



Automating Quantitative Narrative Analysis of News Data

Saatviga Sudhahar
Intelligent Systems Laboratory, University of Bristol

Roberto Franzosi
Department of Sociology/ Program in Linguistics, Emory University

Nello Cristianini
Intelligent Systems Laboratory, University of Bristol



Overview

- Introduction
- Goals
- Related Work
- System Pipeline
- Experiments & Results
- Network Analysis
- Conclusion & Future Work



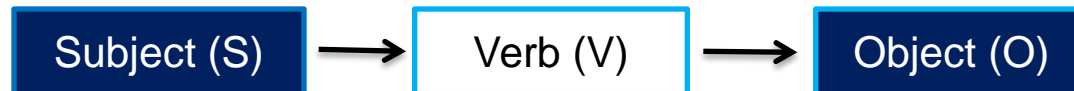
Introduction

- The analysis of news media content is crucial for social science research where sequence of socio-historical events are studied significantly.
- *“An event according to social scientists is an action performed by human beings that can be summed up by a verb or a name of action”* (Roberto Franzosi, 2010).
- An event in Crime news could be,
Police (Actor) arrested (Action) a thief (Actor)
- How could this be represented in computational linguistics?



Introduction

- In linguistics an event can be expressed in the form of a semantic triplet



- Computer assisted Story grammars have been used to analyse narrative text (Quantitative Narrative Analysis)
- The disadvantage is that it is labour intensive



Goals

- We present a working system for large scale quantitative narrative analysis of news corpora experimenting with 100,000 articles about crime from the New York Times corpus
- It automatically extracts SVO triplets out of data
 - By weighting actors we identify the key players in a given domain
 - By analysing the centrality of actors we identify the most influential characters in news narrative
 - By classifying types of actions we further analyse the roles different actors play
 - We detect changes of role in time by analysing the time series of actors
 - We identify the subject/object bias of actors in news

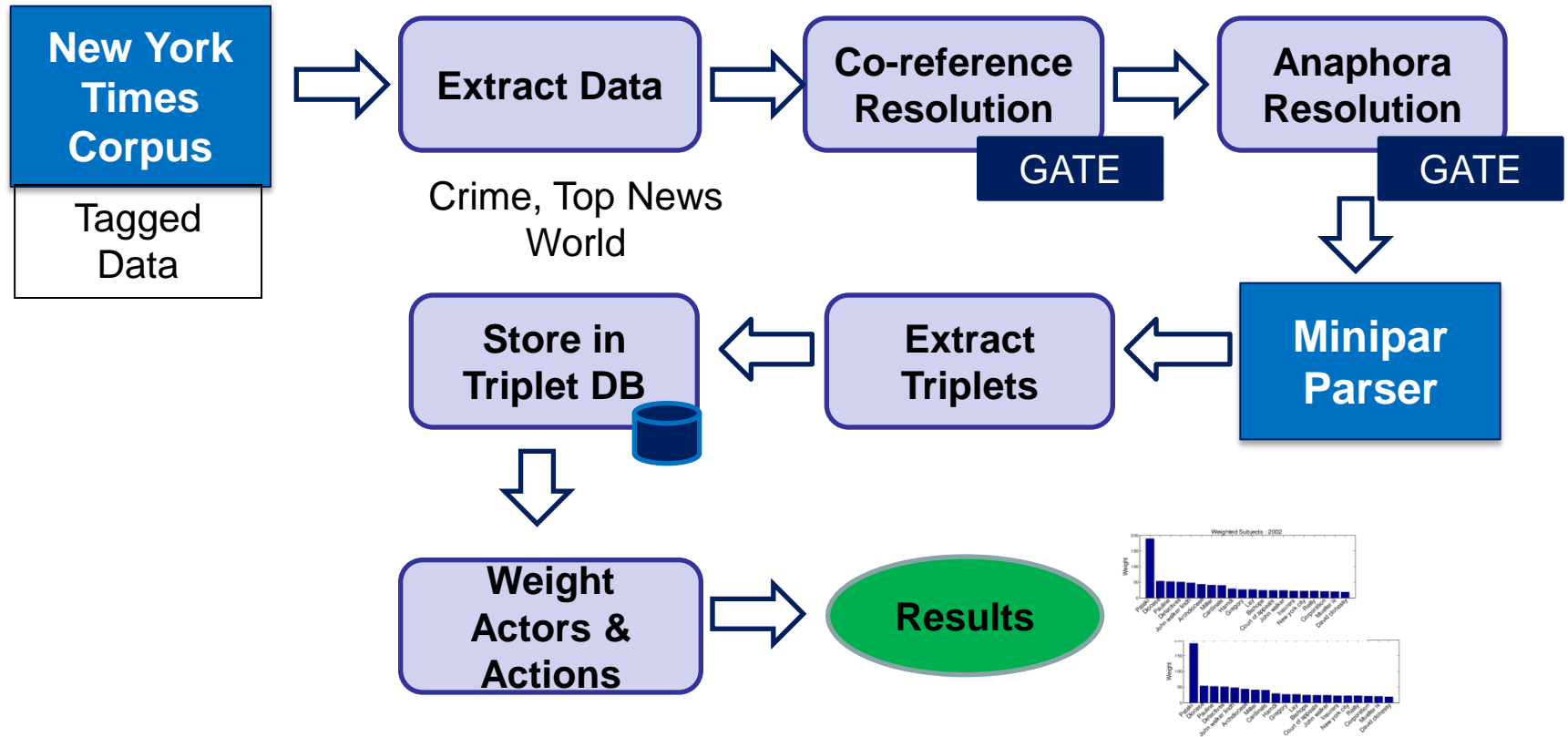


Related Work

- Our approach builds on an idea presented in (Rusu et al, 2007), for purposes of text summarisation and related tasks, and later developed also for the generation of event templates (Trampus, Mladenec, 2011) in information extraction.
- We developed that idea in various ways, in order to address the specific needs of researchers in QNA.
- Our goal is to create a working system that performs key tasks needed in that application area, and that can scale to very large corpora.



System Pipeline



News Article:

Perry criticizes Obama for "dangerous" Mideast policy

Rick Perry accused President Barack Obama on Tuesday of not standing behind Israel as the Texas governor sought to draw Jewish voter support in his bid to win the 2012 Republican presidential nomination. Perry, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that Obama's Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, I have a clear directive to support Israel. Both as an American and as a Christian, I am going to stand with Israel," Perry said.

Perry made his foray into Middle East politics as the Palestinians prepared a unilateral bid for statehood, which they are expected to present as early as Friday to the U.N. Security Council. Perry condemned those Palestinian efforts and called Obama's Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.

(Reuters)



Identifying Co-references

Rick Perry accused President **Barack Obama** on Tuesday of not standing behind Israel as the Texas governor sought to draw Jewish voter support in his bid to win the 2012 Republican presidential nomination. **Perry**, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that **Obama's** Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, I have a clear directive to support Israel. Both as an American and as a Christian, I am going to stand with Israel," **Perry** said.

Perry made his foray into Middle East politics as the Palestinians prepared a unilateral bid for statehood, which they are expected to present as early as Friday to the U.N. Security Council. **Perry** condemned those Palestinian efforts and called **Obama's** Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.



Resolve Co-references

Perry accused President **Obama** on Tuesday of not standing behind Israel as the Texas governor sought to draw Jewish voter support in his bid to win the 2012 Republican presidential nomination. **Perry**, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that **Obama's** Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, I have a clear directive to support Israel. Both as an American and as a Christian, I am going to stand with Israel," **Perry** said.

Perry made his foray into Middle East politics as the Palestinians prepared a unilateral bid for statehood, which they are expected to present as early as Friday to the U.N. Security Council. **Perry** condemned those Palestinian efforts and called **Obama's** Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.



Identifying Pronouns referring each other

Perry accused President Obama on Tuesday of not standing behind Israel as the Texas **governor** sought to draw Jewish voter support in **his** bid to win the 2012 Republican presidential nomination. Perry, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that Obama's Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, **I** have a clear directive to support Israel. Both as an American and as a Christian, I'm going to stand with Israel," **Perry** said.

Perry made **his** foray into Middle East politics as the **Palestinians** prepared a unilateral bid for statehood, which **they** are expected to present as early as Friday to the U.N. Security Council. Perry condemned those Palestinian efforts and called Obama's Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.



Performing Anaphora Resolution

Perry accused President Obama on Tuesday of not standing behind Israel as the Texas governor sought to draw Jewish voter support in **governor's** bid to win the 2012 Republican presidential nomination. Perry, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that Obama's Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, **Perry** have a clear directive to support Israel. Both as an American and as a Christian, I'm going to stand with Israel," Perry said.

Perry made **Perry's** foray into Middle East politics as the Palestinians prepared a unilateral bid for statehood, which **Palestinians** are expected to present as early as Friday to the U.N. Security Council. Perry condemned those Palestinian efforts and called Obama's Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.



Minipar Parser – by Dekang Lin, 1998

- An evaluation with the SUSANNE corpus shows that MINIPAR achieves about 88% precision and 80% recall with respect to dependency relationships.
- Tags each word of the sentences with a grammatical relation
 - i : the relationship between a C clause and its I clause
 - mod : the relationship between a word and its adjunct modifier
 - pnmmod : post nominal modifier
 - pcomp-c : clausal complement of prepositions
 - pcomp-n : nominal complement of prepositions
 - post : post determiner
 - pre : pre determiner
 - pred : predicate of a clause
 - rel : relative clause
 - vrel : passive verb modifier of nouns
 - wha, whn : wh-elements at C-spec positions
 - obj : object of verbs
 - s : surface subject



> (U	*)	
E23 (()		C	E23)	
E15 (()	fin	N	2	s	(gov accuse))
1	(Perry ~				
2	(accused	accuse	V	E15	i (gov fin))
E24 (()	perry	N	2	subj	(gov accuse) (antecedent 1))
3	(President	~	N	2	obj (gov accuse))
4	(Obama~	N	3	person	(gov president))
5	(on ~	Prep	3	mod	(gov president))
6	(Tuesday	~	N	5	pcomp-n (gov on))
7	(of ~	Prep	2	mod	(gov accuse))
E14 (()	vpssc	C	7	pcomp-c	(gov of))
E21 (()	~	N	E14	s	(gov vpssc) (antecedent 1))
8	(not ~	A	9	amod	(gov stand))
9	(standing	stand	V	E14	i (gov vpssc))
E25 (()	perry	N	9	subj	(gov stand) (antecedent E21))
10	(behind ~	Prep	9	mod	(gov stand))
11	(Israel ~	N	10	pcomp-n	(gov behind))
12	(as ~	Prep	11	mod	(gov Israel))
13	(the ~	Det	15	det	(gov governor))
14	(Texas ~	N	15	nn	(gov governor))
15	(governor	~	N	12	pcomp-n (gov as))
16	(sought seek	V	15	vrel	(gov governor))
E26 (()	governor	N	16	obj	(gov seek) (antecedent 15))
E13 (()	inf	C	16	mod	(gov seek))
E20 (()	~	N	E13	s	(gov inf) (antecedent E21))
17	(to ~	Aux	18	aux	(gov draw))
18	(draw ~	V	E13	i	(gov inf))
E27 (()	~	N	18	subj	(gov draw) (antecedent E20))
19	(Jewish ~	A	21	mod	(gov support))
20	(voter ~	N	21	nn	(gov support))
21	(support~	N	18	obj	(gov draw))



Extracting Triplets

Rick Perry accused President Barack Obama on Tuesday of not standing behind Israel as the Texas governor sought to draw Jewish voter support in his bid to win the 2012 Republican presidential nomination. Perry, an evangelical Christian who leads the opinion polls among Republican presidential hopefuls, told several dozen New York Jewish leaders that Obama's Middle East policy was "naive, arrogant, misguided and dangerous. " As a Christian, I have a clear directive to support Israel. Both as an American and as a Christian, I am going to stand with Israel," Perry said.

Perry made his foray into Middle East politics as the Palestinians prepared a unilateral bid for statehood, which they are expected to present as early as Friday to the U.N. Security Council. Perry condemned those Palestinian efforts and called Obama's Middle East policy "appeasement" for contending that the Israelis and Palestinians should use the 1967 borders as the starting point for negotiations.

(Reuters)

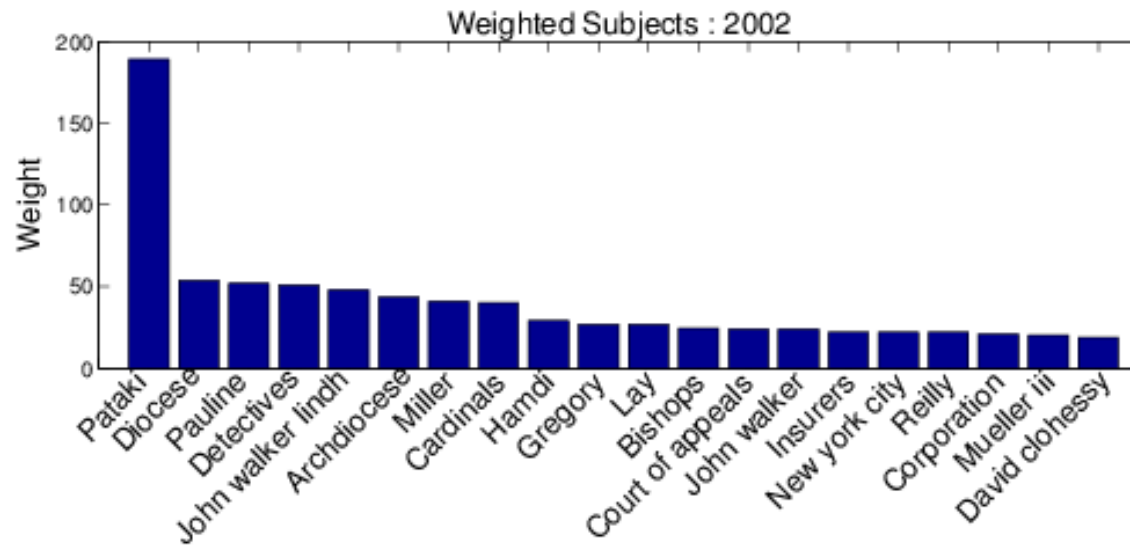
Perry Accuse President Obama
Perry Lead Polls
Perry Have Directive
Perry Make Foray
Palestinians Prepare Bid
Perry Condemn Efforts
Israelis Use Borders

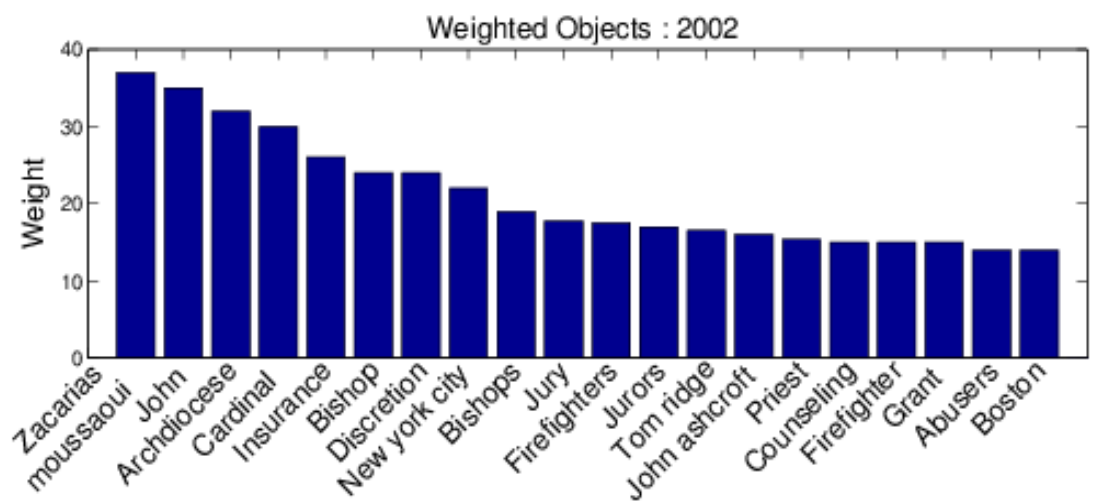
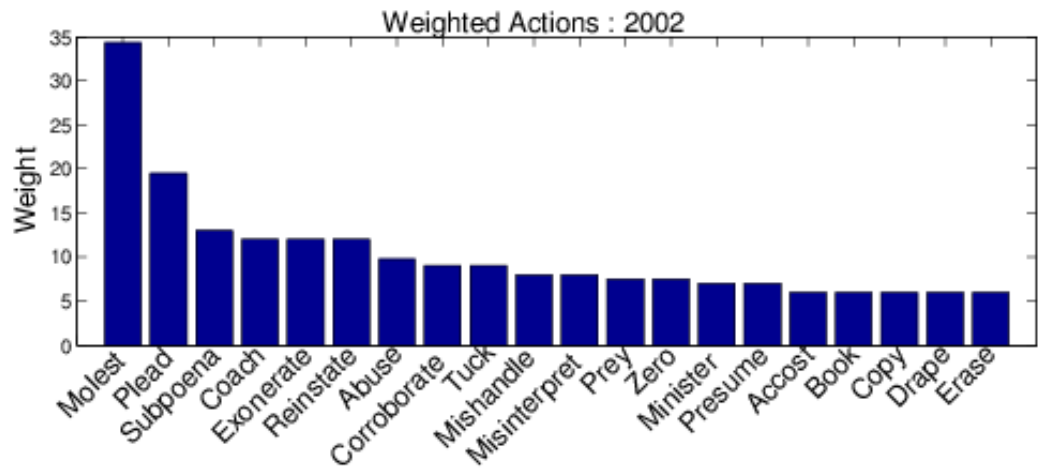


🔥 Identify Key Actors and Actions

- 100,000 articles in Crime and over 180,000 articles in Top News World
- Weighted Actors and Actions

$$\text{Weight } w_{(Actor/Action)} = \frac{f_{(Actor/Action)} \text{ in Crime}}{f_{(Actor/Action)} \text{ in Top News World}}$$



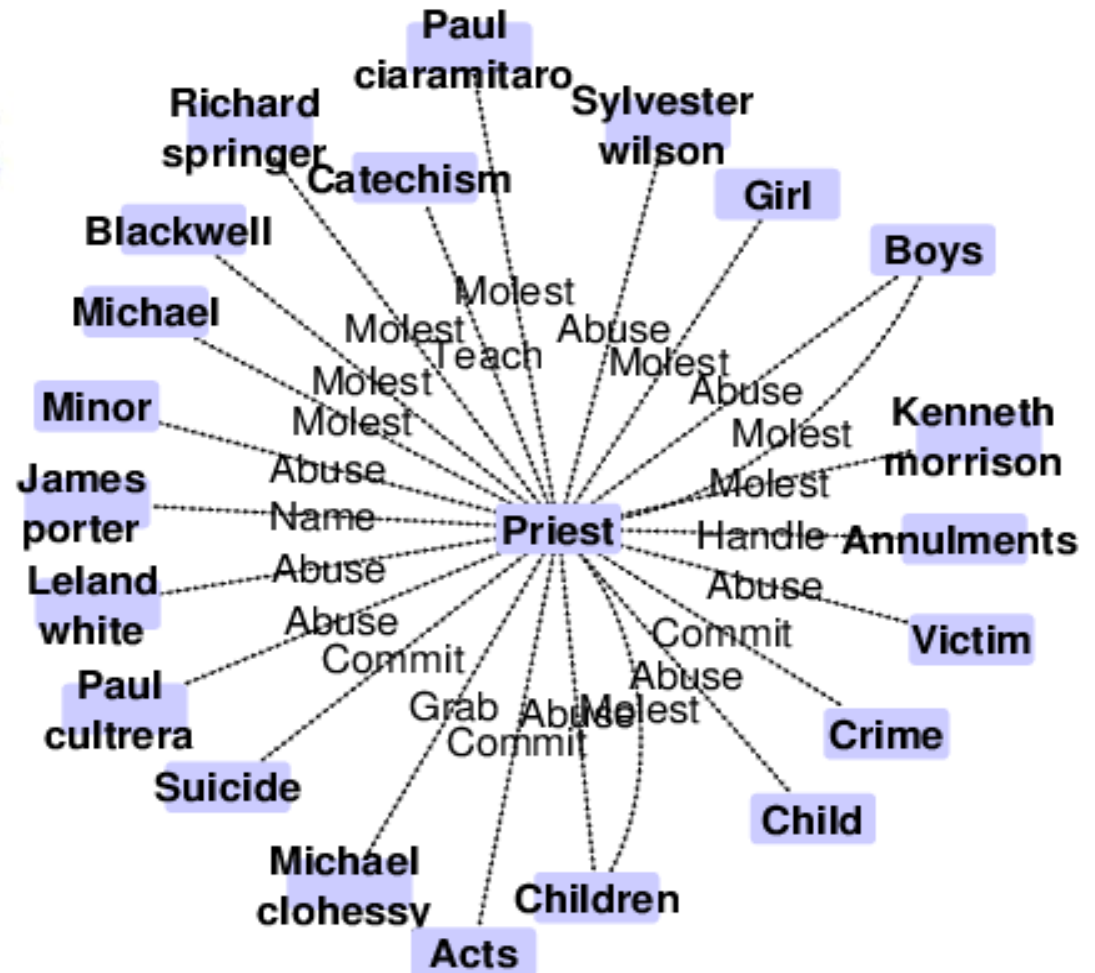


Triplet Networks

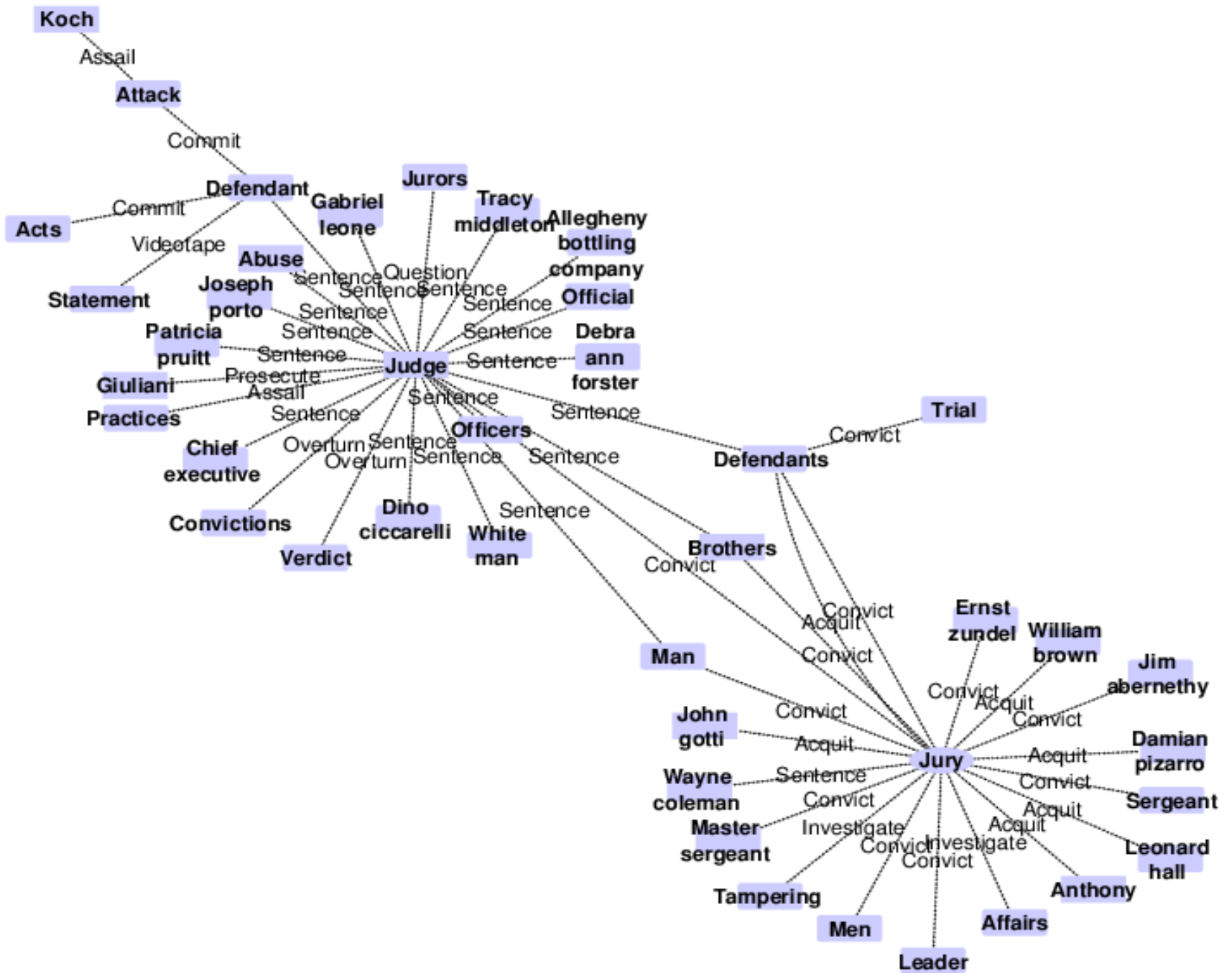
- Network Analysis and its directed graphs provide an ideal tool to map a network of social actors involved in an event
- To identify the most interesting triplets,
 - Filtered the triplets that contained the Top 300 Key actors and actions each year
 - Case1: Key Subject → Key Action → Object
 - Case 2: Subject → Key Action → Key Object
- The most frequent triplets did not reveal interesting relations
- We generated directed networks using Cytoscape
- The networks had subjects and objects as nodes and the verbs linking them as edges.



🌿 Triplet Network : 2002



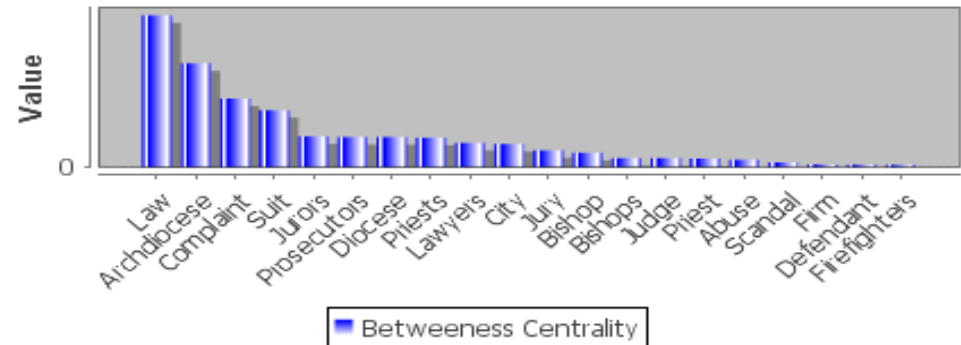




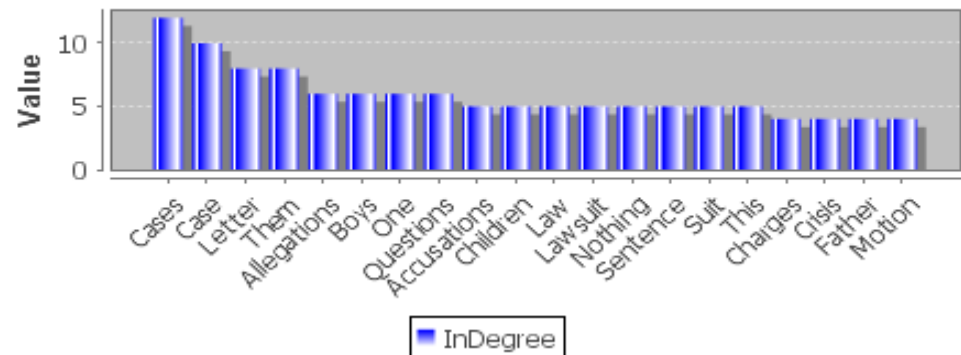
🔥 Network Analysis

- Network Analyser plug-in
 - Basic Network properties
 - Topological properties
 - Betweenness Centrality
 - In degree
 - Out degree
 - Page Rank
 - HITS (Hub & Authority)

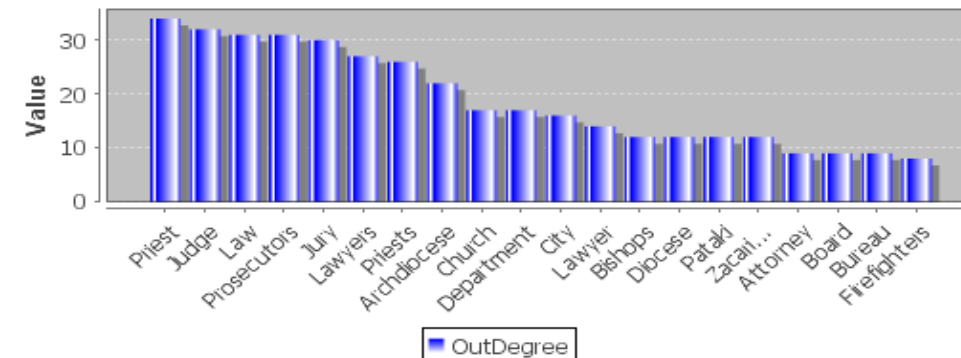
Crime Network: 02



Crime Network: 02



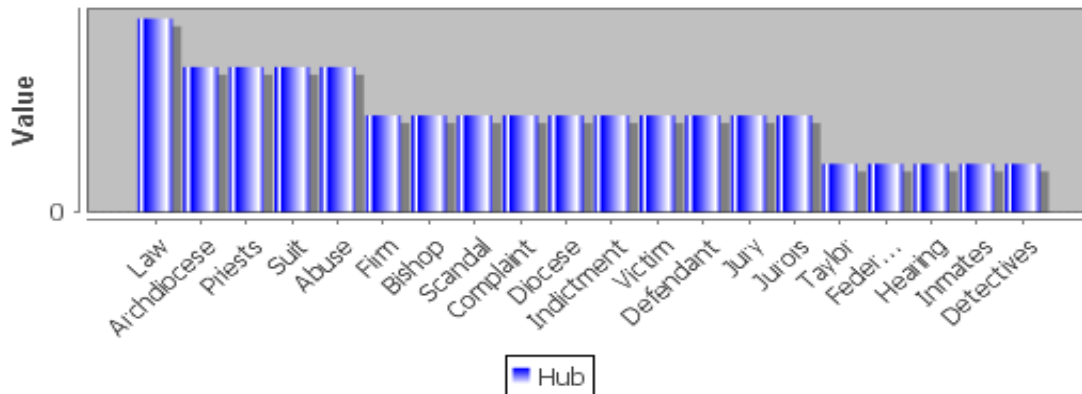
Crime Network: 02



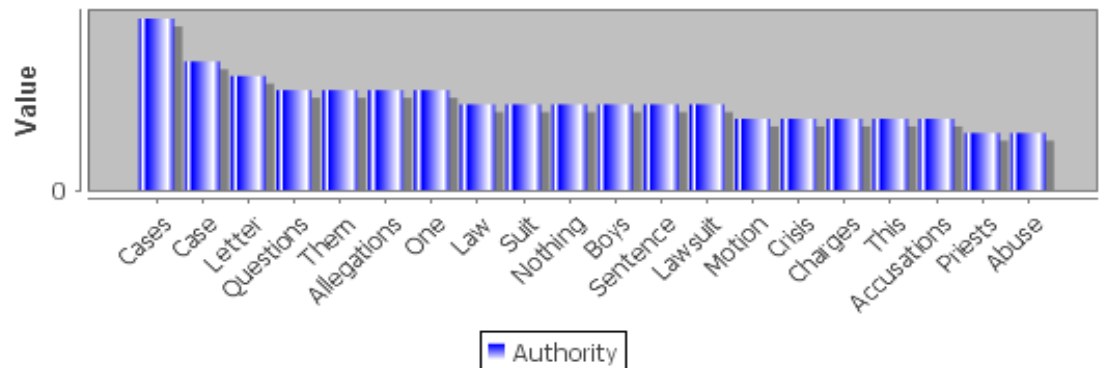
PageRank: 02



Hub: 02



Authority: 02



🔥 Top 10 Most Central Actors : 2002

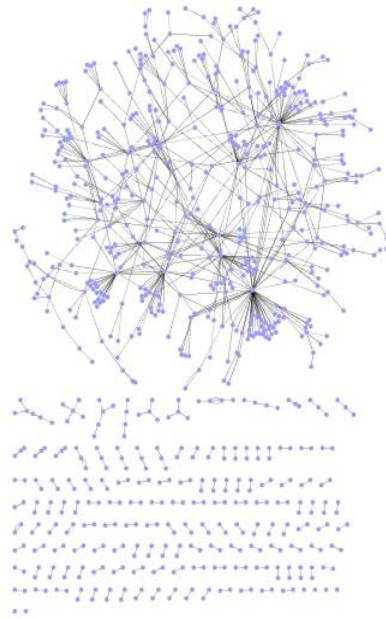
Betweenness Centrality	InDegree	OutDegree	Hub	Authority	Page Rank
Law	Cases	Priest	Law	Cases	Cases
Archdiocese	Case	Judge	Archdiocese	Case	Court
Complaint	Letter	Law	Priests	Letter	Lawsuit
Suit	Allegations	Prosecutors	Suit	Questions	Anyone
Jurors	Boys	Jury	Abuse	Allegations	Nothing
Prosecutors	One	Lawyers	Firm	One	One
Diocese	Questions	Priests	Bishop	Law	Properties
Priests	Accusations	Archdiocese	Scandal	Suit	Play
Lawyers	Children	Church	Complaint	Nothing	Sorts
City	Law	Department	Diocese	Boys	Dying



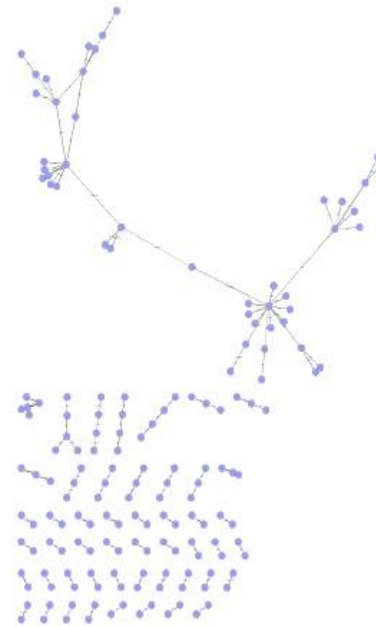
Spheres of Interaction

- In QNA it is also common to investigate separately different “spheres of interaction” between actors (eg: communication, aggression, etc).

Crime against Person : 2002



Crime against Property : 2002



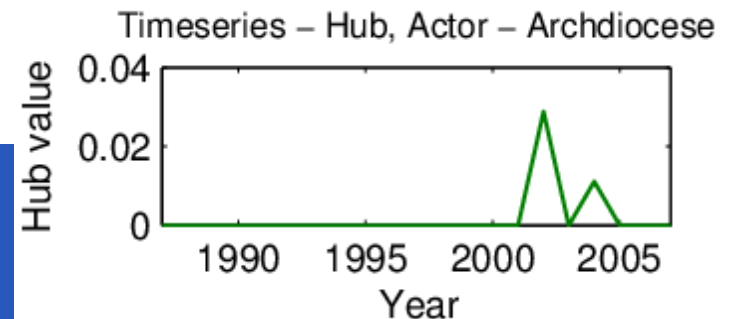
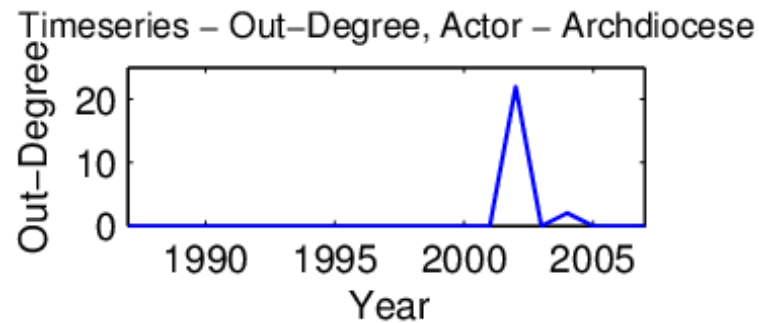
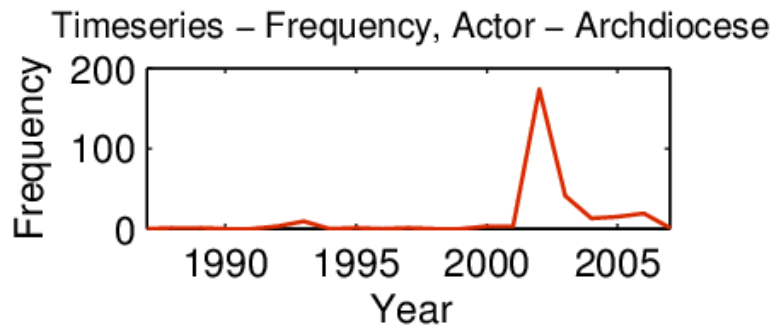
🔥 Top 10 Ranked Subjects/Objects : 2002

Crime against Person		Crime against Property	
Subjects	Objects	Subjects	Objects
Priest	People	Man	Money
Man	Boy	Police	Bank
Troops	Child	Soldiers	Records
Reyes	Girl	Winona Ryder	Millions
Geoghan	Man	Priest	Weapons
Shanley	Woman	People	Wallet
Forces	Jogger	Jason Bogle	Trade Secret
Police	Victim	Investigators	Steven Seagal
United States	Minors	Employee	Most
Others	Me	Agents	Man



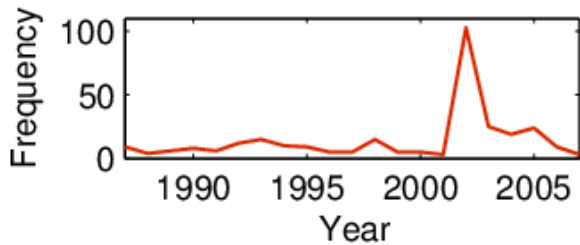
🔥 Time Series Analysis - Actors

- Network measures like outdegree and hub picked up the most central and interesting actors out of the data.

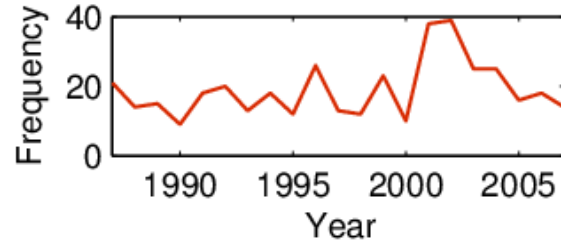


🌟 Time Series Analysis - Actions

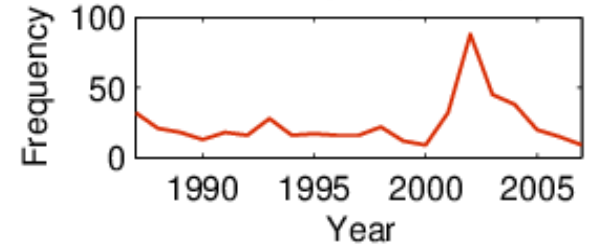
Timeseries - Frequency, Action - Molest



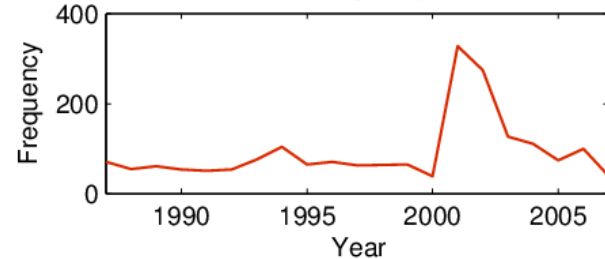
Timeseries - Frequency, Action - Plead



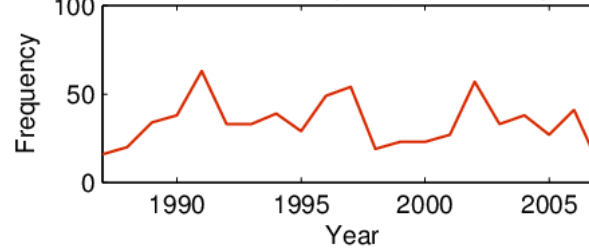
Timeseries - Frequency, Action - Abuse



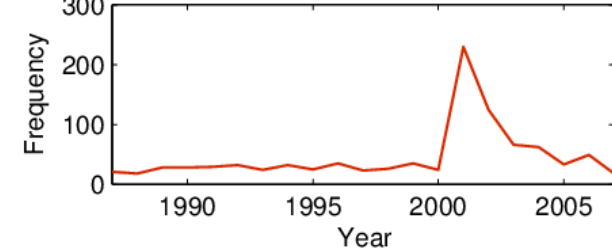
Timeseries - Frequency, Action - Kill



Timeseries - Frequency, Action - Rape



Timeseries - Frequency, Action - Attack



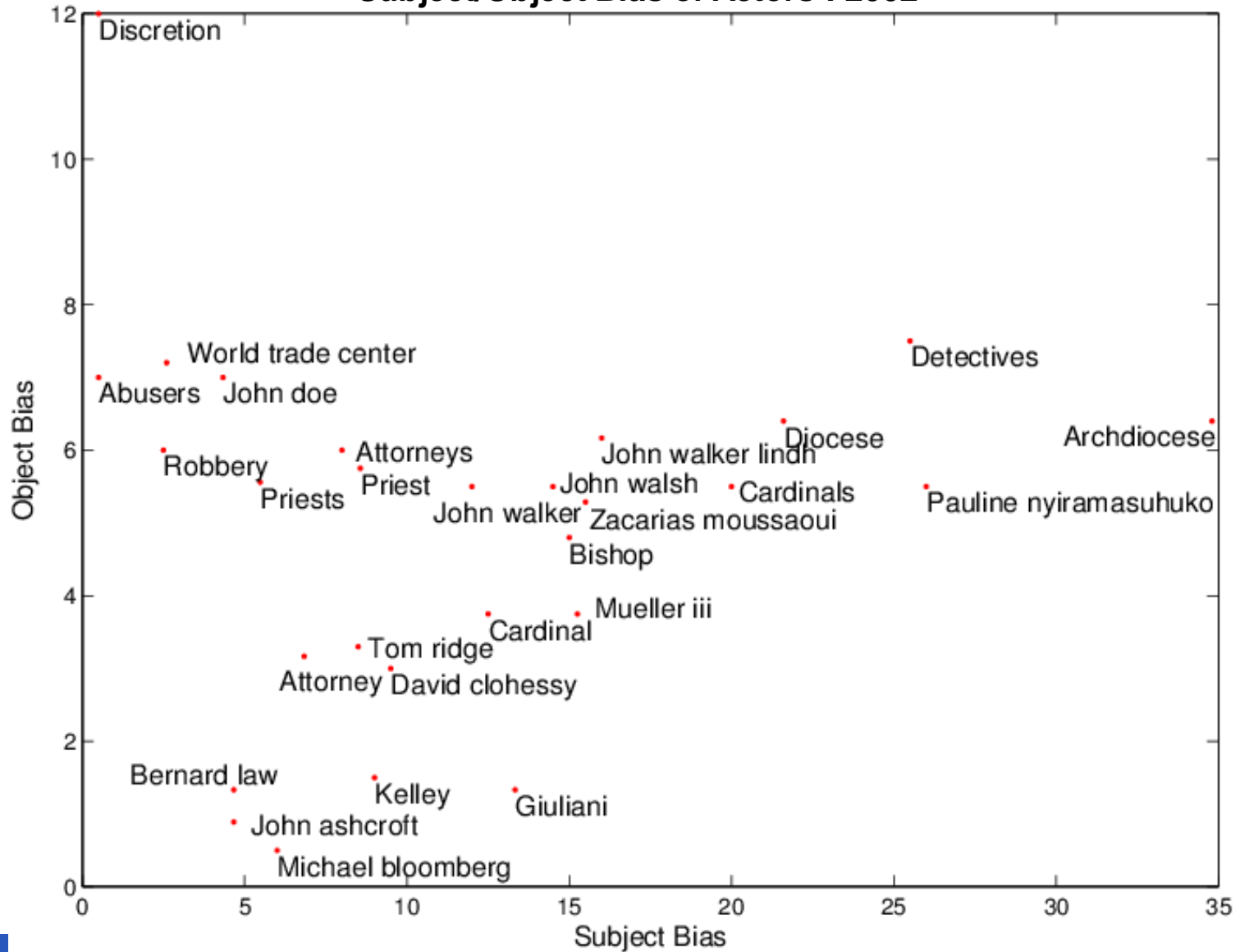
Subject/Object Bias of Actors

- The Subject/Object bias of an actor reveals the role it plays in the news narrative: that is its tendency to be portrayed as an active or passive element in the story.

$$\text{Subj/Obj Bias of Actor}(K) = \frac{f_{\text{Subj/Obj}}^{\text{Crime}}(K)}{f_{\text{Subj}}^{\text{TopNews}}(K) + f_{\text{Obj}}^{\text{TopNews}}(K)}$$



Subject/Object Bias of Actors : 2002



Conclusion & Future Work

- We demonstrated a system for automating narrative analysis of news corpora, a task traditionally labour intensive
- This builds on various recent contributions from the field of Pattern analysis.
- Possible sources of error could be co-reference resolution/pronoun resolution/other steps related to parsing
- Minipar can only parse sentences within a limit of 1024 characters
- Future work will involve both a validation of the performance and a deployment of the system to even larger analysis tasks
- The system can directly feed into existing tools such as PC-ACE (Program for Computer-Assisted Coding of Events, Franzosi, 2010)
- We look forward to integrate this with the larger NOAM infrastructure : News Outlets Analysis and Monitoring System



More to come..

- Analysis of U.S Election news
- Use triplets to infer the existence of certain relations between actors
- Introducing Positive and negative weights to actions in the network
- Automatically identify allegiance of actors to a party and their role in political discourse



 **Thank You**

