FRIENDSHIPS, RIVALRIES, AND TRYSTS: CHARACTERIZING RELATIONS BETWEEN IDEAS IN TEXTS

Chenhao Tan

Dallas Card (CMU), Noah Smith (UW)

RELATIONS BETWEEN IDEAS

pro-choice ←— **rivals** —→ pro-life

undocumented immigrants ←— **rivals** —→ illegal alien

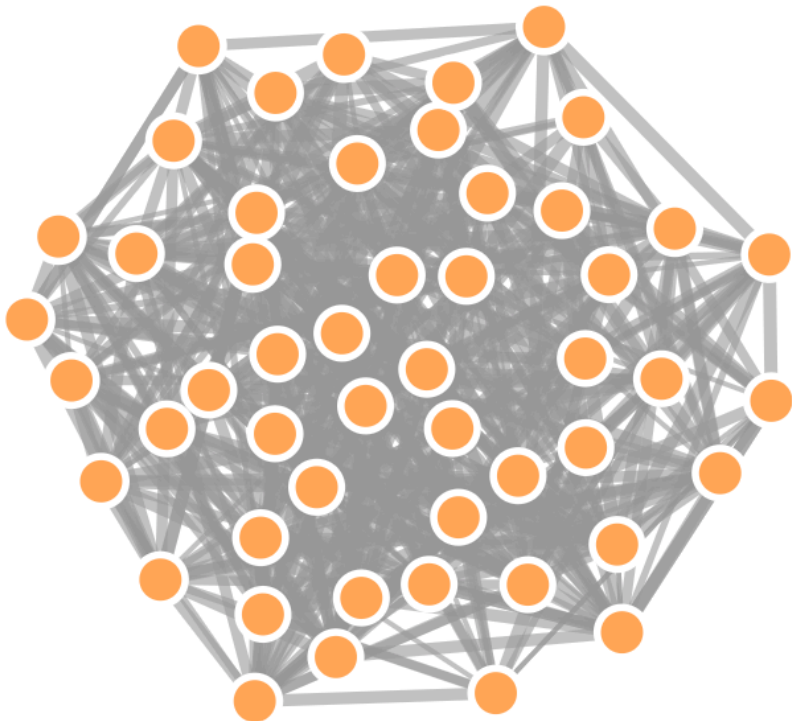small government ←— **friends** —→ free market

word alignment ←— **friends** —→ machine translation

Chong and Druckman, 2007; Dawkins 1976; Entman, 1993; Gitlin, 1980; Lakoff, 2014; Milton 1964

MAIN CONTRIBUTIONS

First **quantitative** framework to systematically describe relations between ideas

Demonstrate **effective explorations** with this framework on a wide range of datasets

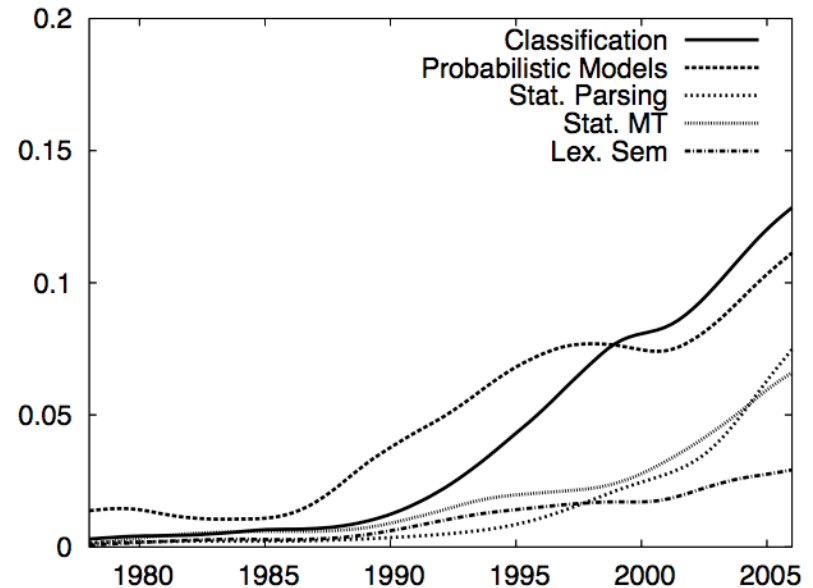undocumented immigrants ← **rivals** → illegal alien

small government ← **friends** → free market
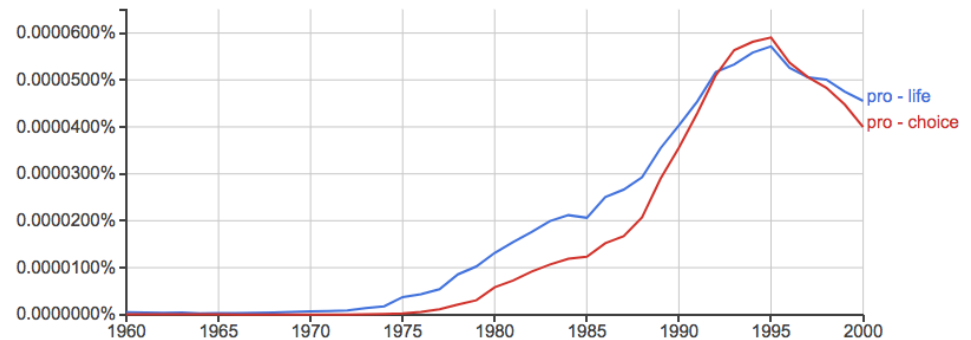
USING TEXT TO TRACE IDEAS

Our focus is on
relations between ideas.

We will use standard approaches
- Topics from latent Dirichlet allocation (Blei et al. 2003)
- Keywords (Monroe et al. 2008)

Hall et al. 2008

Culturomics, Michel et al. 2011

QUANTITATIVELY DESCRIBE RELATIONS BETWEEN IDEAS

- Given a corpus of documents over time, each document consists of a set of ideas

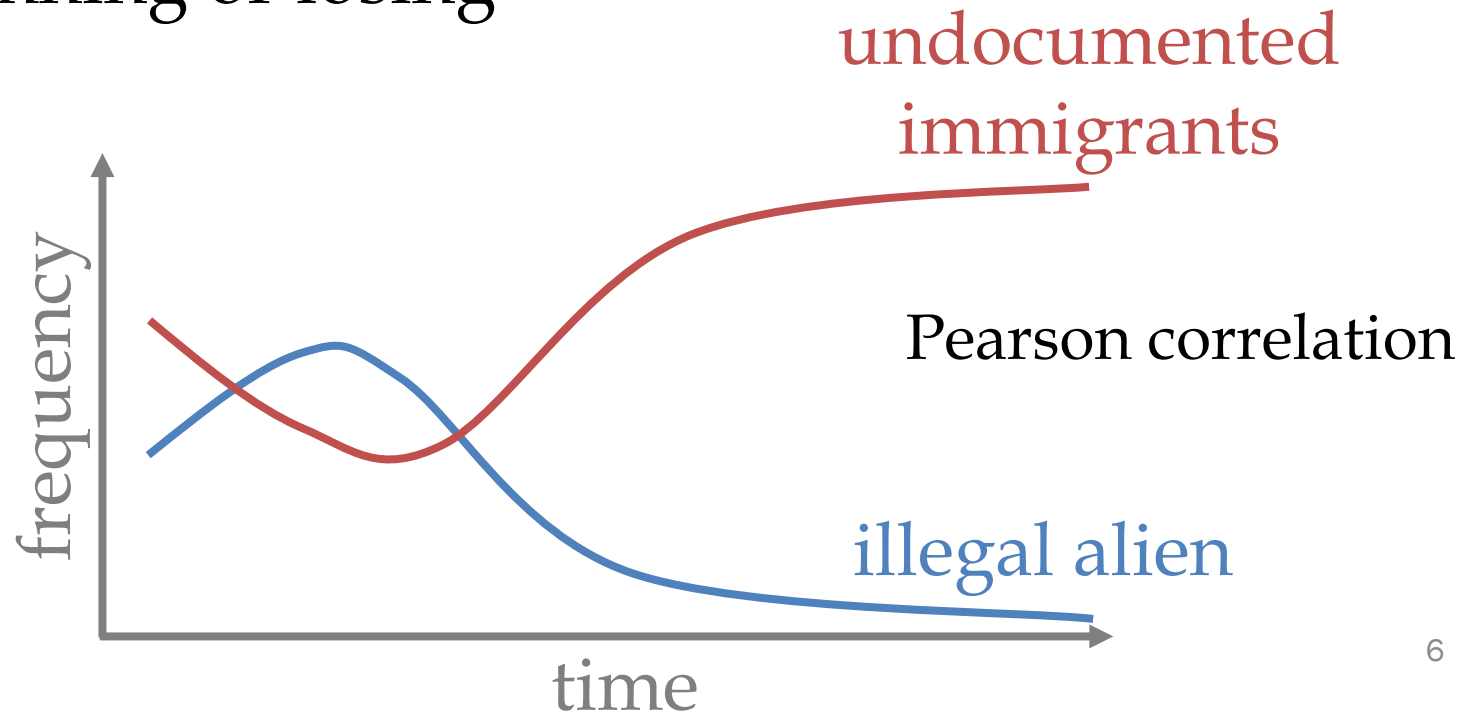undocumented immigrants ←—— rivals ——→ illegal alien

Cooccurrence
  Pointwise mutual information
  [Church and Hanks 1990]

Rarely cooccur

QUANTITATIVELY DESCRIBE RELATIONS BETWEEN IDEAS

- Given a corpus of documents over time, each document consists of a set of ideas
 - Cooccurrence does not capture which is winning or losing

undocumented immigrants

Pearson correlation

illegal alien

frequency

time

QUANTITATIVELY DESCRIBE RELATIONS BETWEEN IDEAS

- Given a corpus of documents over time, each document consists of a set of ideas

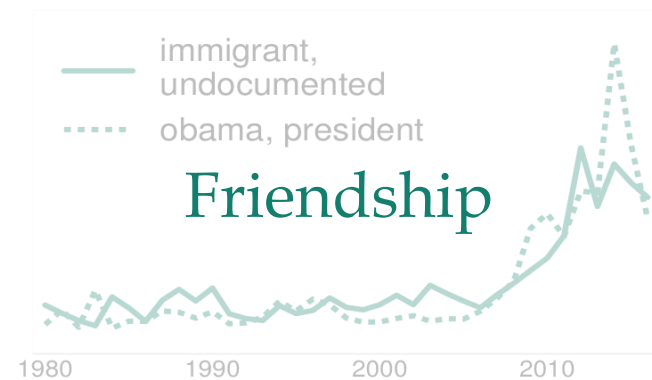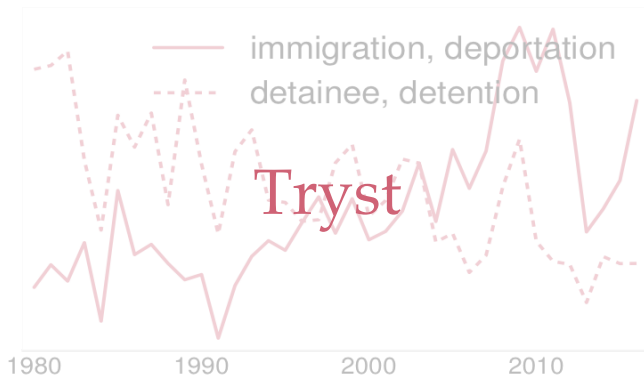Cooccurrence & Prevalence correlation

Within-document

Across-document
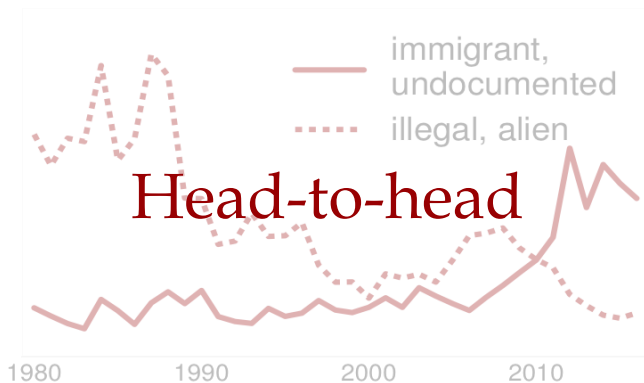
HEAD-TO-HEAD
(ANTI-CORRELATED, RARELY COOCCUR)
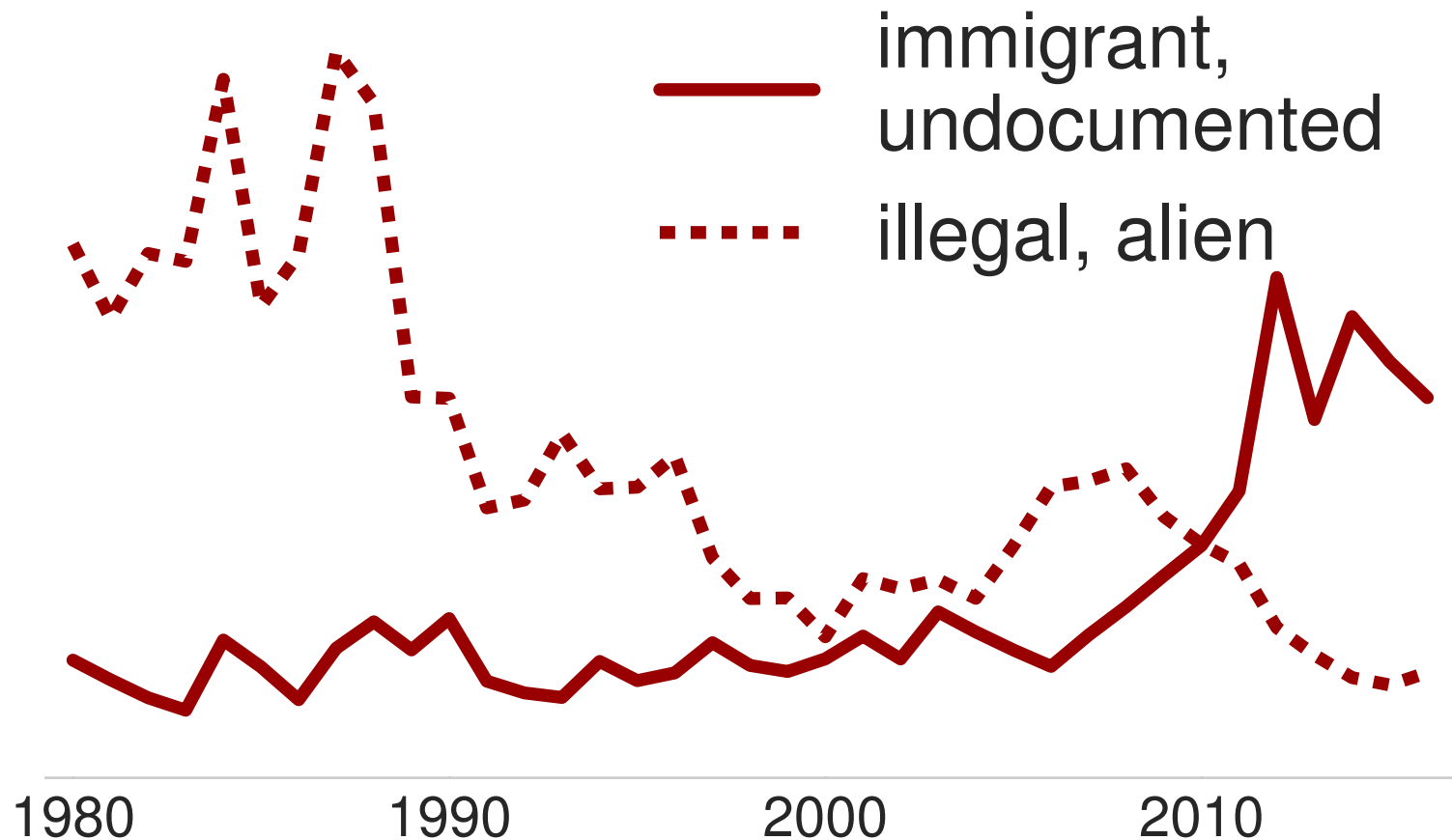
RELATIONS BETWEEN IDEAS
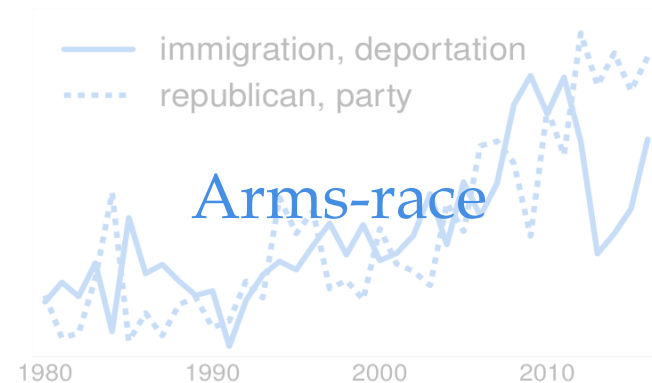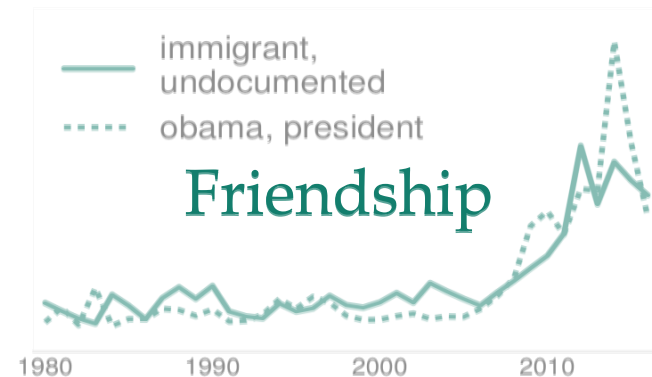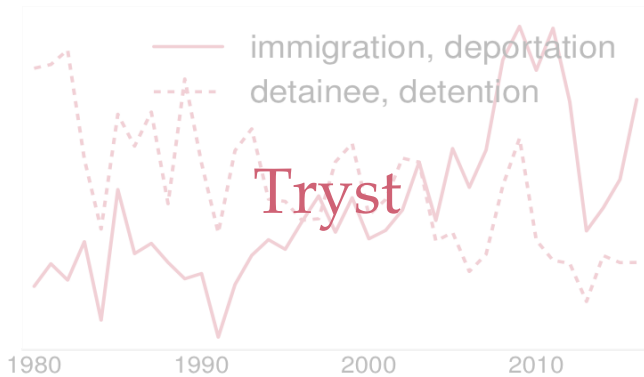

Always cooccur

Anti-correlated — Correlated

Rarely cooccur

Tryst
immigration, deportation
detainee, detention

Friendship
immigrant, undocumented
obama, president

Head-to-head
immigrant, undocumented
illegal, alien

Arms-race
immigration, deportation
republican, party

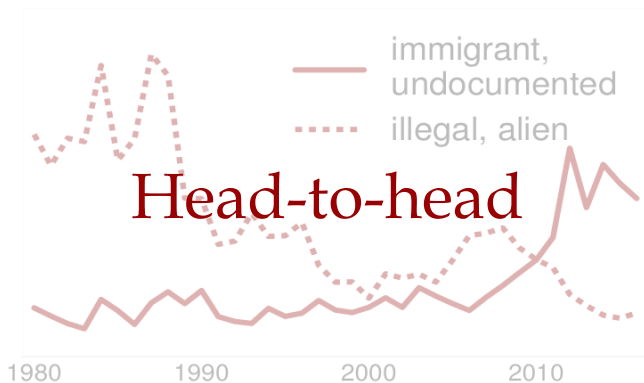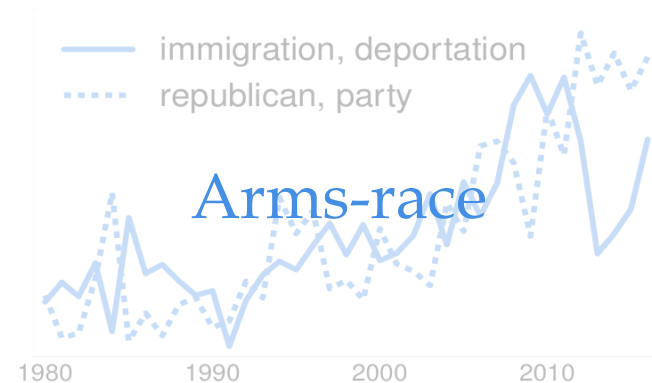# RELATIONS BETWEEN IDEAS

Always cooccur

Anti-correlated

Correlated

immigration, deportation
detainee, detention

Tryst

1980   1990   2000   2010

immigrant,
undocumented
obama, president

Friendship

1980   1990   2000   2010

immigrant,
undocumented
illegal, alien

Head-to-head

1980   1990   2000   2010

immigration, deportation
republican, party

Arms-race

1980   1990   2000   2010
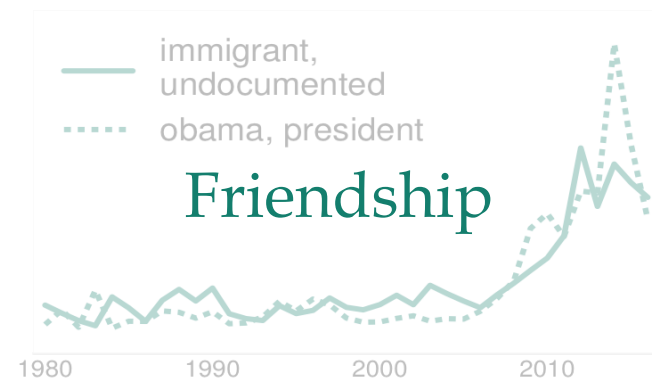
Rarely cooccur

FRIENDSHIP
(CORRELATED, LIKELY TO COOCCUR)
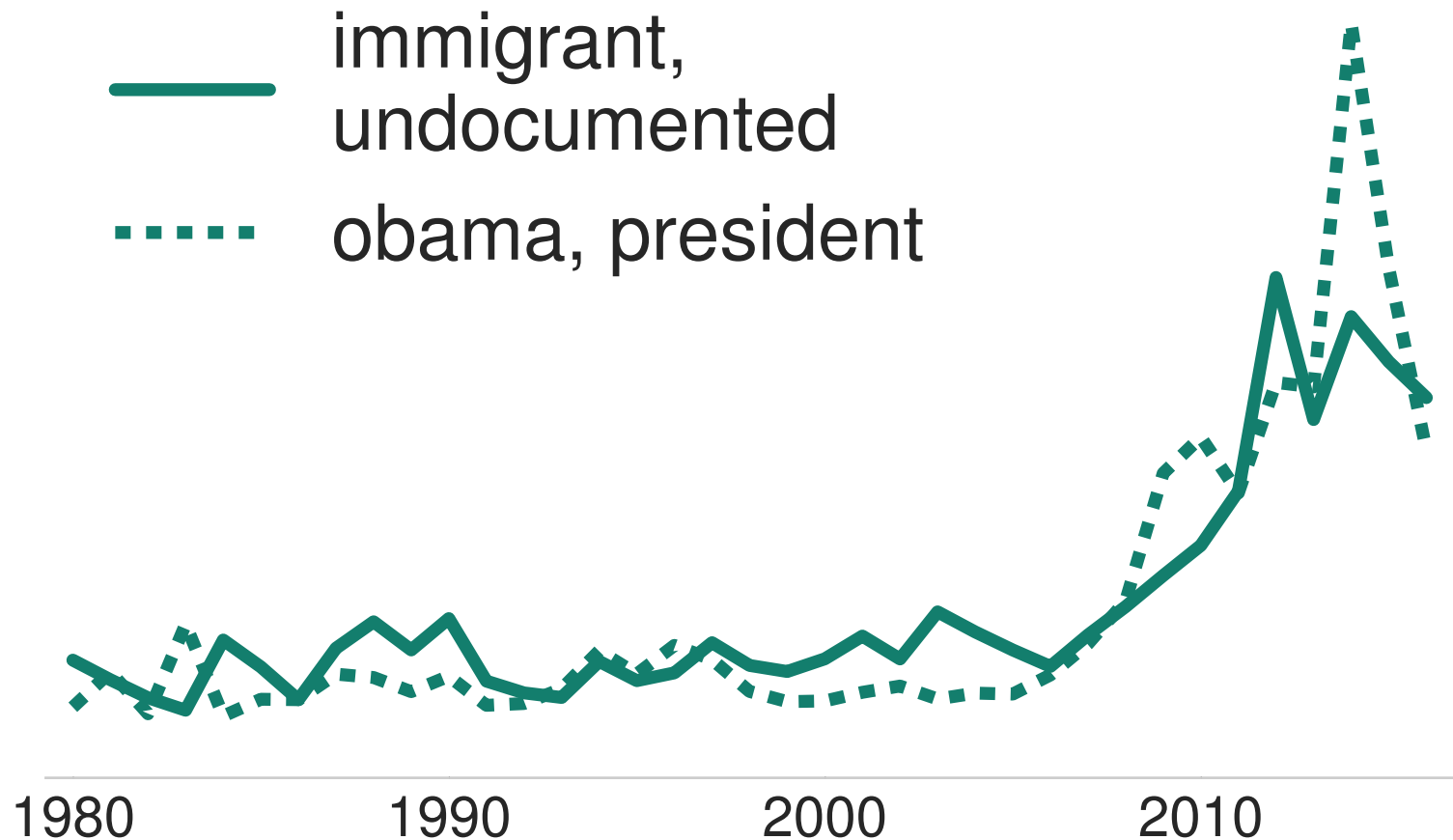
RELATIONS BETWEEN IDEAS

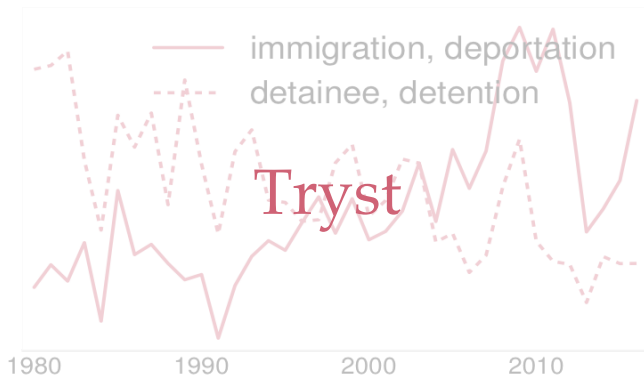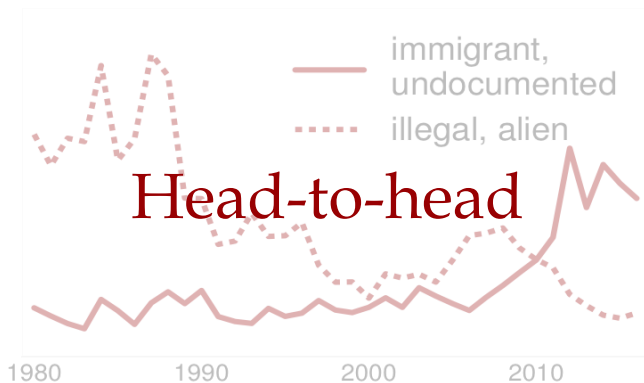
Always cooccur

Anti-correlated — Correlated

Rarely cooccur

Tryst — immigration, deportation / detainee, detention

Friendship — immigrant, undocumented / obama, president

Head-to-head — immigrant, undocumented / illegal, alien

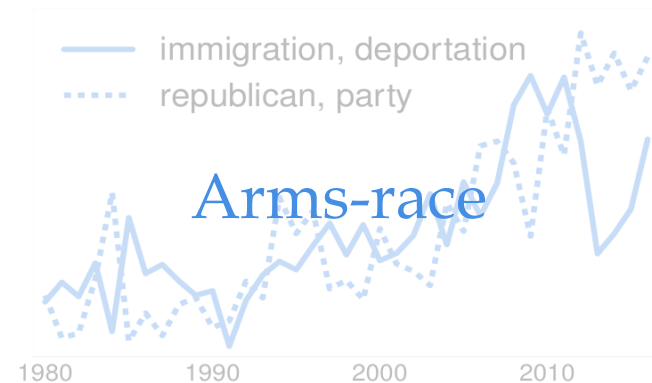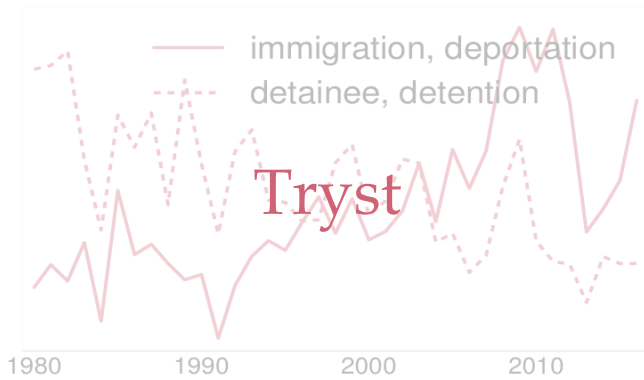**Arms-race** — immigration, deportation / republican, party

RELATIONS BETWEEN IDEAS
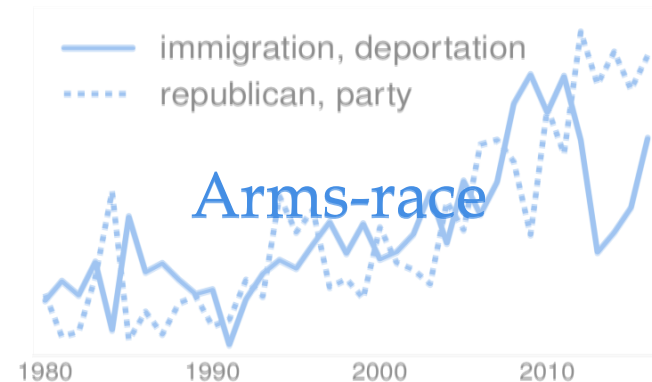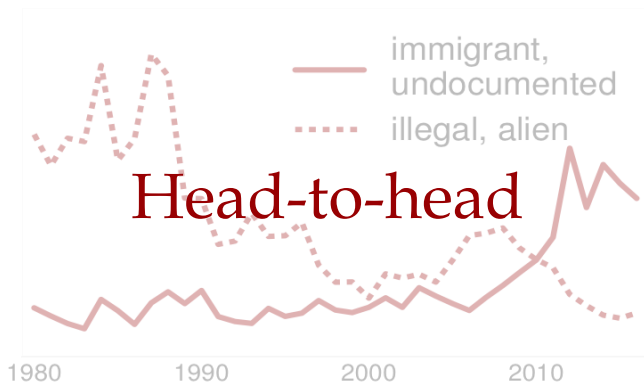

Always cooccur

Anti-correlated

Correlated

Rarely cooccur

Tryst — immigration, deportation / detainee, detention

Friendship — immigrant, undocumented / obama, president

Head-to-head — immigrant, undocumented / illegal, alien
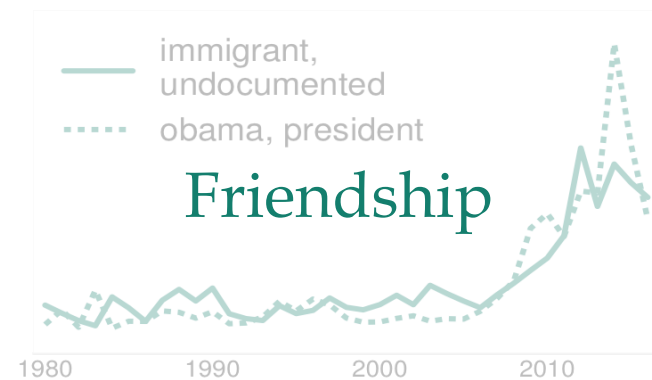
Arms-race — immigration, deportation / republican, party
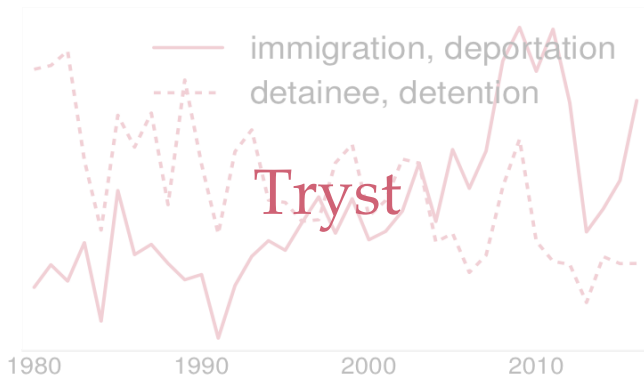
ARMS-RACE
(CORRELATED, RARELY COOCCUR)

— immigration, deportation
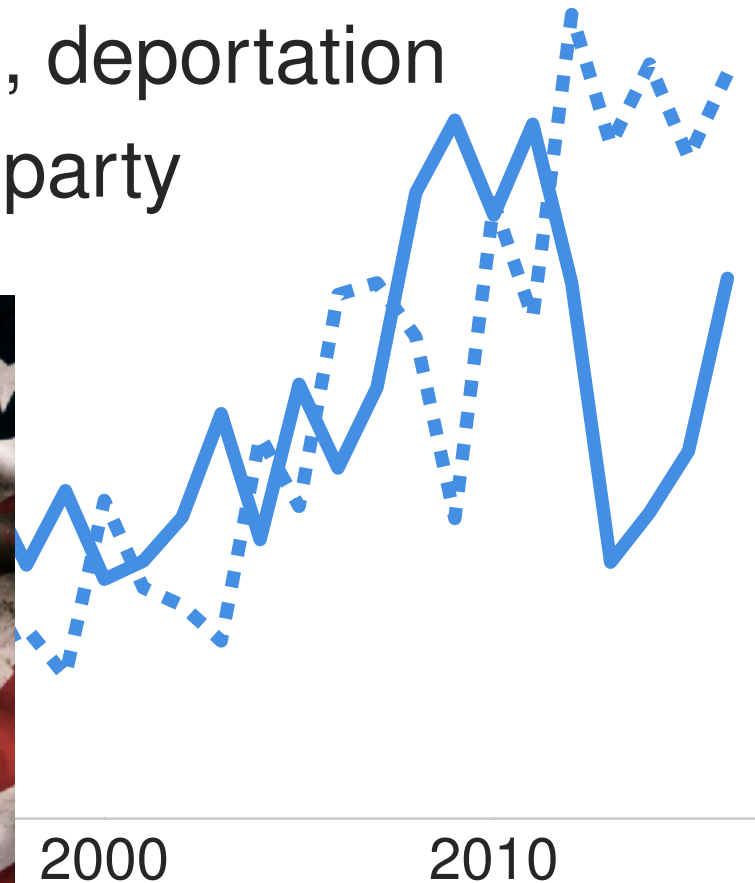
···· republican, party

2000   2010

RELATIONS BETWEEN IDEAS

Always cooccur

immigration, deportation
detainee, detention

Tryst

immigrant,
undocumented
obama, president

Friendship

1980   1990   2000   2010

1980   1990   2000   2010

Anti-correlated   Correlated

immigrant,
undocumented
illegal, alien

Head-to-head

immigration, deportation
republican, party

Arms-race

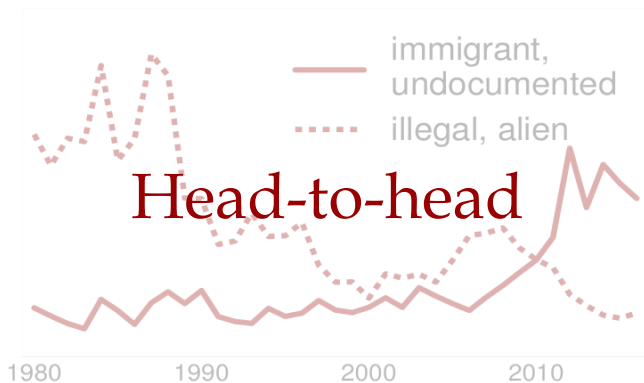1980   1990   2000   2010

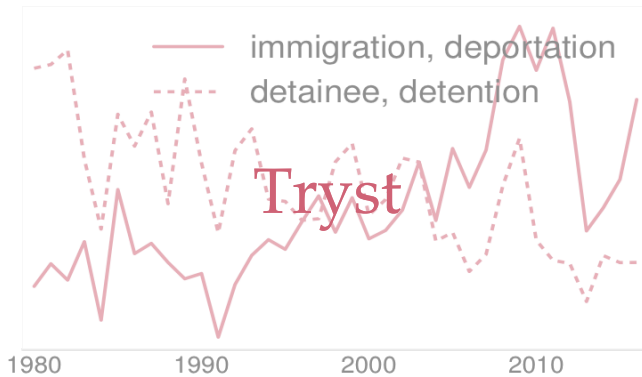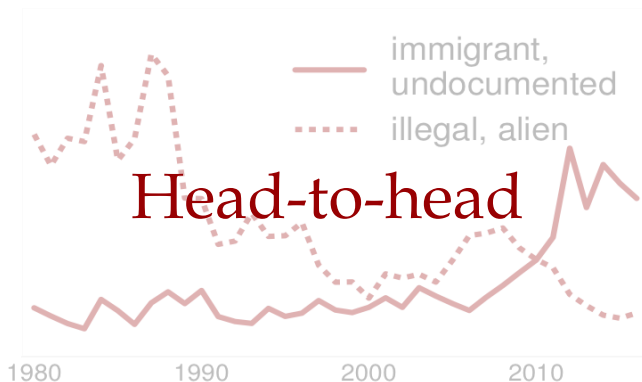1980   1990   2000   2010

Rarely cooccur

RELATIONS BETWEEN IDEAS

Always cooccur

Anti-correlated

Correlated

Rarely cooccur

Tryst

- immigration, deportation
- detainee, detention

Friendship

- immigrant, undocumented
- obama, president

Head-to-head

- immigrant, undocumented
- illegal, alien

Arms-race

- immigration, deportation
- republican, party
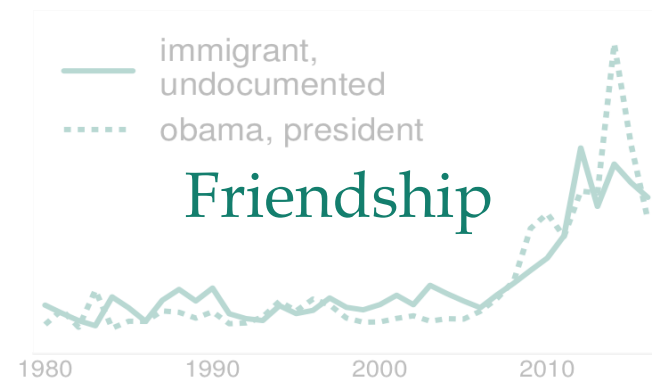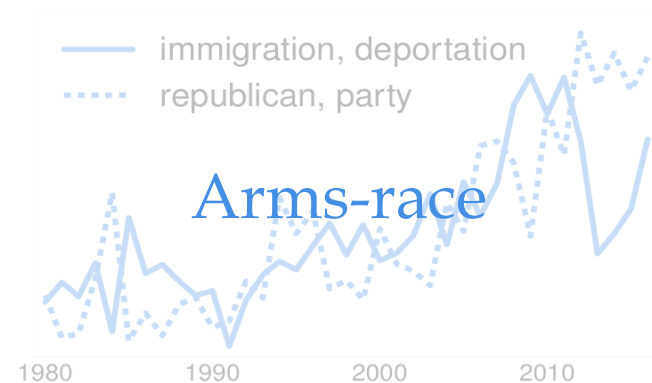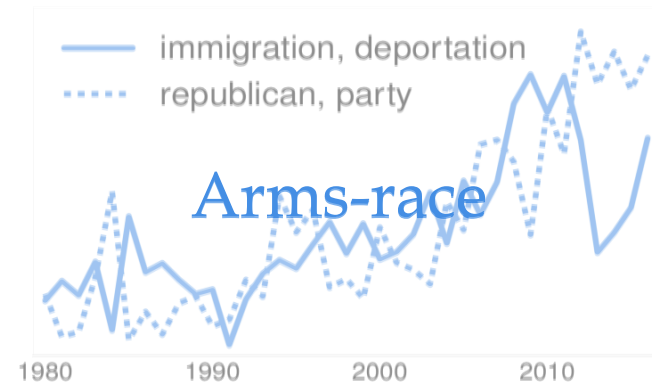
TRYST
(ANTI-CORRELATED, LIKELY TO COOCCUR)

Legend:
- —— immigration, deportation
- ---- detainee, detention

X-axis: 1980, 1990, 2000, 2010

RELATIONS BETWEEN IDEAS

Always cooccur

immigration, deportation
detainee, detention

immigrant,
undocumented
obama, president

Tryst

Friendship
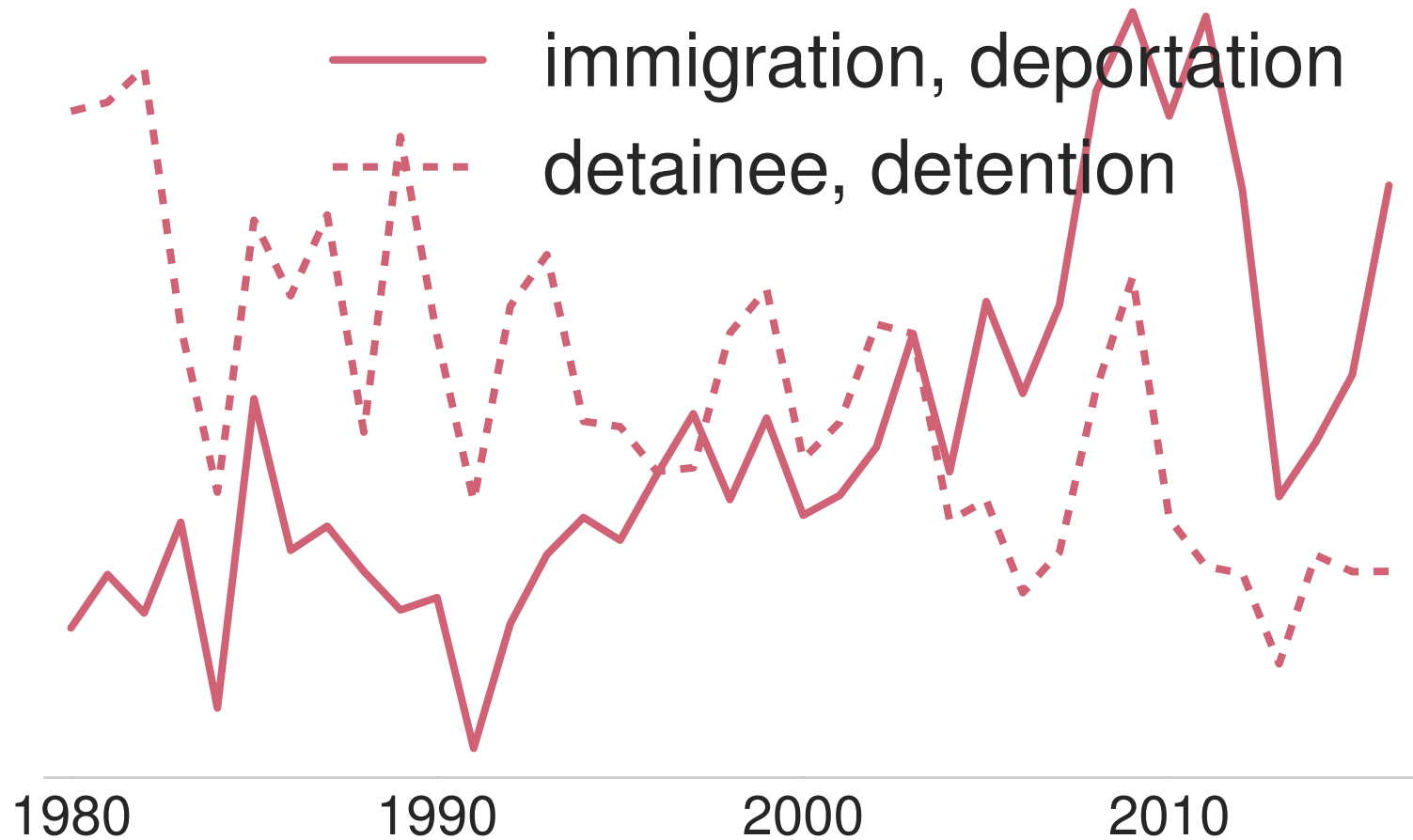
We have shown a framework to quantitatively describe relations between ideas.

Anti-correlated

Correlated

Can we use them to effectively explore relations between ideas?

immigrant,
undocumented
illegal, alien

immigration, deportation
republican, party

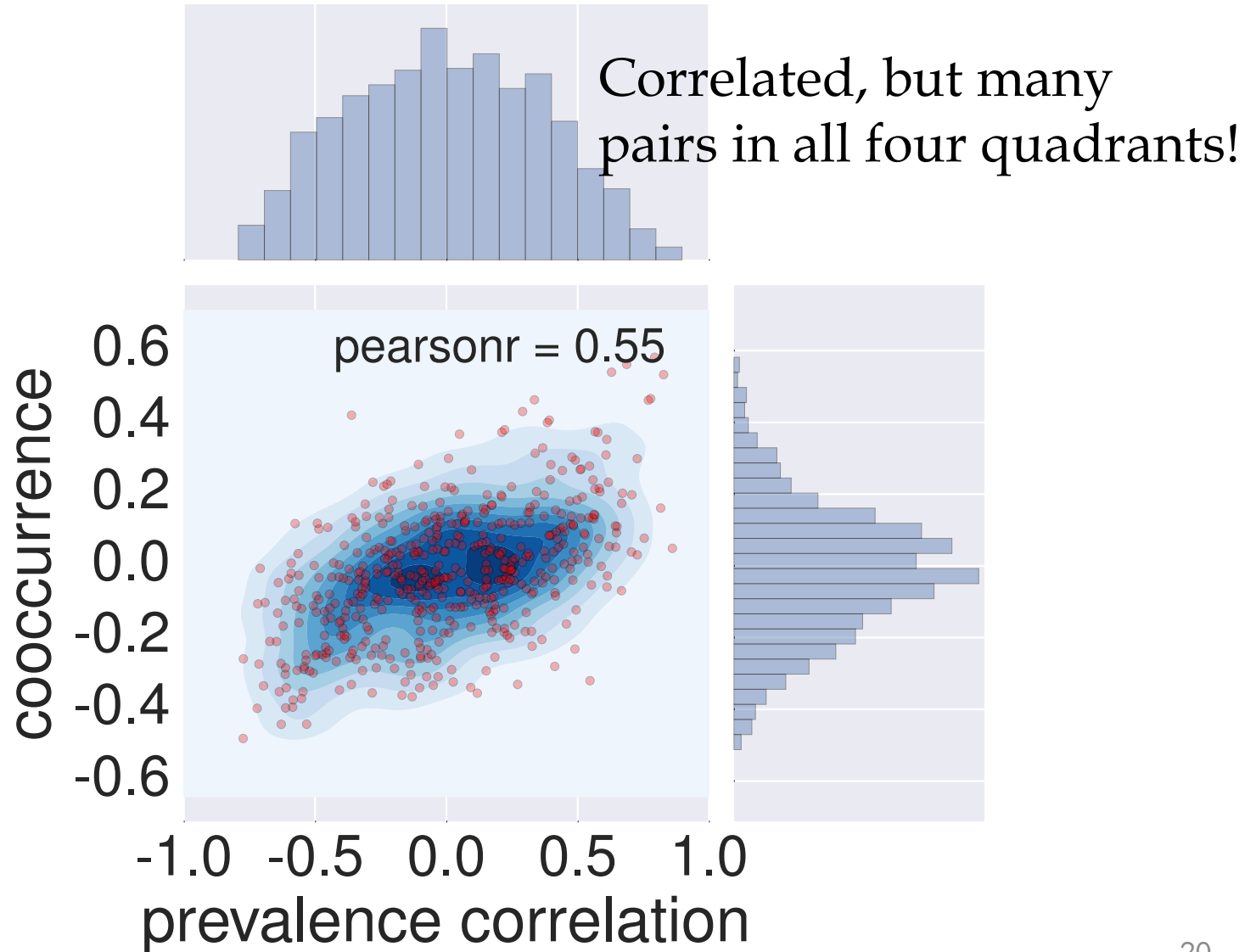Head-to-head

Arms-race
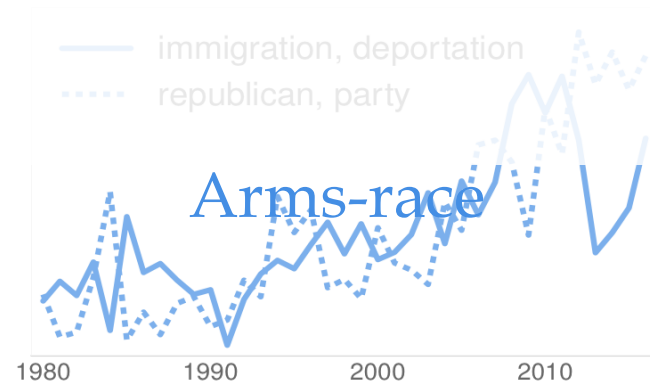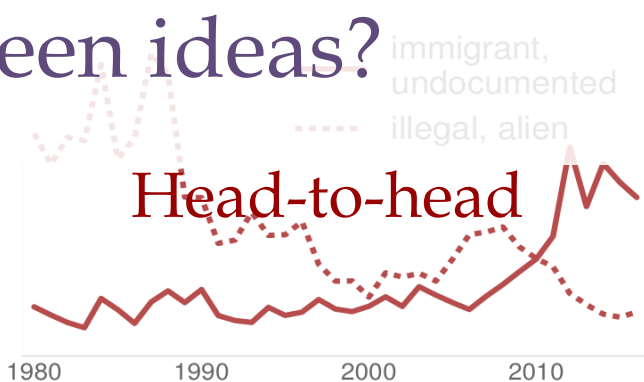
Rarely cooccur

A WIDE RANGE OF DATASETS

- Newspapers and research articles as datasets
  - Immigration
  - Terrorism
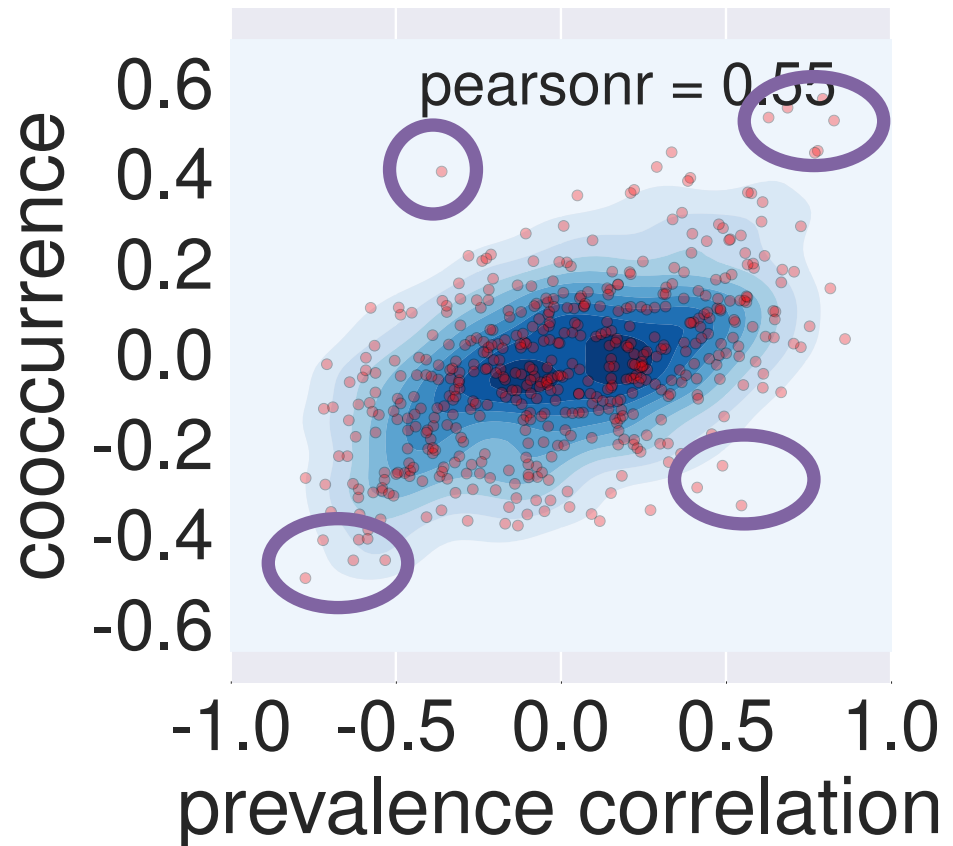 - Same-sex marriage
  - Abortion
  - Tobacco

  - ACL
  - NIPS

JOINT DISTRIBUTIONS

Correlated, but many pairs in all four quadrants!

THE STRENGTH OF RELATIONS

Strength = |PMI| × |correlation|

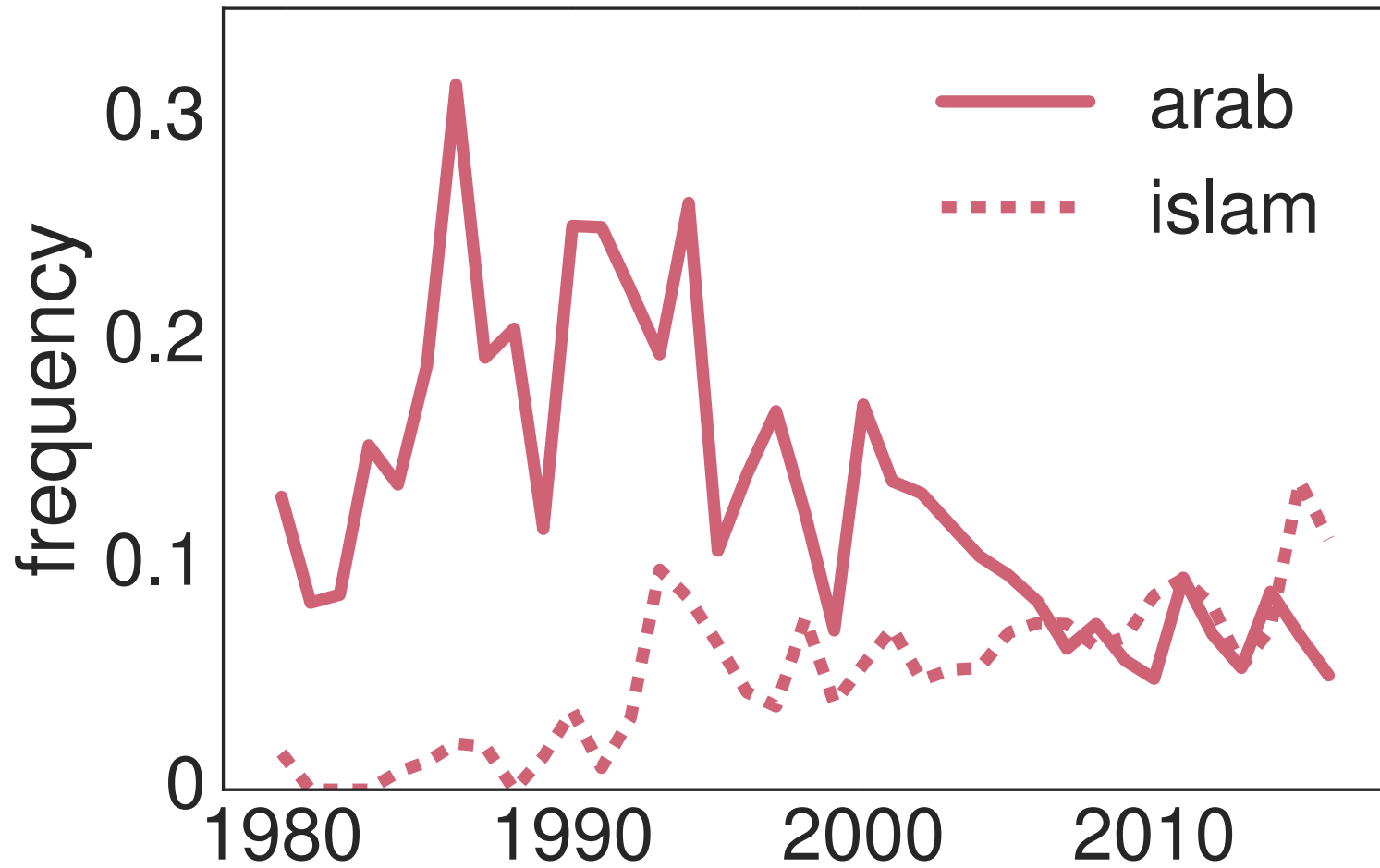Extreme pairs are the interesting ones!

EFFECTIVELY EXPLORE RELATIONS BETWEEN IDEAS

- Terrorism
  - Keywords
  - Topics

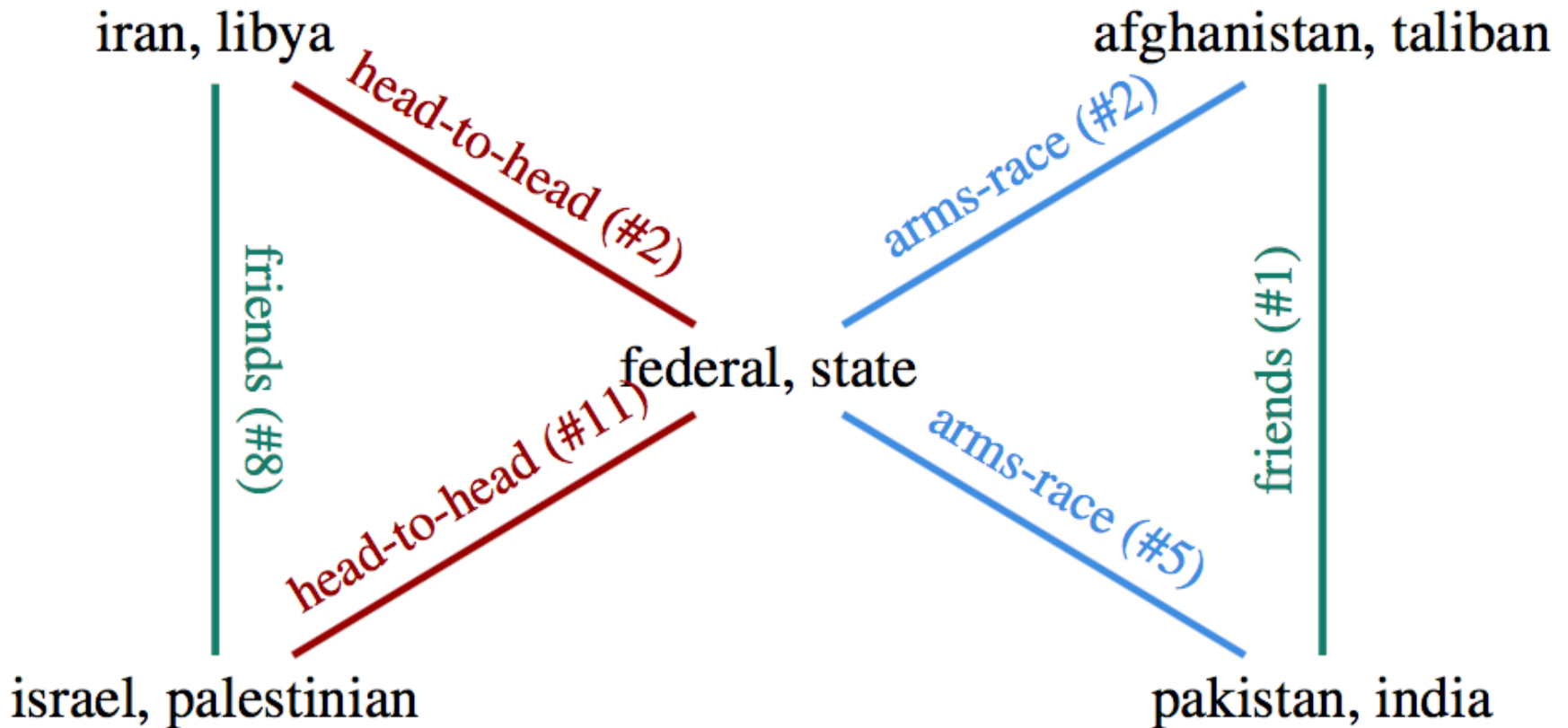# EFFECTIVELY EXPLORE RELATIONS BETWEEN IDEAS

- Terrorism
  - Keywords
  - Topics

#2 IN TRYSTS

EFFECTIVELY EXPLORE RELATIONS BETWEEN IDEAS

- Terrorism
  - Keywords
  - Topics

TOP RELATIONS BETWEEN IDEAS

The relations between these topics are consistent with structural balance theory: the enemy of an enemy is a friend [Cartwright and Harary, 1956; Heider, 1946]

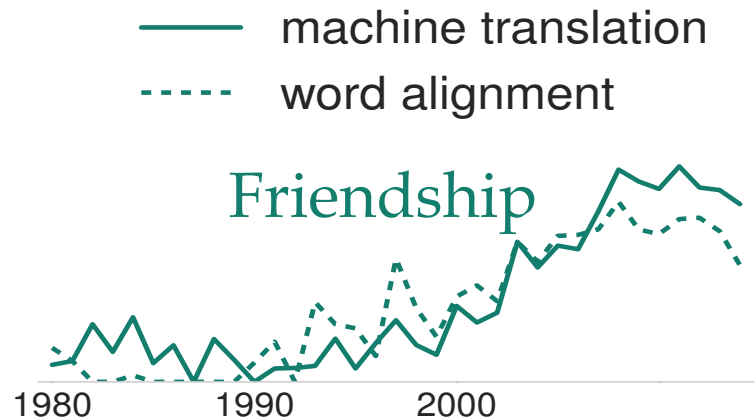# EFFECTIVE EXPLORATIONS
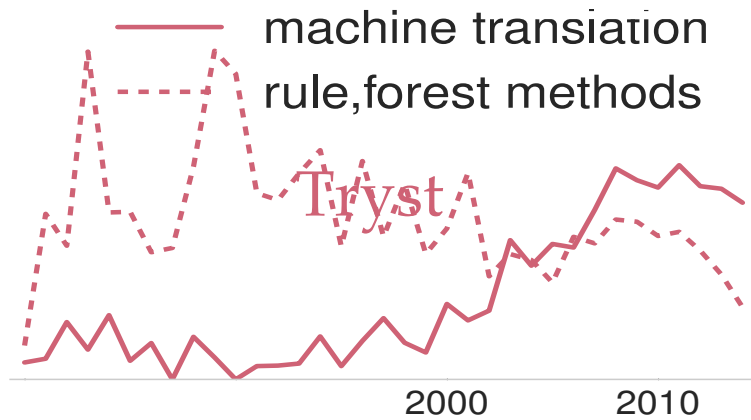
			Rank among all relations		
			PMI	**Correlation**	**Joint**
Keywords	arab	islam	106	1,494	2
Topics	federal, state	afghanistan, taliban	43	99	2
	federal, state	iran, lybia	36	56	2

The "interesting" pair is ranked much higher according to our framework.
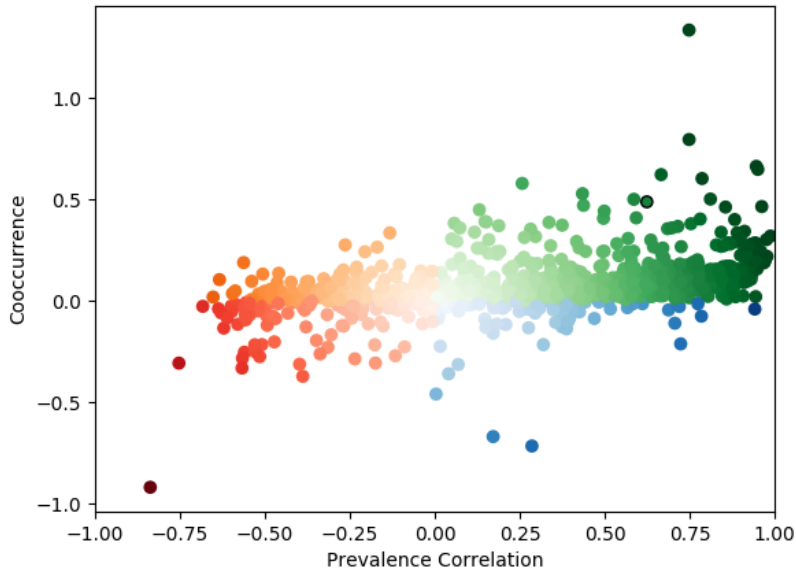
ACL TEASER

Always cooccur

machine translation
rule,forest methods

Tryst

machine translation
word alignment

Friendship

2000   2010

1980   1990   2000

Anti-correlated

Correlated

machine translation
discourse (coherence)

Head-to-head

machine translation
sentiment analysis

Arms-race

1980   1990   2000   2010

1980   1990   2000   2010

Rarely cooccur

28

VISUALIZATION TOOL

https://github.com/nwrush/Visualization

Thank you!

cooccurrence

+

prevalence correlation

⟹

A quantitative way to describe relations between ideas: **friendships, head-to-head, arms-race, tryst**

An effective framework to explore temporal text corpora

chenhao@chenhaot.com,
Twitter: @ChenhaoTan
Data & code:
https://chenhaot.com/papers/idea-relations.html