SAND2017-12619PE

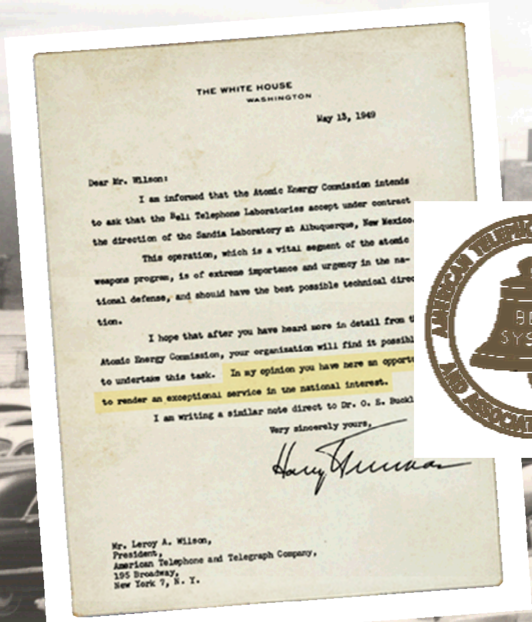
New Horizons in Geospatial Pattern Analysis

Kristina Czuchlewski, Ph.D.

November 14, 2017

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Sandia: An FFRDC for nearly seven decades


THE WHITE HOUSE
WASHINGTON

May 13, 1949

Dear Mr. Wilson:

I am informed that the Atomic Energy Commission intends to ask that the Bell Telephone Laboratories accept under contract the direction of the Sandia Laboratory at Albuquerque, New Mexico.

This operation, which is a vital segment of the atomic weapons program, is of extreme importance and urgency in the na- tional defense, and should have the best possible technical direc tion.

I hope that after you have heard more in detail from t Atomic Energy Commission, your organization will find it possibl to undertake this task. In my opinion you have here an opport to render an exceptional service in the national interest.

I am writing a similar note direct to Dr. O. E. Buckl

Very sincerely yours,

Harry Truman

Mr. Leroy A. Wilson,
President,
American Telephone and Telegraph Company,
195 Broadway,
New York 7, N. Y.

JULY 1945
Los Alamos creates Z Division at Sandia Base

NOVEMBER 1, 1949
Sandia Laboratory is established and managed by AT&T

MARCH 8, 1956
Sandia's California site is established

JULY 26, 1993
Martin Marietta wins first Sandia Corporation contract competition and in 1995 merges with Lockheed to become Lockheed Martin

DECEMBER 16, 2016
The NNSA awards the Sandia prime contract to National Technology and Engineering Solutions of Sandia (NTESS), a subsidiary of Honeywell International

MAY 1, 2017
The new prime contract goes into effect

Sites of Sandia

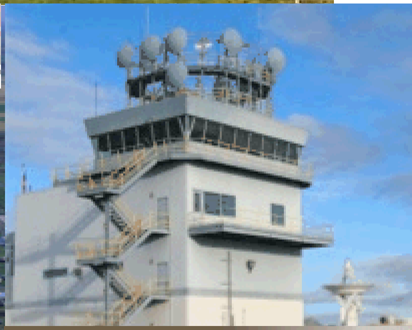
Albuquerque, New Mexico

Livermore, California

Kauai, Hawaii

Waste Isolation Pilot Plant, Carlsbad, New Mexico

Pantex Plant, Amarillo, Texas



Tonopah, Nevada

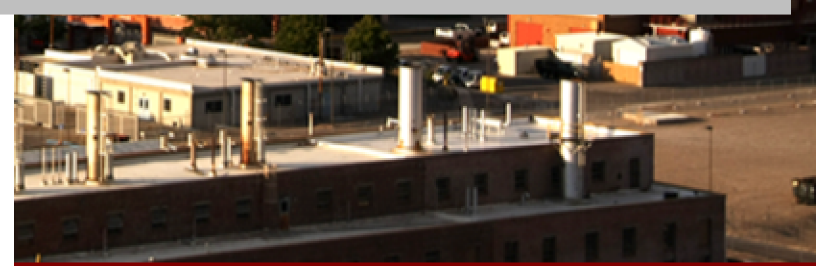
History of Sandia Laboratories

Sandia Corporation

- AT&T: 1949–1993
- Martin Marietta: 1993–1995
- Lockheed Martin: 1995–April 2017
- NTESS May 1, 2017 - present
- Government owned, contractor operated

Federally Funded Research and Development Center (FFRDC) Unique nonprofit entities sponsored and funded by the U.S. government to meet some special long-term research or development need

Sandia is 1 of 39 recognized FFRDCs



ORGANIZATION AND KEY CAPABILITIES

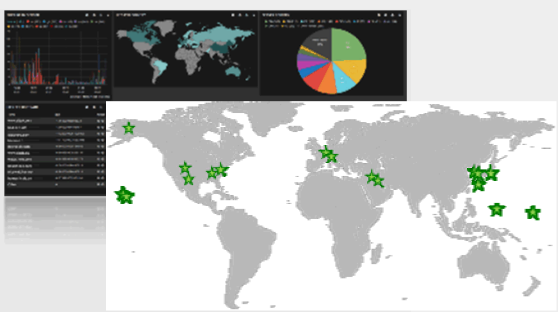
5000 MIKE BURNS
ASSOCIATE LAB DIRECTOR

5800 JIM HUDGENS
CENTER DIRECTOR

5850 KEVIN DIXON
SENIOR MANAGER

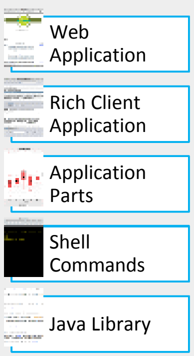
DANIEL GARCIA, MANAGER 5851

- Big-Data Analytics
- High-Speed Data
- Cyber Threat Discovery



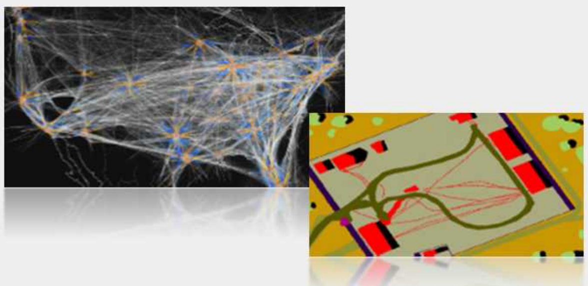
CURTIS JOHNSON, MANAGER 5852

- Text Understanding
- Social-Network Analysis
- Graph Analytics



KRISTINA CZUCHLEWSKI, MANAGER 5853

- Geospatial-Temporal Analytics
- Pattern-of-Life Analytics
- Image Understanding



The changing data landscape presents challenges… and opportunities

- Many important phenomena are below the limit of human perception—in nearly every domain.

  Sensor Science/ Image Processing

- The phenomena are scaling much faster than the ability to observe and process them.

  Computer Science

- Key connections between observables cannot be made.

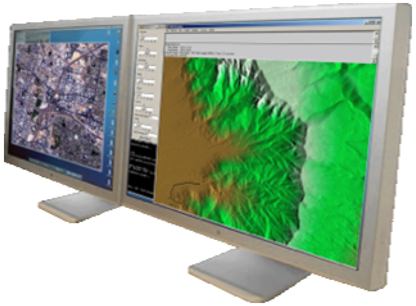
  Information Science

- Overwhelmed operators struggle to use data for predictive and forensic purposes—especially in real time.

  Human Factors

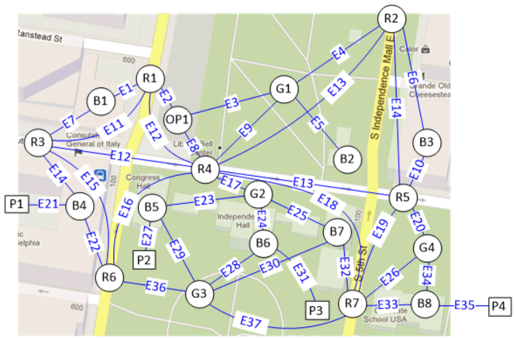
- Data transmission and storage limitations confound the problem.

  Computer Science



These problems demand multi-disciplinary scientific inquiry

Pattern Analytics R&D



Sandia rethinks traditional GIS and geospatial search.

New. Populate graphs with image-derived, geo- and time- tagged features from multiple data sources.

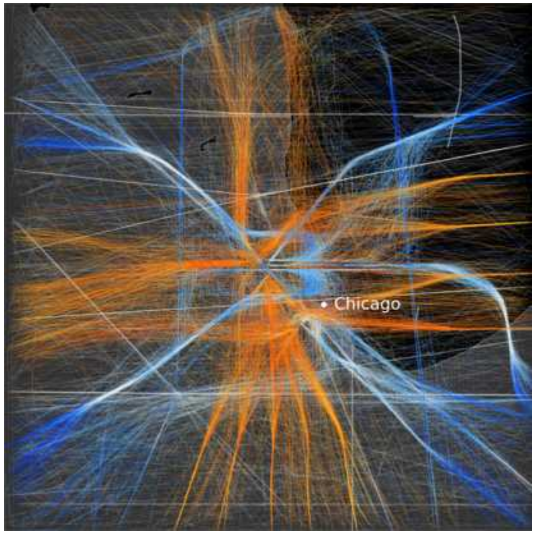
New. Execute graph-algorithm powered search for threat signatures that consist of durable features and activities.

New. Uncertainty and quality of search considerations with real, noisy data.

PANTHER rethinks patterns in motion.

New. Geometric and temporal trajectory analyses represent and compare tracks efficiently and lightning-fast.

New. Discovery of geospatial-temporal relationships and comparison of more than two trajectories.



Three Vignettes

- Rethinking Patterns in Motion
- Rethinking Geographical Information Systems
- Rethinking Search

RETHINKING PATTERNS IN MOTION

Trajectory Analysis

Idea: Analyze trajectories represented by paths of objects in space/time

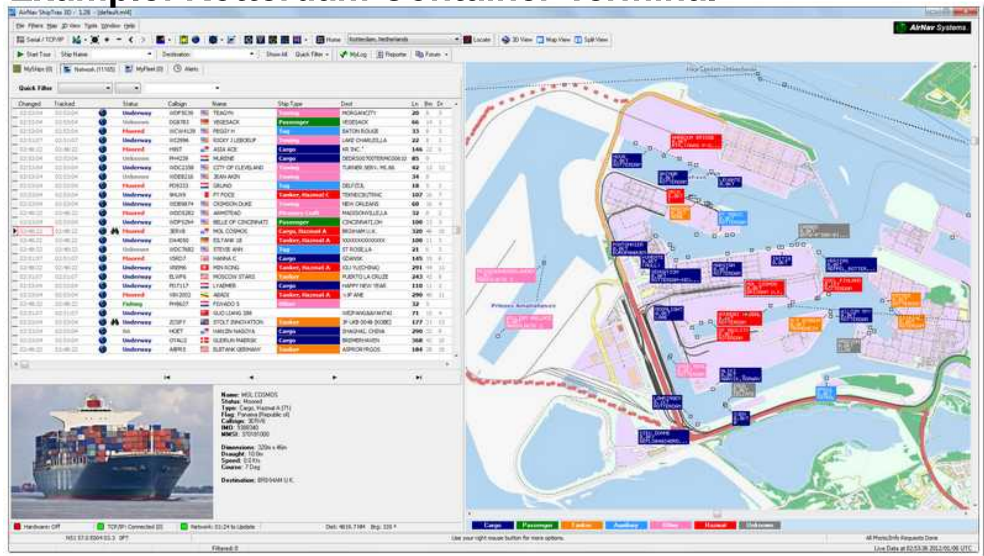
Motivating Questions:

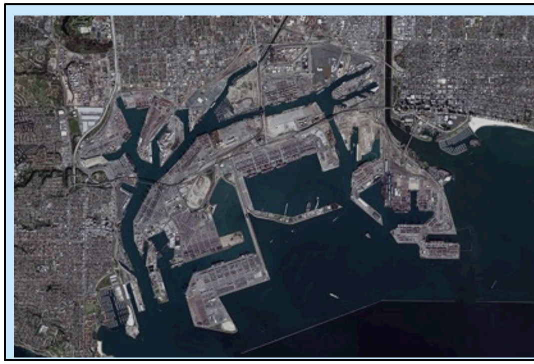
- Where are all of the moving objects?
- What are the moving objects?
- What are they doing now?
- Why are they doing these things?
- What might they do?
- When will we be able to tell?
- Are these things they are doing unusual, or have they done them before?

How it’s done today: Layers of visual data, *weak analytic support*

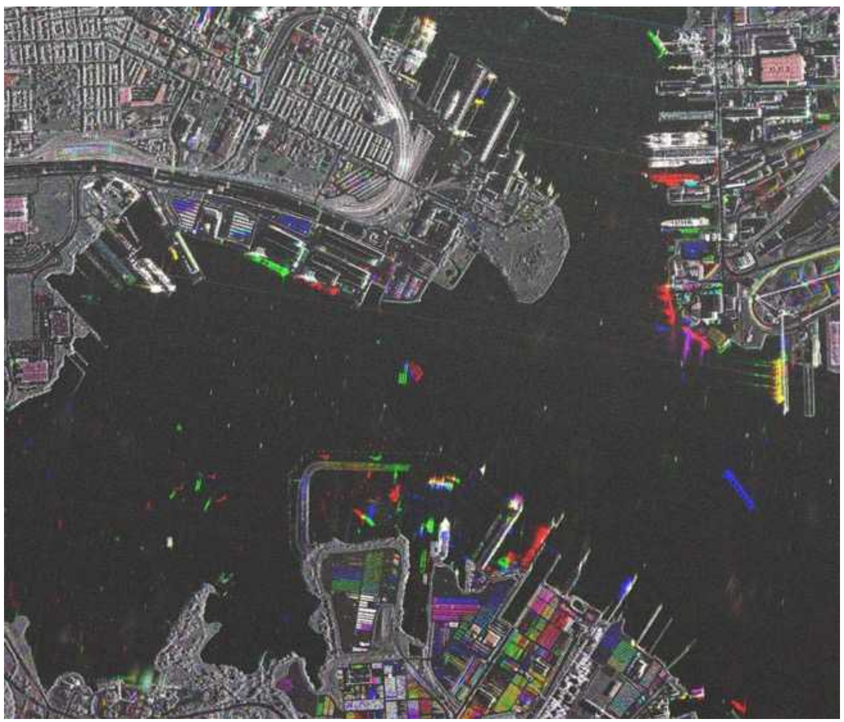
Example: Virtual Port, Port of Long Beach. System for monitoring infrastructure, status of sensitive materials, and exposure of people to hazards.

Example: Rotterdam Container Terminal





Satellite image of Long Beach, Calf. Collected using GE1 on Jan 15, 2013. Photo Credit: DigitalGlobe http://trajectorymagazine.com/defense-intelligence/item/1752-visualizing-vessels.html

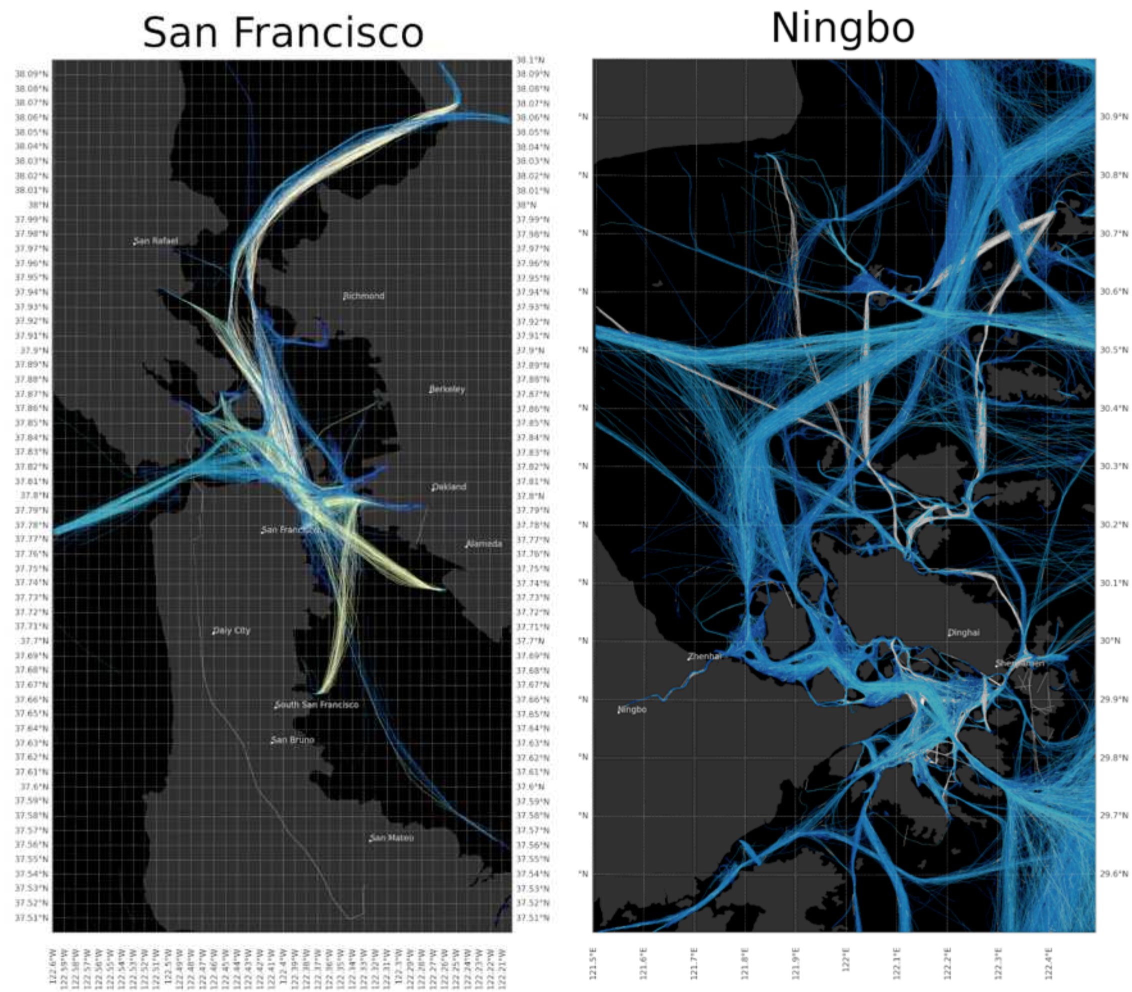


TerraSAR-X based Amplitude Change Detection Map of Baltimore Harbor, Md.

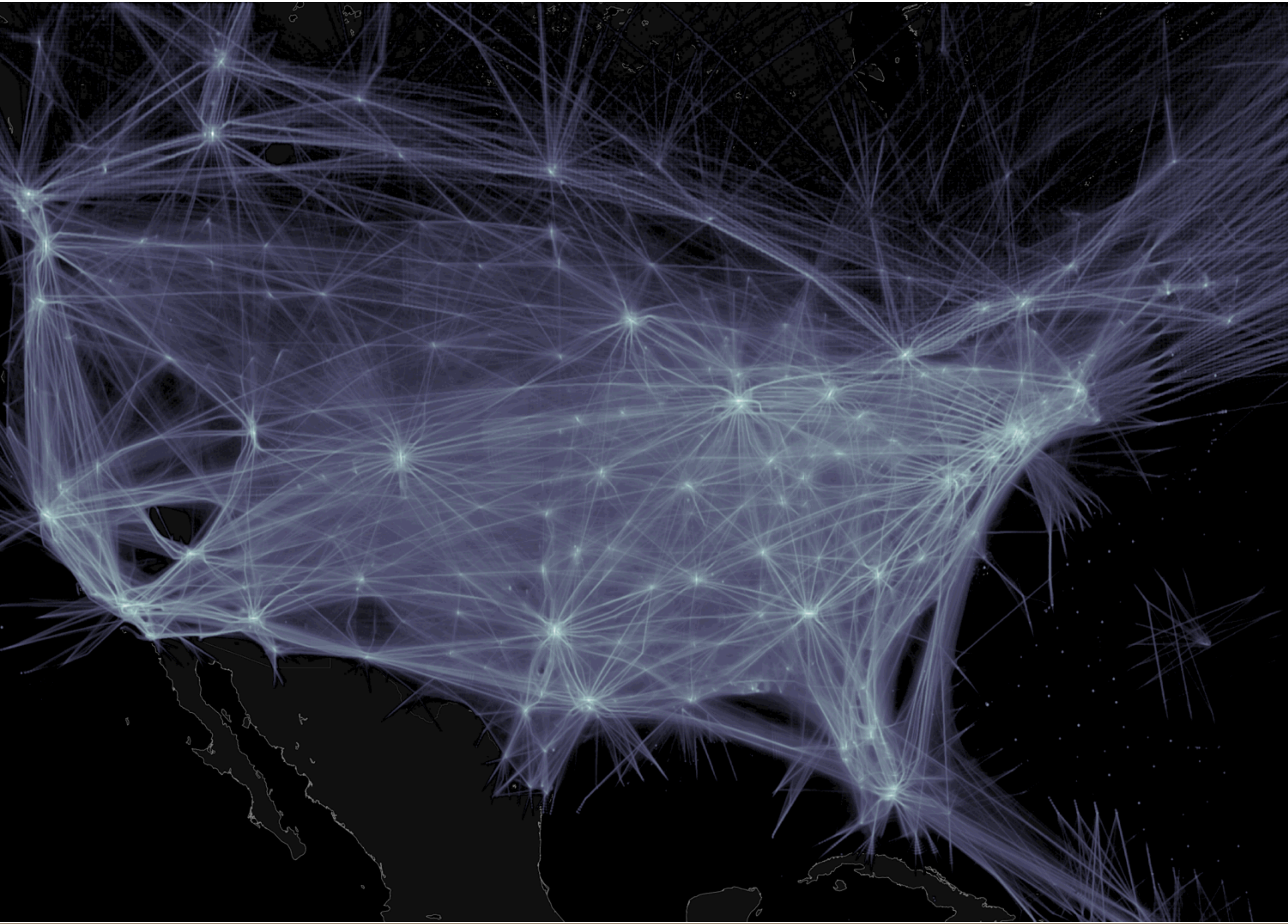
Photo Credit: Airbus DS/Infoterra http://trajectorymagazine.com/defense-intelligence/item/1752-visualizing-vessels.html

AIS Maritime Traffic

- Automatic Identification System
  - Modern collision avoidance
  - Required on int'l ships >300T, all passenger ships
  - Ship-to-ship, ship-to-shore
    - We have traffic within ~80km of land
  - Broadcast interval varies by size, speed
- Broad diversity of traffic and behavior
  - Lumbering herds of supertankers
  - Water ballet in and around ports



From: UUR SAND 2014 18850 PE

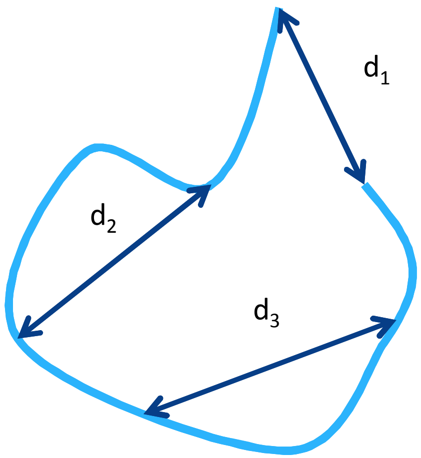


Geometric Descriptors

Key Idea:

- A trajectory can be characterized by a *feature vector*
    - $V = (f_1, f_2, \dots, f_n)$
- Examples: total curvature, path length, length between end-points, etc.

Impact: Can apply generic techniques to analyze the vector space of geometric features



Note:

- Intra-trajectory distances can be used as features to capture trajectory shapes
- This approach supports rigid-transform invariant shape search
- We can efficiently find nearby trajectories using R-Trees
- Time (flight length, start/stop times, day of week) can also be used as a feature allowing us to find patterns in time

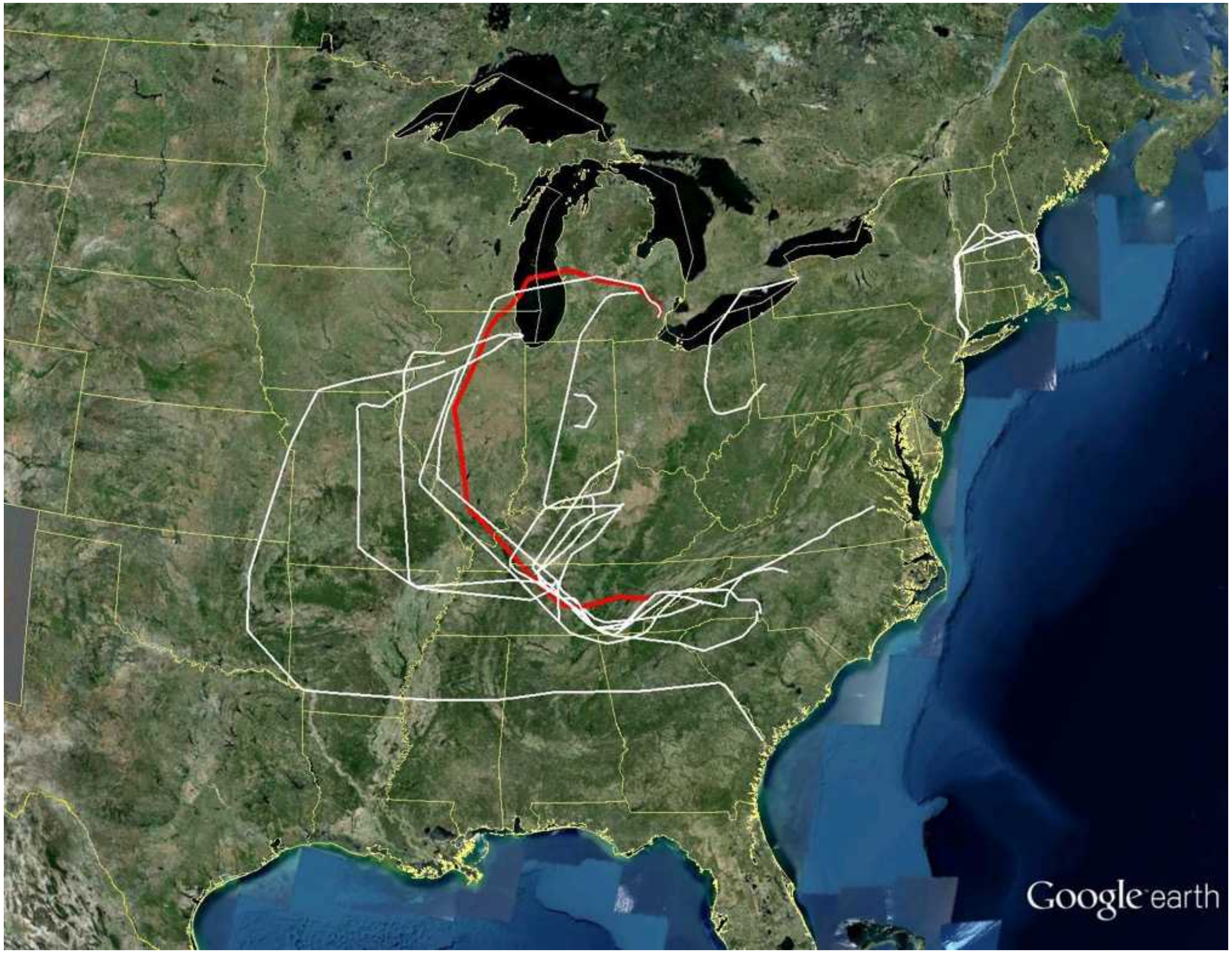
Trajectory Queries: Flight Example

Now, finding flights similar to a given flight

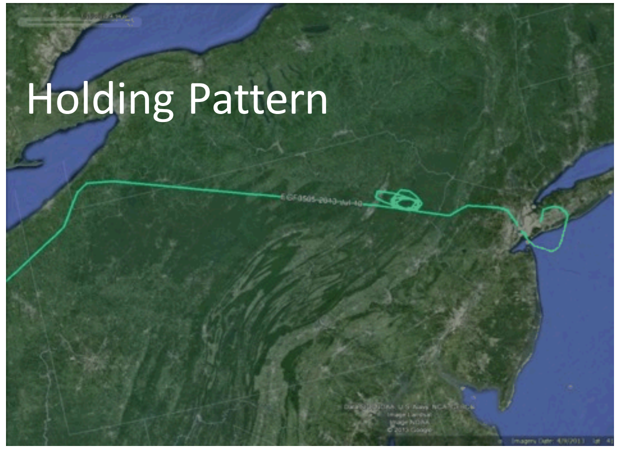
- $O(\log_M n)$ search

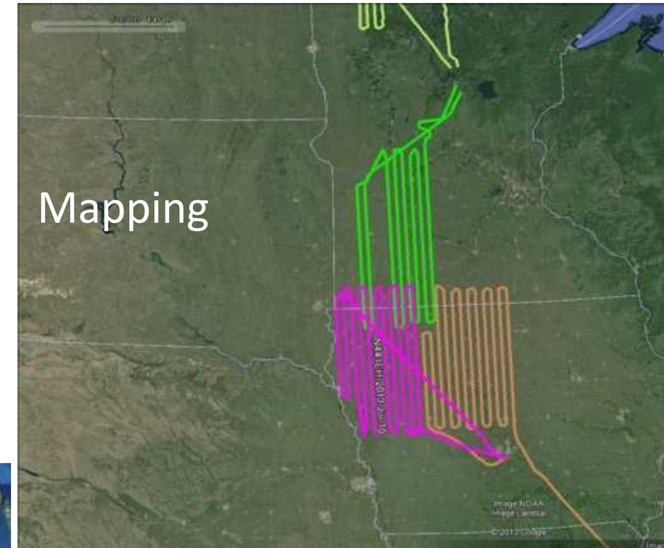
FAA's ASDI Data

- ~5M points/day, ~1GB/day

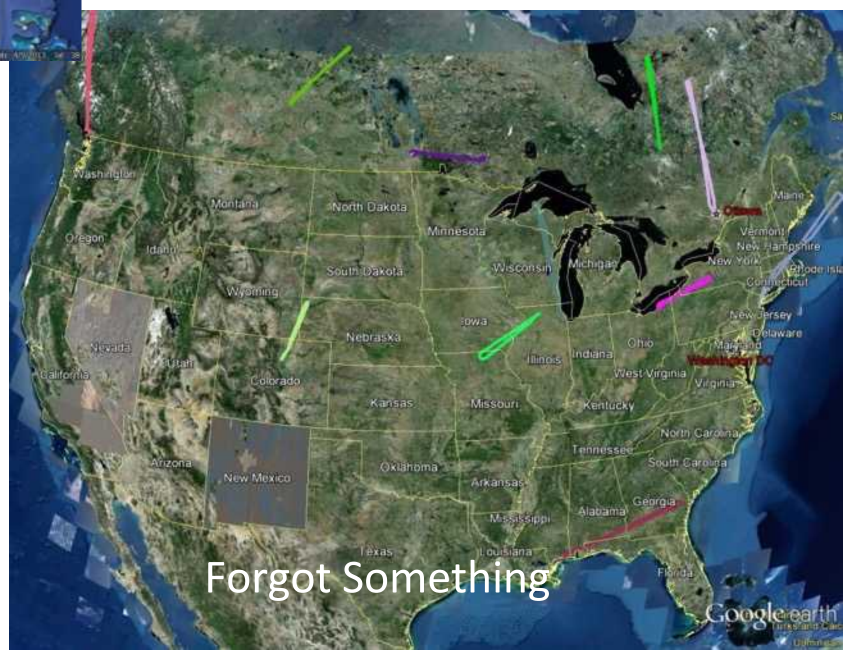


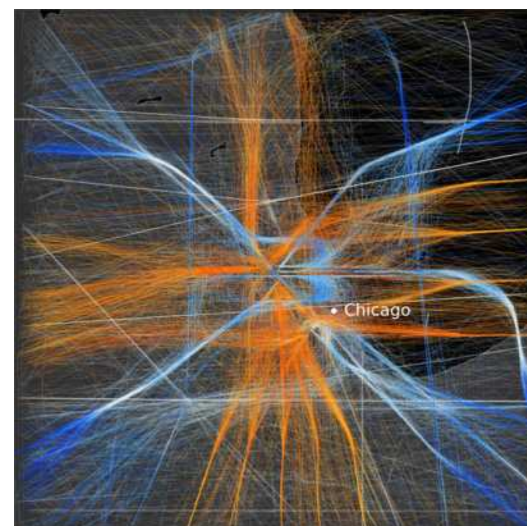
Discovery of Flight Patterns





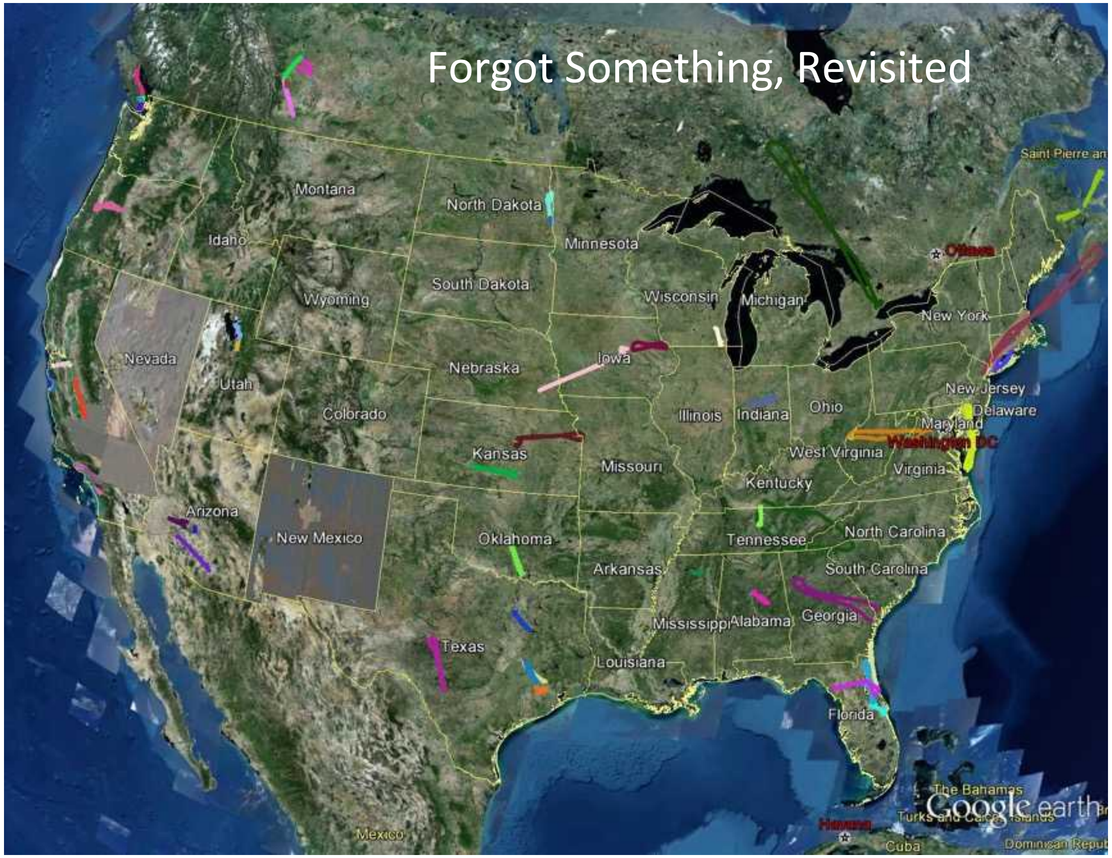
Collections of geometric descriptions can describe a trajectory. Extensions: impact of time.





Big Feature Space Advantage

- Clustering, leading to *unsupervised learning* techniques
- Previous examples showed searching the space for a specific pattern or a specific volume of the feature space
- But, with clustering, the computer can group the different patterns in the feature space without knowing *a priori* what they are.
- Perhaps most importantly, many clustering algorithms specifically identify outliers in the feature space that correspond to odd behaviors



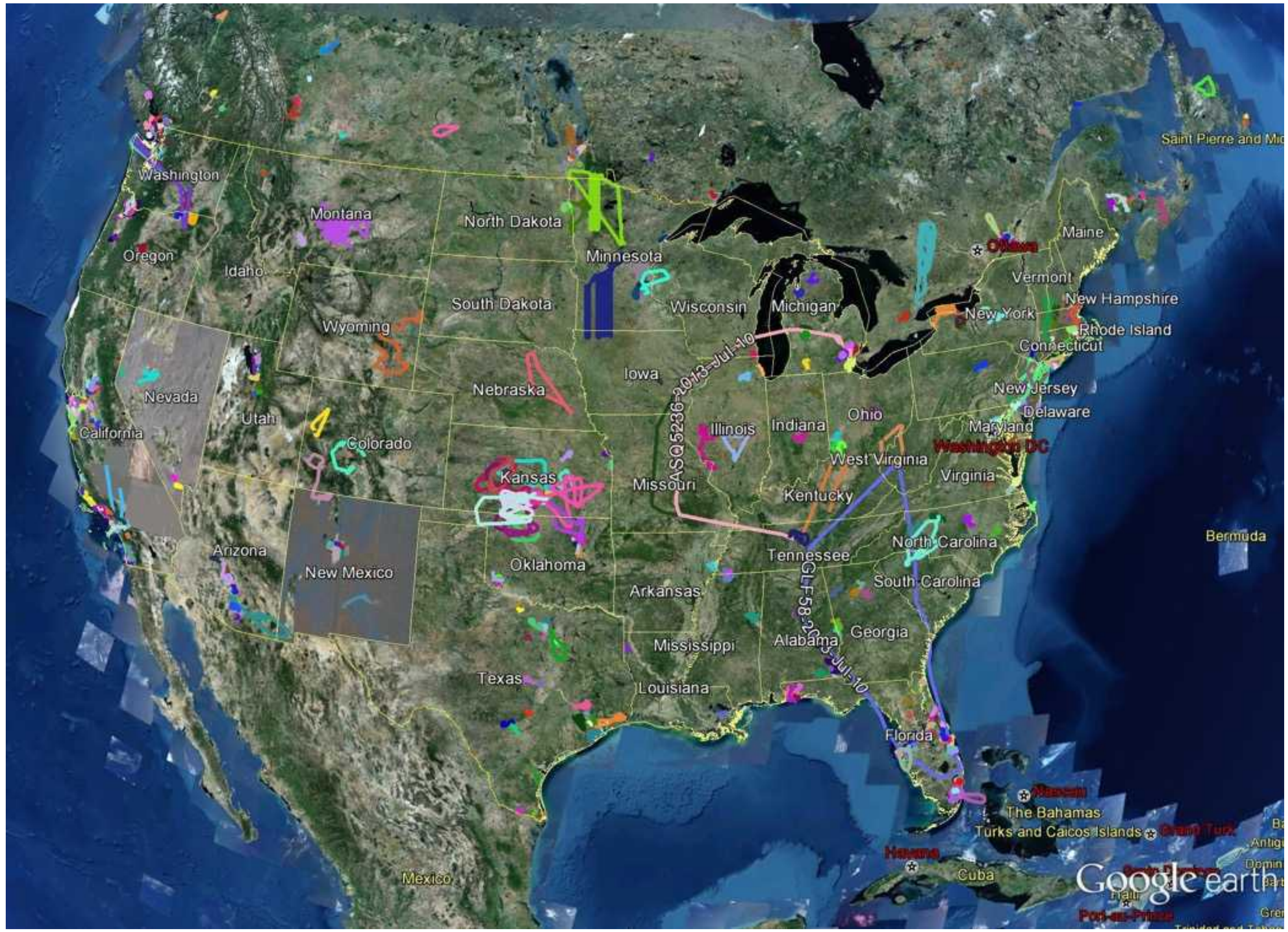
Did not specify “find this,” only told routine to “make groups of similar flights.”
This was one of many clusters that had distinctive shapes

Discovery of Odd flights

Clustering done based on geometric features

Many clusters found, but what remains is…

Note: we have ~5M points/day, ~1GB/day, currently >300GB

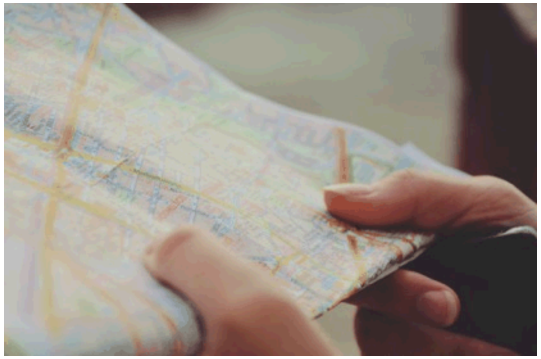


Represents approximately 700 out of a total of 50,000 flights from one day

RETHINKING GEOGRAPHICAL INFORMATION SYSTEMS

Search in Geospatial Semantic Graphs

Idea: Search for patterns in geospatial data using semantic graphs to represent object relationships in space and time

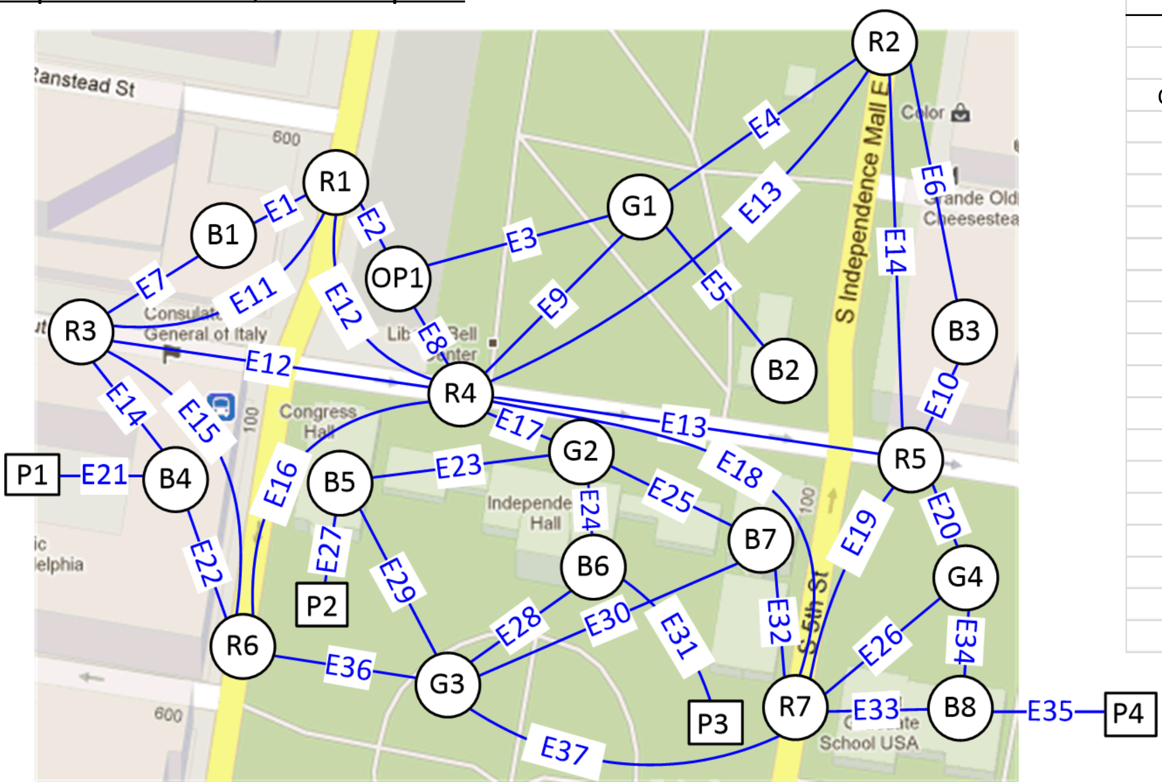


Motivating Questions:

- Where are chemical processing plants?
- Where are active businesses?
- Did someone arrive in a car and enter a building? Which one(s)?
- Where has new construction occurred?
- What previous events are geospatially correlated with current activity?

Example Geospatial Semantic Graph

Independence Hall, Philadelphia:



Region node table:

id	type	area	centroid x	centroid y
B1	building	3200	-75.14900	39.94939
R1	road	1800	-75.14910	39.94949
OP1	paved	4700	-75.14935	39.94934
G1	grass	22000	-75.15010	39.94944
R2	road	1900	-75.15060	39.94999
R3	road	1100	-75.14885	39.94934
R4	road	2200	-75.14980	39.94924
B2	building	780	-75.15045	39.94931
B3	building	6000	-75.15075	39.94944
B4	building	12000	-75.14895	39.94884
B5	building	2100	-75.14920	39.94899
G2	grass	7700	-75.14990	39.94906
R5	road	870	-75.15065	39.94896
B6	building	2000	-75.15000	39.94889
B7	building	3150	-75.15040	39.94884
G4	grass	15300	-75.15080	39.94869
R6	road	1970	-75.14905	39.94844
G3	grass	25000	-75.14960	39.94829
R7	road	1810	-75.15050	39.94834
B8	building	2700	-75.15090	39.94819

Edge table:

edge_id	node_1	node_2
E1	B1	R1
E2	R1	OP1
E3	OP1	G1
E4	G1	R2
E5	G1	B2
E6	R2	B3
E7	R3	B1
E8	OP1	R4
E9	R4	G1
.	.	.
.	.	.
.	.	.

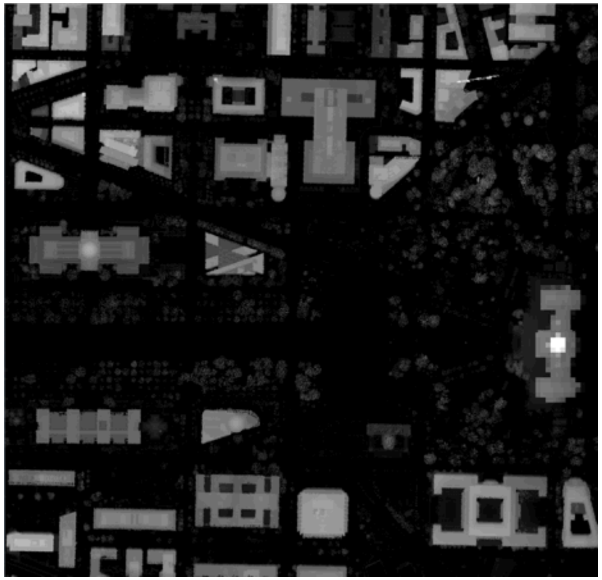
Point node table:

id	Name	Address	Latitude	Longitude
P1	Consulate of Italy	150 S. Independent Mall West #1026	-75.14895	39.94884
P2	Congress Hall	41 N 6th Street	-75.14920	39.94899
P3	Independence Hall	520 Chestnut Street	-75.15000	39.94889
P4	Graduate School USA	150 S. Independence Mall West #674	-75.15090	39.94819

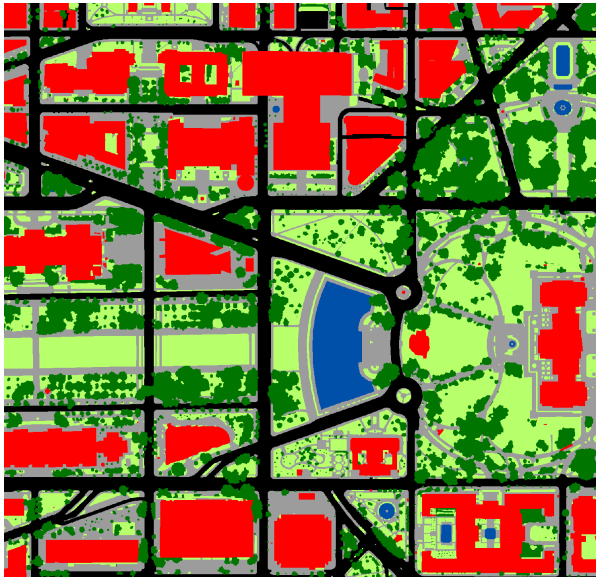
Populating the Graph: Example from Washington, DC

Zoomed in:

RGB+IR Optical Image



LiDAR Height Map (nDSM)

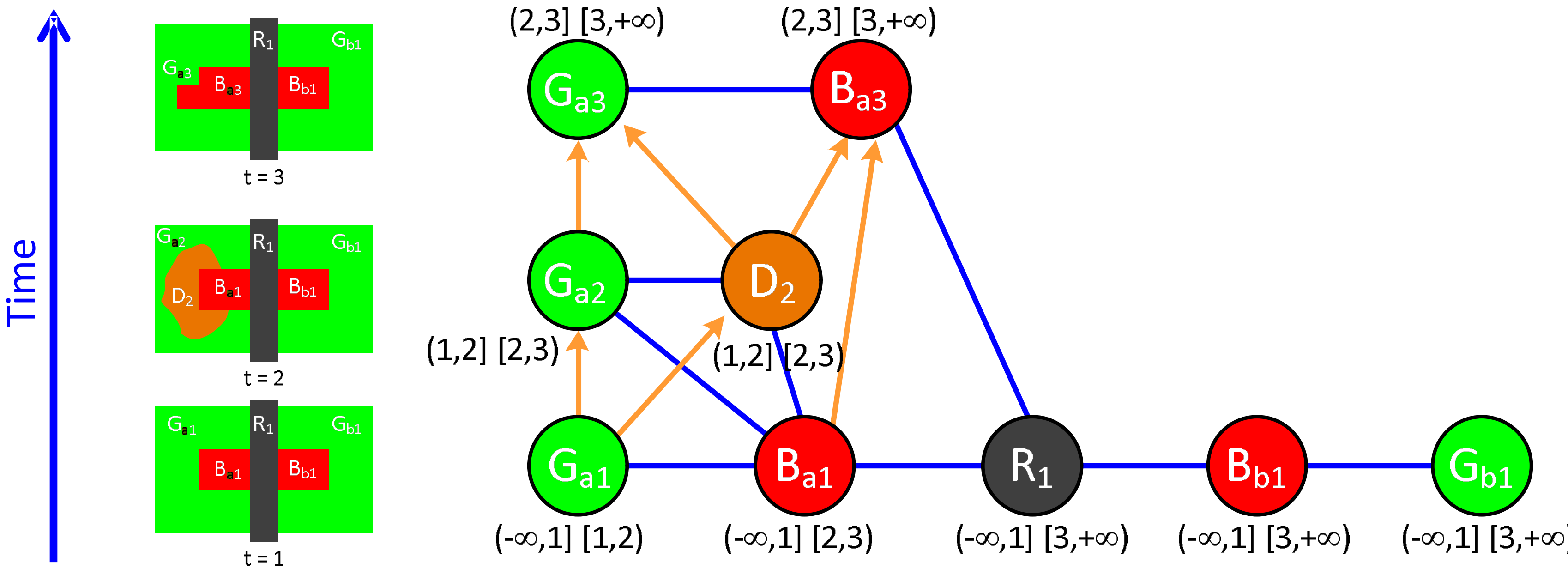


Posterized Land Cover

All of our wide-area data sets include this level of detail (roughly).

Durable Change Representation

StoredGraph:



Legend:

G Durable nodes

Adjacency edges

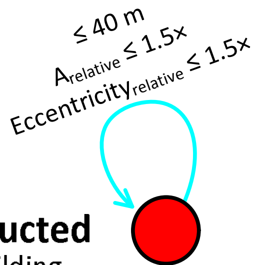
Change edges

Data semantics:

Building	B
Grass	G
Dirt	D
Road	R

New Complexes

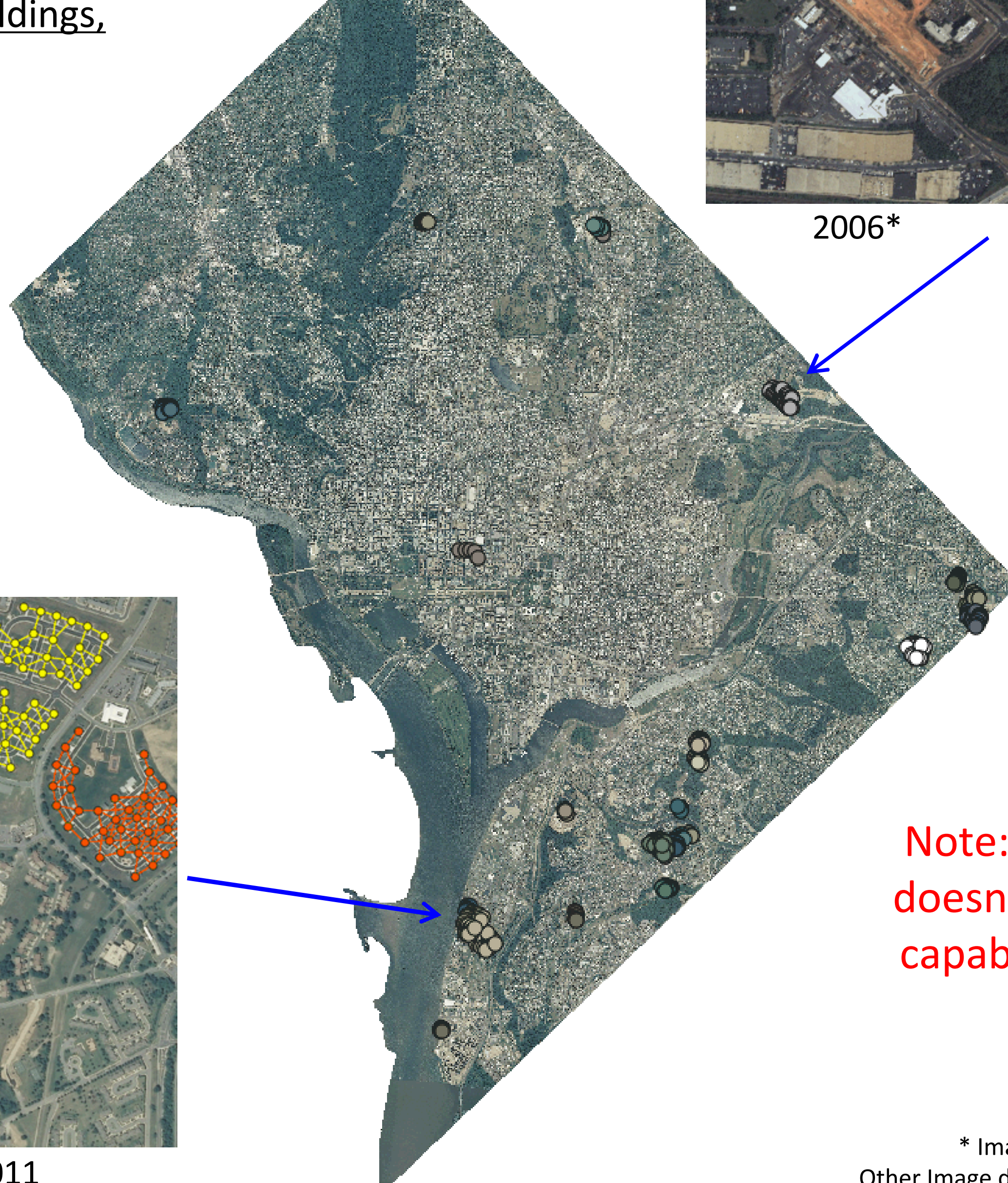
<u>Seek complexes of new buildings, across the entire city:</u>



Constructed
Data: Building
Exists now
$A \geq 100\ m^2$
New, Extended, Changed
$n \in [5, \infty]$

2006*

2011



2006*

2011

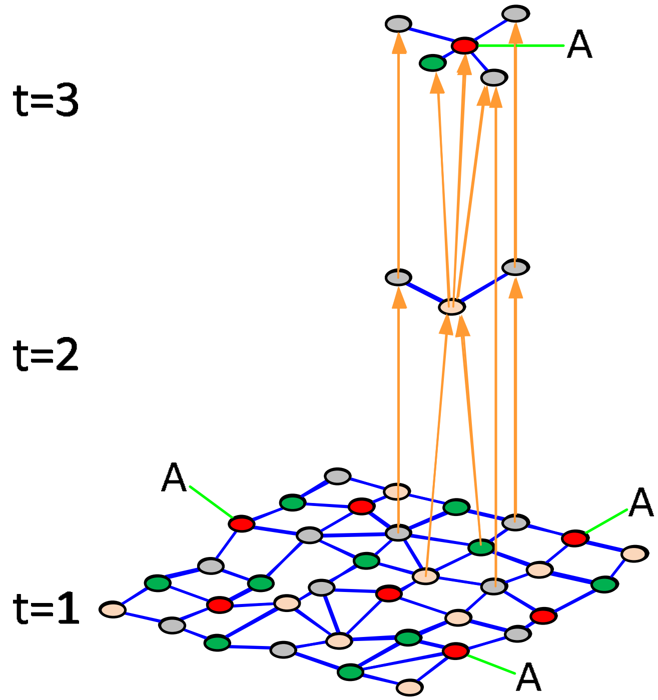
Note: This example doesn't use "activity capability" of graph

From: *UUR SAND 2014-19280 C*

* Image from DigitalGlobe.
Other Image data provided by UVM.

GeoSpatial Semantic Graph Representations of Features & Activity

- Graph includes activity:
  - @ t=1, the graph includes objects with location
  - From t=1 to t=2, the graph encodes change
  - Nodes for static features, ephemeral features and activity events.
  - Node attributes include time observed.
  - No persistence required.
  - Spatial and temporal relationship edges.



Why Temporal Analysis?

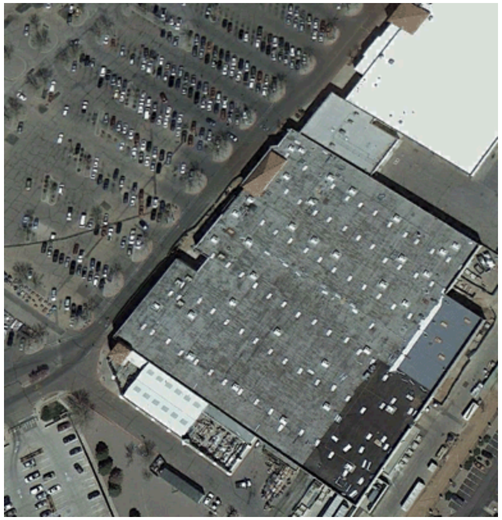
- Time is often crucial to understanding what's happening.

Successful?

On Sunday:

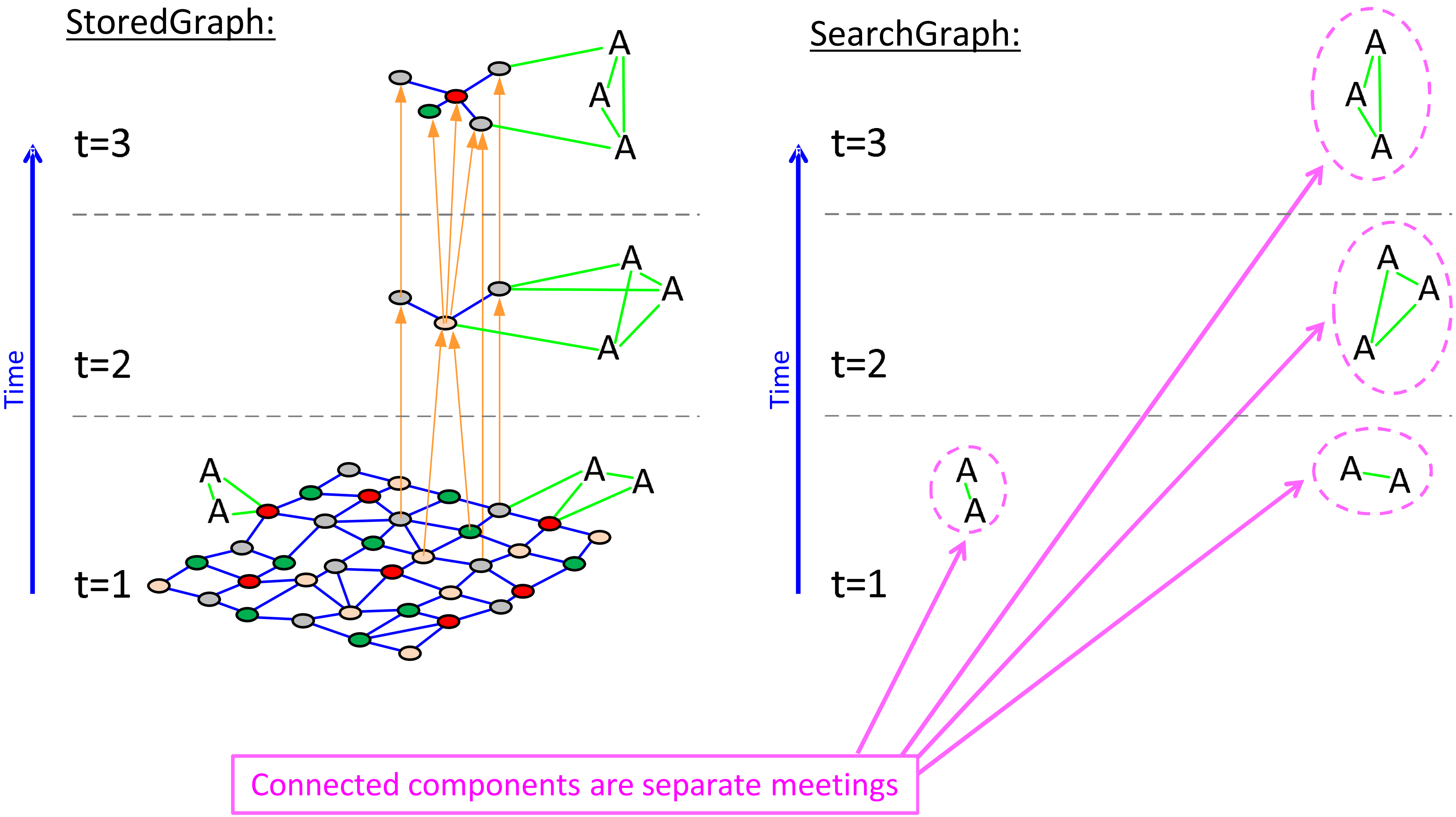
Successful?

On Sunday:

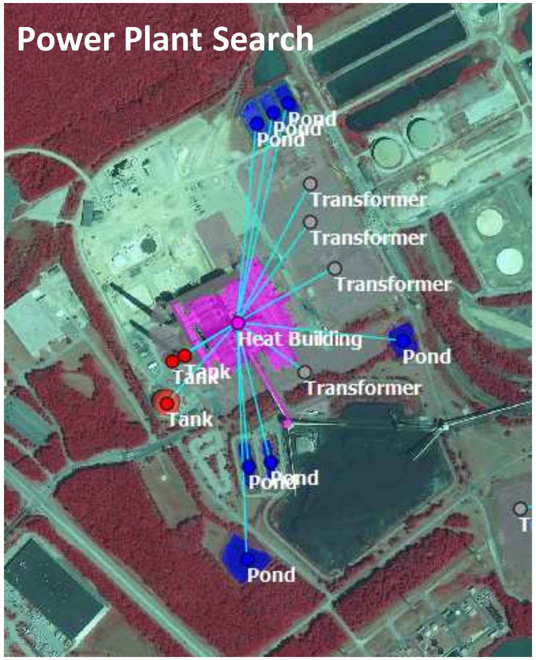


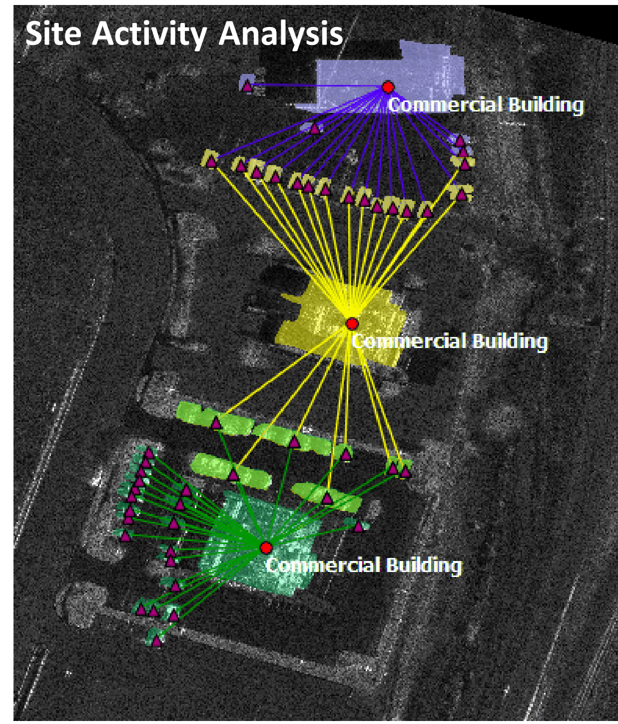
StoredGraph with Multi-Time Activity

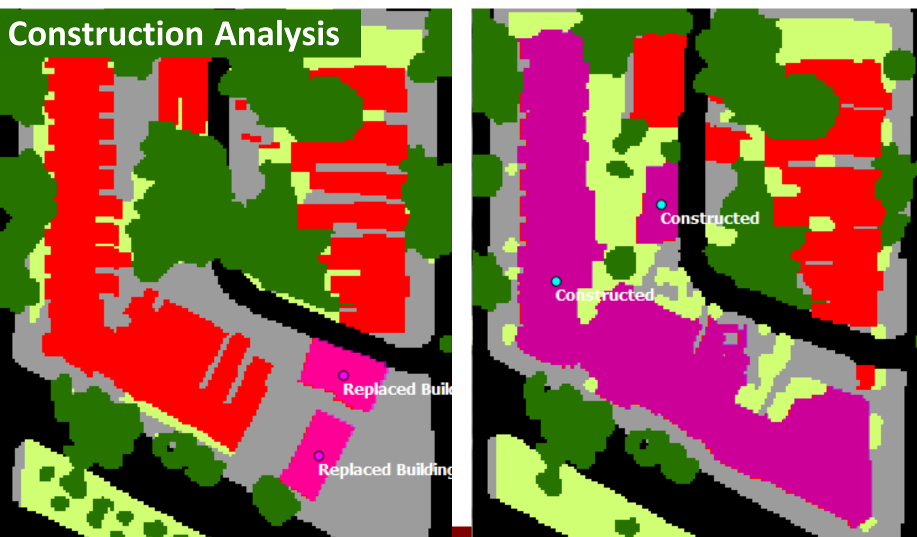
Question: Have I seen this type of activity before?



Diversity of Problems







All of these were solved by the same code.

Advantages of a Graph

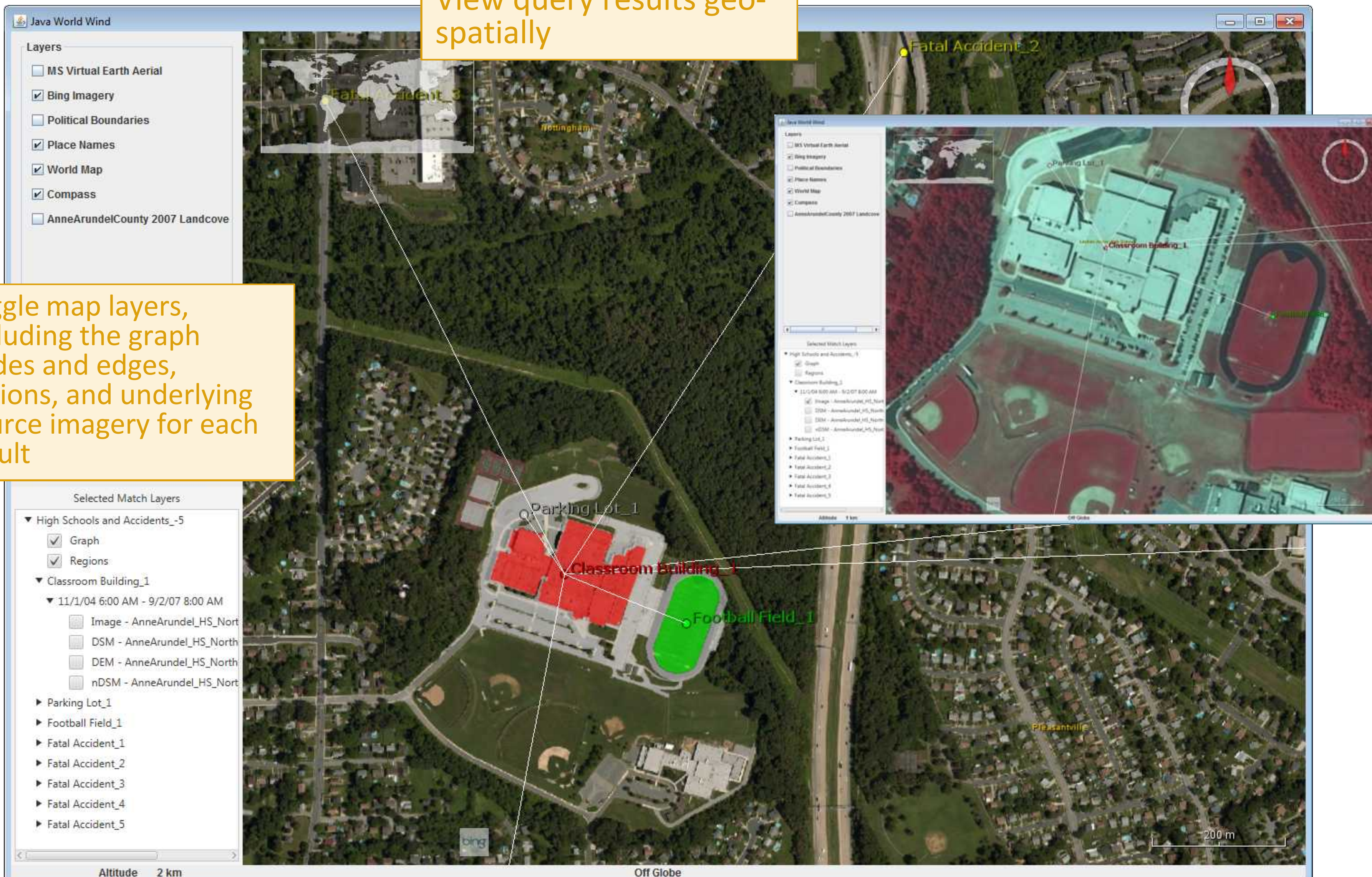
- Efficient representations in time (only store change).
- Relationship, change, and temporal analysis and heterogeneous spatial ensembles in the same query.
  - Change detection.
  - Activity characterization.
  - Different aspects of temporal analysis…
- Combines direct analysis of geospatial imagery, database query filtering, and graph search algorithms within one framework.
  (e.g, SQL queries cannot solve graph transitive closure operations.)
- Able to take full advantage of graph topology search, enhanced by geospatial-temporal semantics.
- Feature-based analysis:
  - Multi-modality, in a single search representation.
  - Sensor agnostic.

Serves as powerful tool for analysts to *remember*

UI to Render query results

View query results geo-spatially

Toggle map layers, including the graph nodes and edges, regions, and underlying source imagery for each result



The science of foraging and filtering…

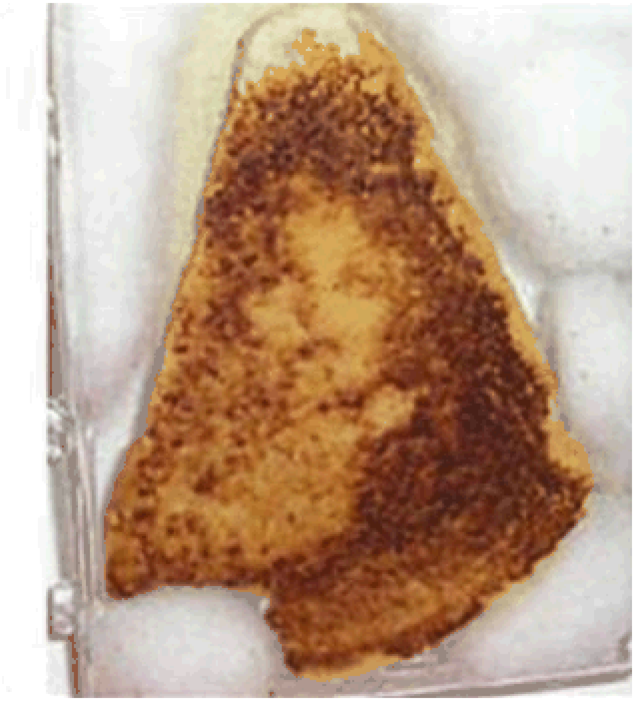
RETHINKING SEARCH

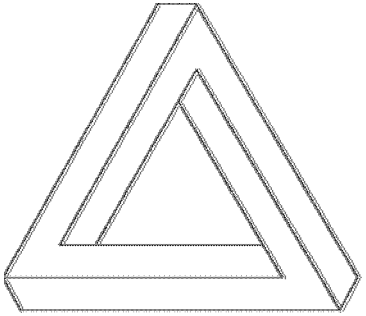
Visual Cognition Basics

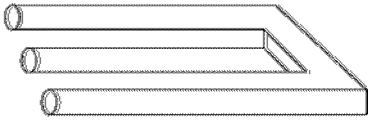
- The human visual system is VERY good at:
  - Finding patterns
  - Making inferences

- Perceptual systems are constantly receiving ambiguous information and trying to make sense of it

- Draws on both perceptual cues and conceptual knowledge (bottom-up and top-down processing)
  - Relatively little is understood about top-down processing

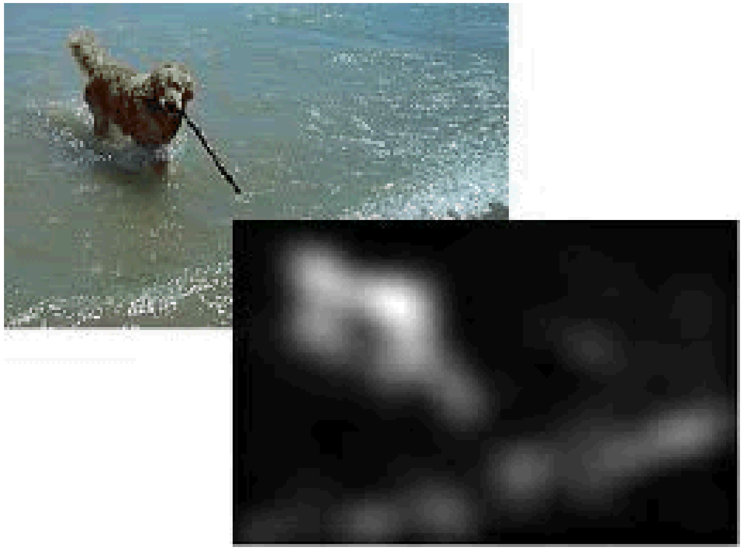




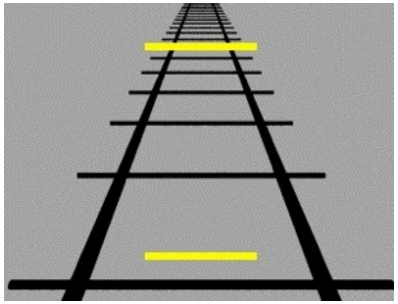


Visual Attention

- Bottom-up
  - Driven by properties of stimulus
    - ***Visual salience*** (contrast between features of a stimulus and the features of its neighbors) captures attention
  - Parameters are well understood and can be modeled



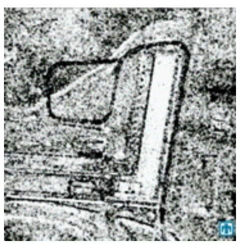
- Top-down
  - Driven by viewer's goals
  - Affected by cognitive load, working memory, past knowledge and experience
  - Has a very powerful influence on bottom-up perception
  - Parameters are NOT well understood



Can we model top-down visual saliency for domain experts?

- What bottom-up features capture attention in non-optical imagery?
- How does domain experience influence visual search/inspection?
- How can top-down visual attention be modeled?
- Do people with expertise in one domain perform differently on domain-general tasks?

SAR
CCD Products

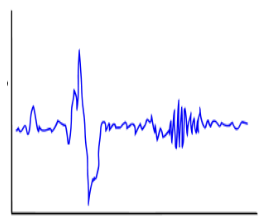


Intended to make important features more salient

NGA
Satellite Imagery

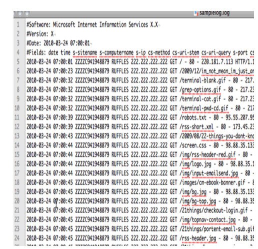
Similar to optical imagery

NG Engineers
Waveforms



Visualizations of raw data

Cyber
Log Files



Raw data

TSA
False color X-rays

Intended to make important features more salient

Novices

Experienced with optical imagery only

All participants completed a battery of domain-general tasks and a domain-specific tasks

How do humans process and filter visual clutter? Empirical Analysis of Top-down Modeling

- The first model spiral tested our ability to predict expert fixation patterns for a given image, search goal and previously identified goal-relevant regions
- We investigate how top-down elements could be applied to the output of a bottom-up model as filters or amplifiers of modeled fixation patterns
- Results: A simple mask removing salience of shadows dramatically improved match between salience and eyetracking-derived gaze maps

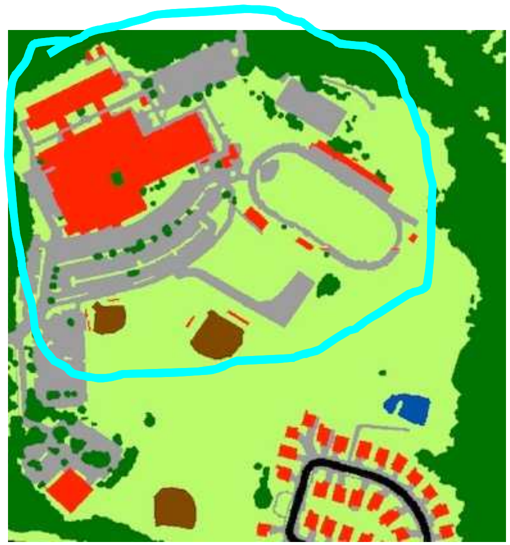
CRADA with Eyetracking, Inc.

Future: Rethinking Search Queries

Idea: Enable analysts to generate queries from examples – enable intuition.

Challenges:

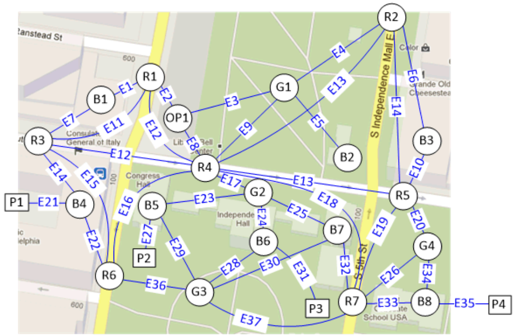
- Interactive selection of salient image features
- How do human-machine systems handle visual clutter?
- Infer queries from images
- Query refinement



Human Analytics:

- Visual search and attention
- Reasoning under uncertainty

Pattern Analytics R&D



Sandia rethinks traditional GIS and geospatial search.

New. Populate graphs with image-derived, geo- and time- tagged features from multiple data sources.

New. Execute graph-algorithm powered search for threat signatures that consist of durable features and activities.

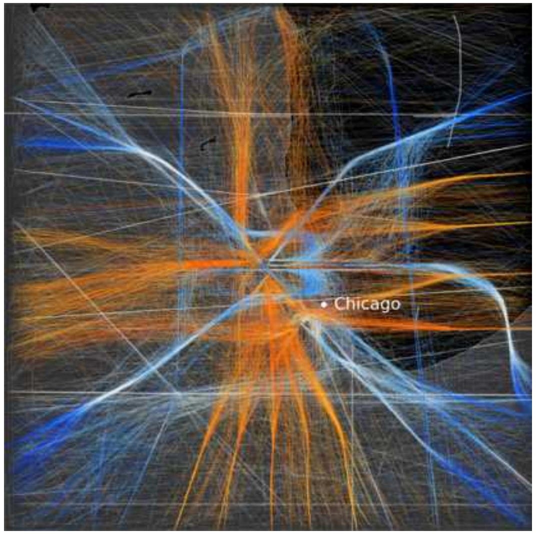
New. Uncertainty and quality of search considerations with real, noisy data.

PANTHER rethinks patterns in motion.

New. Geometric and temporal trajectory analyses represent and compare tracks efficiently and lightning-fast.

New. Discovery of geospatial-temporal relationships and comparison of more than two trajectories.

Elevate the analysts ability to discover and disambiguate.



CREDITS: PANTHER Leadership Team

- PI: Kristina Czuchlewski
  - krczuch@sandia.gov
- PM: Bill Hart
  - wehart@sandia.gov

- Team Leads:
  - Randy Brost
  - Jim Chow
  - Laura McNamara
  - Danny Rintoul
  - David Stracuzzi

CREDITS: PANTHER Members

- M. Carroll
- K. Cole
- J. Coram
- S. Dauphin
- J. Goold
- M. Haass
- P. Kegelmeyer
- M. Koch
- R. Malinas
- L. Matzen
- W. McLendon
- S. McMichael
- L. McNamara
- D. Morrow
- M. Moya
- D. Perkins
- C. Phillips
- D. Patterson
- M. Peterson
- A. Pound
- R. Riley
- D. Robinson
- T. Quach
- T. Shead
- S. Stevens-Adams
- C. Valicka
- D. West
- A. Wilson
- D. Woodbridge