Forecasting conflict and tracking terrorist group interaction using open source intelligence

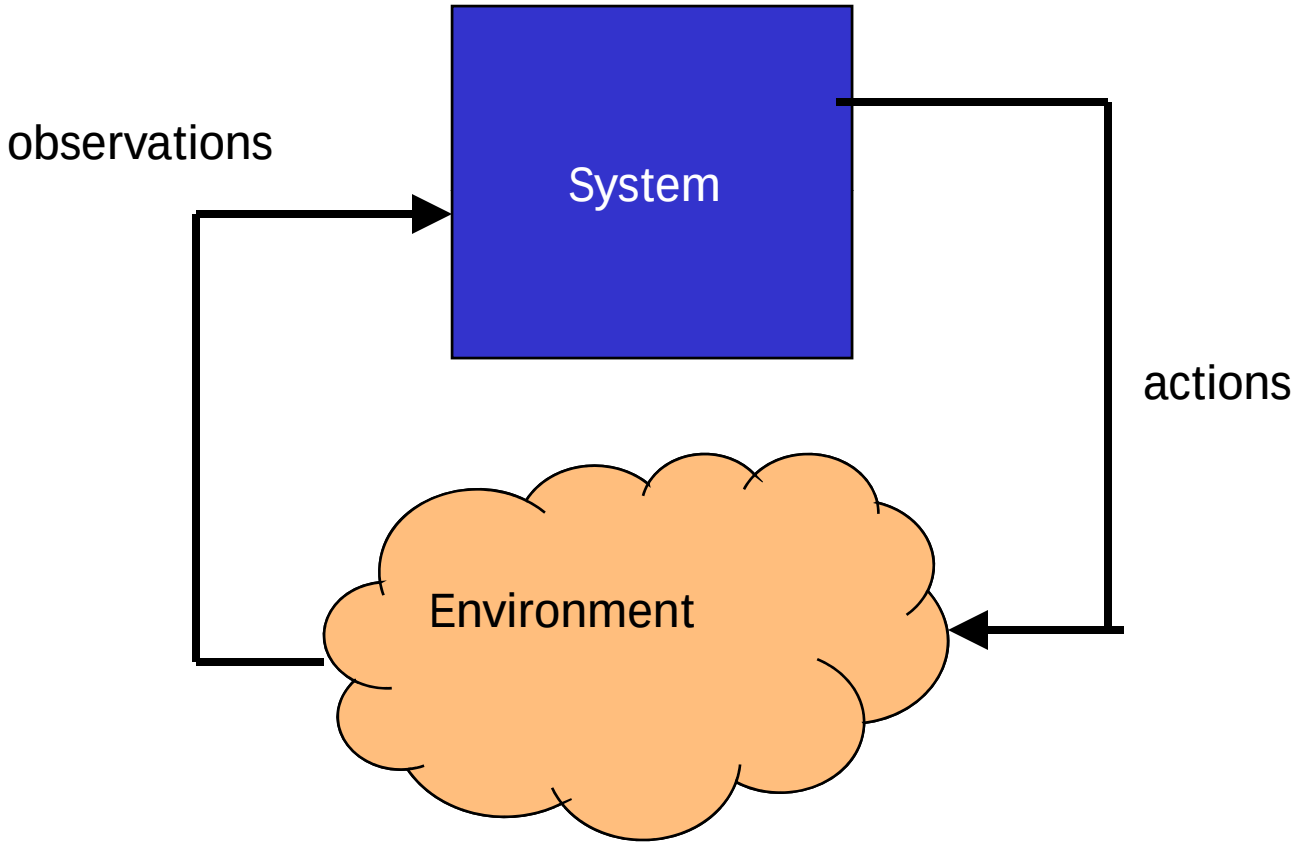
Devika Subramanian
Department of Computer Science
Rice University
http://www.cs.rice.edu/~devika

RICE

(

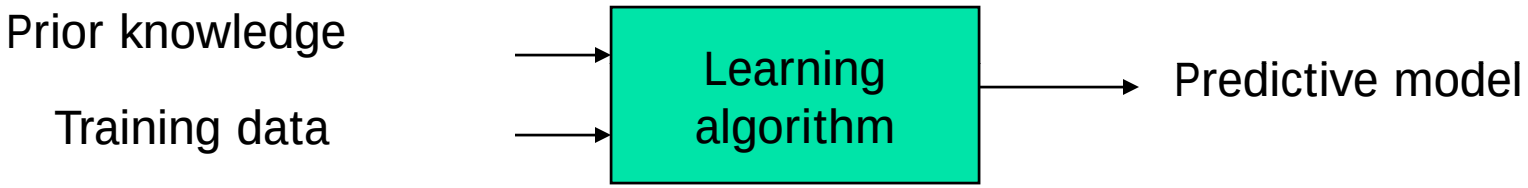
What can COMPUTING do for you?

Machine learning nanotutorial (1/3)

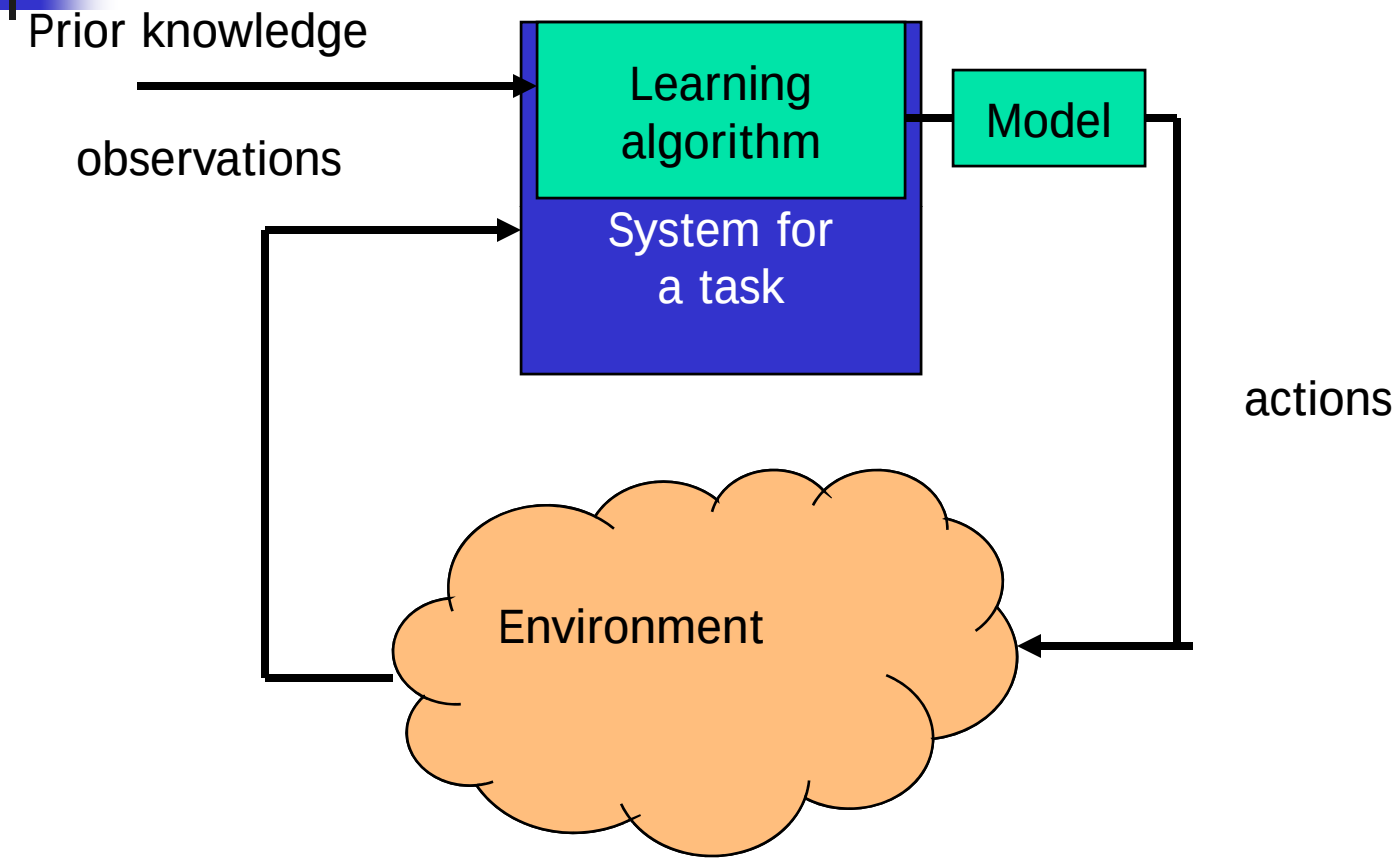


Observation driven, task-specific decision making

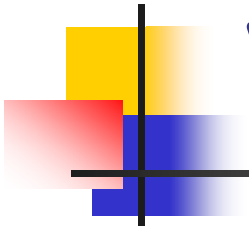
ML Nanotutorial (2/3)



Ml Nanotutorial (3/3)



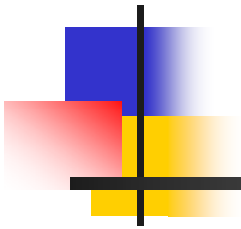
Calculate decisions on the basis of learned models of systems



Why embed learning?

- We cannot calculate and implement an action-choice/decision-making strategy for the system at design time.
  - System dynamics are unknown/partially known.
  - System dynamics change with time.
  - A one-size-fits-all solution is not appropriate – customization is needed.

Forecasting conflict by extracting events from wire news stories

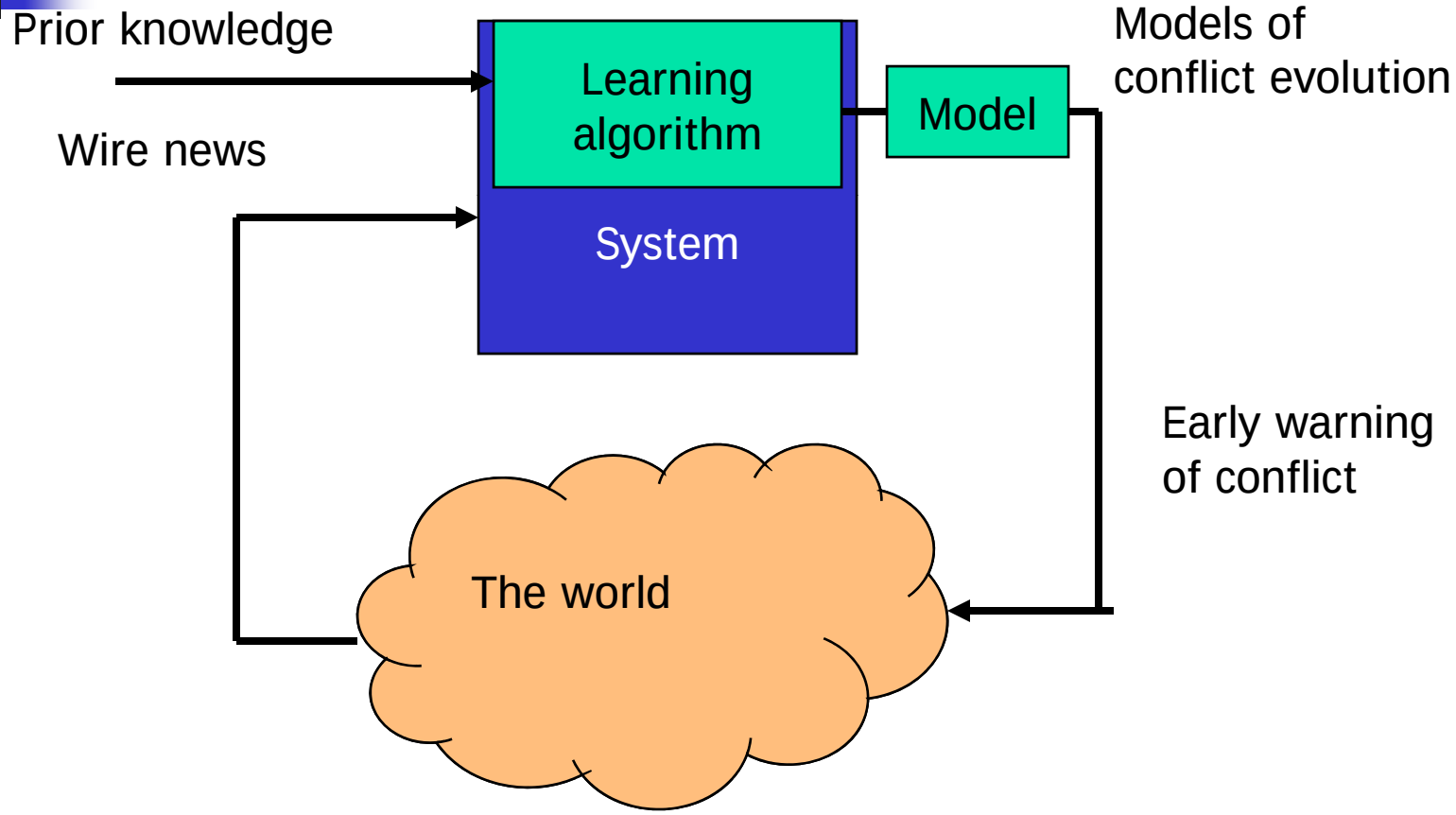


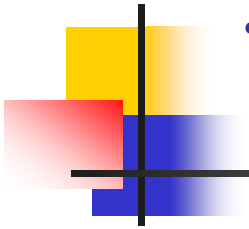
Devika Subramanian

Ric Stoll

Rice University

Adaptive Systems



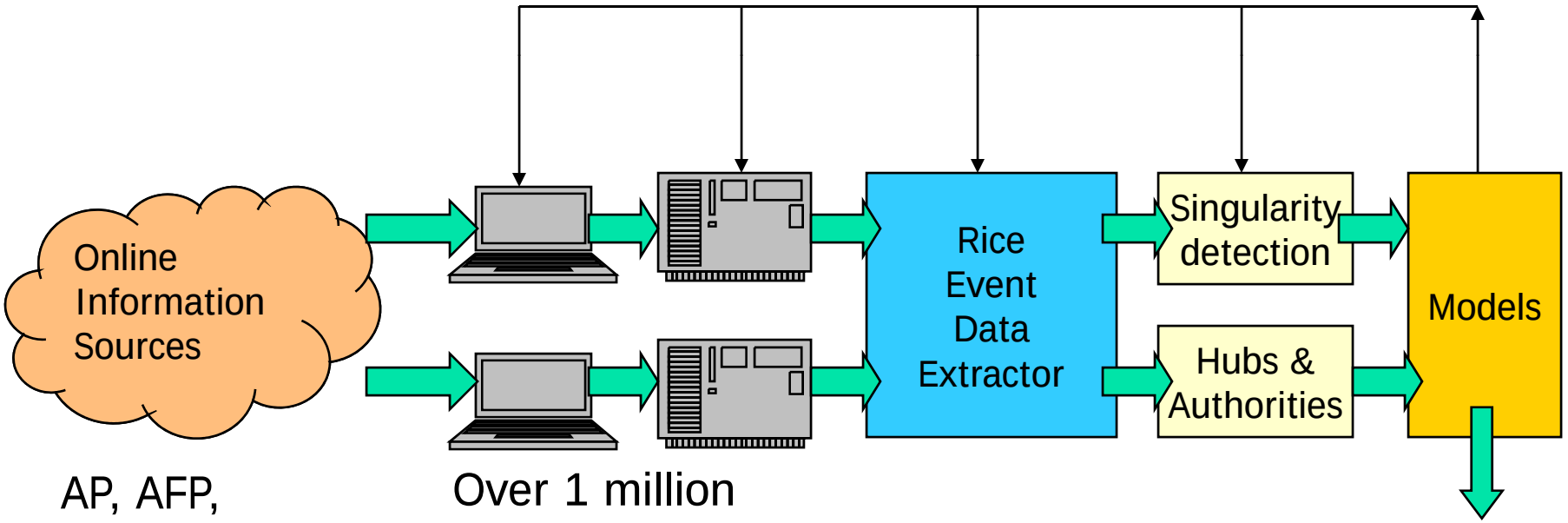


Task Question

- Is it possible to monitor news media from regions all over the world over extended periods of time, extracting low-level events from them, and piece them together to automatically track and predict conflict in all the regions of the world?

The Ares project

http://ares.cs.rice.edu



AP, AFP,
BBC, Reuters,
...

Over 1 million articles on the Middle East from 1979 to 2005 (filtered automatically)

Analysis of wire stories

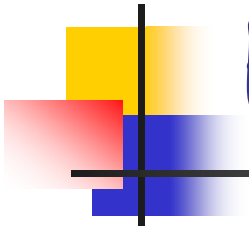
"President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein but declined to say whether that meant he would forego immediate air strikes on Iraq."

Relevance filter

Date	Actor	Target	Weis Code	Wies event	Goldstein scale
790415	ARB	ISR	223	(MIL ENGAGEMENT)	-10
790415	EGY	AFD	194	(HALT NEGOTIATION)	-3.8
790415	PALPL	ISR	223	(MIL ENGAGEMENT)	-10
790415	UNK	ISR	223	(MIL ENGAGEMENT)	-10
790415	ISR	EGY	31	(MEET)	1
790415	EGY	ISR	31	(MEET)	1
790415	ISRMIL	PAL	223	(MIL ENGAGEMENT)	-10
790415	PALPL	JOR	223	(MIL ENGAGEMENT)	-10
790415	EGY	AFD	193	(CUT AID)	-5.6
790415	IRQ	EGY	31	(MEET)	1
790415	EGY	IRQ	31	(MEET)	1
790415	ARB	CHR	223	(MIL ENGAGEMENT)	-10
790415	JOR	AUS	32	(VISIT)	1.9
790415	UGA	CHR	32	(VISIT)	1.9
790415	ISRGOV	ISRSET	54	(ASSURE)	2.8

Singularity detection on aggregated events data

Hubs and authorities analysis of events data



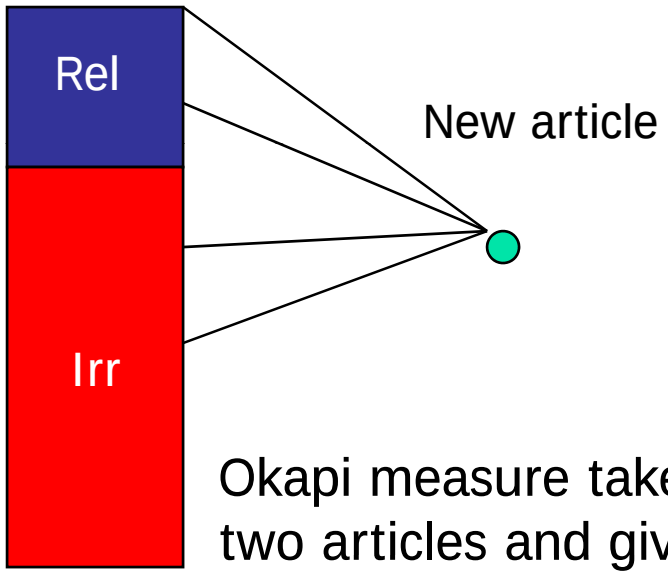
Embedded learner design

- Representation
  - Identify relevant stories, extract event data from them, build time series models and graph-theoretic models.
- Learning
  - Identifying regime shifts in events data, tracking evolution of militarized interstate disputes (MIDs) by hubs/authorities analysis of events data
- Decision-making
  - Issuing early warnings of outbreak of MIDs

Identifying relevant stories

- Only about 20% of stories contain events that are to be extracted.
  - The rest are interpretations, (e.g., op-eds), or are events not about conflict (e.g., sports)
- We have trained Naïve Bayes (precision 86% and recall 81%), SVM classifiers (precision 92% and recall 89%) & Okapi classifiers (precision 93% and recall 87%) using a labeled set of 180,000 stories from Reuters.
- Surprisingly difficult problem!
  - Lack of large labeled data sets;
  - Poor transfer to other sources (AP/BBC)
  - The category of "event containing stories" is not well-separated from others, and changes with time

Okapi classifier



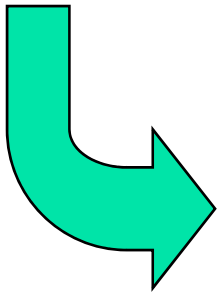
Okapi measure takes two articles and gives the similarity between them.

- Reuters data set: relevant categories are GVIO, GDIP, G13; irrelevant categories: 1POL, 2ECO, 3SPO, ECAT, G12, G131, GDEF, GPOL

Decision rule: sum of top N Okapi scores in Rel set > sum of top N Okapi scores in Irr set then classify as rel; else irr

Event extraction

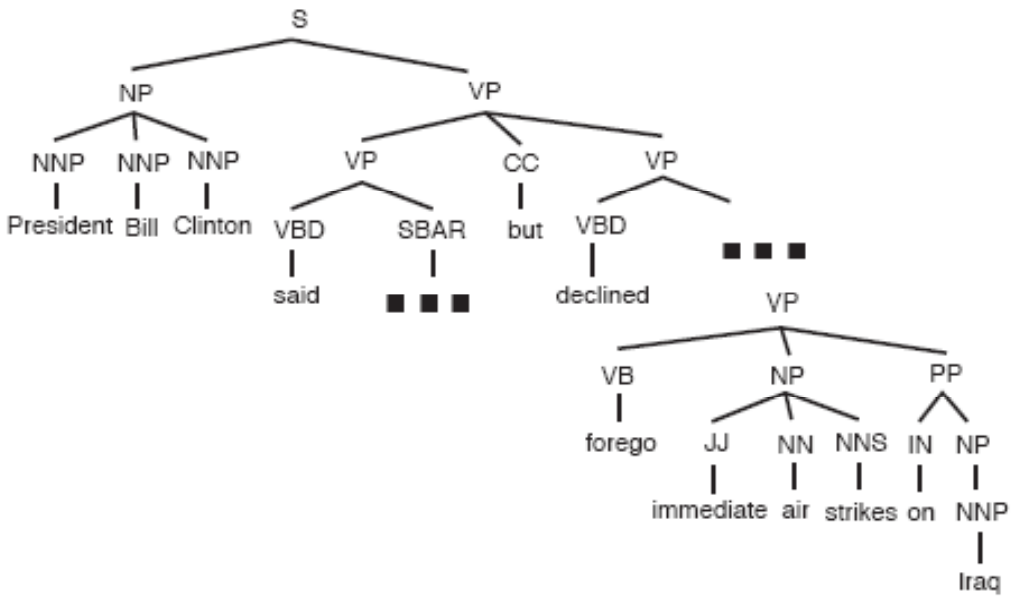
"President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein but declined to say whether that meant he would forego immediate air strikes on Iraq."



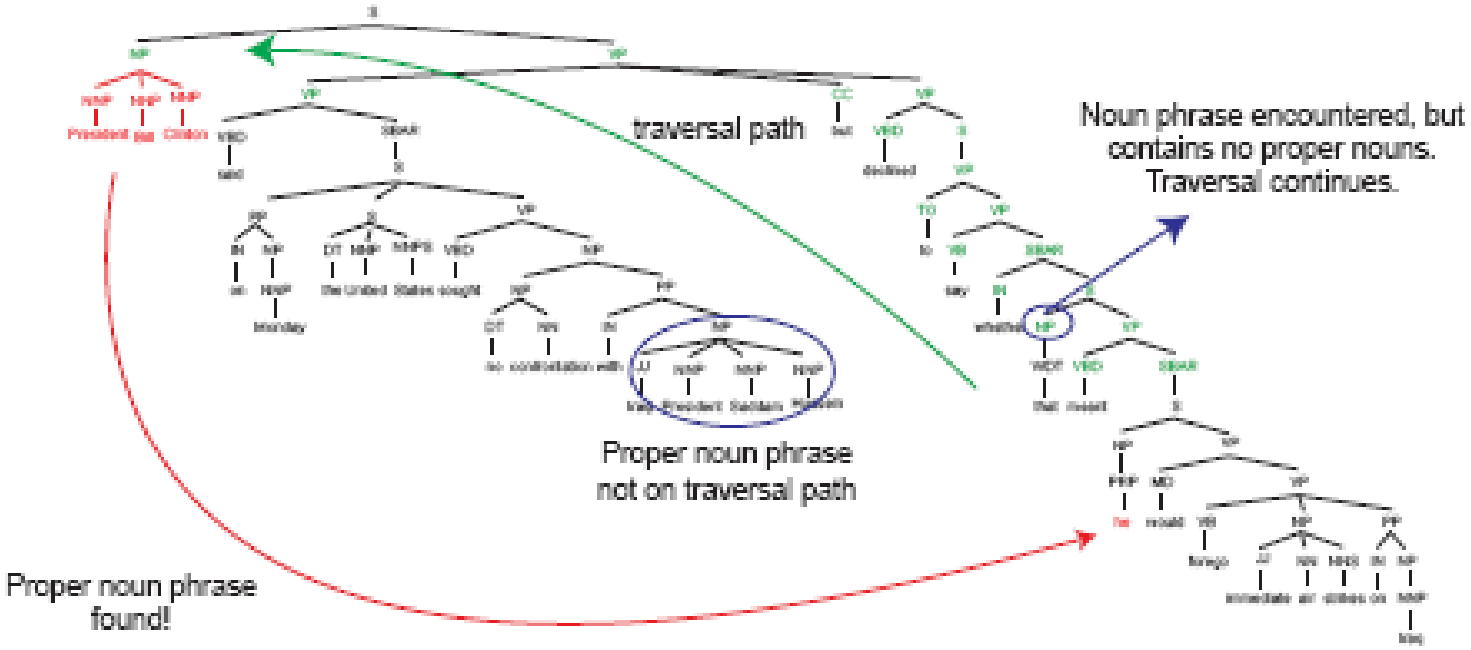
Fragment	Event data
President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein	USA Comment USA
the United States sought no confrontation with Iraqi President Saddam Hussein	USA Deny IRQ
President Bill Clinton declined to say whether that meant President Bill Clinton would forego immediate air strikes on Iraq	USA Comment USA
President Bill Clinton would forego immediate air strikes on Iraq	not part of event phrase (did not actually happen)

Parse sentence

1. *President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein but declined to say whether that meant he would forego immediate air strikes on Iraq.*

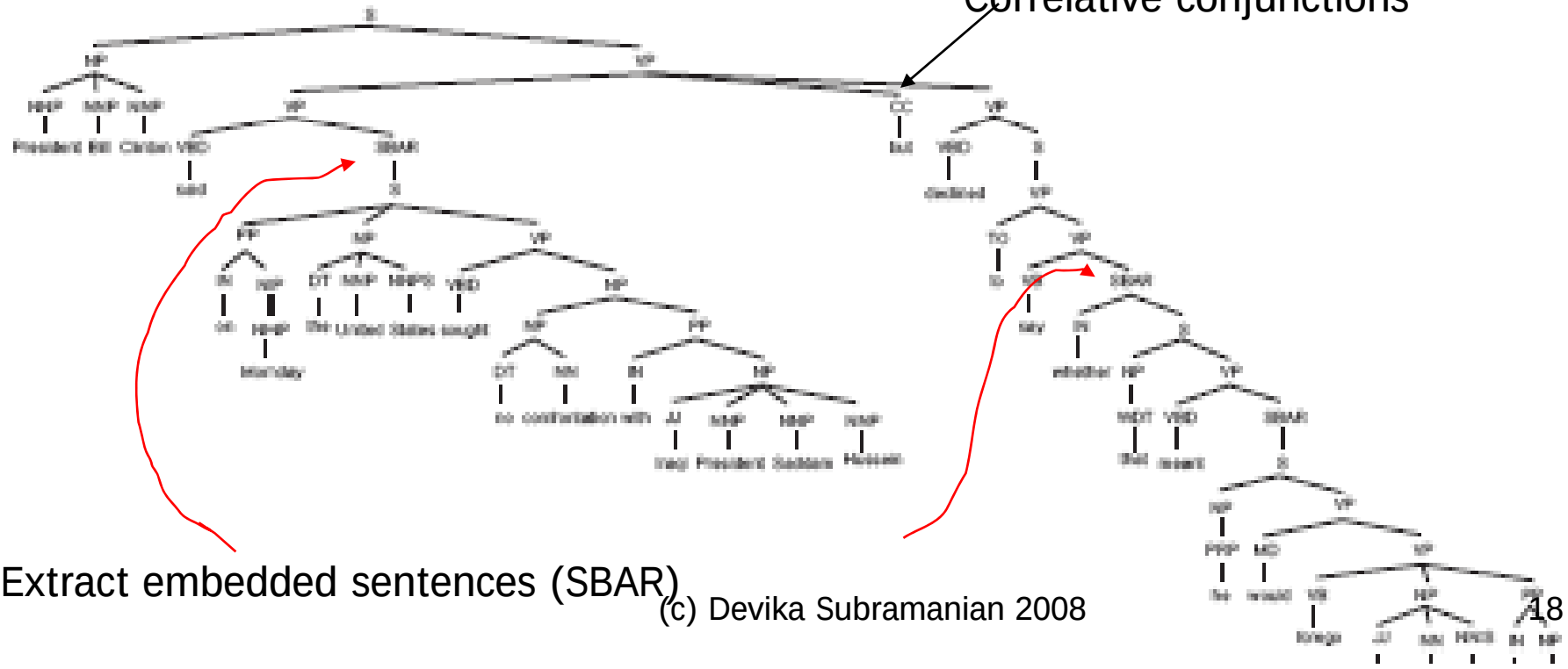


Pronoun de-referencing



Sentence fragmentation

"President Bill Clinton said on Monday the United States sought no confrontation with Iraqi President Saddam Hussein but declined to say whether that meant he would forego immediate air strikes on Iraq."



Extract embedded sentences (SBAR)

Conditional random fields

We extract who (actor) did what (event) to whom (target)

Actor & target labels	Event category labels
-countries, capitals, nationalities	-specific event keywords
-words in proper noun phrases	-words in main verb phrase
-actors occur before main verb	-specific parts of speech
-actors at higher levels of tree	-not modified by negative words
-targets occur after main verb	-part of event phrase
-targets at lower levels of tree	

Not exactly the same as NER

Results

TABARI is state of the art coder in political science

Table 1: Results for 22 (Force) category

Coder	Accuracy	Recall	Precision
TABARI	22%	7%	50%
TABARI with frag	20%	8%	83%
CRF	72%	70%	91%

Table 2: Results for 02 (Comment) category

Coder	Accuracy	Recall	Precision
TABARI	81%	31%	67%
TABARI with frag	88%	54%	93%
CRF	89%	96%	68%

200 Reuters sentences; hand-labeled with actor, target, and event codes (22 and 02).

Stepinksi, Stoll, Subramanian 2006

Events data

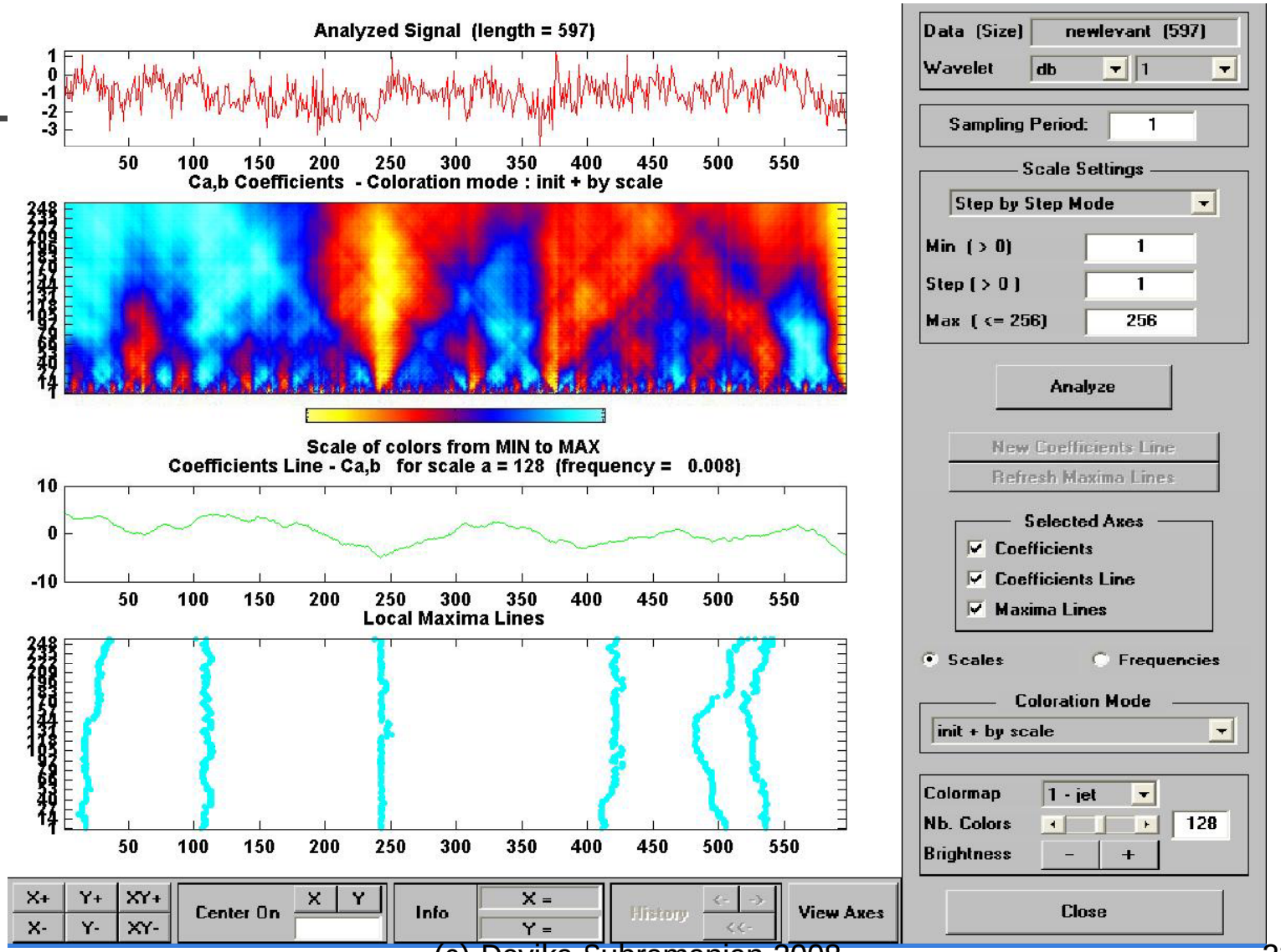
Date	Actor	Target	Weis Code	Wies event	Goldstein scale
790415	ARB	ISR	223	(MIL ENGAGEMENT)	-10
790415	EGY	AFD	194	(HALT NEGOTIATION)	-3.8
790415	PALPL	ISR	223	(MIL ENGAGEMENT)	-10
790415	UNK	ISR	223	(MIL ENGAGEMENT)	-10
790415	ISR	EGY	31	(MEET)	1
790415	EGY	ISR	31	(MEET)	1
790415	ISRMIL	PAL	223	(MIL ENGAGEMENT)	-10
790415	PALPL	JOR	223	(MIL ENGAGEMENT)	-10
790415	EGY	AFD	193	(CUT AID)	-5.6
790415	IRQ	EGY	31	(MEET)	1
790415	EGY	IRQ	31	(MEET)	1
790415	ARB	CHR	223	(MIL ENGAGEMENT)	-10
790415	JOR	AUS	32	(VISIT)	1.9
790415	UGA	CHR	32	(VISIT)	1.9
790415	ISRGOV	ISRSET	54	(ASSURE)	2.8

177,336 events from April 1979 to October 2003 in Levant data set (KEDS).

What can be predicted?

"IF THESE NUMBERS ARE CORRECT, THEN EVERYTHING IS GOING TO HAPPEN AT ONCE TOMORROW MORNING AT 10:35."

Singularity detection



Stoll and Subramanian, 2004, 2006

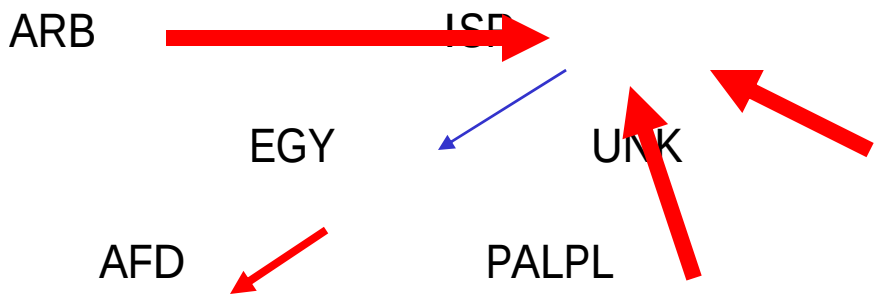
Singularities = MID start/end

biweek	Date range	event
17-35	11/79 to 8/80	Start of Iran/Iraq war
105-111	4/83 to 7/83	Beirut suicide attack, end of Iran/Iraq war
244	1/91 to 2/91	Desert Storm
413-425	1/95 to 7/95	Rabin assassination/start of Intifada
483-518	10/97 to 2/99	US/Iraq confrontation via Richard Butler/arms inspectors
522-539	4/99 to 11/99	Second intifada Israel/Palestine

Interaction graphs

- Model interactions between countries in a directed graph.

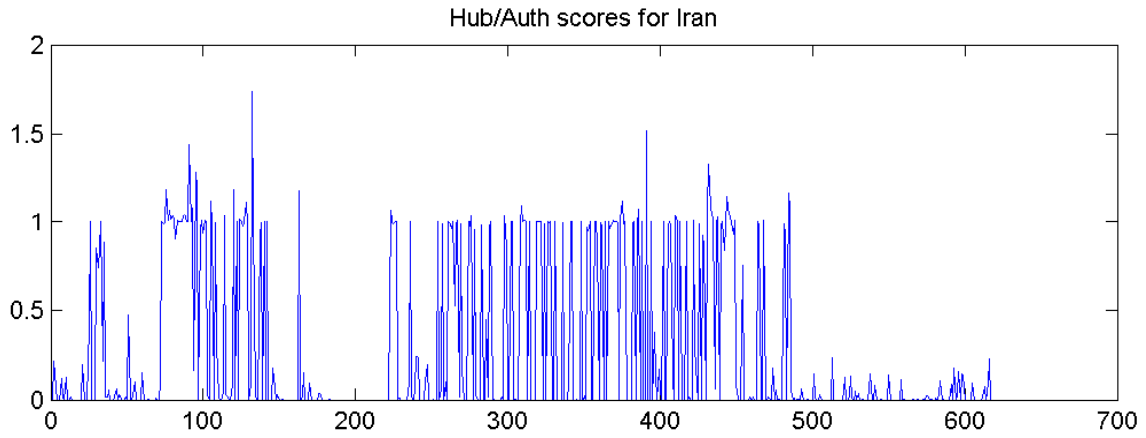
Date	Actor	Target	Weis Code	Wies event	Goldstein scale
790415	ARB	ISR	223	(MIL ENGAGEMENT)	-10
790415	EGY	AFD	194	(HALT NEGOTIATION)	-3.8
790415	PALPL	ISR	223	(MIL ENGAGEMENT)	-10
790415	UNK	ISR	223	(MIL ENGAGEMENT)	-10
790415	ISR	EGY	31	(MEET)	1

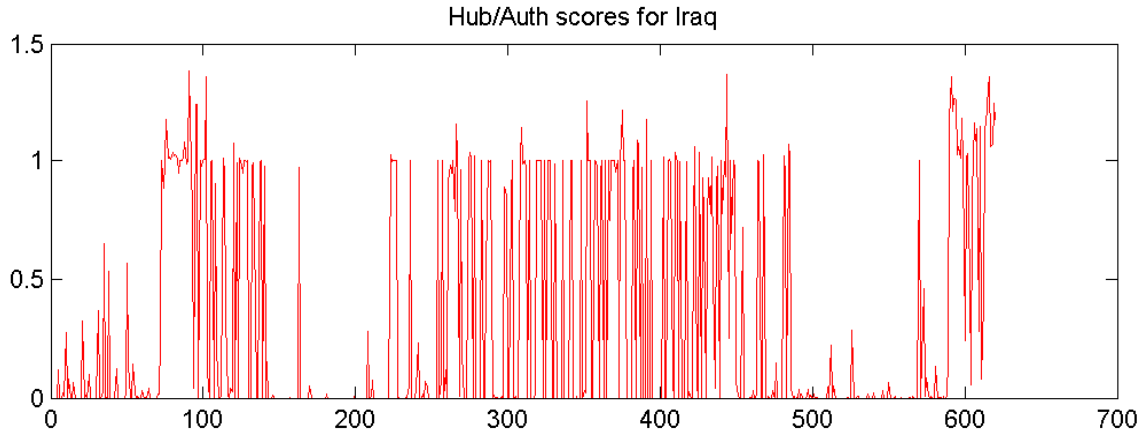


Hubs and authorities for events data

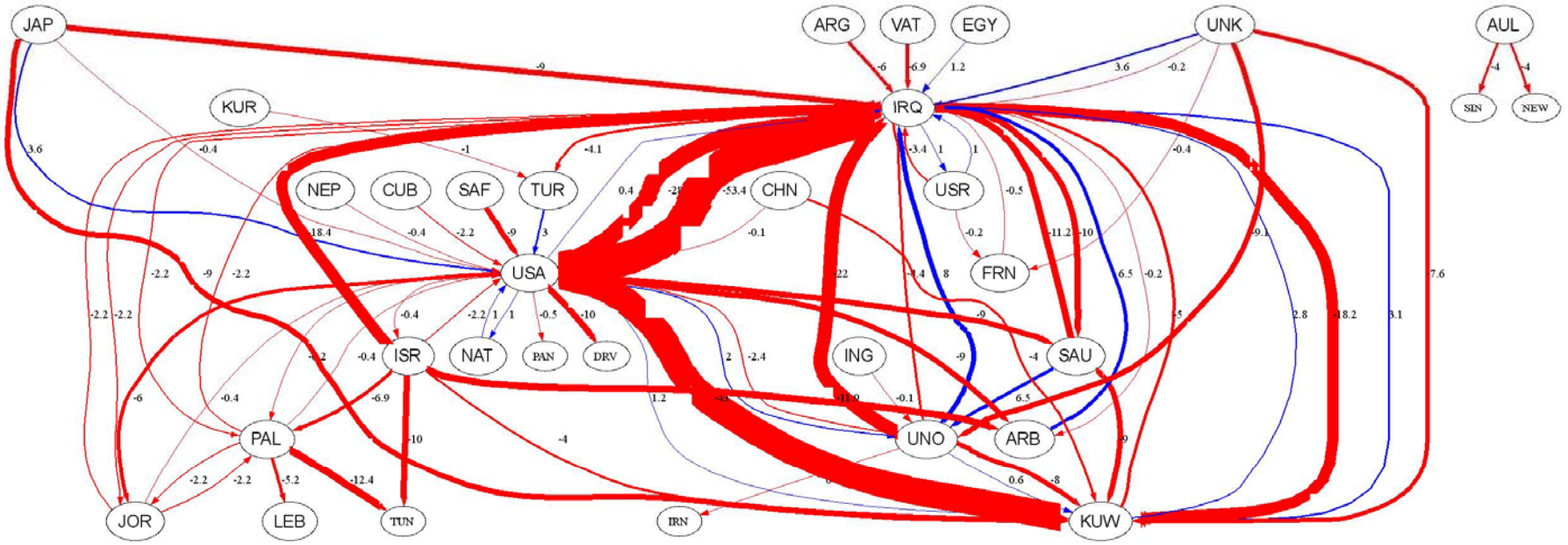
- A hub node is an important initiator of events.
- An authority node is an important target of events.
- Hypothesis:
  - Identifying hubs and authorities over a particular temporal chunk of events data tells us who the key actors and targets are.
  - Changes in the number and size of connected components in the interaction graph signal potential outbreak of conflict.

Hubs/Authorities picture of Iran Iraq war





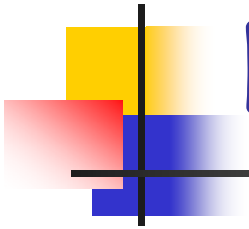
2 weeks prior to Desert Storm



Validation using MID data

- Number of bi-weeks with MIDS in Levant data: 41 out of 589.
- Result 1: Hubs and Authorities correctly identify actors and targets in impending conflict.
- Result 2: Simple regression model on change in hubs and authorities scores, change in number of connected components, change in size of largest component 4 weeks before MID, predicts MID onset.
- Problem: false alarm rate of 16% can be reduced by adding political knowledge of conflict.

"HERE'S THE GROUND RULE: DON'T TELL ME WHAT I SHOULD HAVE DONE."



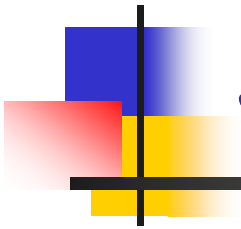
New direction

- Extracting economic events along with political events to improve accuracy of prediction of both economic and political events.

Publications

- An OKAPI-based approach for article filtering, Lee, Than, Stoll, Subramanian, 2006 Rice University Technical Report.
- Hubs, authorities and networks: predicting conflict using events data, R. Stoll and D. Subramanian, International Studies Association, 2006 (invited paper).
- Events, patterns and analysis, D. Subramanian and R. Stoll, in Programming for Peace: Computer-aided methods for international conflict resolution and prevention, 2006, Springer Verlag, R. Trappl (ed).
- Four Way Street? Saudi Arabia's Behavior among the superpowers, 1966-1999, R. Stoll and D. Subramanian, James A Baker III Institute for Public Policy Series, 2004.
- Events, patterns and analysis: forecasting conflict in the 21st century, R. Stoll and D. Subramanian, Proceedings of the National Conference on Digital Government Research, 2004.
- Forecasting international conflict in the 21st century, D. Subramanian and R. Stoll, in Proc. of the Symposium on Computer-aided methods for international conflict resolution, 2002.

The research team



Analysis of events in the SHSU terrorism database

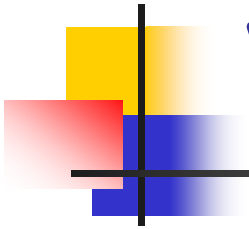
Devika Subramanian

Rice University

Joint work with Chris Bronk
SHSU team provided the data

The SHSU database

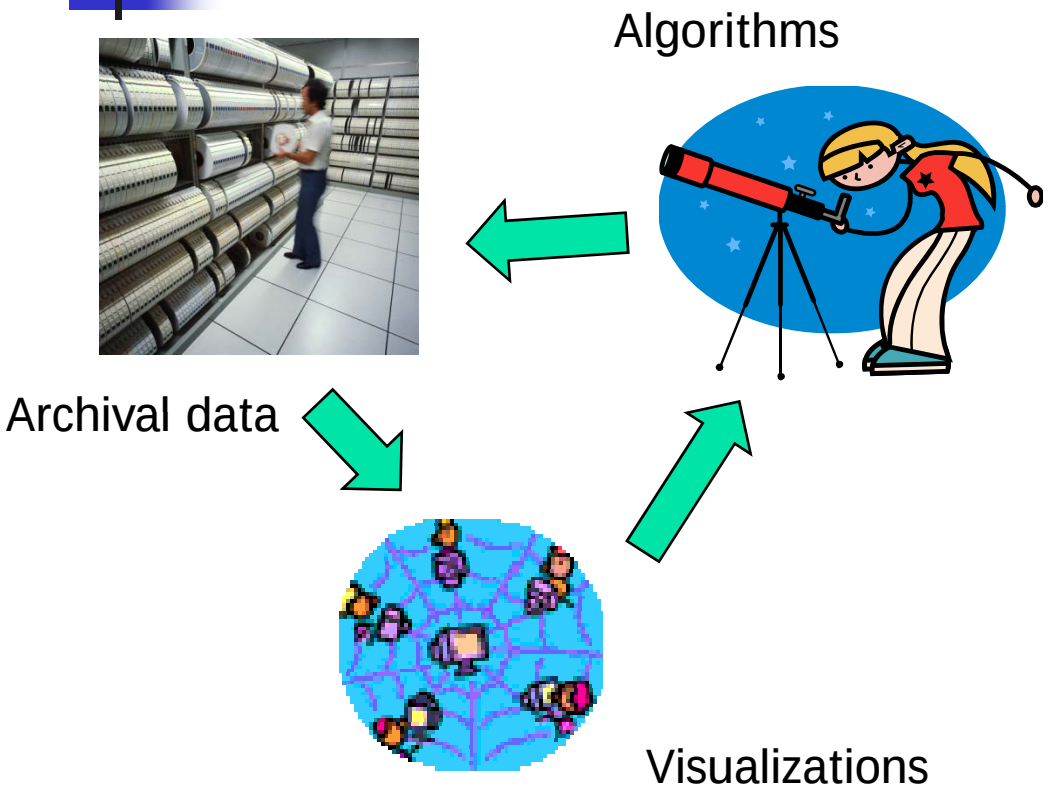
- A human curated database of global terrorist incidents from 1/22/1990 to 12/31/2007
  - 31,199 incidents
  - 1257 groups
- Very detailed information on incidents (e.g. weapons used, fatalities, etc) and some information on the groups.



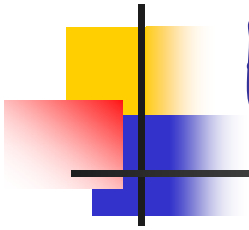
What can we do this data?

- A unique resource.
- Potential for answering substantive questions in how terror networks form and evolve, trace weapon use to link terror groups, …

Approach

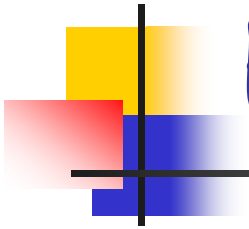


- Start with a substantive question or hypothesis.
- Train our computational instruments on the data, and extract statistical evidence to support/falsify the hypothesis.
- Rinse, repeat!



Example

- Step 0: Substantive question: What was the impact of the Bali bombings on the interaction between terror groups around the world?
- Step 1: Extract relevant data
  - Get incident data (incident,group) in two chunks: 90 days prior to Bali event, 90 days after Bali event



Example (contd.)

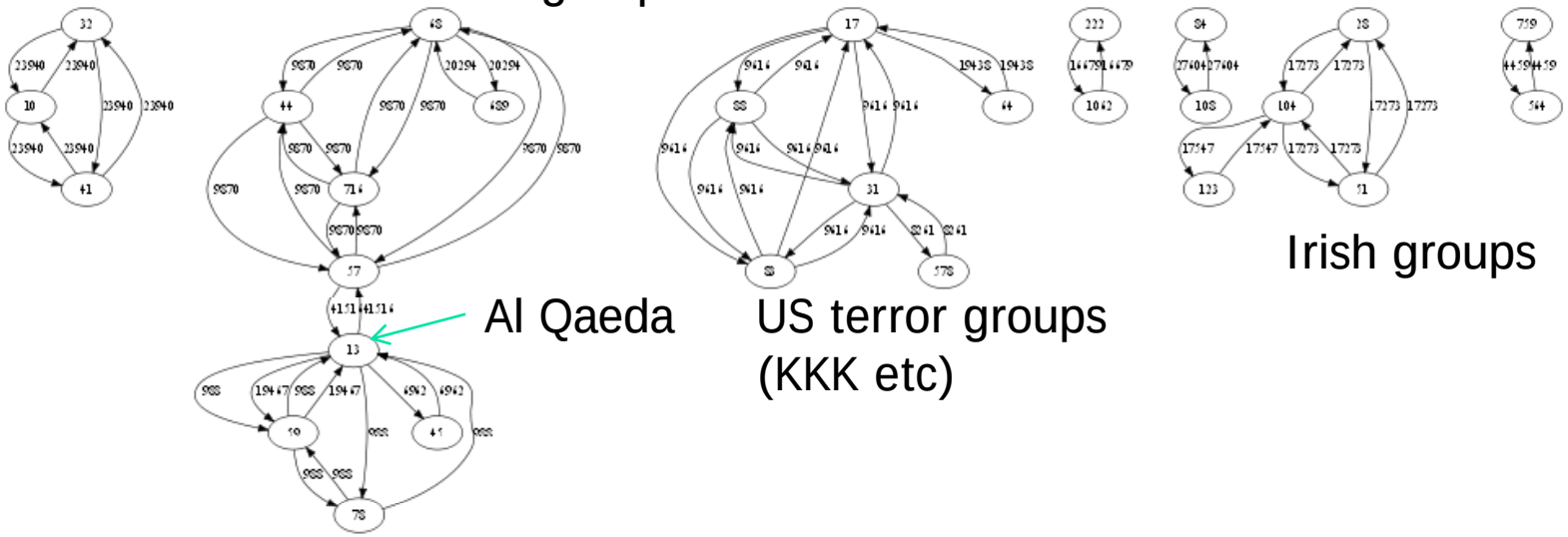
- Step 2: Analyze extracted data
  - Calculate which groups cooperated in these two incident chunks.
  - Two groups are related if they both participate in an incident.
- Step 3: Visualize the data and perform statistical hypothesis tests

Pre-Bali network

Palestine groups

Kashmir groups

Columbia



Irish groups

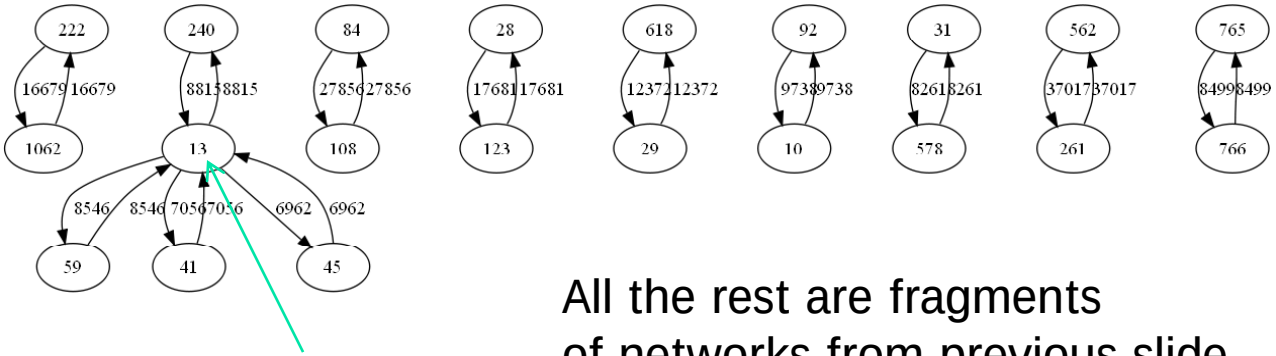
Al Qaeda

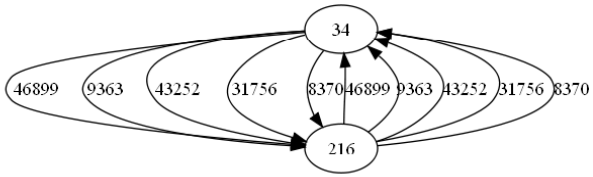
US terror groups
(KKK etc)

Philippines, Indonesian groups
Hamas

Post Bali network

Bangladesh



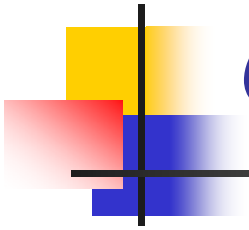


Al Qaeda

All the rest are fragments
of networks from previous slide

US environmental
Terror groups

Splintering of the terror network into smaller, more decentralized pieces



Current work

- A more systematic way to construct these networks, to trace their evolution through time from 1990 to 2007.
- Need more substantive questions to derive analysis.
- Need to link to weapons usage and other criteria to cluster groups better.