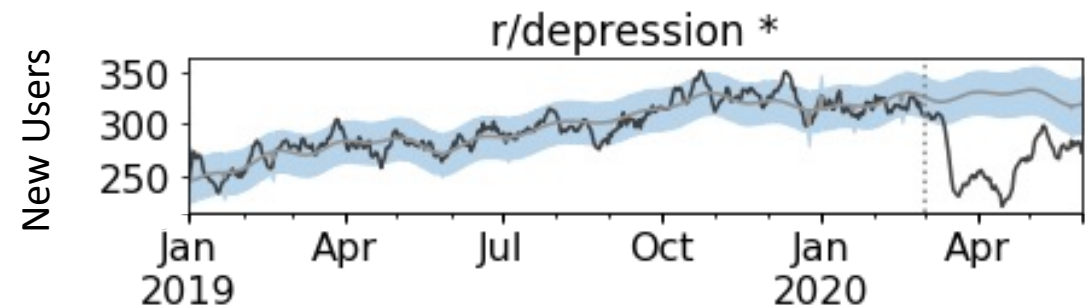
DIFFERENCE BETWEEN THEORY AND PRACTICE

Conflicting Results

- Sensitivity to the target metric (e.g., language-based estimates of mental health status vs. activity patterns)
- Platform-specific outcomes
- Variation due to parameterization choice (same underlying methodology/data)

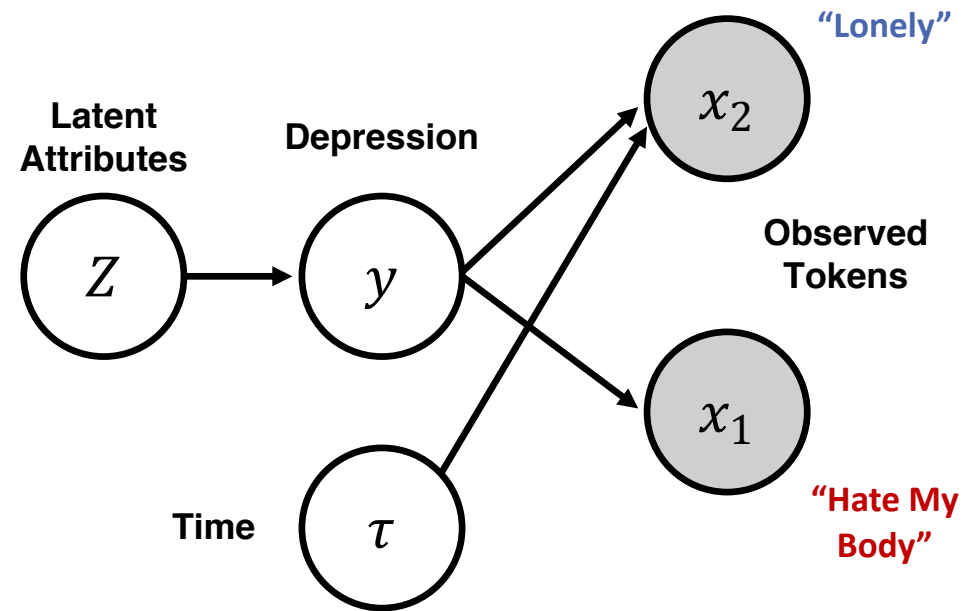
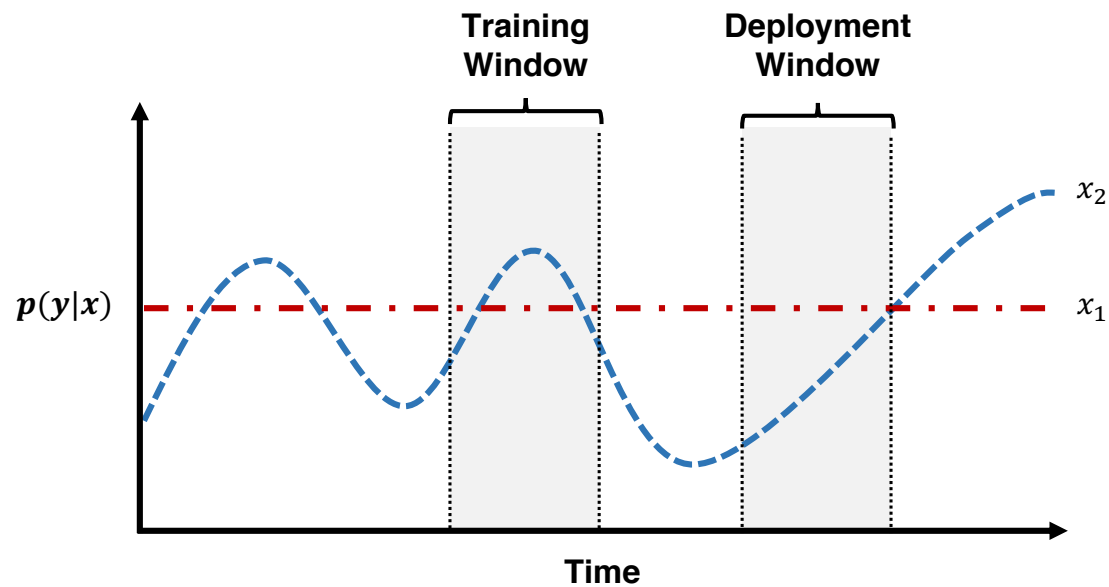


Fine et al. (2020)



Biester et al. (2020)

THE PROBLEM OF SEMANTIC SHIFT



Language meaning *evolves* from one domain (e.g., time period) to another

THE PROBLEM OF SEMANTIC SHIFT

“Looks like another day goes by **without leaving the house**”

“This is a good reminder to reach out to your **vulnerable** friends”

“These **panics** are really starting to interfere with my daily life...”

“Is it weird that I kind of like being **isolated**?”

OUR FOCUS

Technical Aim: Improve Generalization

Can we improve predictive generalization over time using semantically stable vocabularies?

Research Aim: Understand Practical Effects of Semantic Shift

How does semantic shift affect downstream outcomes measured through longitudinal analyses of social media?



IMPROVING GENERALIZATION



EXPERIMENTAL DESIGN

Hypothesis: Classifiers trained using semantically stable vocabularies perform as well as (or better than) alternative feature selection approaches

Method: Compare classification performance on future temporal data as a function of vocabulary size and vocabulary/feature selection technique

Platform	Dataset	Time Range	# Users	# Posts
Twitter	CLPsych 2015 Shared Task	2012 to 2014	732	1.4M
	Multi-Task Learning	2012 to 2016	2,547	7.8M
Reddit	Topic-Restricted Text	2016 to 2020	8,726	9.1M
	Self-Reported Mental Health Diagnoses (SMHD)	2013 to 2018	11,664	2.9M

Annotated Datasets. Task is to infer whether user is part of Depression or Control group.

MEASURING SEMANTIC SHIFT

“Simple, Interpretable, and Stable Method for Detecting Words with Usage Change across Corpora” (Gonen et al., 2020)

1. Fit word embeddings within each independent domain
2. Identify Top-K neighbors of each word in shared vocabulary
3. Measure set overlap between neighborhoods

$$S(w; \mathcal{P}, \mathcal{Q}) = \frac{\text{nb}_{\mathcal{P}}^{(k)}(w) \cap \text{nb}_{\mathcal{Q}}^{(k)}(w)}{k},$$

VOCABULARY SELECTION

Selection Type	Method	Description
Naïve Baselines	Cumulative	Frequency > 100 in the training time period
	Intersection	Frequency > 100 in training <i>and</i> test time periods
	Frequency	Top K Most Frequent Terms
	Random	Randomly Selected Sample of Size K
Statistical Baselines	Chi Squared	Top K Terms with largest deviation from expected distribution
	Coefficient	Top K Terms with highest absolute task coefficient in the training time period
Semantically Aware	Overlap	Top K Terms with smallest semantic shift from training to deployment time periods
	Weighted Overlap	50/50 weighted combination of overlap and coefficient selection scores

	Dataset	Train	Test	Naïve			Statistical		Semantic	
				Cumulative	Intersection	Frequency	Chi-Squared	Coefficient	Overlap	Weighted
Twitter	CLPysch	2012-2013	2013-2014	0.656	0.677	0.676	0.687	0.677	0.715*	0.696
	Multi-Task Learning	2012-2013	2013-2014	0.746	0.759	0.759	0.761	0.757	0.779*	0.772
			2014-2015	0.703	0.760	0.760	0.762	0.758	0.778*	0.765
			2015-2016	0.699	0.775	0.773	0.777	0.772	0.783*	0.772
		2012-2014	2014-2015	0.778	0.779	0.778	0.781	0.783	0.781	0.786
			2015-2016	0.788	0.787	0.787	0.789	0.792	0.789	0.791
	2012-2015	2015-2016	0.799	0.800	0.800	0.800	0.806	0.802	0.806	
Reddit	Topic Restricted Text	2016-2017	2017-2018	0.659	0.662	0.661	0.660	0.660	0.661	0.661
			2018-2019	0.670	0.669	0.668	0.668	0.668	0.666	0.668
			2019-2020	0.670*	0.665	0.663	0.665	0.667	0.666	0.667
		2016-2018	2018-2019	0.667	0.672	0.671	0.671	0.669	0.674	0.669
			2019-2020	0.672	0.674	0.674	0.674	0.674	0.675	0.674
		2016-2019	2019-2020	0.667	0.668	0.669	0.668	0.668	0.674*	0.670
Reddit	SMHD	2013-2014	2014-2015	0.799	0.798	0.803	0.799	0.799	0.799	0.799
			2015-2016	0.801	0.800	0.800	0.805	0.801	0.802	0.802
			2016-2017	0.792	0.792	0.793	0.798	0.797	0.792	0.799
			2017-2018	0.799	0.800	0.800	0.803	0.804	0.804	0.808
		2013-2015	2015-2016	0.797	0.795	0.798	0.799	0.798	0.801	0.799
			2016-2017	0.786	0.785	0.787	0.790	0.790	0.788	0.791
			2017-2018	0.796	0.796	0.802	0.799	0.804	0.804	0.807
		2013-2016	2016-2017	0.790	0.790	0.791	0.792	0.793	0.792	0.794
			2017-2018	0.798	0.796	0.804	0.798	0.804	0.806	0.808
		2013-2017	2017-2018	0.799	0.797	0.804	0.800	0.803	0.808	0.810

Bold = Best Performer | * = Significant Improvement



UNDERSTANDING PRACTICAL EFFECTS



EXPERIMENTAL DESIGN

Hypothesis: Classifiers will produce outcomes that vary non-trivially as a function of the underlying vocabulary

Method: Estimate change in prevalence of depression using classifiers trained on vocabularies with varying levels of semantic stability

Platform	Dataset	Time Range	# Users	# Posts
Twitter	1% Research Stream	Jan. 2019 to July 2020	25,379	10.6M
Reddit	Pushshift.io	Jan. 2019 to July 2020	40,671	30.9M

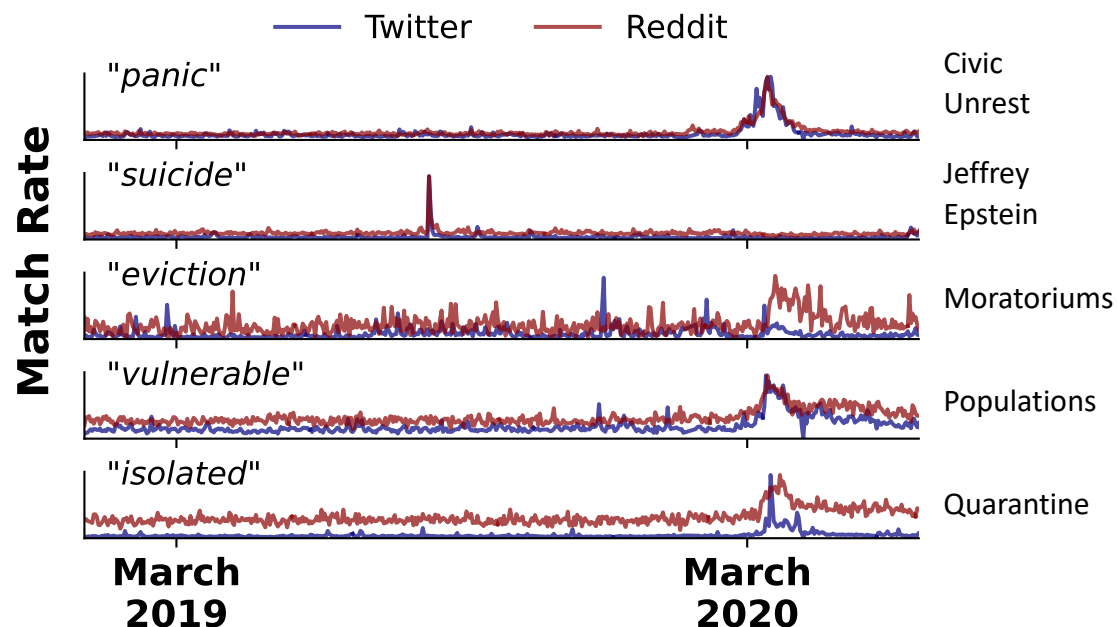
Unlabeled datasets. Minimum activity thresholds and English-language filtering applied to sample cohort.

IDENTIFYING SEMANTIC SHIFT

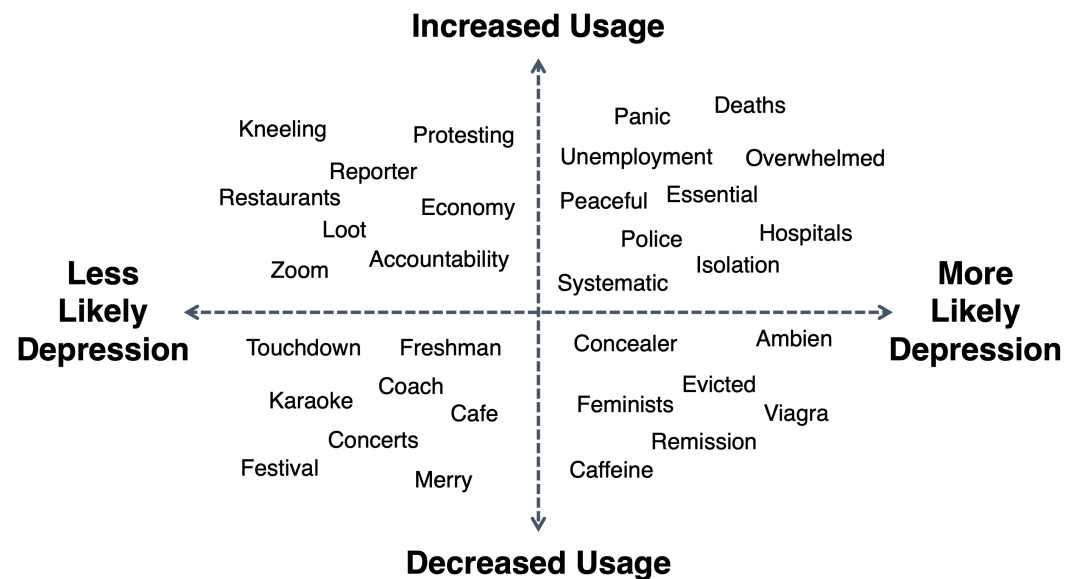
Term	2019 Context	2020 Context
Panic	Emotion (i.e., Fear) rage, meltdown, anxiety, anger, barrage, migraine, phobia, outrage, manic, rush, asthma	Panic Buying, Misinformation hysteria, chaos, fear, misinformation, confusion, frenzy, paranoia, mayhem, insanity, fearmongering
Cuts	Physical cut, jumps, runs, cutting, pulls, moves, bounces, falls, turns, burns, drags, dips, breaks, bursts, rips, goes, bumps	Economic cut, cutting, subsidies, budgets, deductions, revenues, checks, payments, breaks, deals, figures, loans, deposits, gains
Isolated	Feeling Detached unpleasant, unstable, detached, unsafe, populated, invasive, unknown, confined, endangered, absent, vulnerable, insulated	Quarantine quarantined, isolating, separated, enclosed, insulated, infectious, confined, active, populated, autonomous, vulnerable, detached
Strain	Discomfort/Pressure inflammation, deficiency, dose, stress, pressure, calcium, medication, concentration, tissue, nausea, receptors, doses, acne	Virus disease, illness, infections, symptom, mutation, virus, outbreak, pneumonia, infection, strains, influenza, epidemic
Vulnerable	Emotion susceptible, dangerous, prone, unstable, aggressive, hostile, disruptive, detrimental, receptive, fragile, damaging, sensitive	At-risk Populations susceptible, dangerous, immunocompromised, infectious, isolating, elderly, disadvantaged, contagious, tolerant, likely, isolated, symptomatic
Doctors	Personal Experience psychiatrists, medically, clinic, cps, accountants, police, miscarriages, malpractice, abortions, prescribe, counseling, procedure	Frontline Workers, PPE midwives, #nurses, #doctors, epidemiologists, emts, front-line, #coronawarriors, frontliners, virologists, masks, ppe, respirators, heroes

WHY DO WE CARE ABOUT SEMANTIC SHIFT?

Keywords

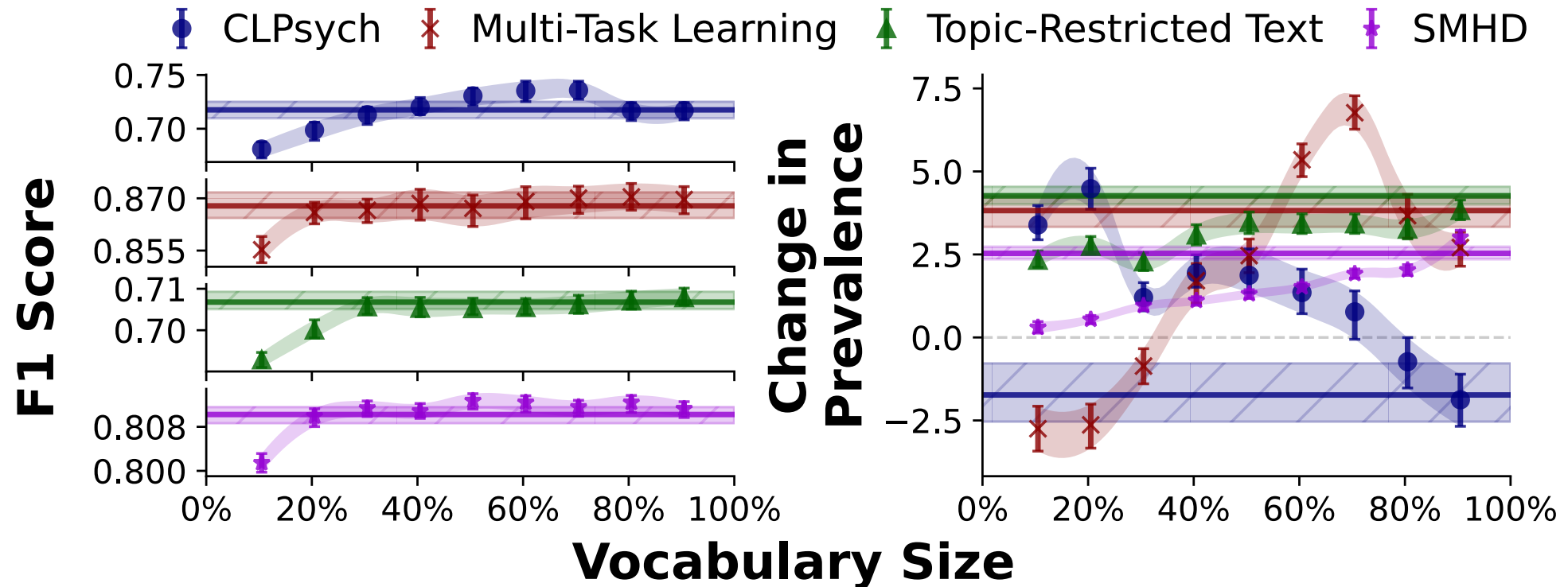


Model Coefficients



Without social context, statistical classifiers may misinterpret changes in usage

EFFECT ON DOWNSTREAM OUTCOMES



Significant variability in prevalence estimates despite insignificant changes in historical predictive performance



CLOSING THOUGHTS



ETHICAL CONSIDERATIONS

Discrimination

- A concern with any modeling approach used to understand personal attributes
- Can have severe financial and social consequences
 - Ostracization from communities, loan rejections

Measurement Error

- Improper assessment of uncertainty in measurements can lead decision makers astray
- Proper estimation of confidence in inferences can mitigate issues with traditional surveillance techniques

Can we trust the models that are already deployed?

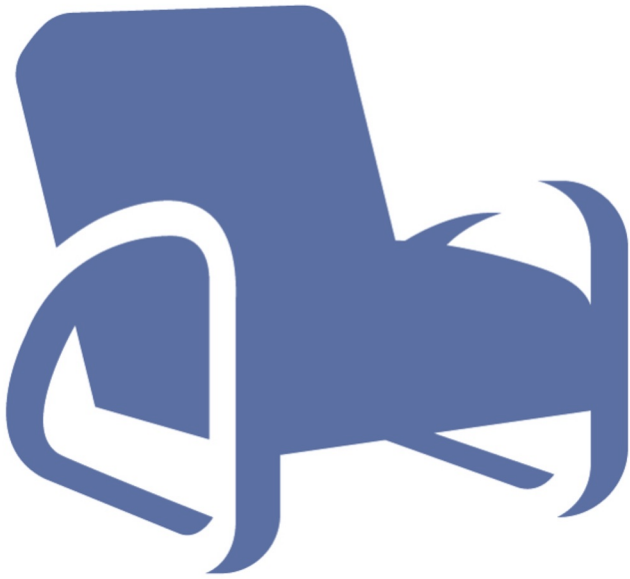
CONCLUSIONS

What we *are not* saying

- Gonen et al.'s method for measuring semantic stability is the best approach for reducing effects of semantic shift on downstream analyses
- Depression prevalence has increased or decreased amongst the population

What we *are* saying

- More attention needs to be paid to the possibility that semantic shift affects longitudinal social media analyses
- Practitioners should consider running similar analyses to our own to understand whether their own work may be susceptible to semantic shift



“Just as war makes every citizen into an amateur geographer and tactician, a pandemic makes epidemiologists of us all.”

“Whether the objective is prediction or inference, it is essential to account for how behavior and other factors contributing to transmission – and observation – may change over time.”

Zelner, J., Riou, J., Etzioni, R., & Gelman, A. (2021). Accounting for uncertainty during a pandemic.

CONTACT US

Keith Harrigian, MSE



Email: kharrigian@jhu.edu

Website: <https://kharrigian.github.io>

Mark Dredze, PhD

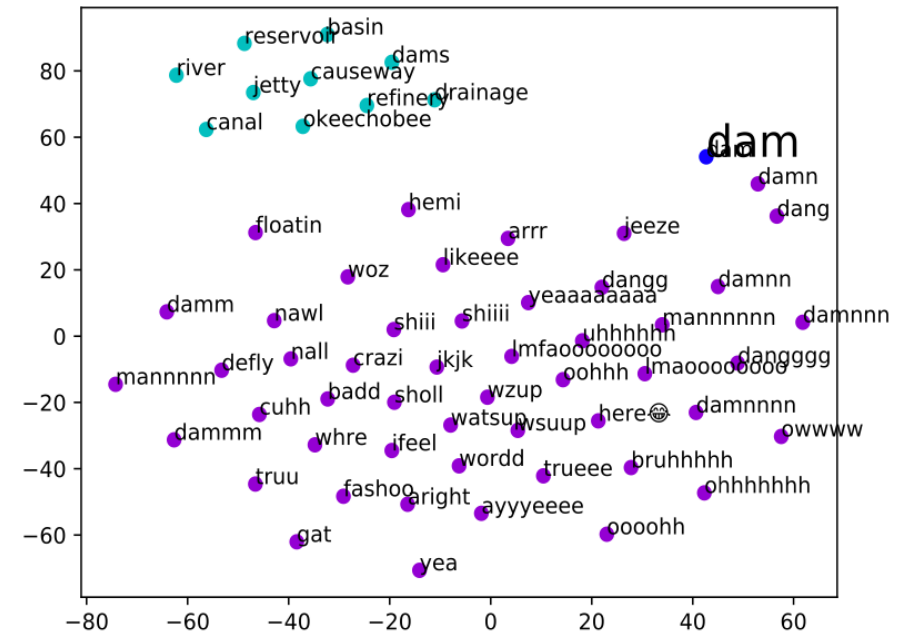


Email: mdredze@cs.jhu.edu

Website: <https://www.cs.jhu.edu/~mdredze/>

MEASURING SEMANTIC SHIFT

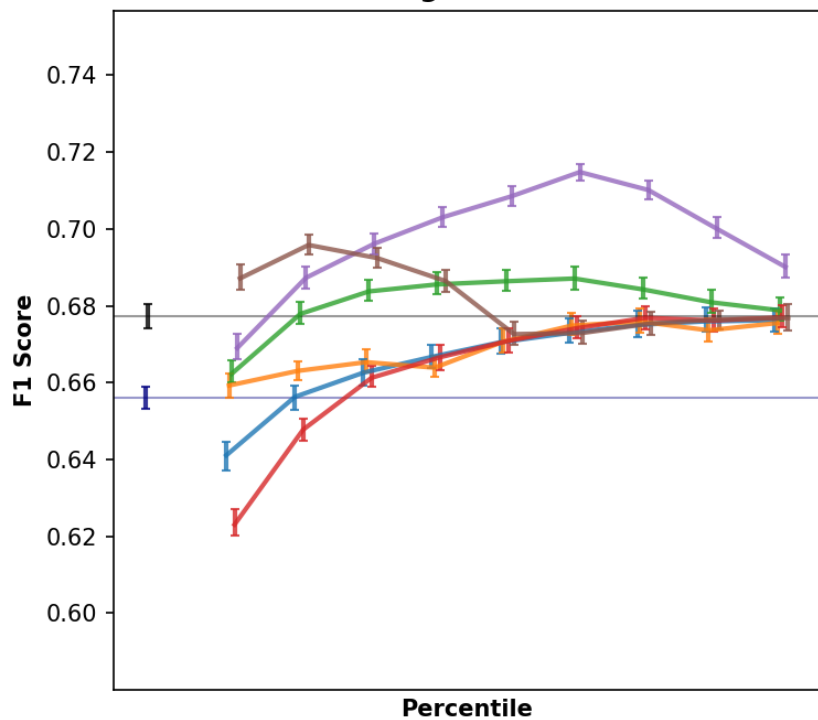
- **Simplicity:** easy to implement and tune using open-source libraries
- **Stability:** robust to stochasticity from embedding training procedure
- **Interpretability:** intuitive to understand
- **Locality:** removes dependency on projection which implicitly considers entire vector space



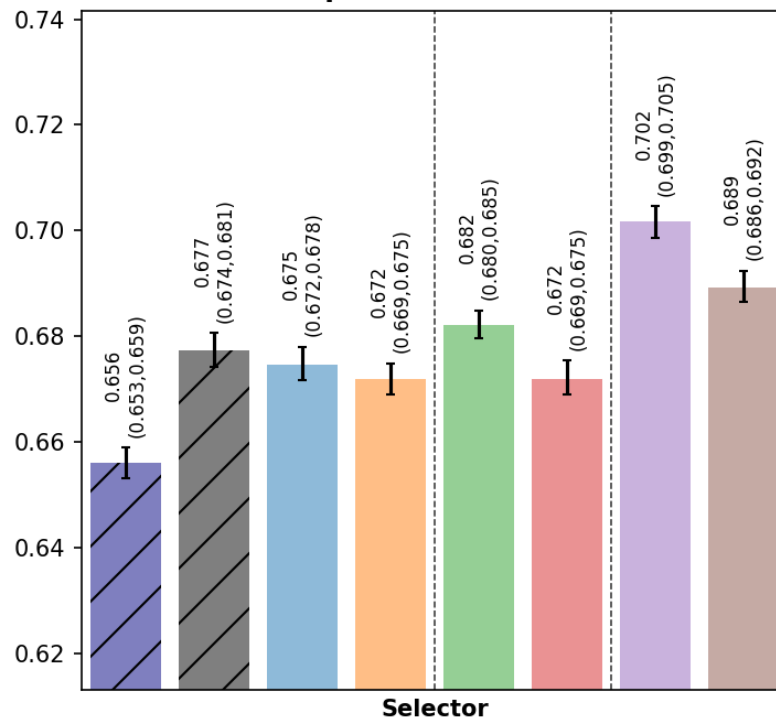
(a) Young space

RESULTS (PARAMETER SWEEP)

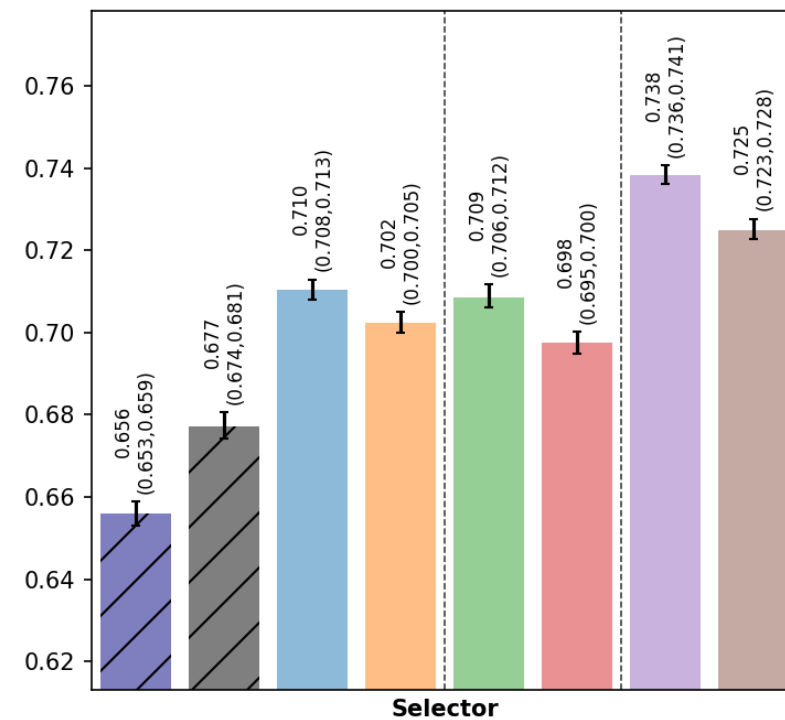
Average F1 Score



Optimal F1 Score

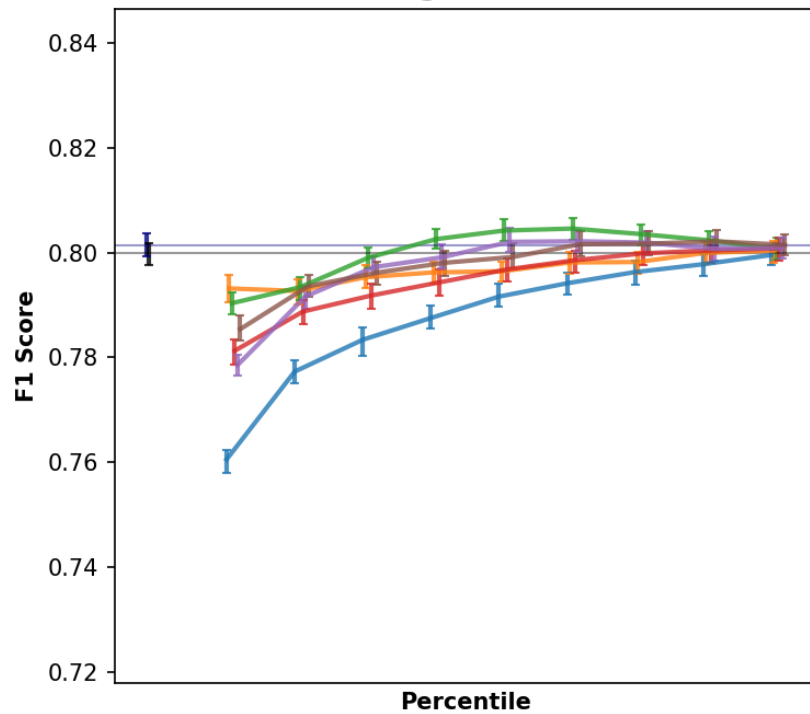


Oracle F1 Score

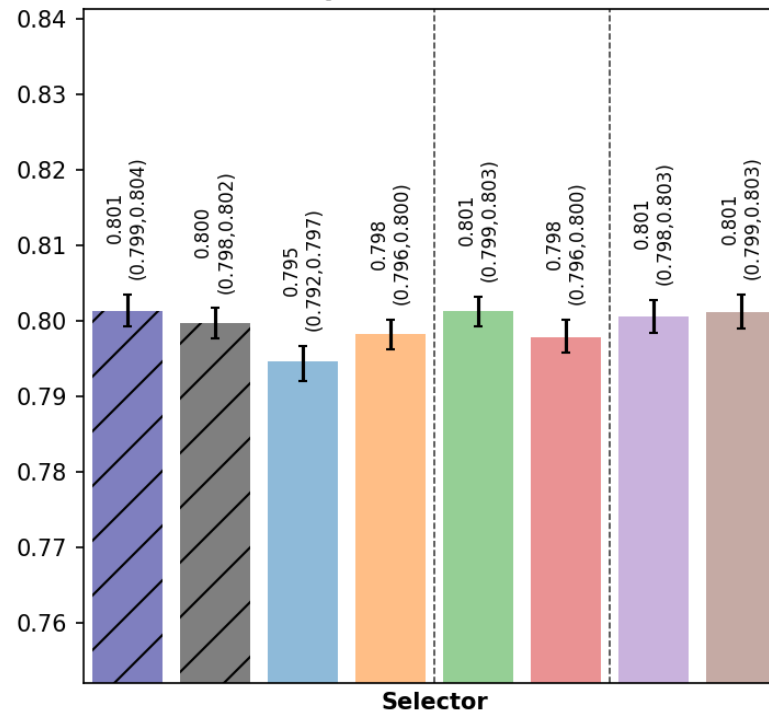


RESULTS (PARAMETER SWEEP)

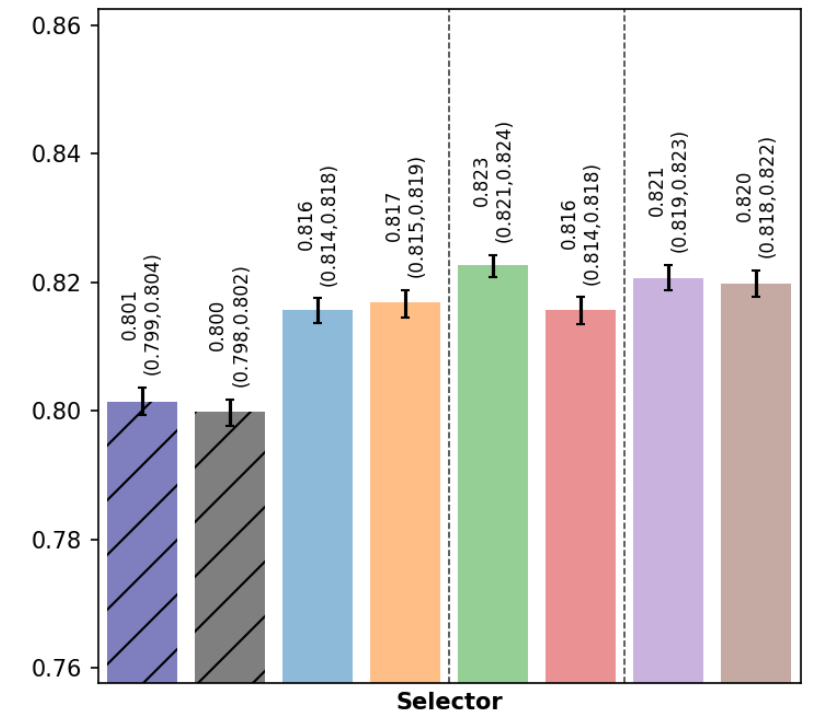
Average F1 Score



Optimal F1 Score



Oracle F1 Score



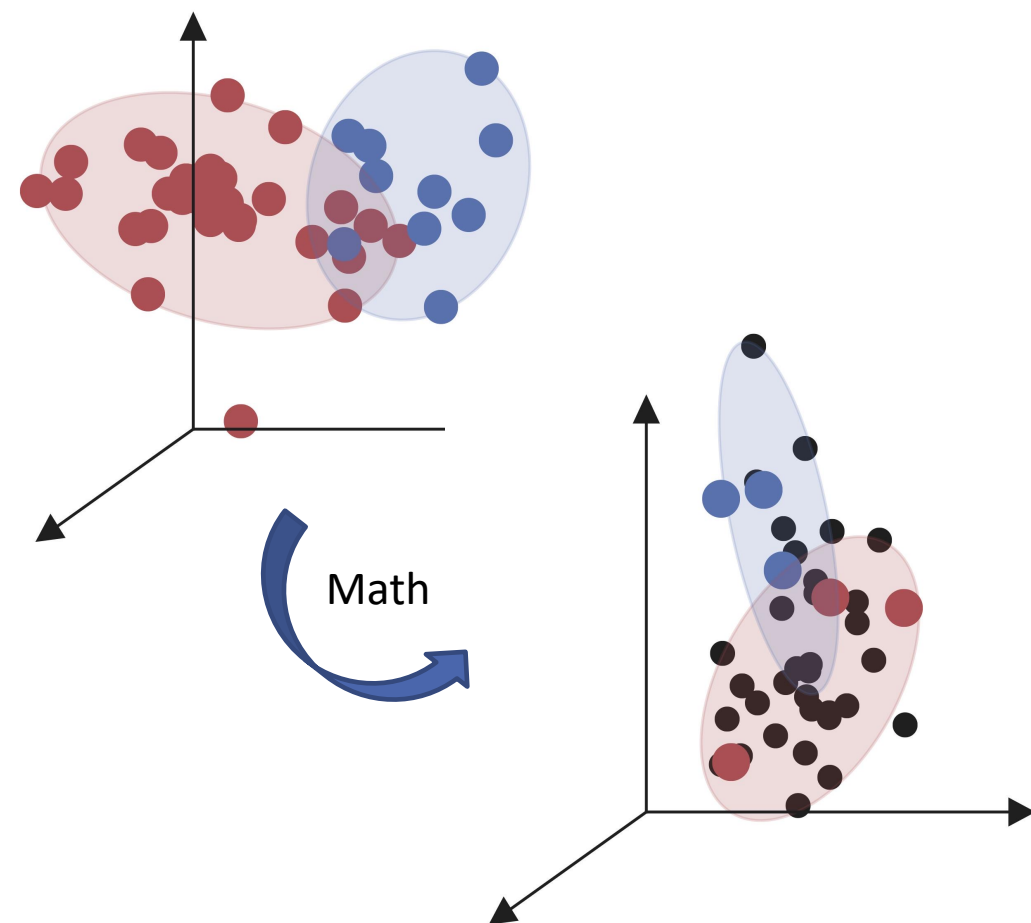
MEASURING SEMANTIC SHIFT (TRADITIONAL APPROACH)

Predominant Approach

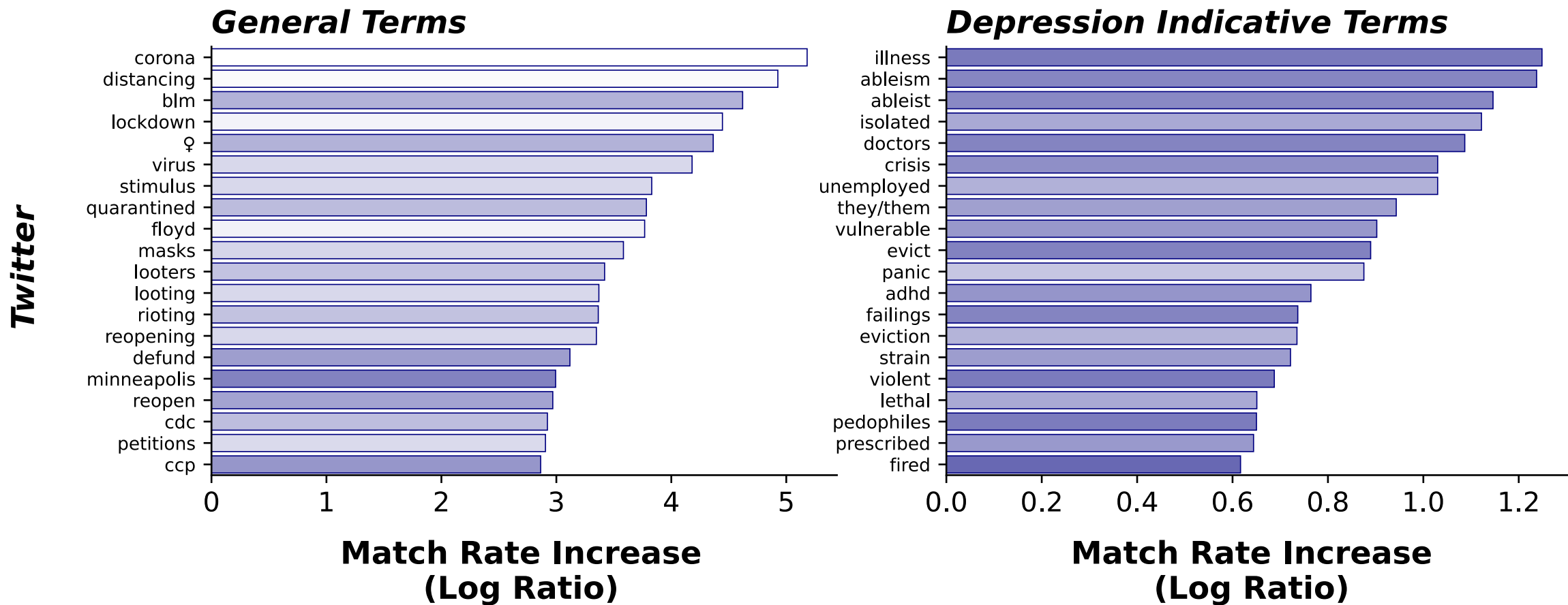
- Fit word embeddings for each domain (i.e. platform, time period)
- Align the embedding vector spaces
- Measure vector similarity (e.g. cosine distance)

Shortcomings

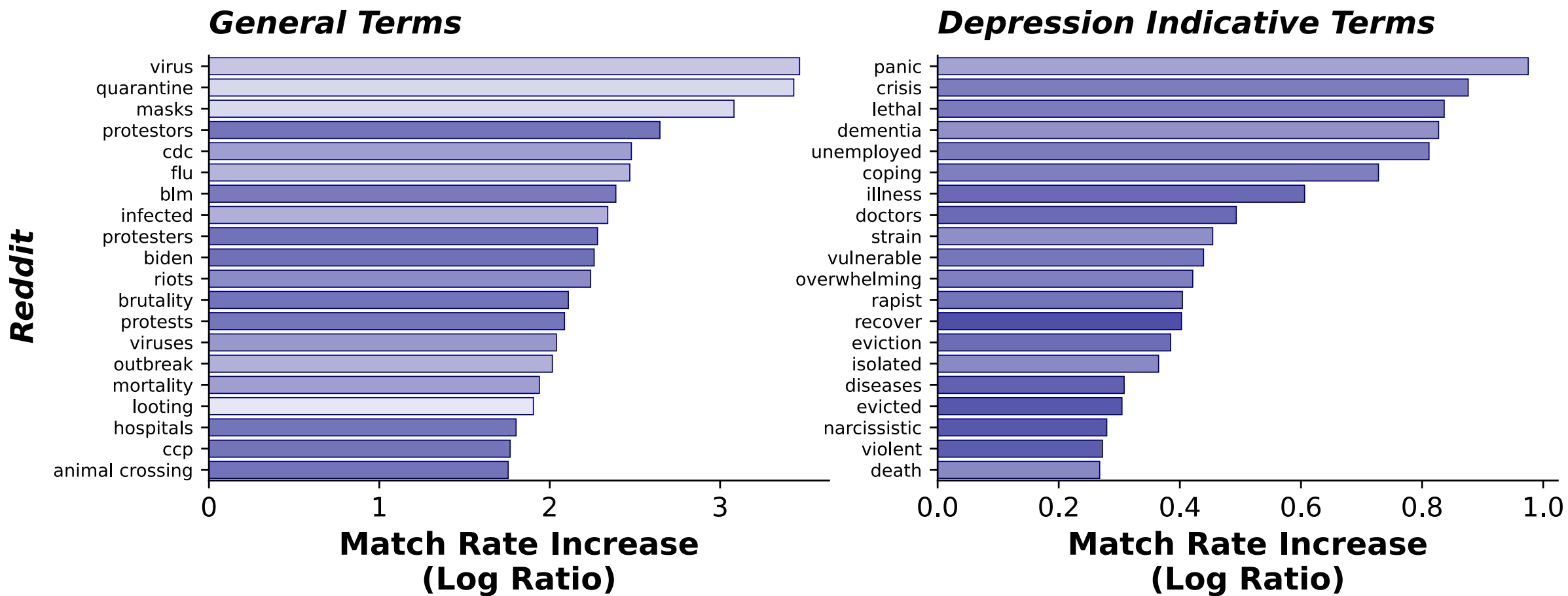
- Self-contradicting objective
- Requires non-trivial filtering
- Stochastic Instability



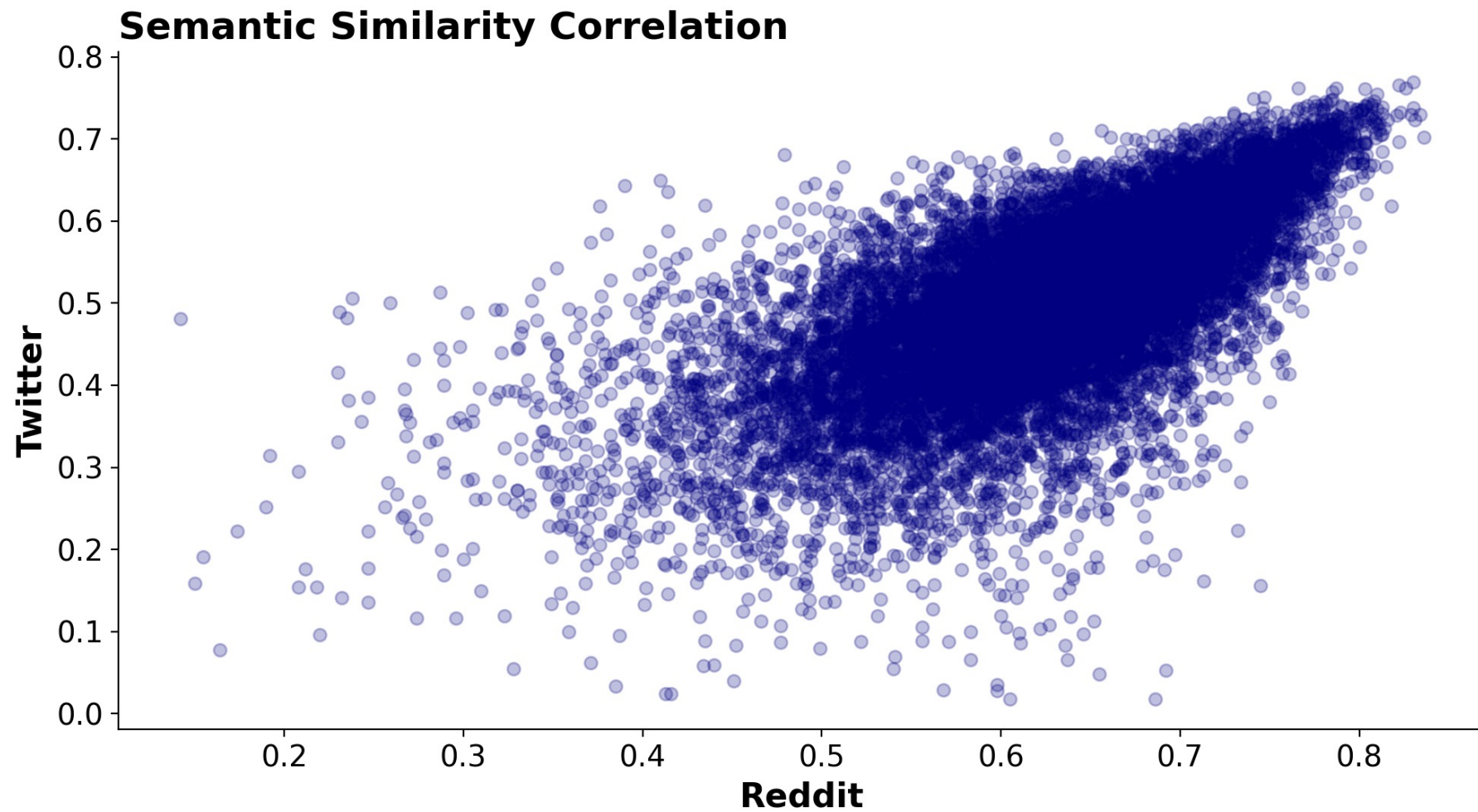
EDA: SEMANTIC SHIFT VS. FREQUENCY CHANGE



EDA: SEMANTIC SHIFT VS. FREQUENCY CHANGE



EDA: CROSS-PLATFORM SEMANTIC SIMILARITY



EDA: POTENTIAL IMPACT OF SEMANTIC SHIFT

