

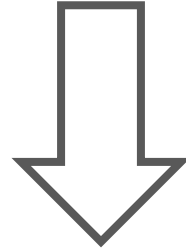
CHARACTERIZING THE ECOSYSTEM OF IDEAS IN TEXTS

Chenhao Tan

University of Washington

Joint work with Lada Adamic, Dallas Card,
Adrien Friggeri, Noah Smith

An unprecedented amount of textual data



Extracting ideas from these data

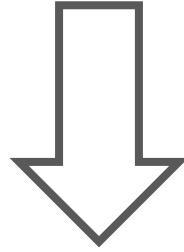
TOPICS IN IMMIGRATION

undocumented
immigrants

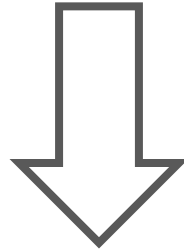
illegal alien

How to characterize this
intuitively competing relation?

An unprecedented amount of textual data



Extracting ideas from these data



Exploring relations between ideas

HOW TO QUANTIFY RELATIONS BETWEEN IDEAS

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

undocumented
immigrants

illegal alien

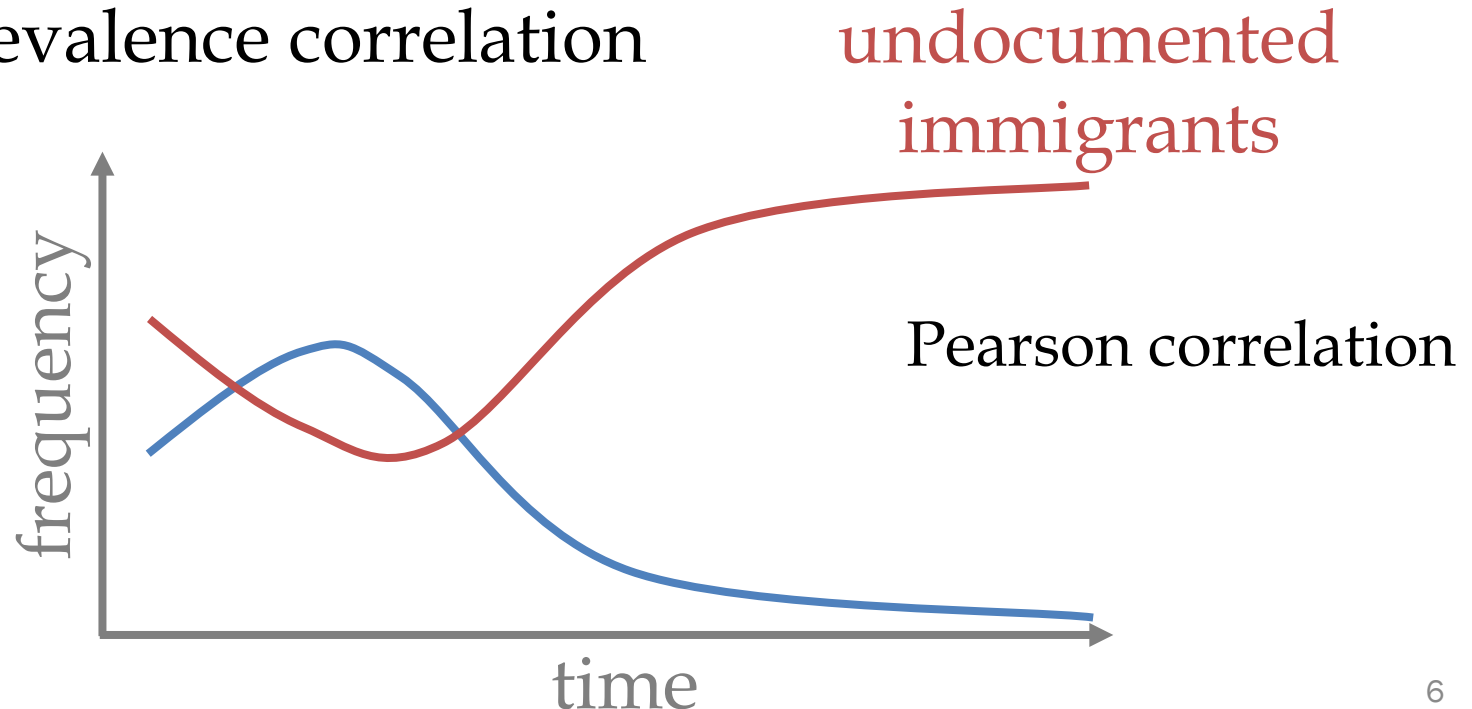
Rarely cooccur

Cooccurrence

Pointwise mutual information
[Church and Hanks 1990]

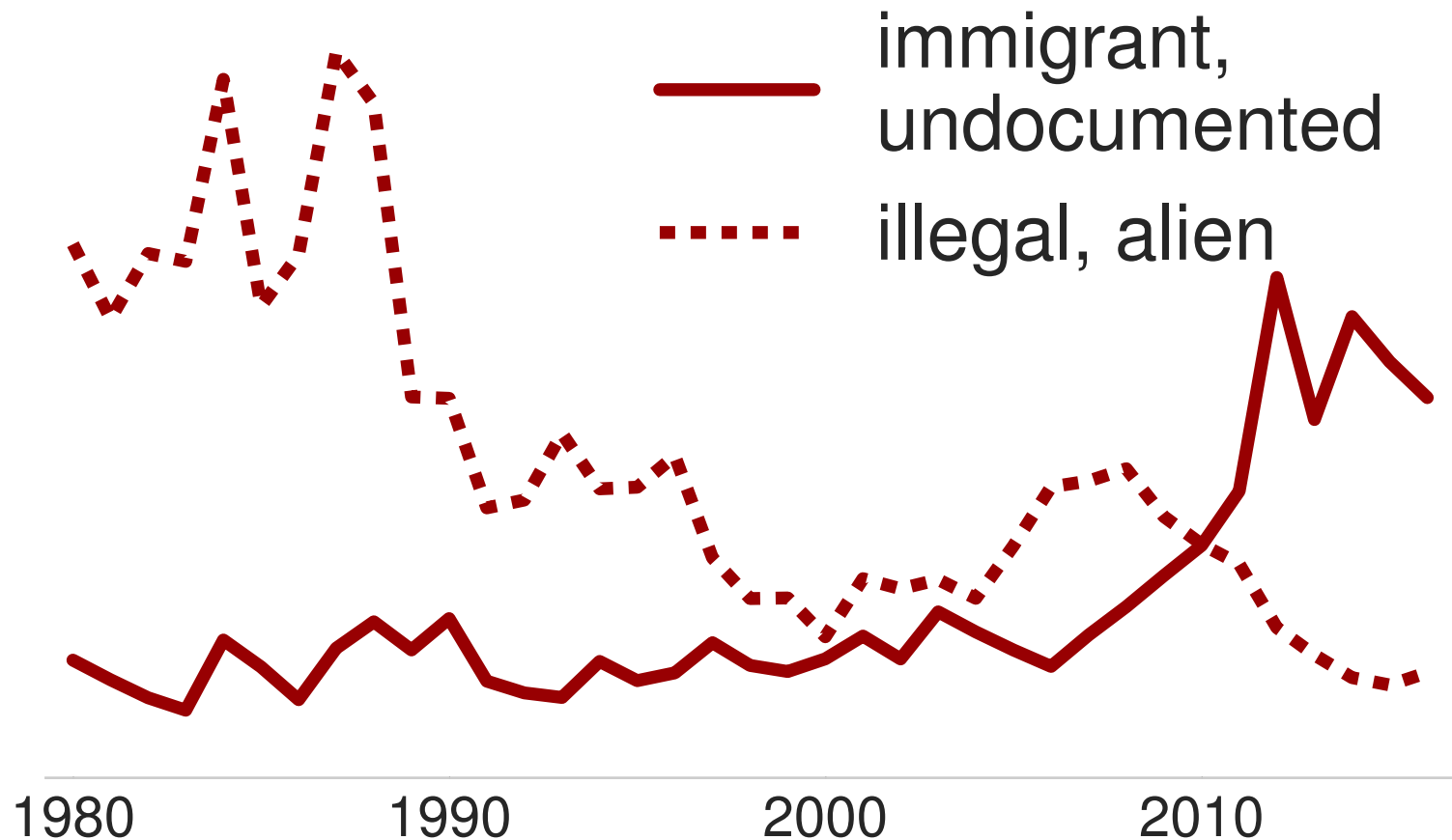
HOW TO QUANTIFY RELATIONS BETWEEN IDEAS

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



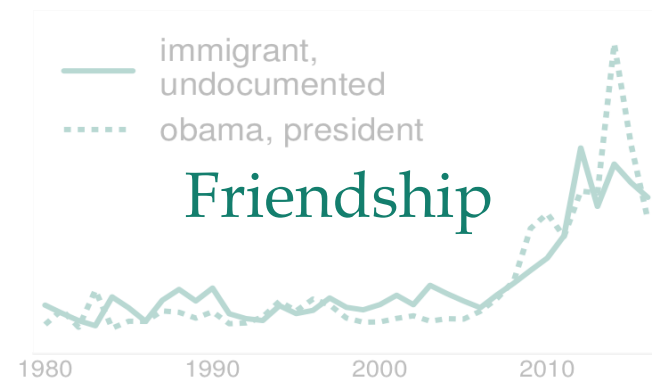
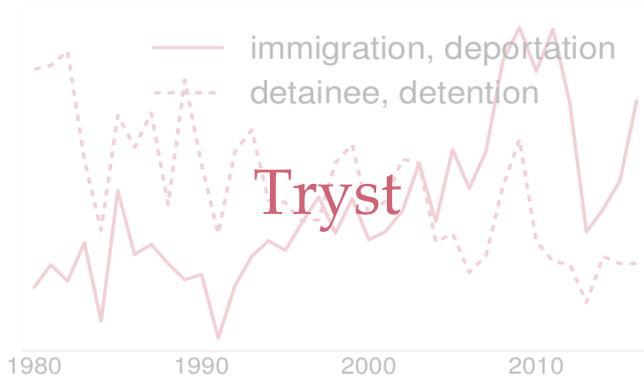
HEAD-TO-HEAD

(ANTI-CORRELATED, RARELY COOCCUR)

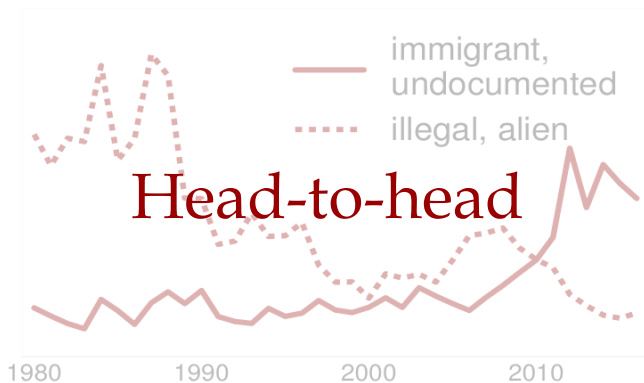


RELATIONS BETWEEN IDEAS

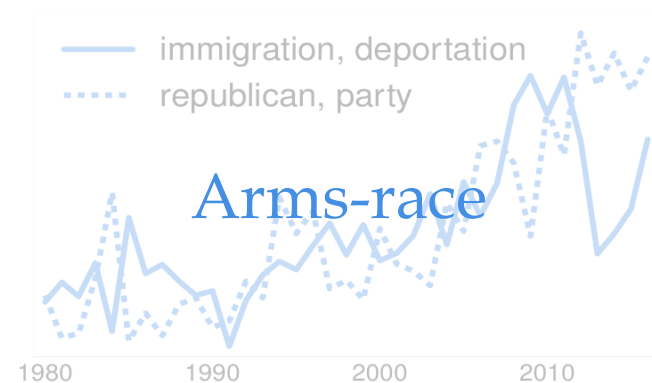
Always cooccur



Anti-correlated



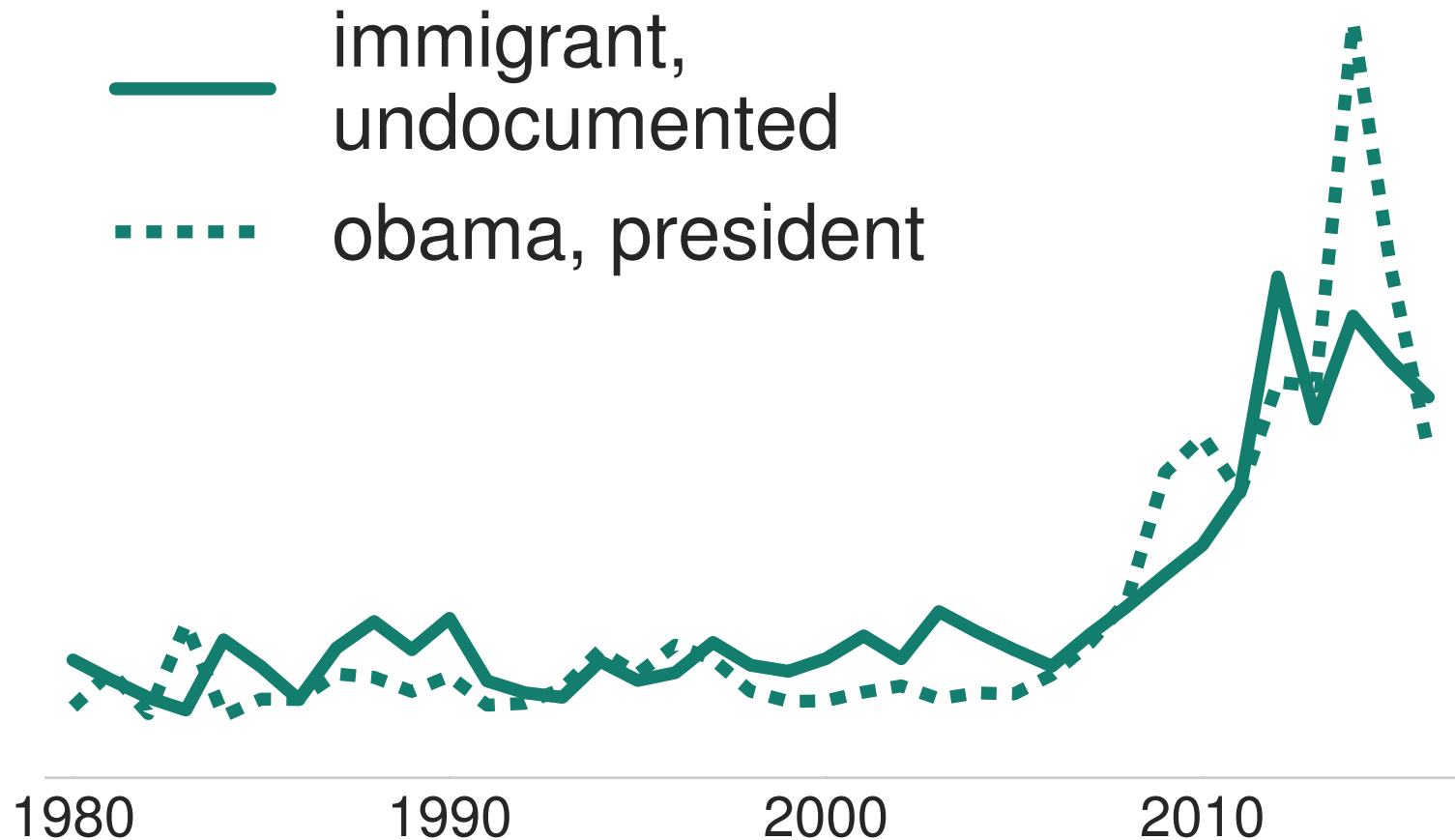
Correlated



Rarely cooccur

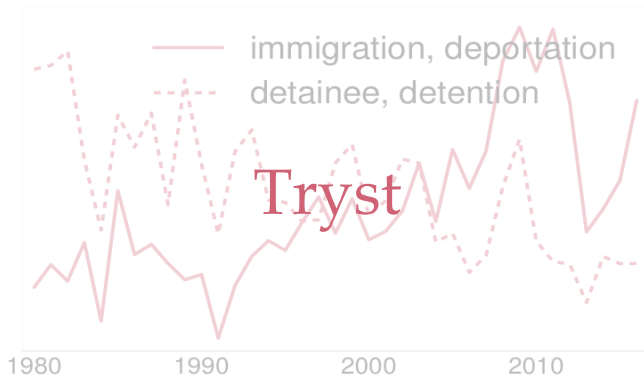
FRIENDSHIP

(CORRELATED, ALWAYS COOCCUR)

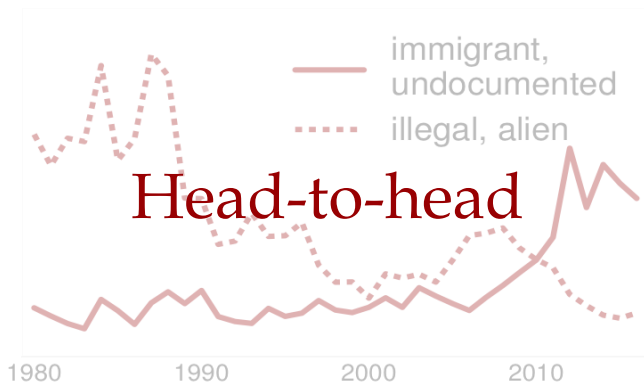


RELATIONS BETWEEN IDEAS

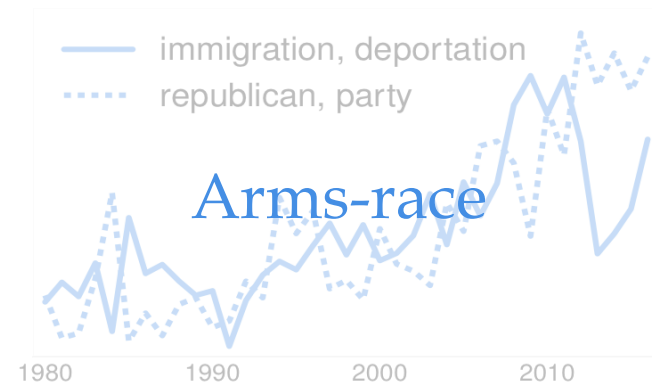
Always cooccur



Anti-correlated



Correlated

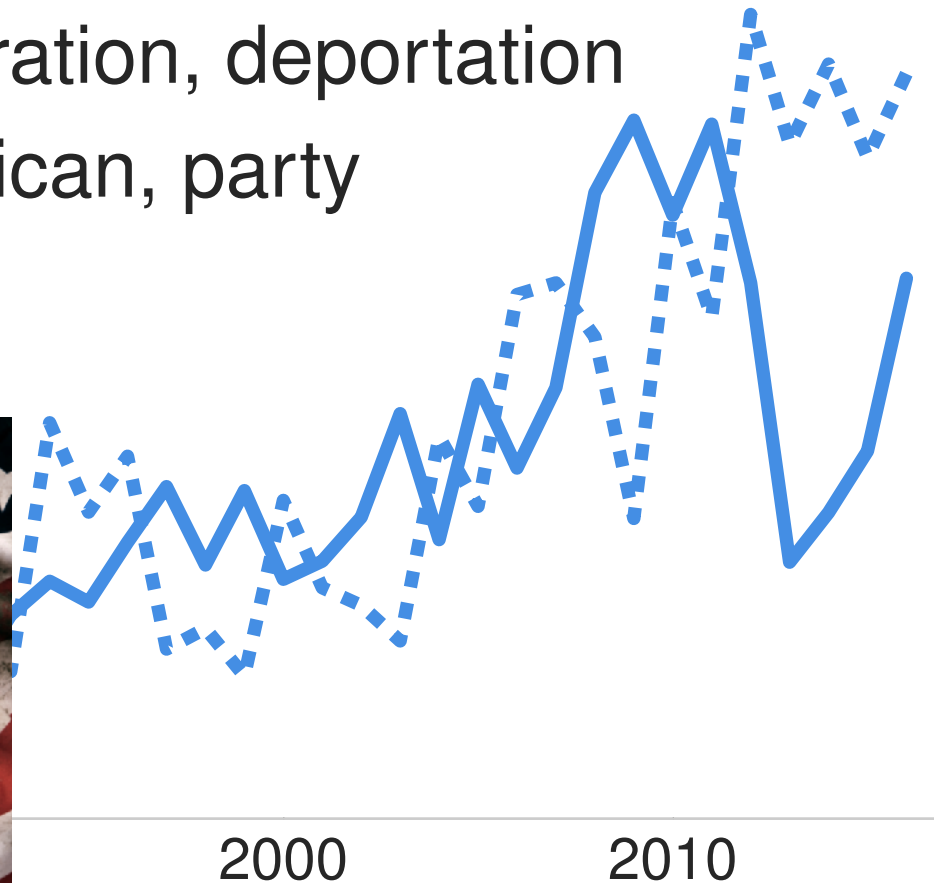


Rarely cooccur

ARMS-RACE

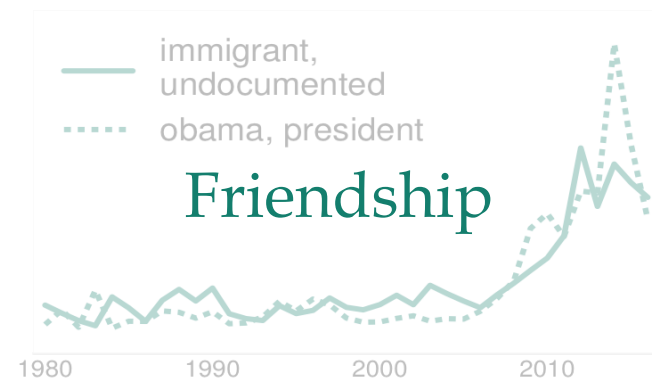
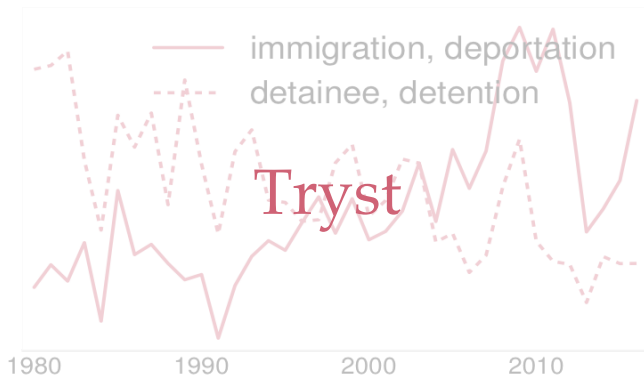
(CORRELATED, RARELY COOCCUR)

— immigration, deportation
- - - republican, party

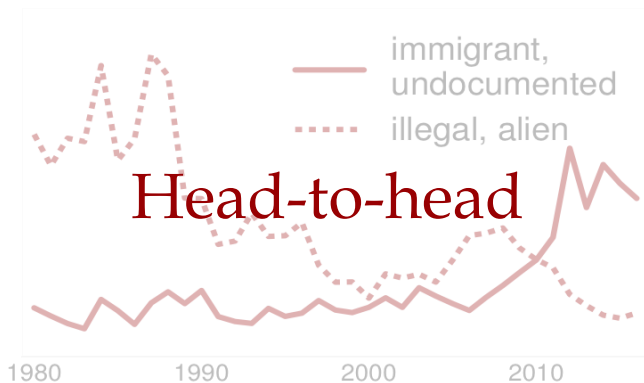


RELATIONS BETWEEN IDEAS

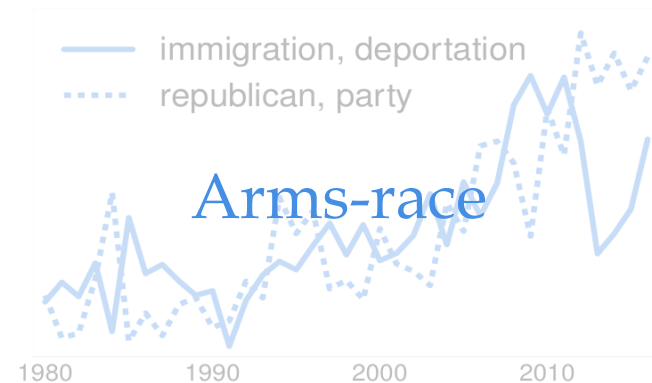
Always cooccur



Anti-correlated



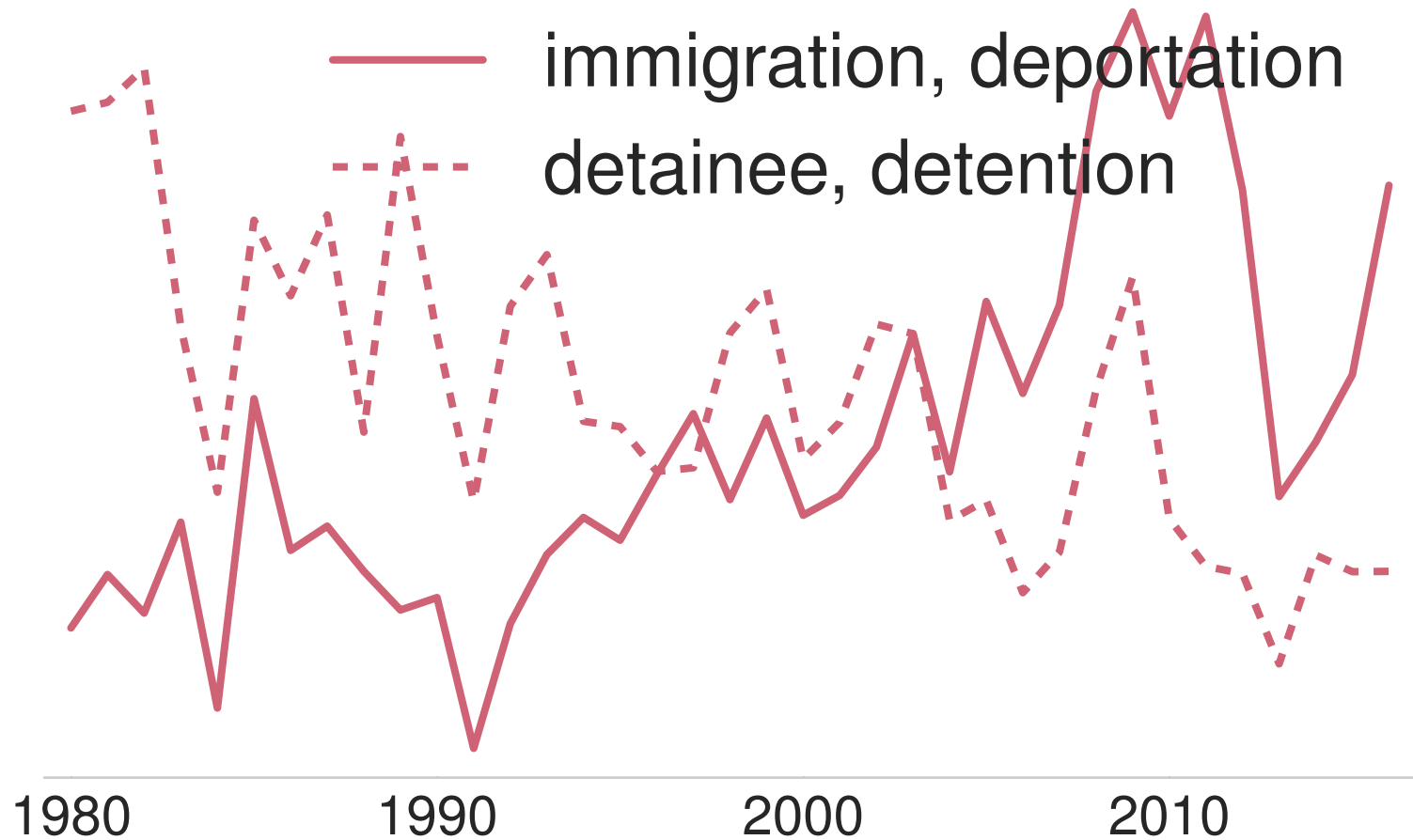
Correlated



Rarely cooccur

TRYST

(ANTI-CORRELATED, **LIKELY TO COOCCUR**)



RELATIONS BETWEEN IDEAS

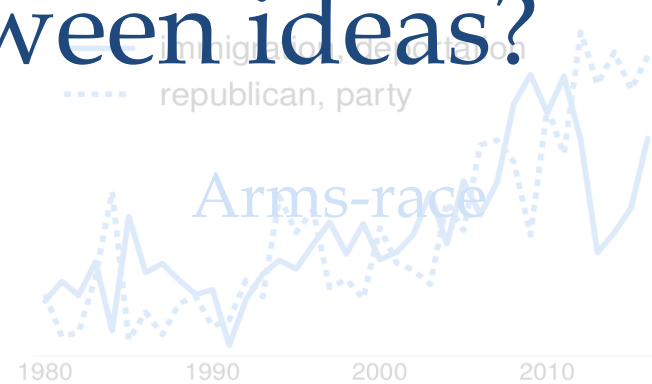
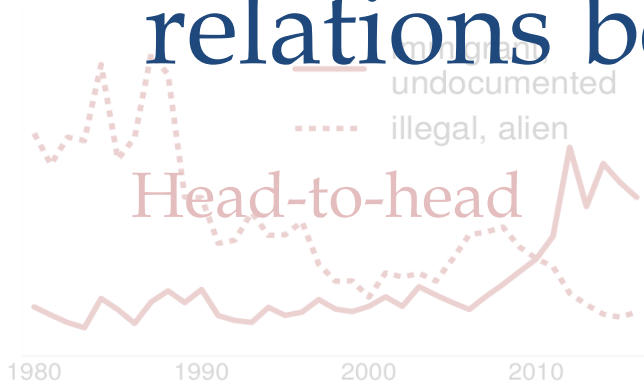
Always cooccur



Do they actually exist?

Can we use them to effectively explore

relations between ideas?



Rarely cooccur

DO THESE RELATIONS ACTUALLY EXIST?

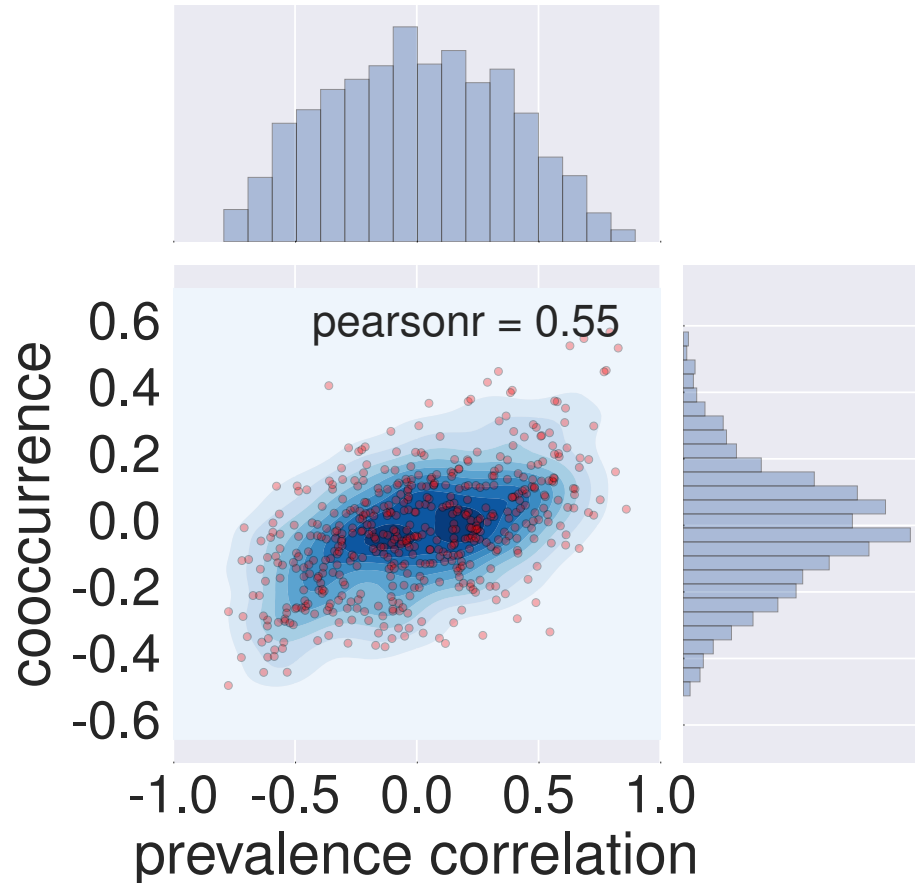
- Newspapers and research articles as datasets
 - Immigration
 - Terrorism
 - Same-sex marriage
 - Abortion
 - Tobacco

 - ACL
 - NIPS

JOINT DISTRIBUTIONS

- Cooccurrence: unimodal
- Prevalance correlation: unimodal
- Cooccurrence and prevalence correlation: positively correlated

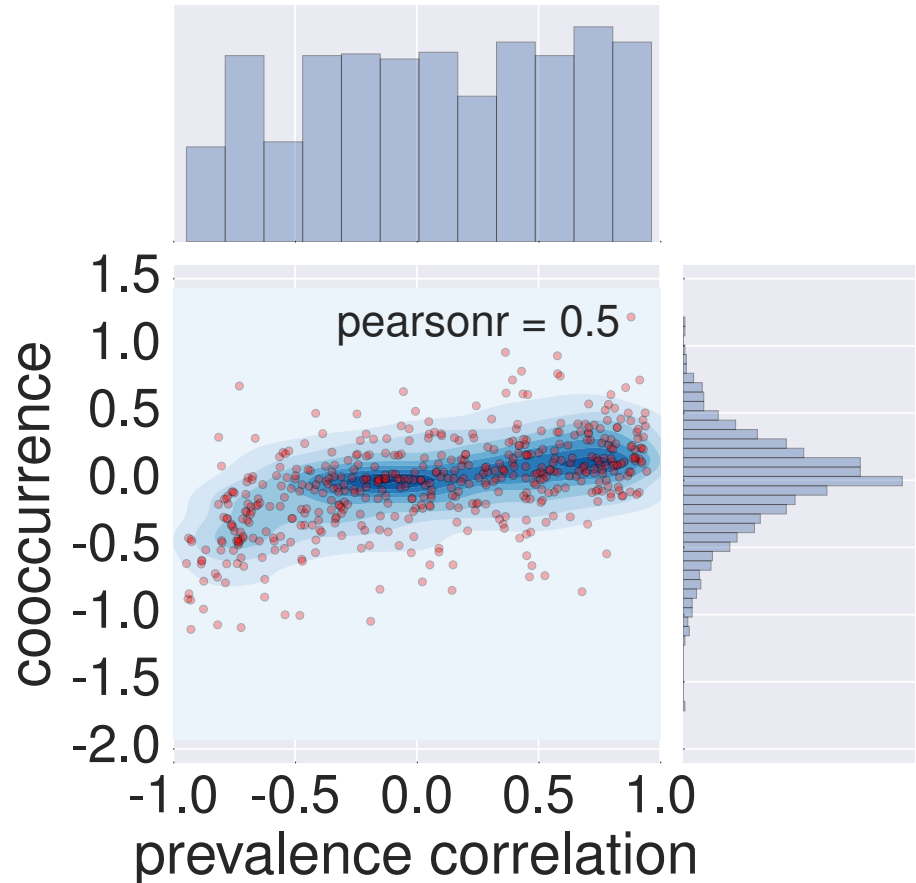
Immigration topics



JOINT DISTRIBUTIONS

- Cooccurrence: unimodal
- Prevalance correlation: complicated, not always unimodal
- Cooccurrence and prevalence correlation: positively correlated

ACL topics

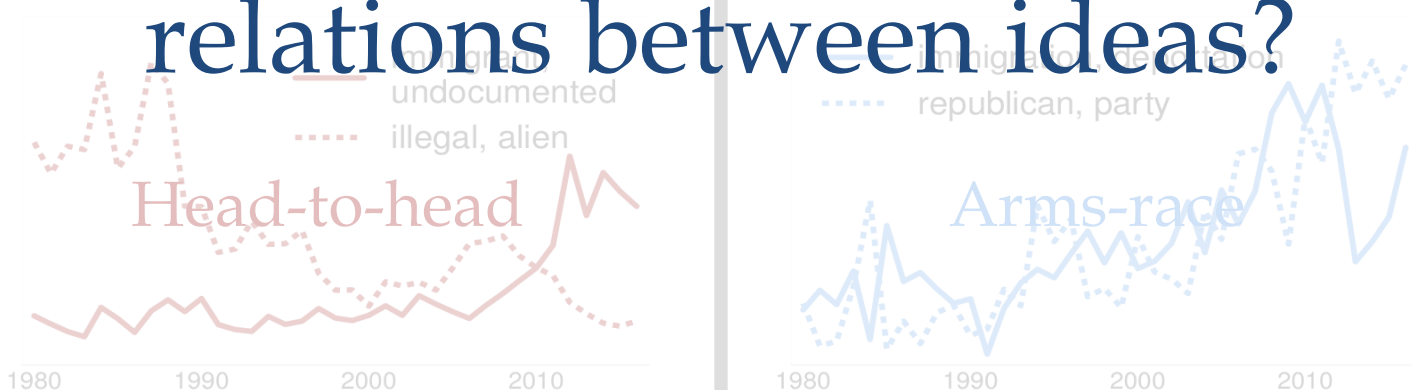


RELATIONS BETWEEN IDEAS

Always cooccur



Can we use them to effectively explore relations between ideas?

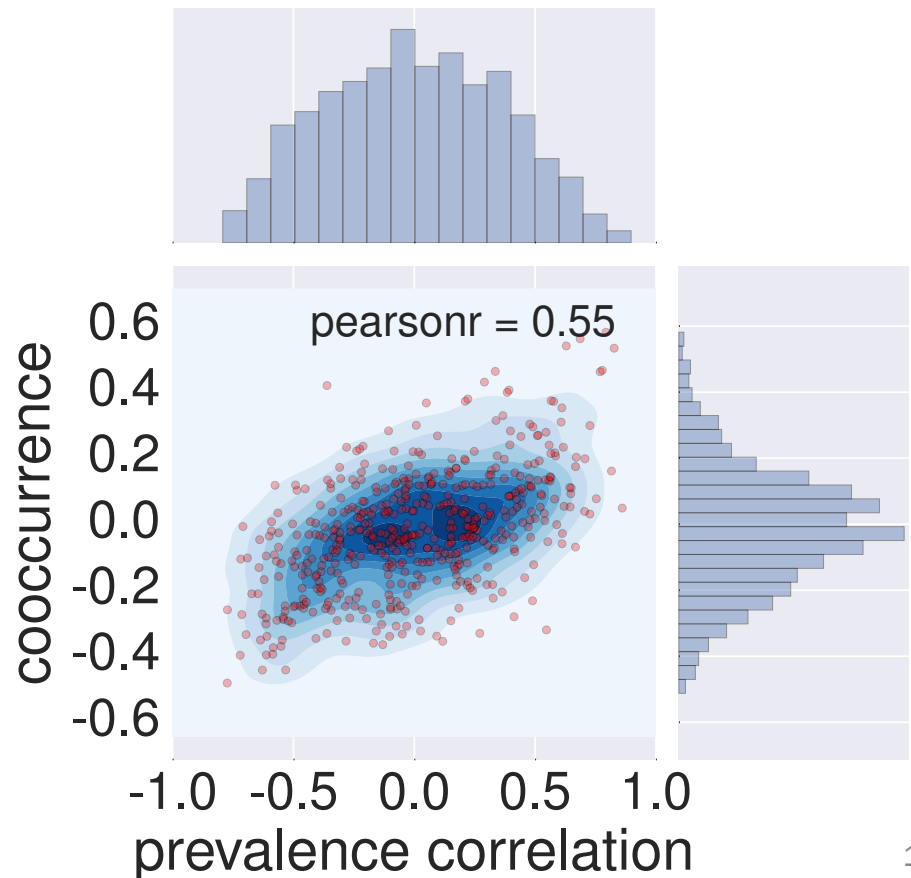


Rarely cooccur

THE STRENGTH OF RELATIONS

$$\text{Strength} = |\text{PMI}| \times |\text{correlation}|$$

Extreme pairs are
the interesting ones!



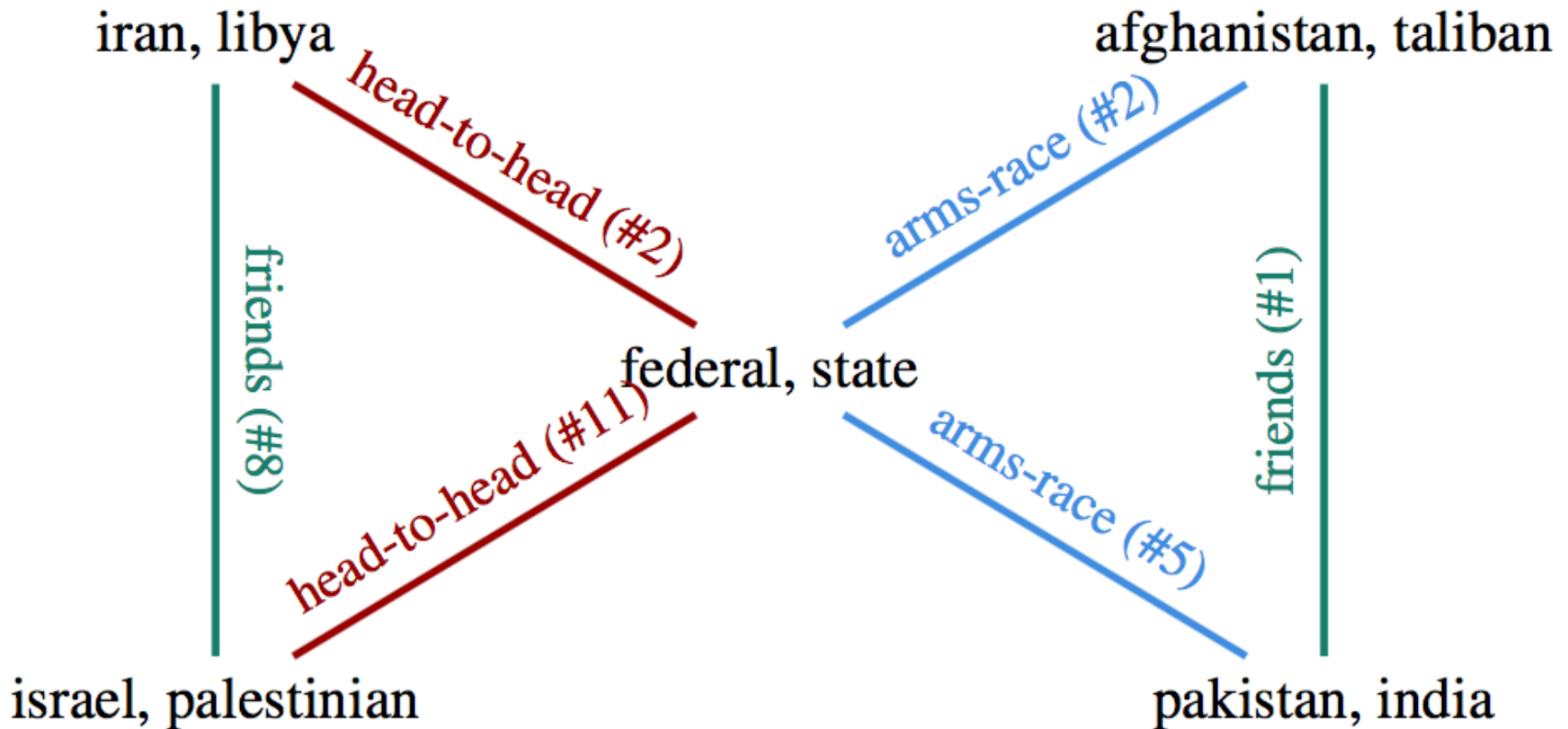
EXPLORE RELATIONS BETWEEN IDEAS

- Terrorism
 - Topic models (Blei et al. 2003)
 - Keywords (Monroe et al. 2008)

EXPLORE RELATIONS BETWEEN IDEAS

- Terrorism
 - Topics
 - Keywords

EXPLORE RELATIONS BETWEEN IDEAS

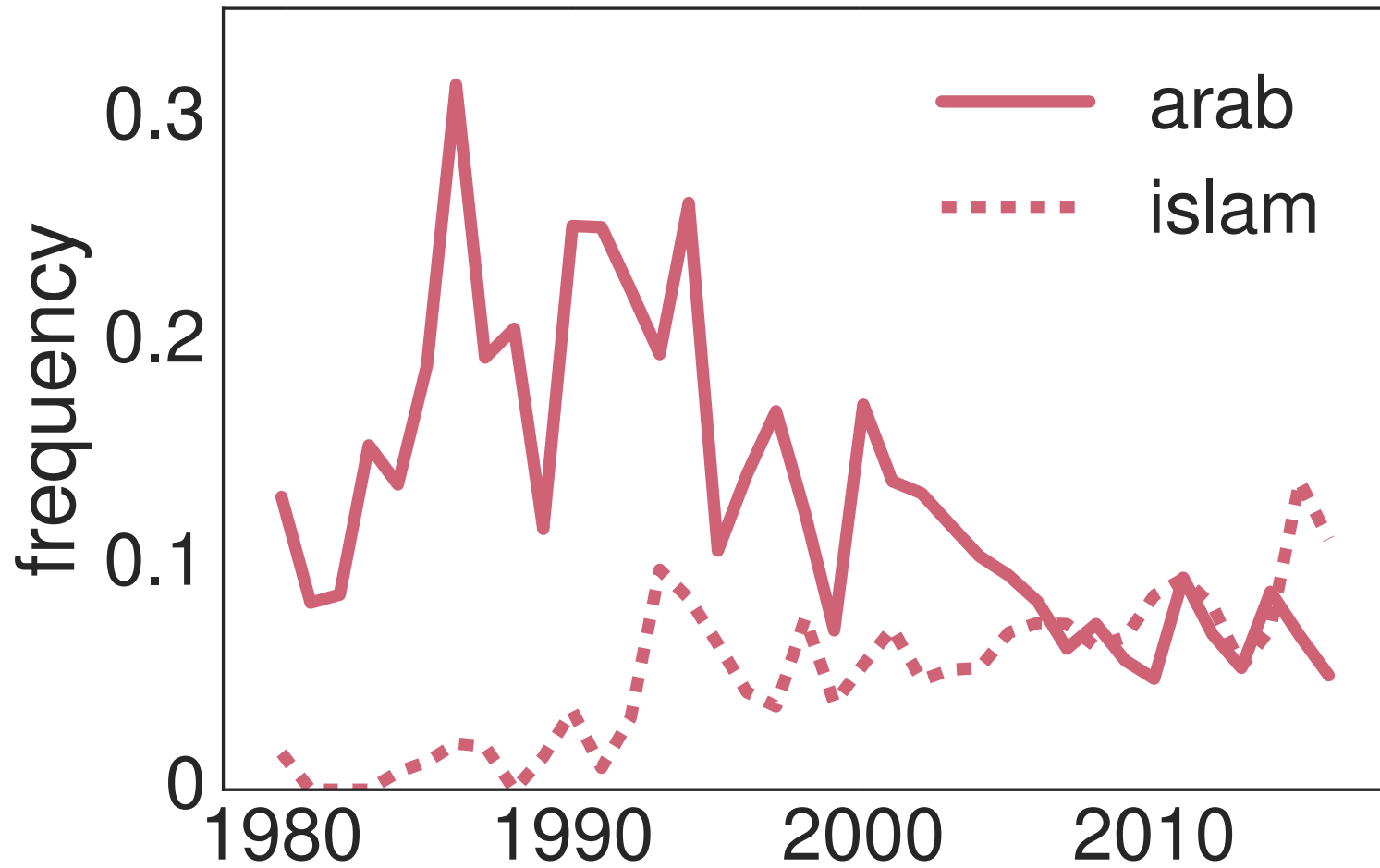


The relations between these topics are consistent with structural balance theory: the enemy of an enemy is a friend [Cartwright and Harary, 1956; Heider, 1946]

EXPLORE MACRO RELATIONS BETWEEN IDEAS

- Terrorism
 - Topics
 - Keywords

#2 IN TRYSTS

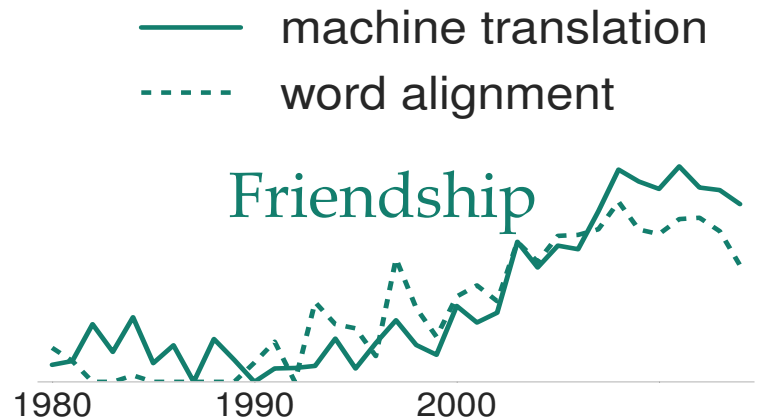
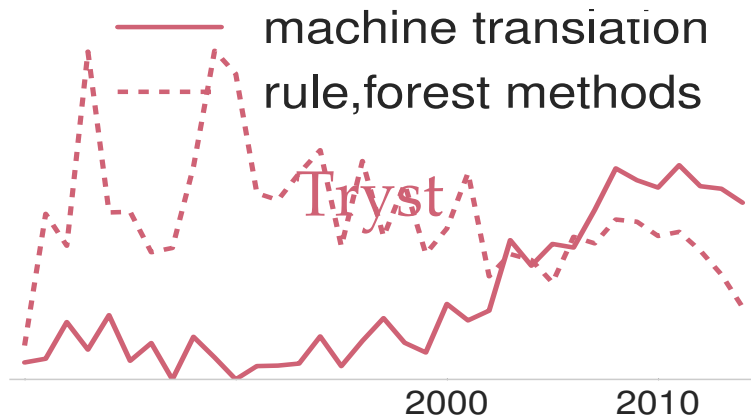


EFFECTIVE EXPLORATIONS

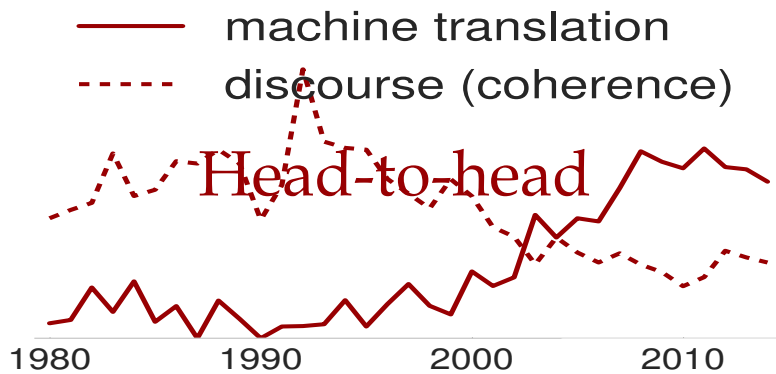
			PMI	Correlation	Joint
Topics	federal, state	afghanistan, taliban	43	99	2
	federal, state	iran, lybia	36	56	2
Keywords	arab	islam	106	1,494	2

ACL TEASER

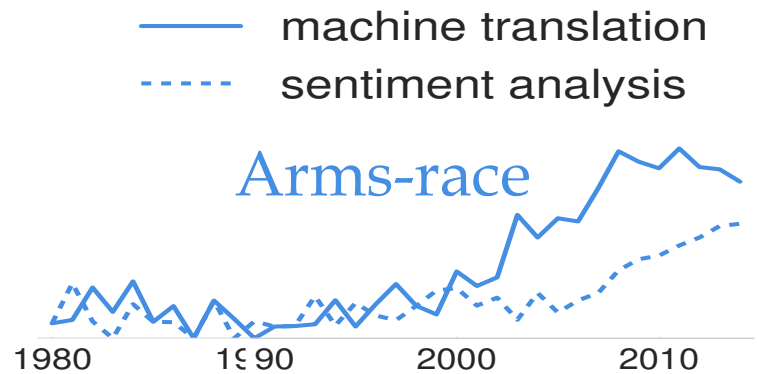
Always cooccur



Anti-correlated



Correlated



Rarely cooccur

RELATIONS BETWEEN IDEAS

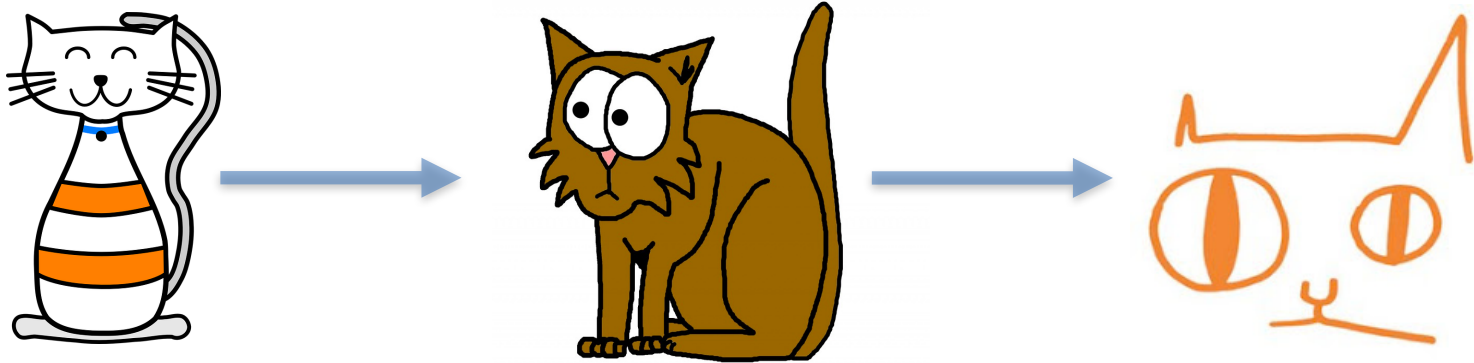
- Competition and collaboration
 - Natural selection [Dawkins 1976]
 - Marketplace of ideas [Milton 1644; Mill 1859]



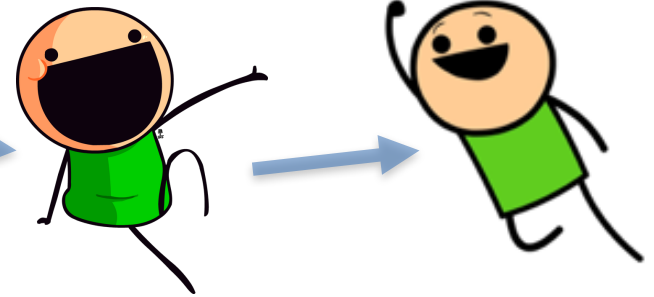
THE ECOSYSTEM OF IDEAS

- Competition and collaboration
 - Natural selection [Dawkins 1976]
 - Marketplace of ideas [Milton 1644; Mill 1859]
- Evolution of ideas
 - Telephone game

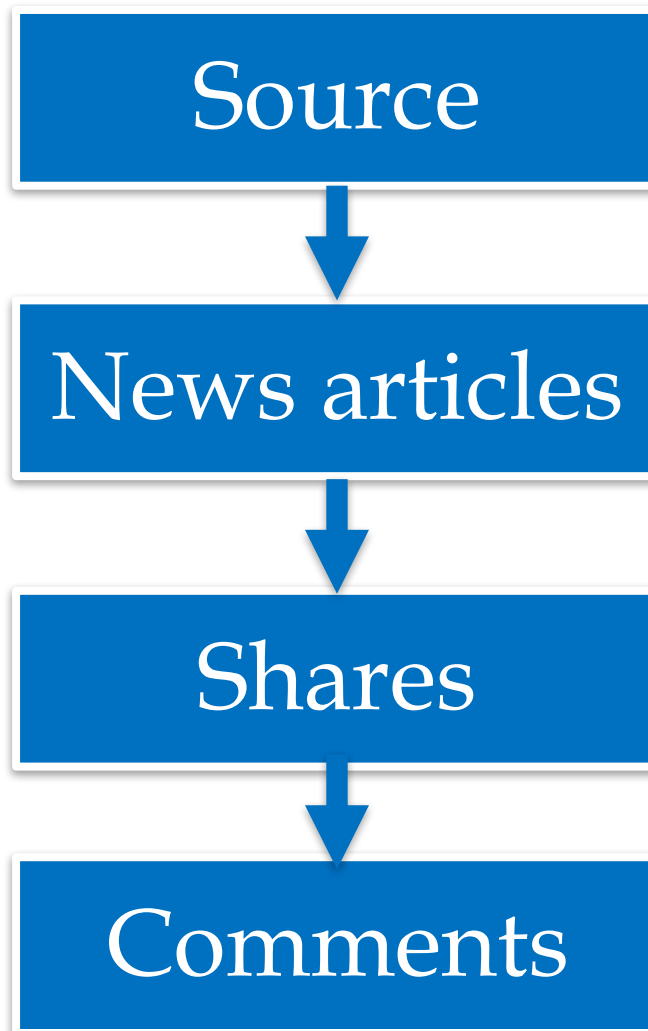
IDEAS REACH US VIA DIFFERENT WAYS



→ **The New York Times**
The Washington Post

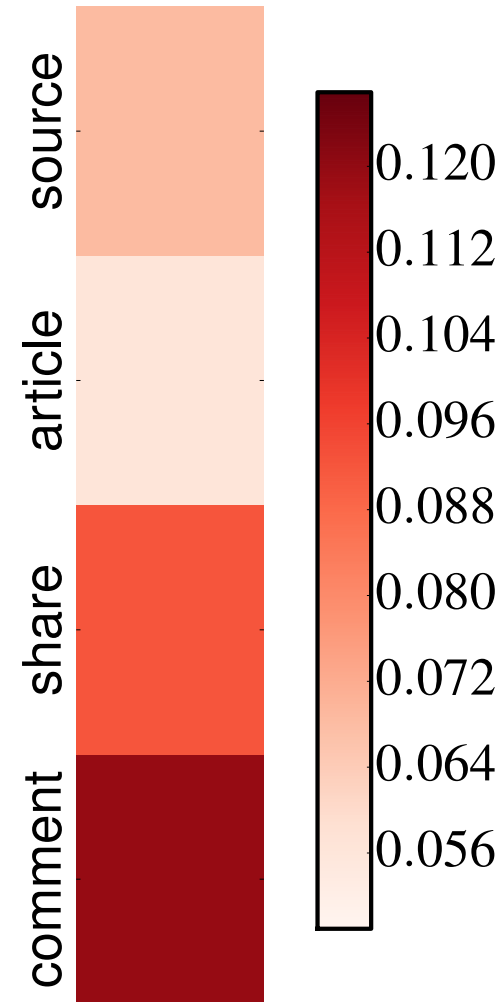


EVOLUTION OF IDEAS



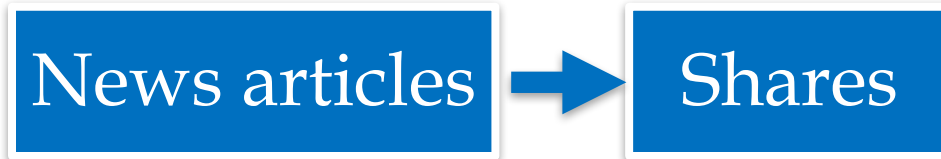
SENTIMENT EVOLUTION

- Subjectivity defined as the fraction of emotional words
- News articles are less subjective than sources, but shares and comments get more and more subjective



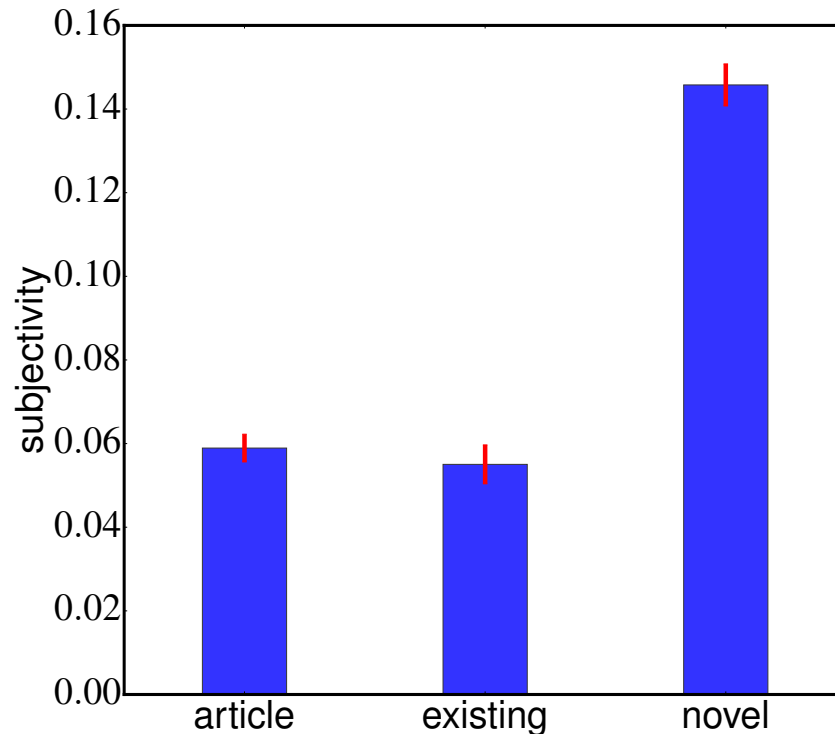
Red: more subjective₃₃

WHY DOES SUBJECTIVITY INCREASE?



words already in news articles
(*magnifier*)

novel words added by individuals
(*creator*)



THE ECOSYSTEM OF IDEAS

- Competition and collaboration
 - Natural selection [Dawkins 1976]
 - Marketplace of ideas [Milton 1644; Mill 1859]
Tan, Card, Smith, ACL'17
- Evolution of ideas
 - Telephone game
Tan, Friggeri, Adamic, ICWSM'16

CHARACTERIZING THE ECOSYSTEM OF IDEAS

- better understand existing ideas in the data
- better generate the ideas for the future