

Visual Analysis of Conflicting Opinions

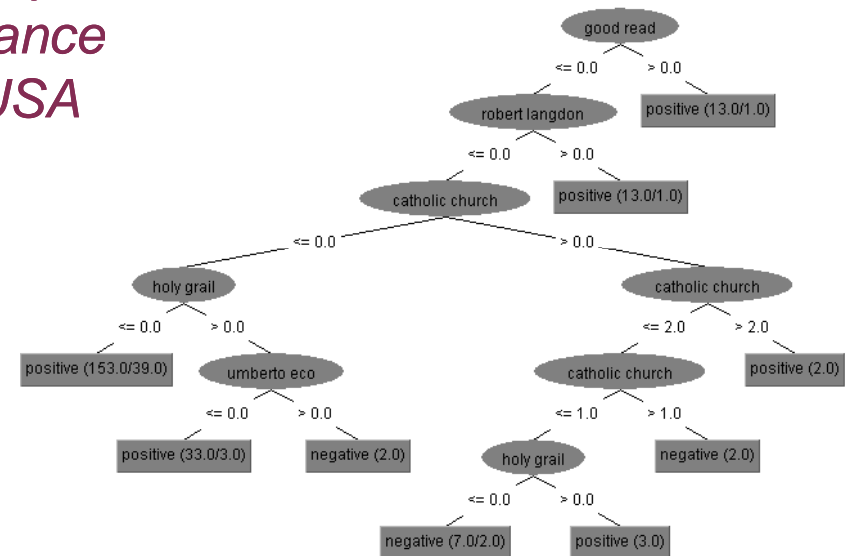
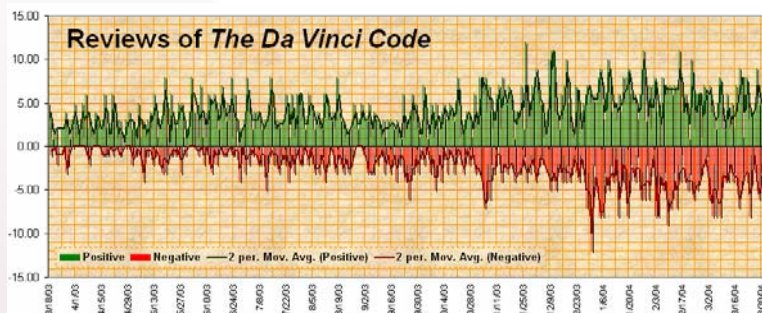


Chaomei Chen, Drexel University USA

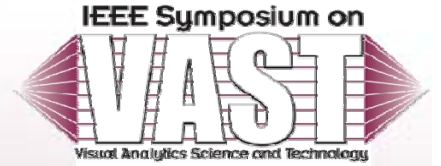
Fidelia Ibekwe-SanJuan, Université de Lyon, France

Eric SanJuan, Université d'Avignon, France

Chris Weaver, Penn State University, USA



Understanding the Dynamics of Knowledge

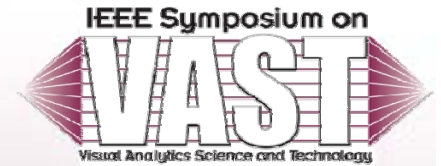


1. What is the hottest topic at time T_0 ?
2. What are the hot topics between time T_a and T_b ? (A Timeline - Temporal)
3. What are the major turning points between time T_a and T_b ? (Turning points - Structural)
4. How did knowledge associated with these turning points spread? (Diffusion - Spatial)
5. **How can we differentiate various perspectives and views based on a large volume of inputs? (Sense-Making - Semantics)**

A quick CiteSpace demo

Topic: **Terrorism**

Time span: **1996-2003**



Q1: Current hot topic?

1. What is the most active topic in the area of **TERRORISM RESEARCH**?

Q2: Previous hot topic?

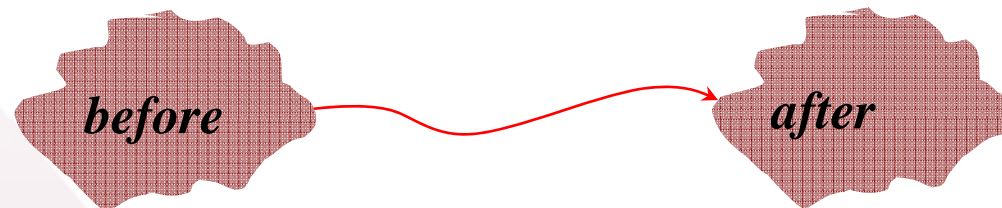
2. What was the previously predominant topic?

Q3: Turning point?

3. What **EVENT/WORK** led to the shift of focus?

Q4: Transition path?

4. What is such a transition path made of?



CiteSpace v. 2.0. Release 10b - (c) 2003-2006 Chaomei Chen ~ Home: D:\CiteSpace\deploy\current

File Edit Project Data Network Analyze Geospatial Demo Preferences Help

Web of Science PubMed CiteSeer

Projects

New Edit Delete Terrorism 1990-2006

Project Home: C:\CiteSpaceProjects\Terroris--1990-2006

Data Directory: C:\DATA\Terroris- 1990-2006.3.21

GO! Stop Reset JVM Memory (K) 509927 Used (%) 51

Space Status

1-year slices	c cc ccv	space	nodes	links
Pruning configuration:				
1996-1996	5 3 0.1	1851	1	0
1997-1997	5 3 0.12	2546	5	10
1998-1998	5 3 0.15	2010	1	0
1999-1999	5 3 0.17	2117	15	39
2000-2000	5 3 0.2	3059	7	10
2001-2001	5 3 0.2	4930	38	189
2002-2002	5 3 0.2	9188	30	65
2003-2003	5 3 0.2	18171	80	317

Process Reports

download4401-4700.txt
download4701-5000.txt
download5001-5300.txt
download5301-5500.txt
download5501-5698.txt
download601-900.txt
download901-1200.txt

Run time: 22256 milliseconds.

Merged network: Nodes=122, Links=551

Exclusion List: 0

Network modeling ends at Wed Oct 25 20:10:04 EDT 2006.

Time Slicing

From 1996 To 2003 # Years per Slice 1

Modeling

Nodes

☐ Citing Authors ☐ Citing Terms
☐ Cited Authors ☒ Cited Documents ☐ Cited Journals

Term Sources

☒ Title ☒ Abstract ☒ Descriptors ☒ Identifiers

Term Selection

☐ None ☐ Noun Phrases ☒ Burst Terms Find Burst Terms

Links

Strength Cosine Scope Within Slices

Thresholding (c, cc, ccv)

5 3 20 5 3 20

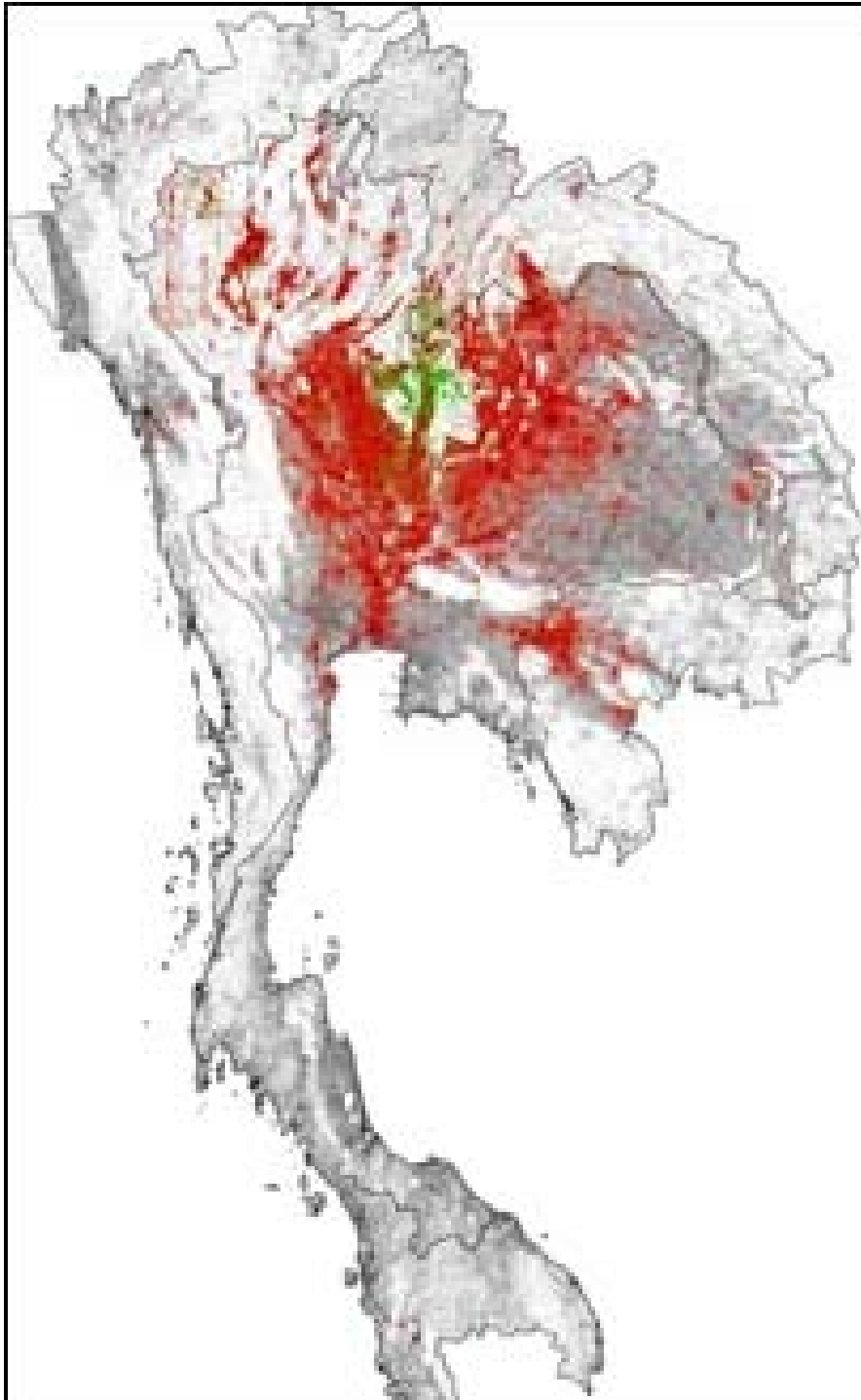
Pruning sliced networks
Minimum Spanning Tree
Pruning the merged network

Visualization

☒ Cluster View (Default) ☐ Time-zone View

☐ Show Networks by Time Slices
☒ Show Merged Network

CiteSpace WebStart



Bird flu spread in Thailand

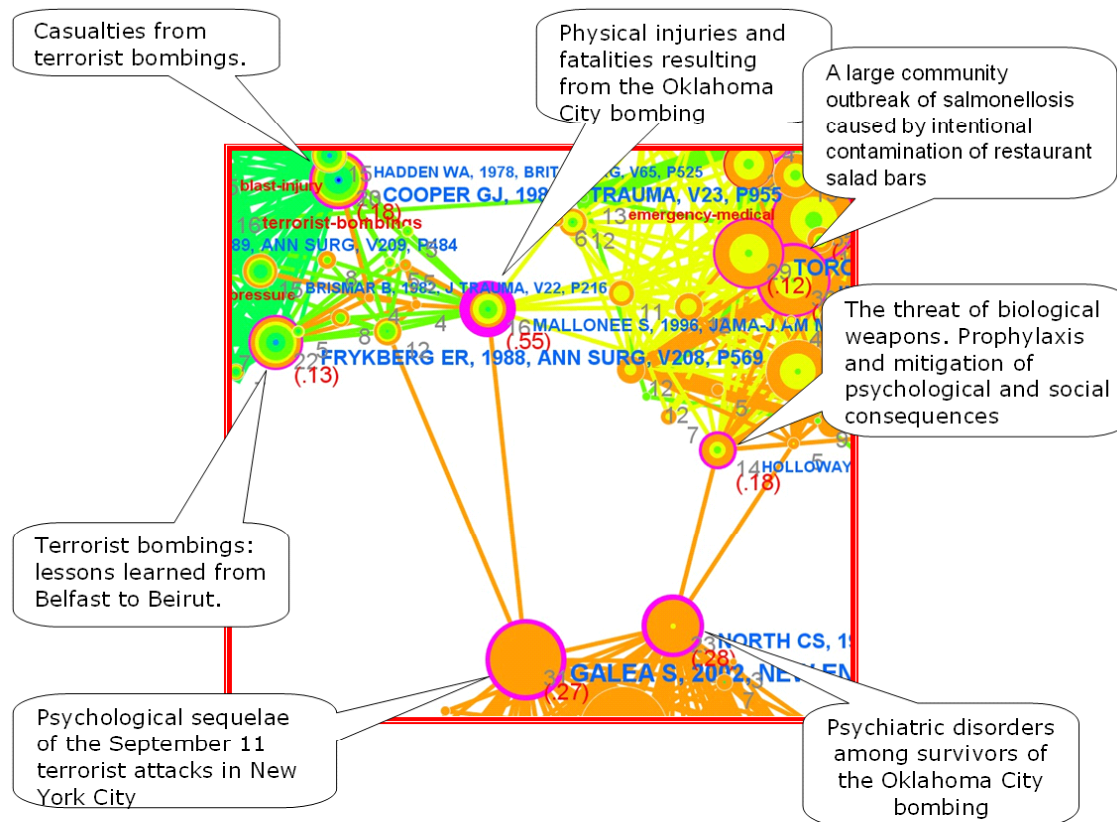
Computer model map of new cases (red) and where the epidemic has finished (green) 60-90 days after an uncontrolled outbreak



Geospatial patterns of terrorist events

Geospatial patterns of relevant research

The Impact of a Pivotal Point



► Glea, S. et al. (2002)

How does the citations to this paper spread geographically?

2002

USA

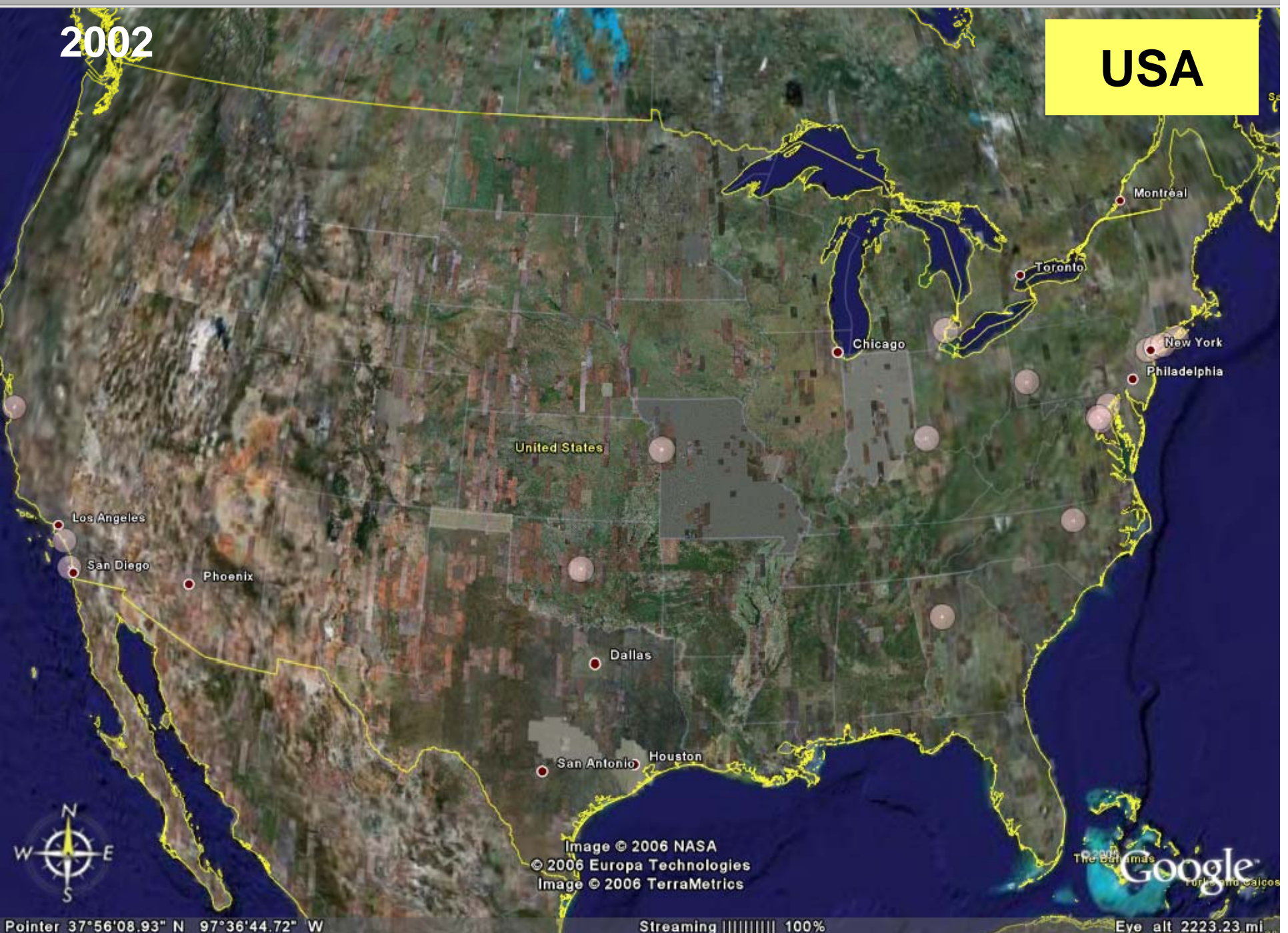


Image © 2006 NASA
© 2006 Europa Technologies
Image © 2006 TerraMetrics

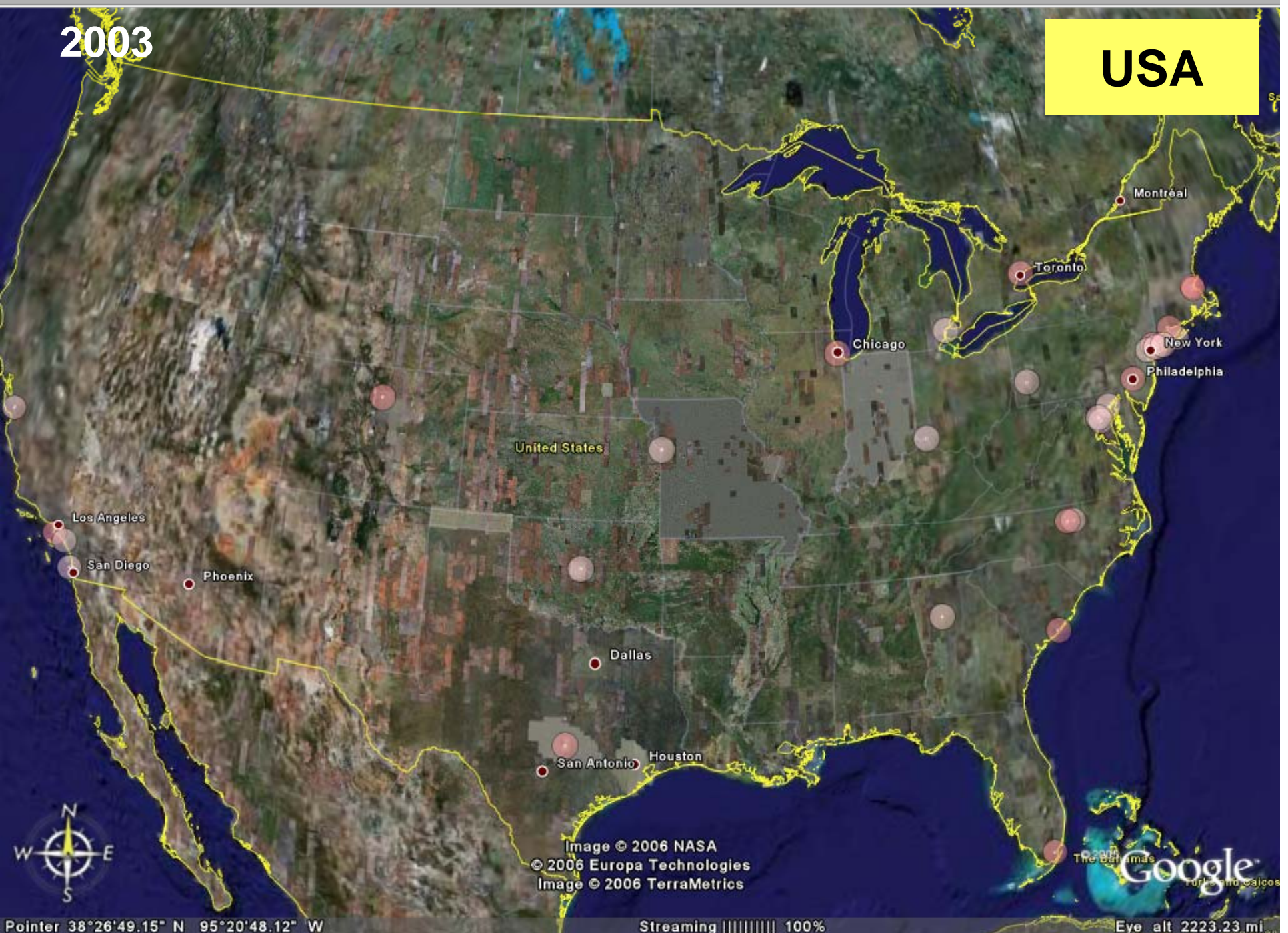
Pointer 37°56'08.93" N 97°36'44.72" W

Streaming ||||| 100%

Eye alt 2223.23 mi

2003

USA



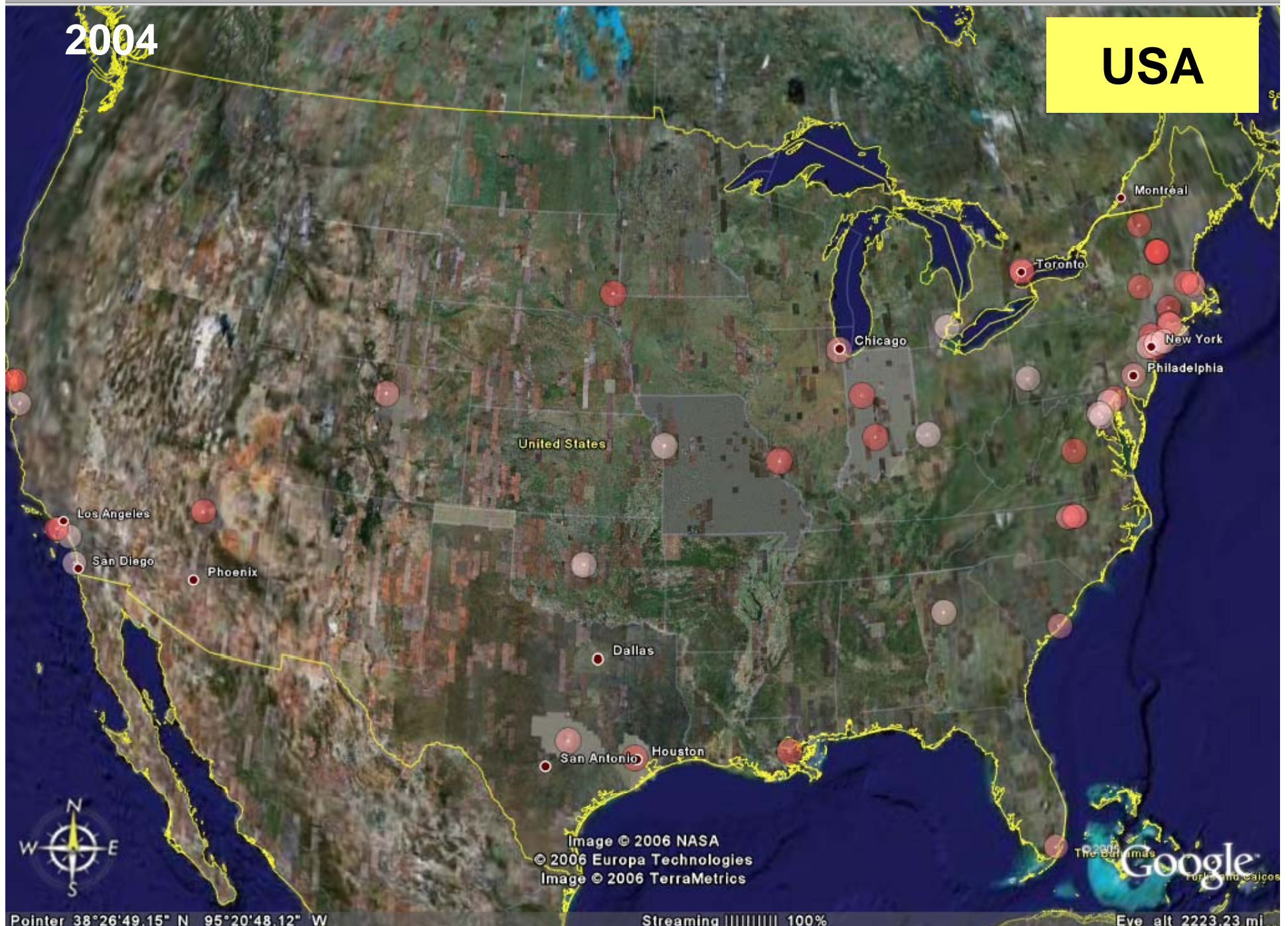
Pointer 38°26'49.15" N 95°20'48.12" W

Streaming ||||| 100%

Eye alt 2223.23 mi

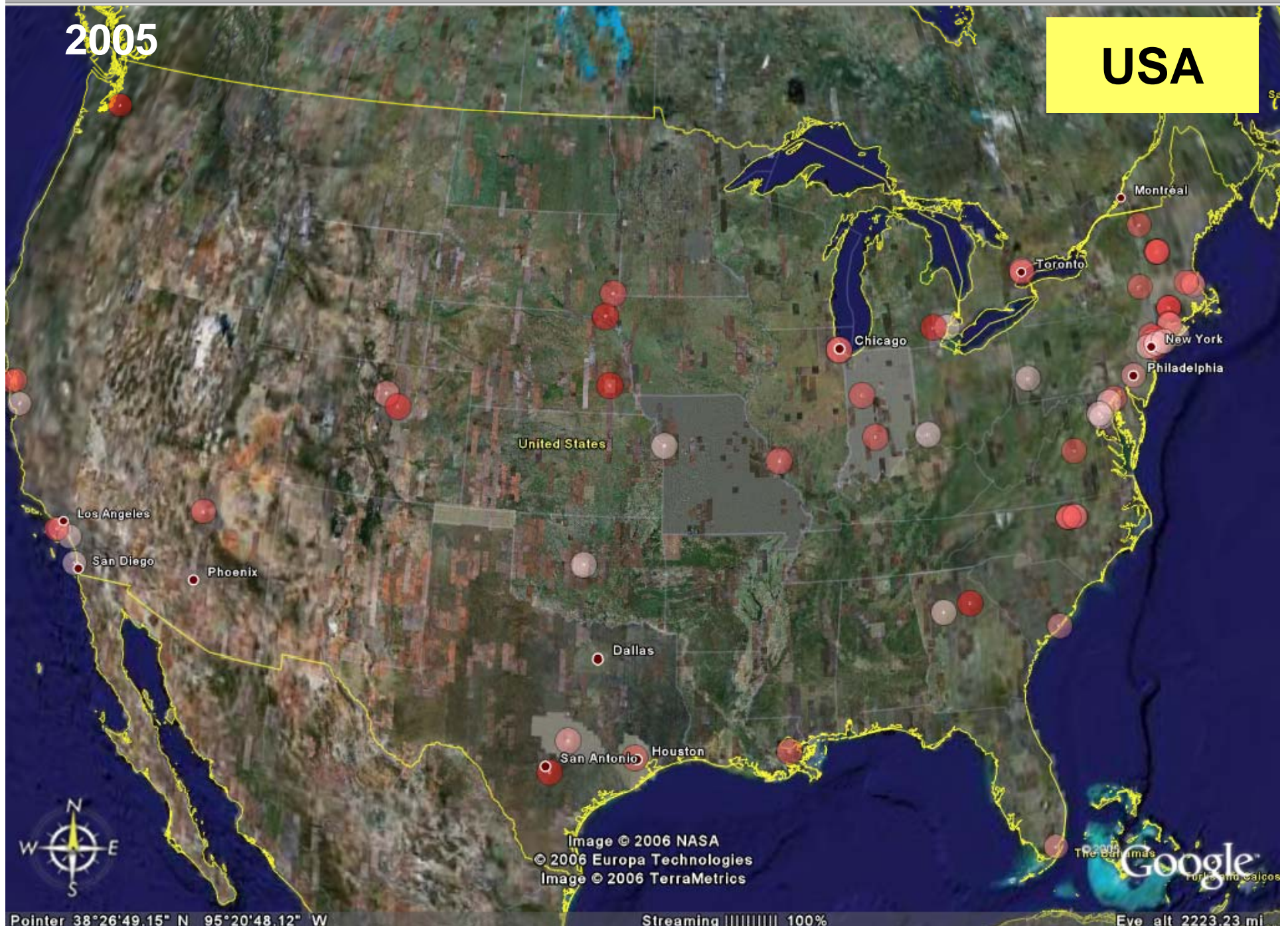
2004

USA



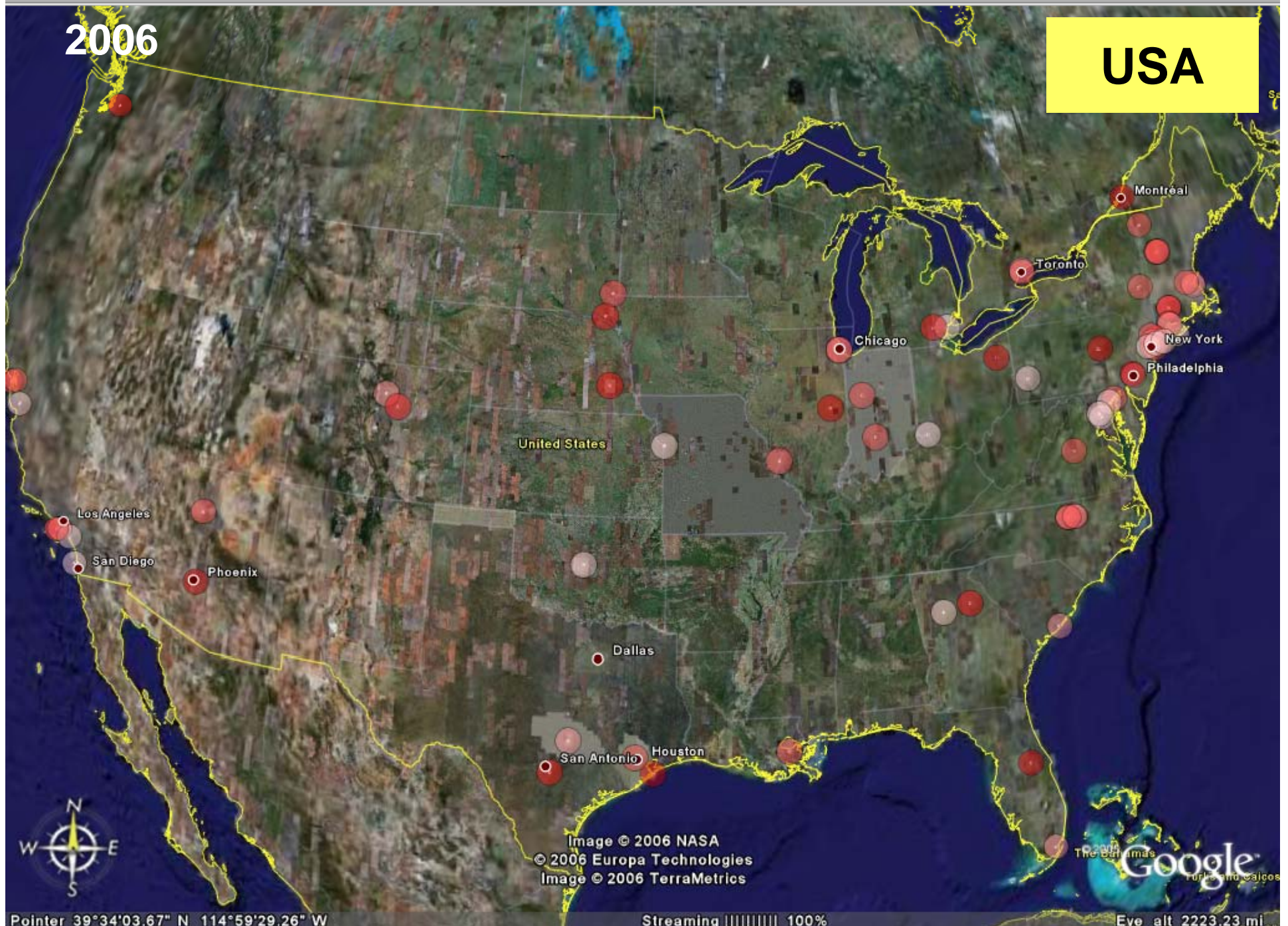
2005

USA



2006

USA



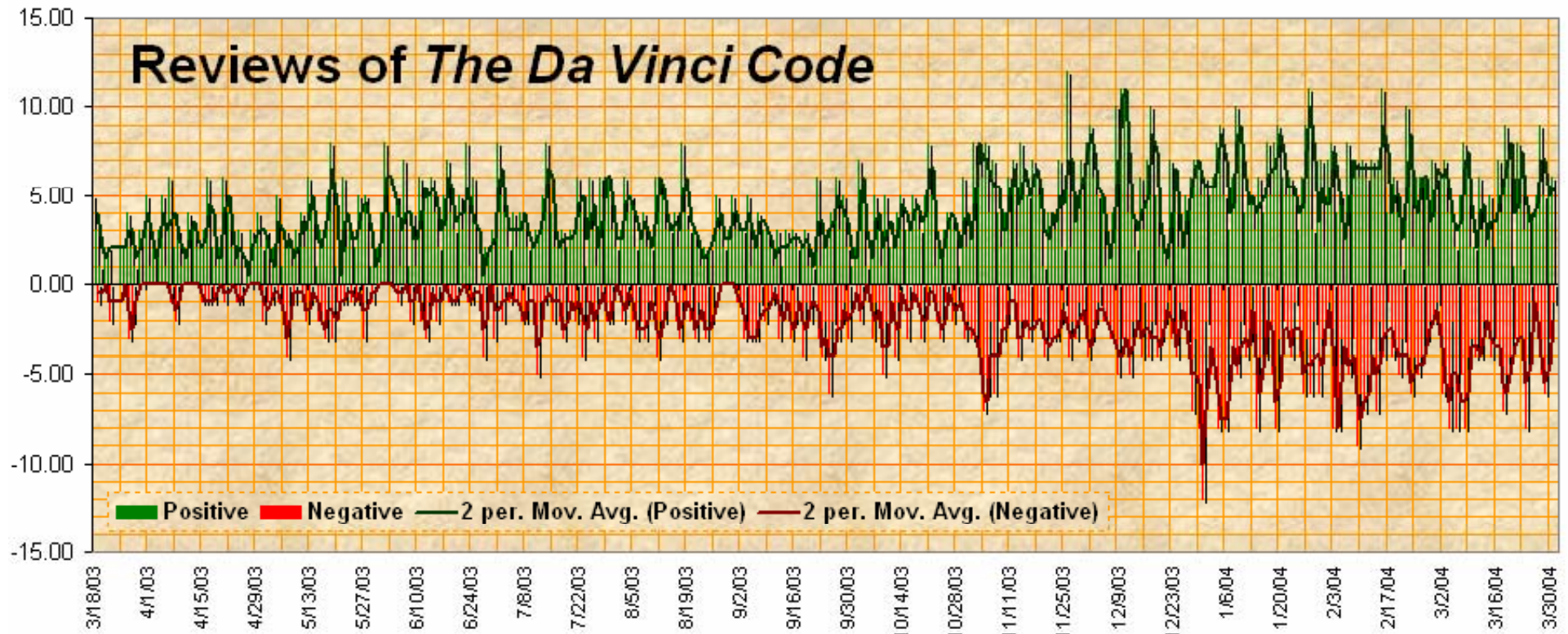
Citers to Galea et al. (2002)



Differentiating Conflicting Views

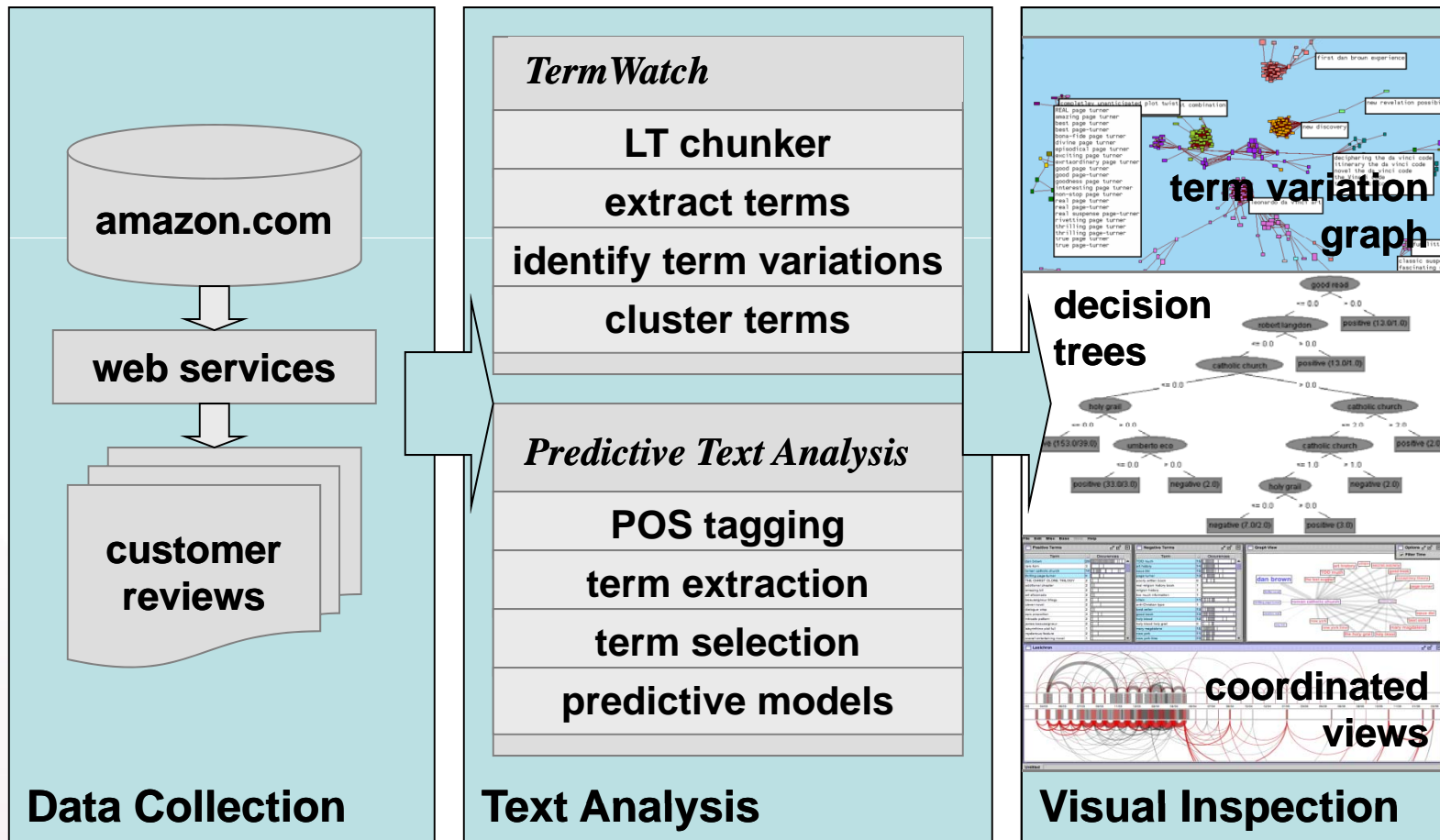
1,738 positive reviews (=4, 5)

918 negative reviews (=1, 2)



The Role of Rare Terms

- ▶ To differentiate conflicting opinions, it is necessary to identify unique perspectives represented by low frequency terms
 - Term variations
 - TermWatch
 - Log likelihood tests
 - Decision Trees



Term Variations

modifier

head word

Operations	Term	Term Variation
Syntactic (adding a modifier)	secret society	<u>ancient</u> secret society
Syntactic (adding a head word)	clever plot	clever plot <u>twist</u>
Syntactic (changing a modifier)	<u>renowned</u> Harvard professor	<u>famous</u> Harvard professor
Syntactic (changing a head word)	secret <u>book</u>	secret <u>agenda</u>
Semantic (synonymous)	<u>ingenious</u> plot	<u>clever</u> plot

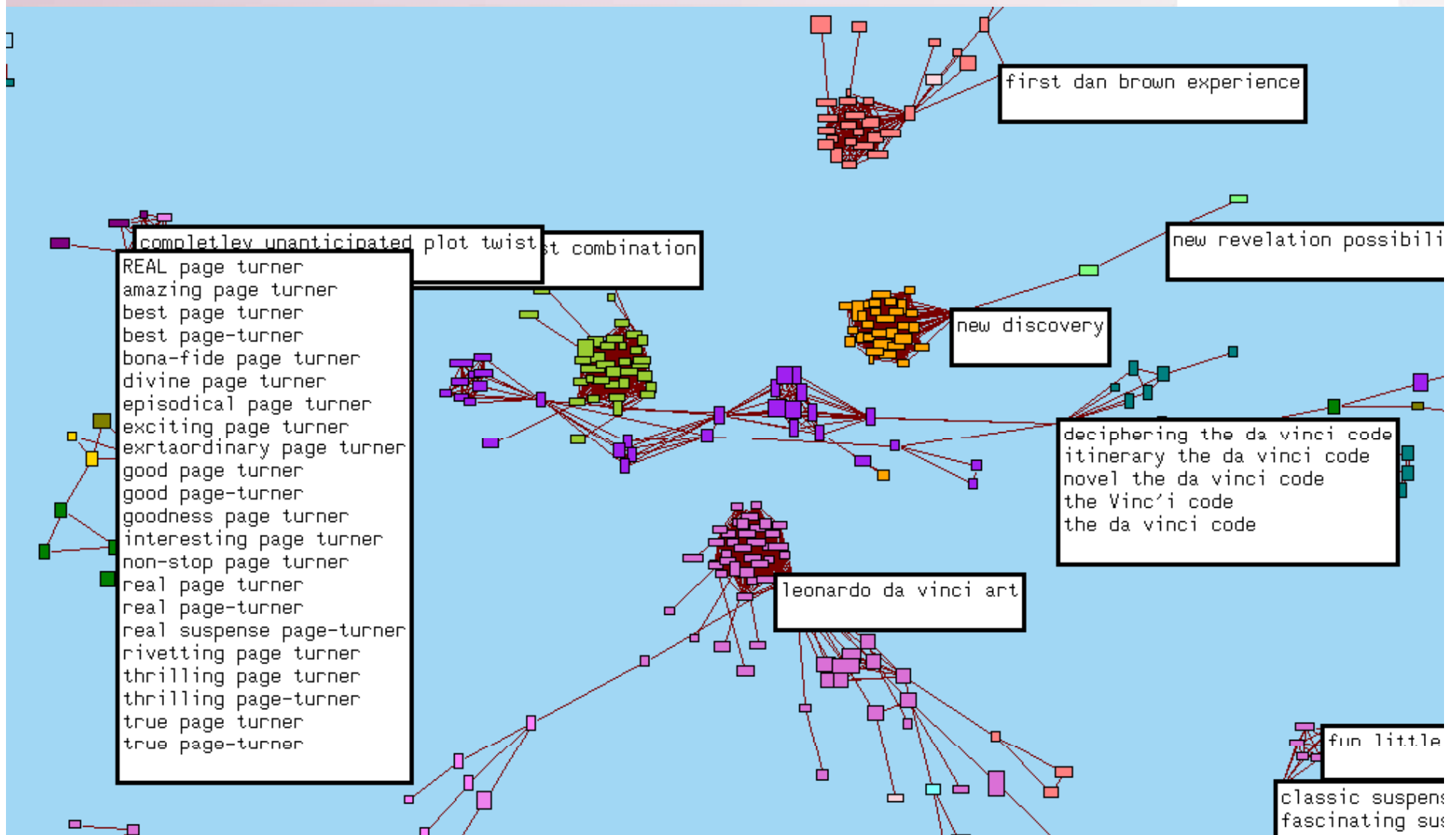
Term Variations

- ▶ Head substitution
age-old religious **secret** → age-old religious **mystery**
- ▶ Expansion
action thriller → **perfect suspense** action thriller
amusing speculation → amusing speculation **collection**
roller coaster → **electrifying** roller coaster **ride**
- ▶ Association
mind-binding puzzle ↔ 5-star entertainment

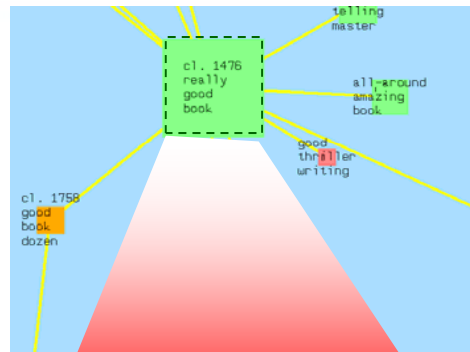


Positive Reviews

[illegible][illegible]



Drilling down the “really good book” cluster



Other
really good book
s by excellent authors are Diamond Age , Snow Crash



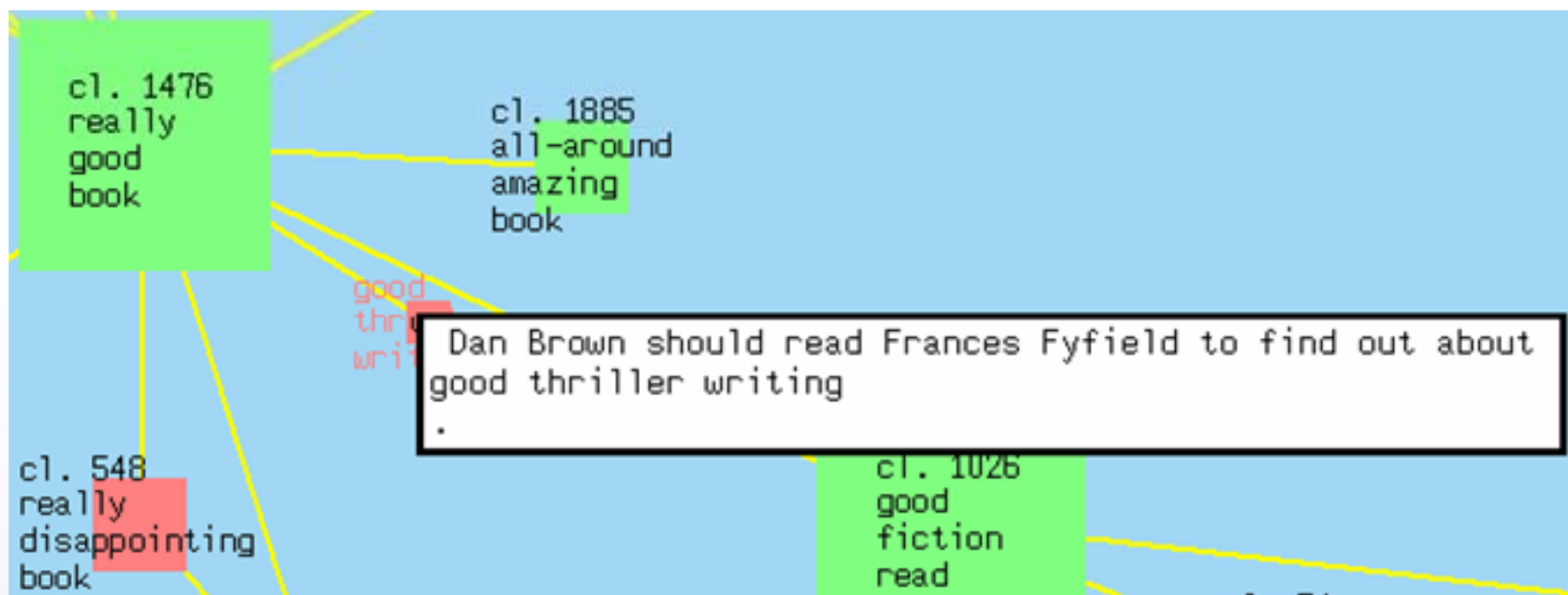
Having said that, I enjoyed this book tremendously - it's a
really good work
of fiction.

I recommend this book, and I also recommend other Dan Brown
really good sci-fi
and cyberpunk like Altered Carbon , Ringworld , Ringworld



If you want to read a
really good thriller
of this type of genre read Paul L.

When “good thriller writing” is used negatively



Document Frequency vs. Log Likelihood ($p=0.5$)



Term	DF	LogLikelihood
book	2456	2.99
story	697	14.25
reader	571	0.23
character	561	59.32
da vinci code	559	10.85
novel	539	0.00
fiction	536	4.89
time	512	21.17
author	504	53.04
plot	499	17.25



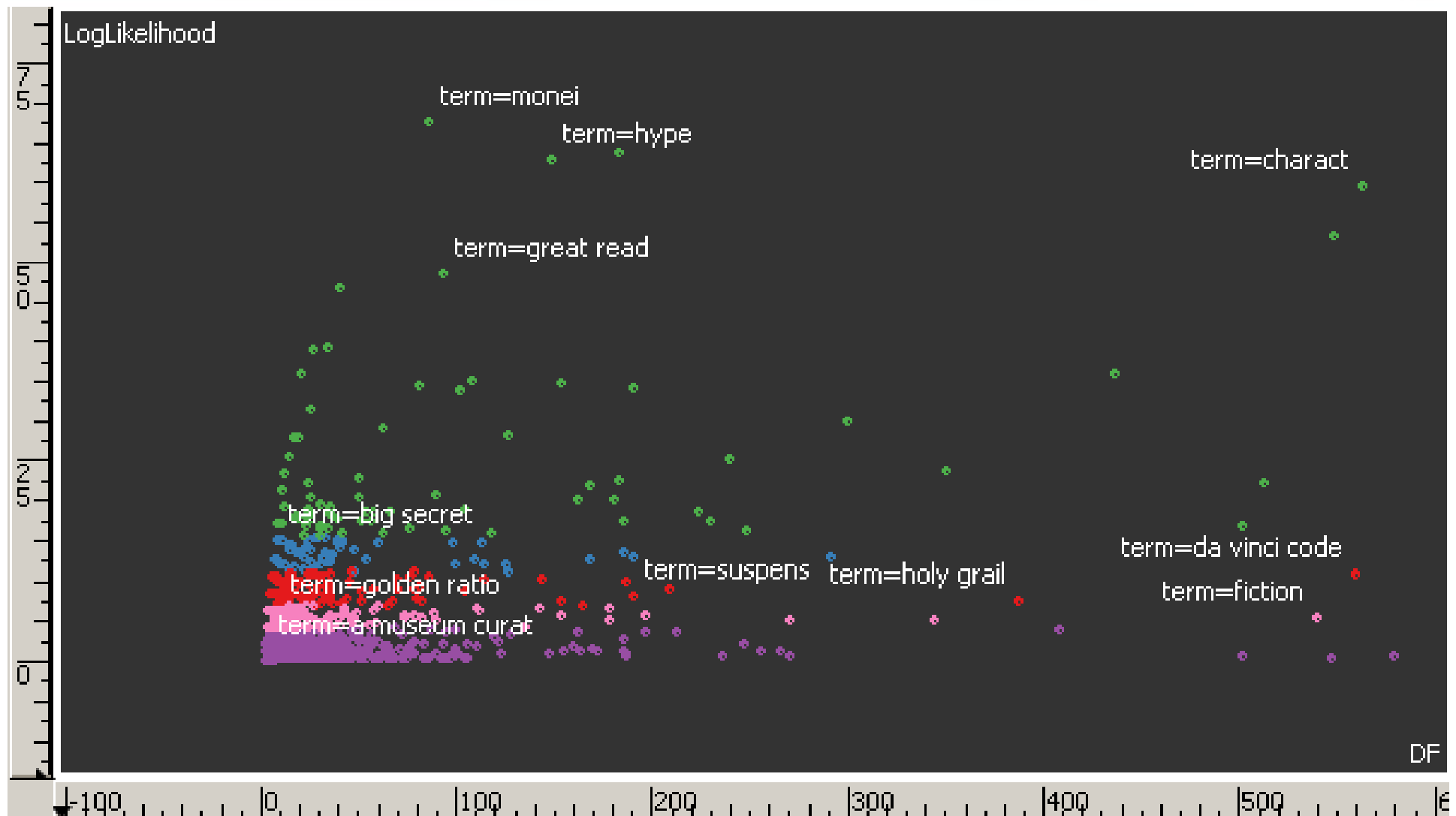
Term	DF	LogLikelihood
money	83	68.27
write	179	66.61
hype	146	61.37
character	561	59.32
author	504	53.04
great read	92	48.40
couldn't	135	48.10
disappoint	39	46.77
waste	33	39.26
don't waste	22	37.63

29,837 Features

Customer Rating	Reviews
1	574
2	344
3	412
4	481
5	1,257
Total	3,068

The Power of Rare Terms

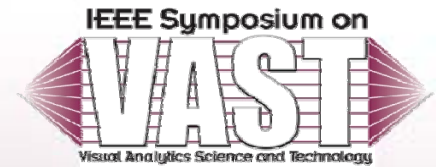
green ($p < 0.001$), blue ($p = 0.001$), red ($p = 0.01$), pink ($p = 0.5$)



Feature Selection: Doc Freq vs. Log Likelihood

Classification accuracy based on 10-fold cross-validation:

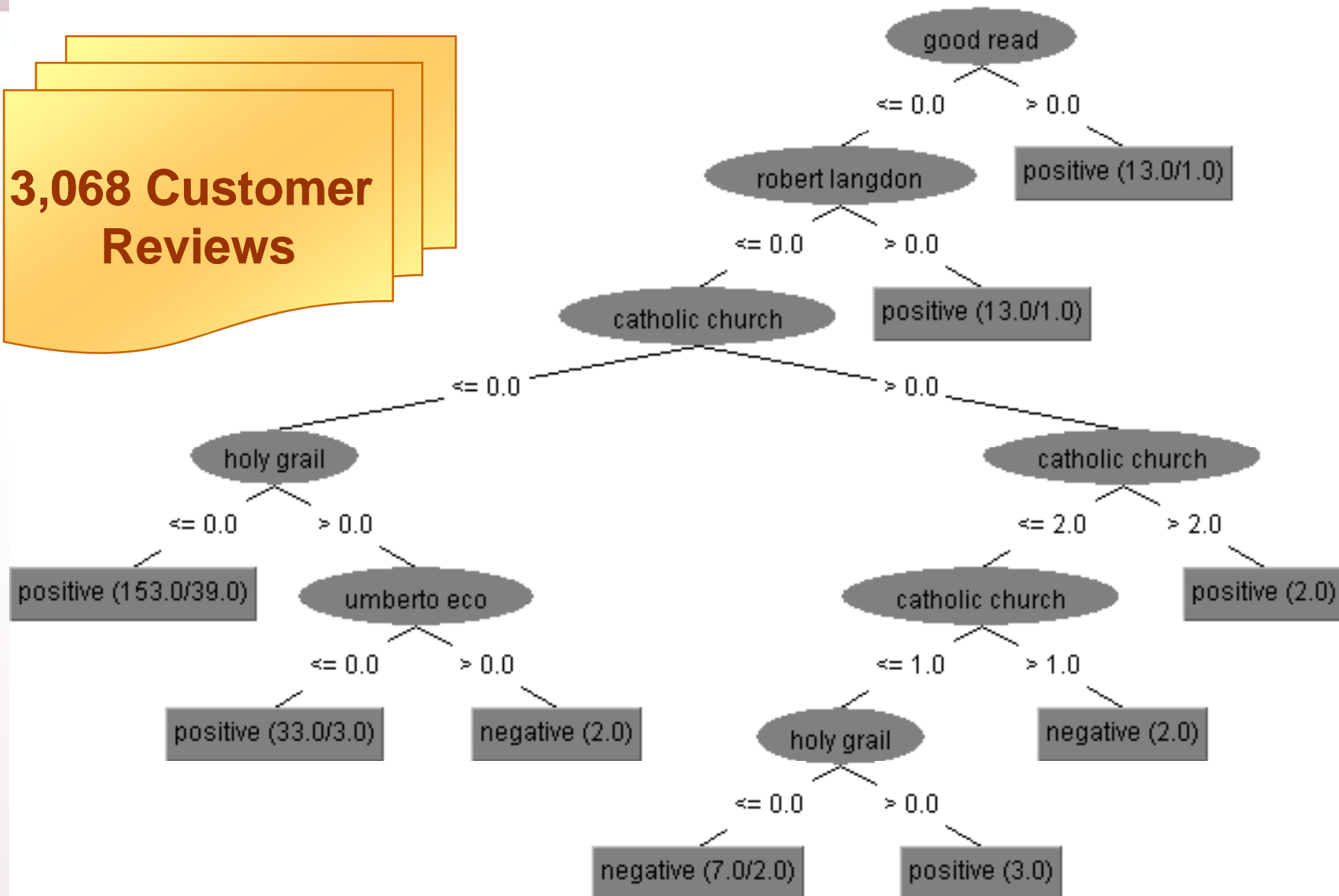
C4.5 << NaiveBayes << SVM



Log Likelihood (p-level)	Selected Terms	C4.5	NaiveBayes	SVM
0.05	1,666	70.26	77.54	84.59
0.01	360	71.67	76.67	83.14
0.001	146	70.01	75.74	81.72
Doc Freq (≥ 2)	6,881	68.89		

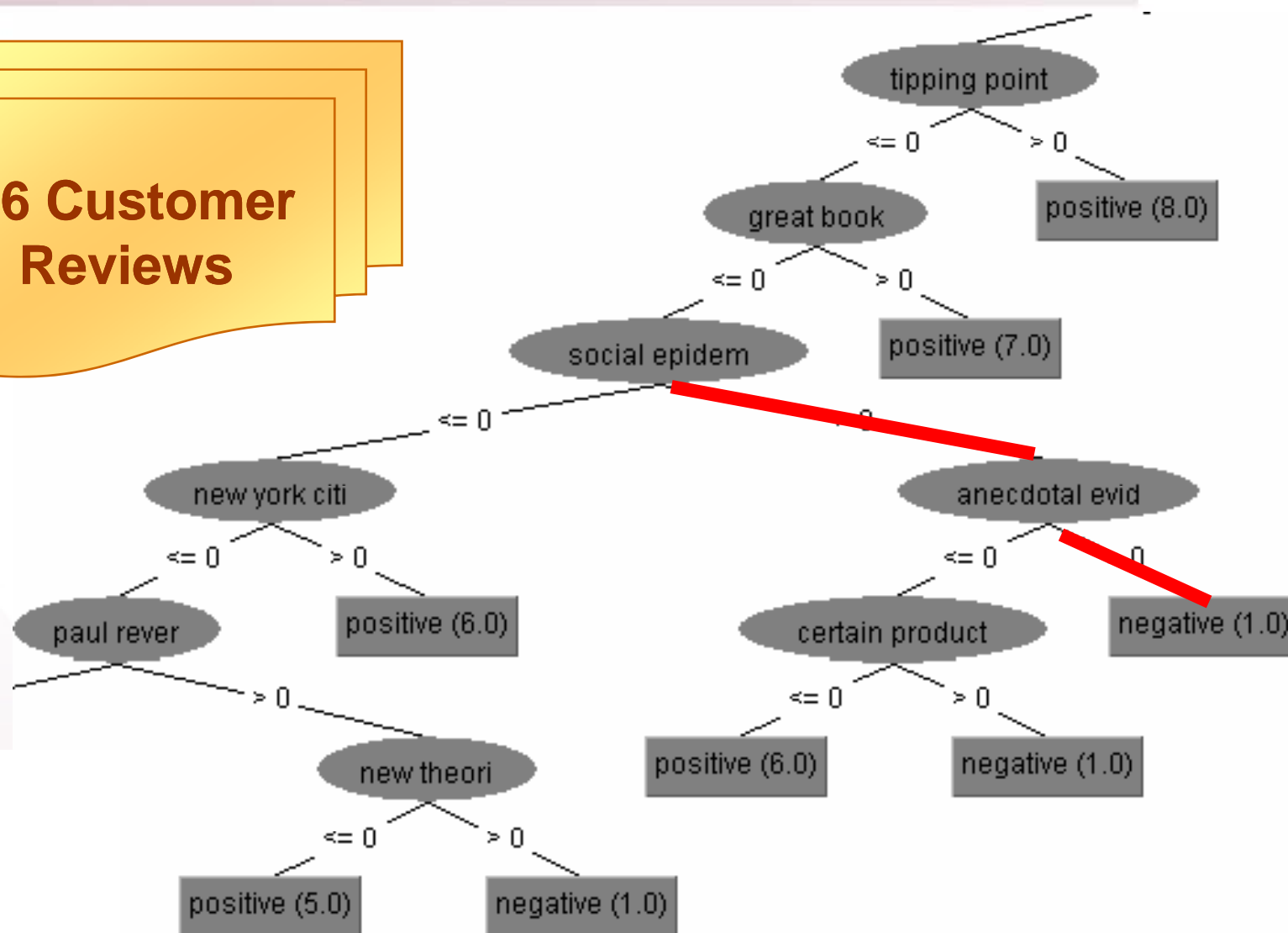
The Da Vinci Code (Amazon Sales Rank: 7)

3,068 Customer
Reviews



The Tipping Point (Amazon Sales Rank: 12)

**596 Customer
Reviews**



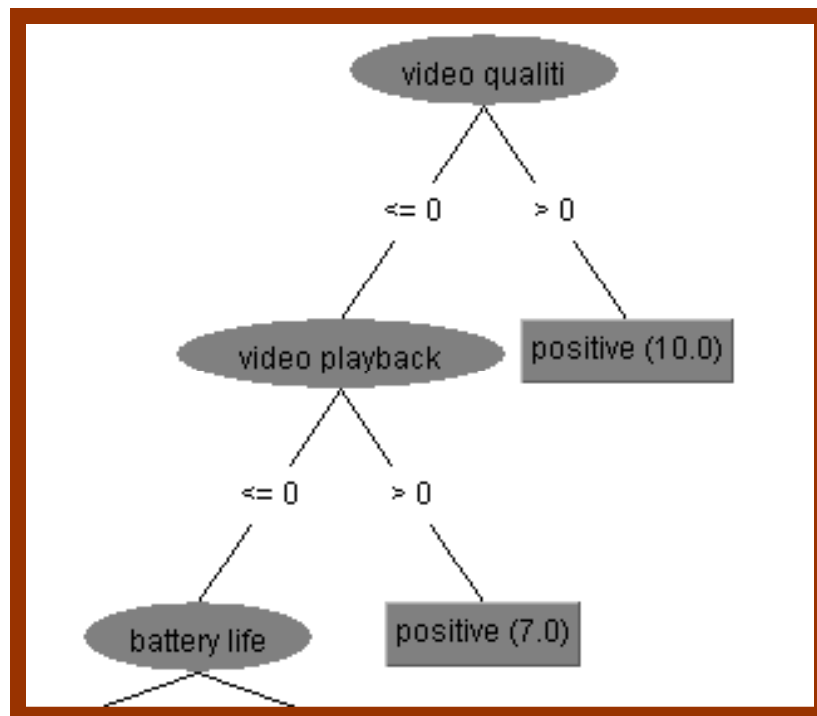
Document Frequency vs. Log Likelihood



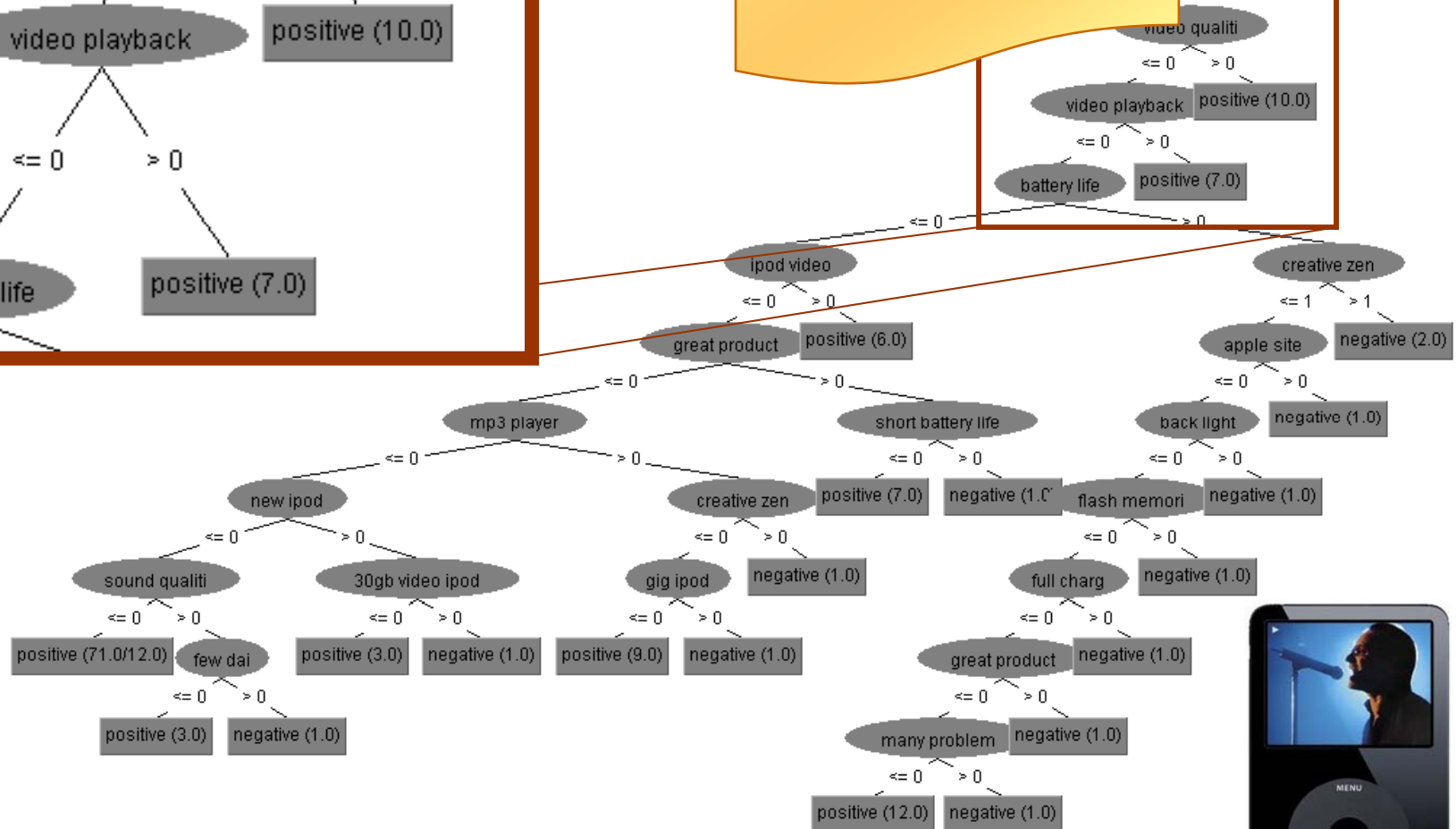
Term	DF	LogLikelihood
ipod	486	82.66
video	275	70.45
music	206	46.21
apple	195	6.41
song	163	33.05
screen	162	29.06
hour	143	9.95
itunes	140	14.70
time	126	4.8
battery life	118	24.34



Term	DF	LogLikelihood
ipod	486	82.66
video	275	70.45
music	206	46.21
love	60	33.30
song	163	33.05
screen	162	29.06
battery	116	26.40
sound	57	24.63
battery life	118	24.34
picture	53	23.12



686 Customer Reviews



Video iPod (accuracy 91.49%)

Amazon.com Sales Rank: **#12** in Electronics



2006



2002

25	new york city
22	posttraumatic stress disorder
11	risk factor
10	world trade center
10	health effect
9	ptsd symptom
9	world trade
9	national comorbidity survey
7	drug user
7	disaster site
7	vietnam veteran
7	adult resident
6	confidence interval
6	odds ratio
6	main outcome measure

2003

38	posttraumatic stress disorder
27	new york city
20	world trade center
14	va patient
13	world trade
12	odds ratio
12	risk factor
9	veterans affair
9	telephone survey
9	health care
9	ptsd symptom
8	vietnam veteran
8	substance use
8	national comorbidity survey
8	panic attack
8	acute stress disorder
8	vehicle accident
7	traumatic stress disorder
6	symptom severity

2004

39	posttraumatic stress disorder
25	world trade center
14	mental health service
14	new york city
14	world trade
14	traumatic stress disorder
12	risk factor
10	mental health problem
10	fine particulate matter
10	mental health impact
9	acute stress disorder
8	language use
8	medication use
7	health service use
6	mental health visit
6	health care system
6	vehicle accident
6	college student
6	minority statu
6	media coverage
6	ptsd symptom

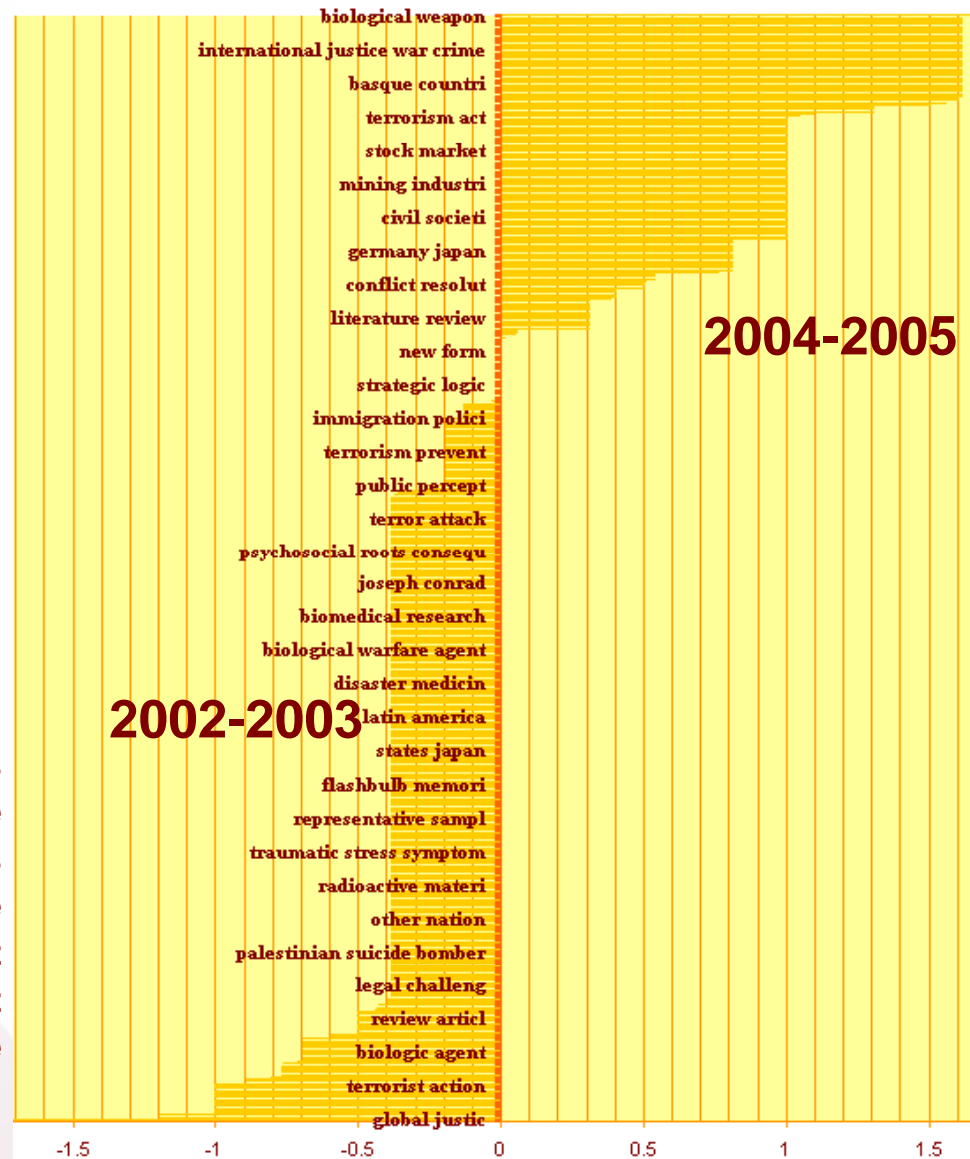
2005

32	world trade center
26	traumatic stress disorder
21	new york city
13	posttraumatic stress disorder
13	probable posttraumatic stress disorder
10	media exposure
10	representative sample
9	drug use
9	ptsd symptom
8	mental health service
8	new york city resident
6	risk factor
6	alcohol consumption

2006

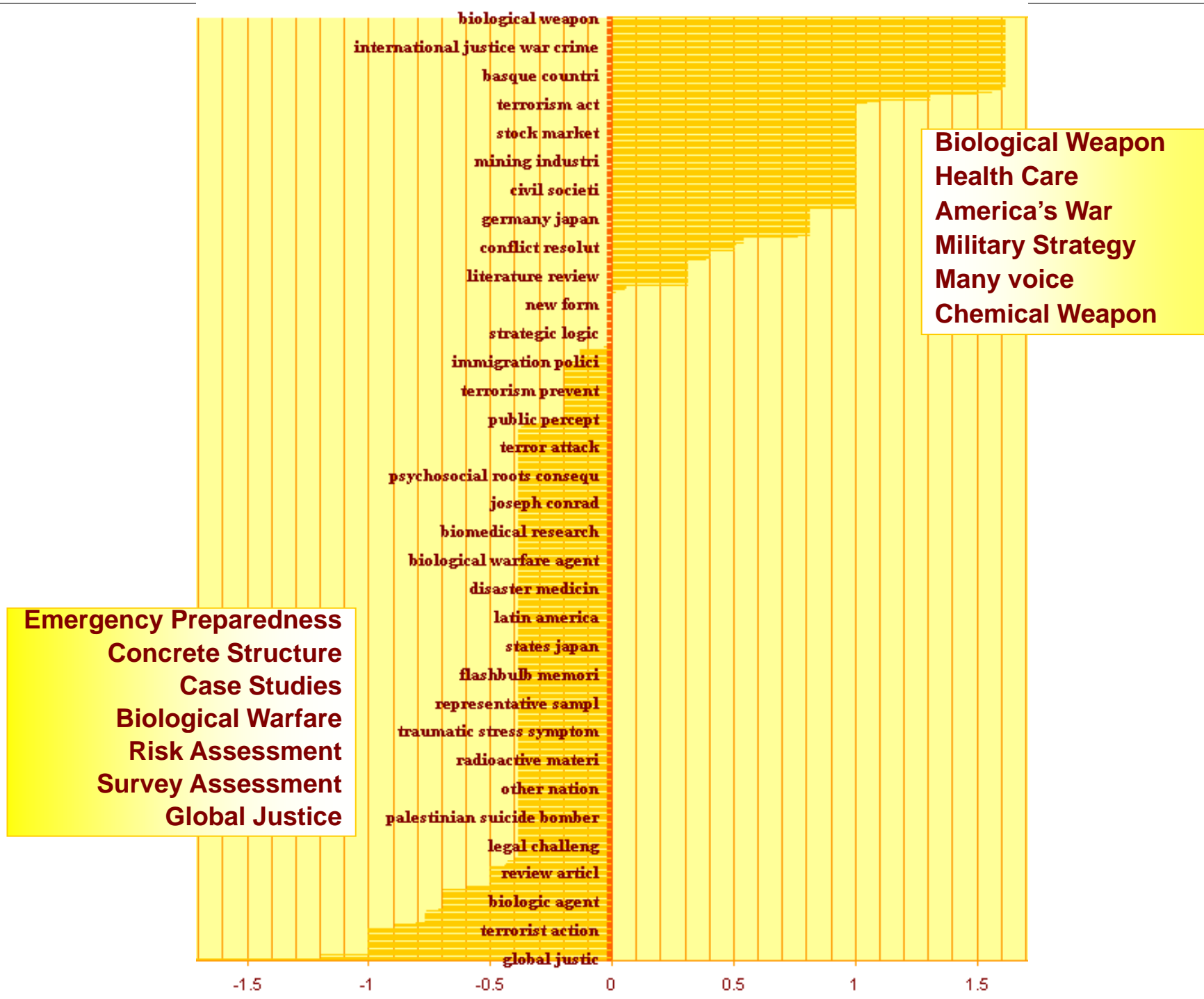
40	posttraumatic stress disorder
38	world trade center
33	new york city
14	project liberty counseling
14	alcohol use
14	traumatic stress disorder
12	project liberty
12	project liberty service
11	mental health service
11	world trade
11	family member
10	project liberty crisis
9	community disaster
9	negative life event
8	service recipient
8	alcohol dependence
8	binge drinking
8	wave panel study
8	mental health statu
8	telephone interview
7	new york city adult
7	project liberty client
7	risk factor
7	center disaster
6	odds ratio
6	health service
6	indirect trauma exposure
6	ptsd symptom

Emergent themes over time on terrorism (2001-2005) Support Vector Classifier



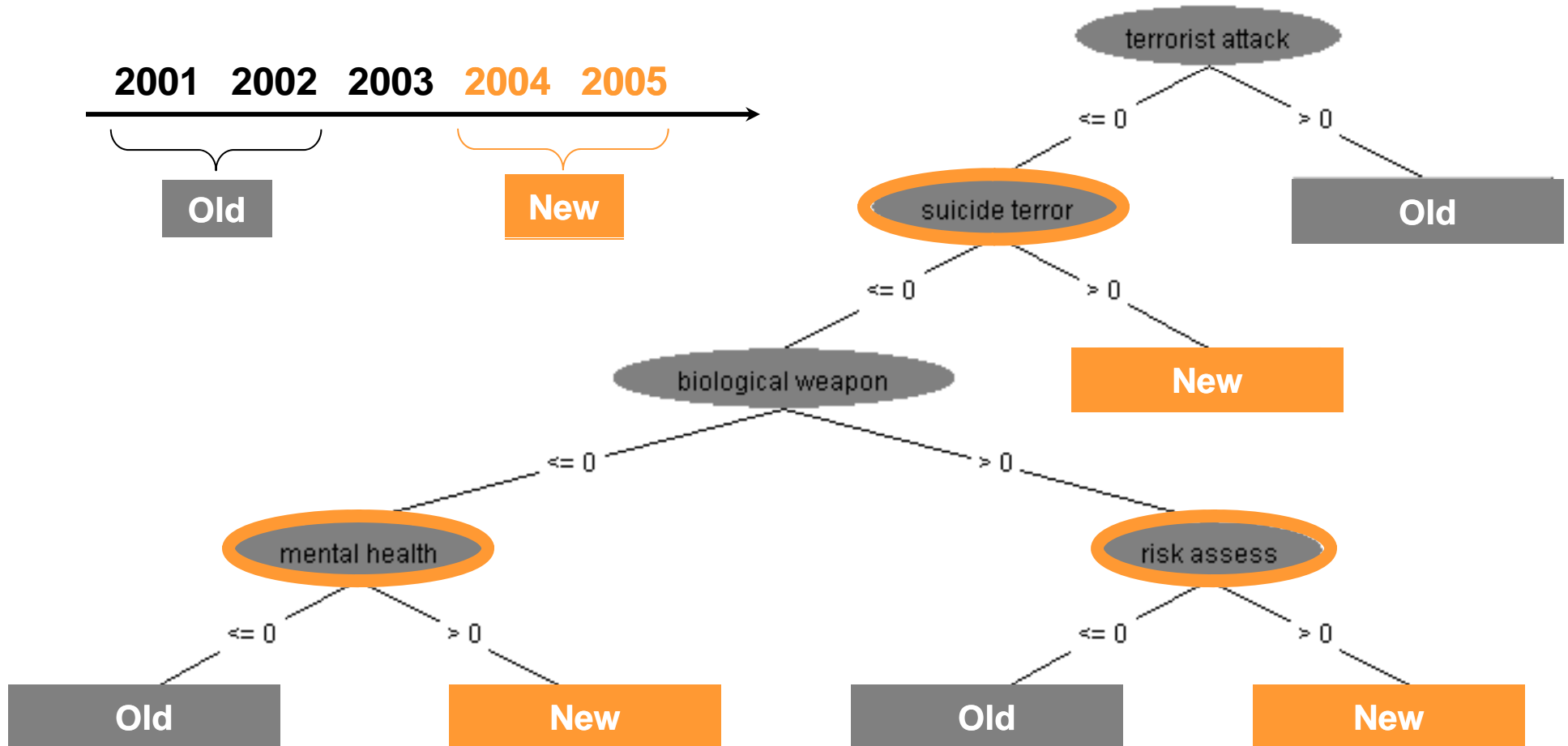
Biological Weapon
Health Care
America's War
Military Strategy
Many voice
Chemical Weapon

Emergency Preparedness
Concrete Structure
Case Studies
Biological Warfare
Risk Assessment
Survey Assessment
Global Justice

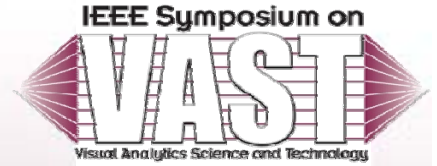


Emergent themes over time on terrorism (2001-2005)

A decision tree derived from 3,944 abstracts

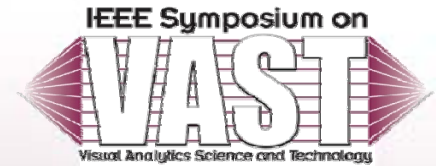


Summary



- ▶ Visual Analysis of Conflicting Views
 - Feature selection and handling low-frequency but critical features and interrelationship
 - Term variation patterns
 - Statistical tests of associations
 - Facilitate sense making tasks
 - What information is the most effective to understand different opinions?
 - How could emergent thematic patterns be visualized?
- ▶ Future Work
 - Integration with more interactivity
 - Comprehensive evaluations of alternative approaches
 - Investigations of the applicability in a wider range of visual analytic tasks

Acknowledgements



- ▶ **National Visualization and Analytics Center (NVAC)**
 - **Northeast Visualization and Analytics Center (NEVAC)**

- ▶ **NSF IIS Award #0612129**
 - **SEI: Coordinated Visualization and Analysis of Sky Survey Data and Astronomical Literature**