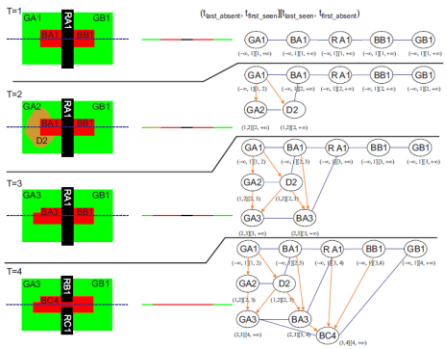
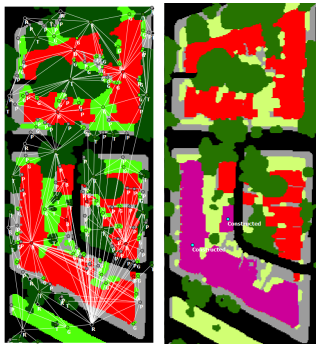
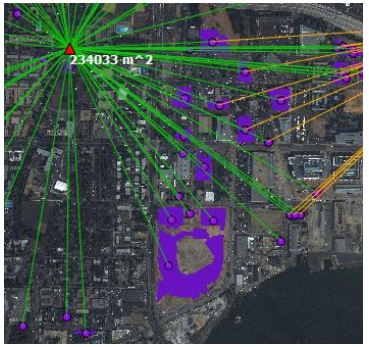
SAND2014-19280C

*Exceptional service in the national interest*

Sandia National Laboratories






# A Computational Framework for Ontologically Storing and Analyzing Very Large Overhead Image Sets

Randy C. Brost, Ph.D.

November 4, 2014

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

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.

# Acknowledgements


Co-Authors:

- Will McLendon
- Ojas Parekh
- Danny Rintoul
- David Strip
- Diane Woodbridge

Colleagues:

- Michelle Carroll
- David Perkins
- Jarlath O-Neil-Dunne, UVM

Support:

- DOE NA-22 Office of Non-Proliferation.
- Sandia LDRD Office.

# Overview

Sandia National Laboratories

- Motivation.
- Computation.
- Results.
- Discussion.

# Motivation

Sandia National Laboratories

GeoGraphy

Motivation:

- Overwhelming remote sensing data.
- Wide-area search is tedious, error-prone.
- Reasoning over time is even more difficult.
- We don't want to miss important items (proliferation, weapons production…).

General approach:

- Automatically find items of potential interest.
- "Cue" to user for review.

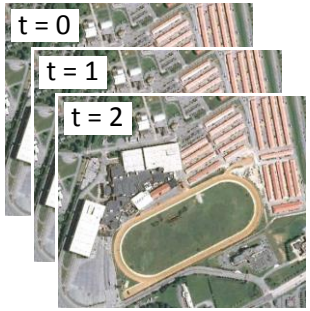
Types of questions:

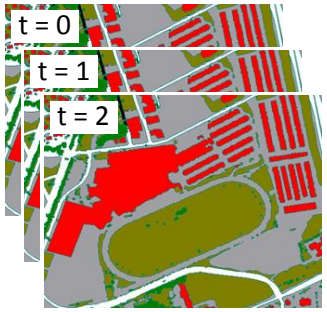
- Geospatial: Find all power plants.
- Temporal: Find all changes.
- Geospatial-temporal: Find all power plants that changed.
- Multi-modality: Find new construction near points of interest.
  Find industrial facilities with unusual emissions.

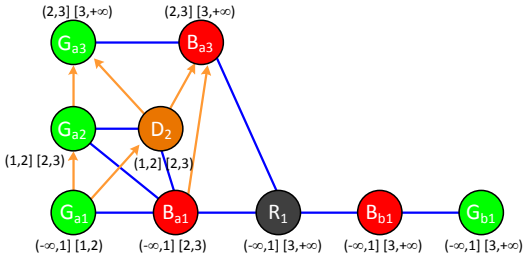
# Overview

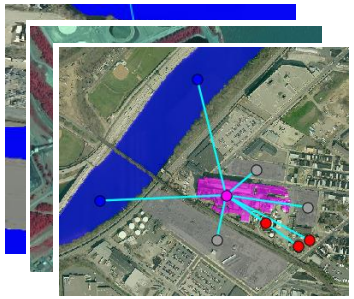
- Motivation.
- Computation.
- Results.
- Discussion.

# Geospatial-Temporal Graph Data Flow



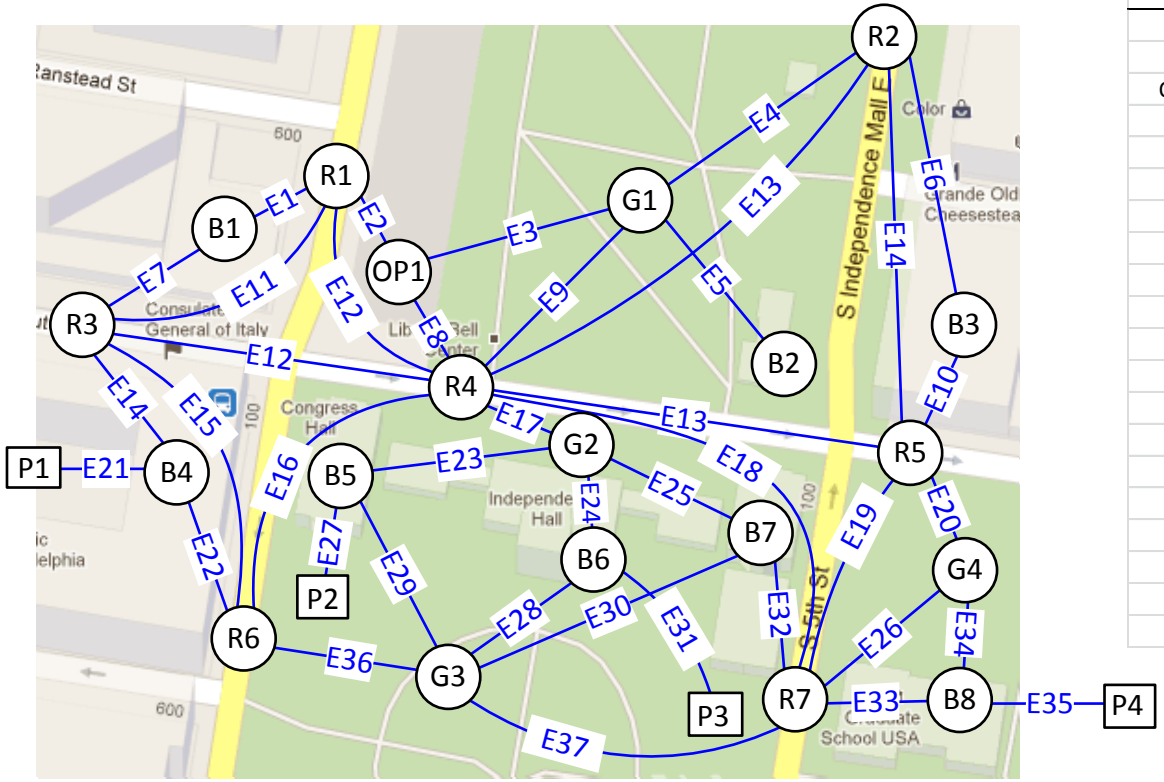
Remote Sensing → Primitive Recognition (U. Vermont) → Spatiotemporal Graph → Spatial, Temporal, Spatiotemporal → Matches

Query:
- Spatial parameters
- Temporal parameters
- Subgraph pattern
- Algorithm control

→ **Change Detection**

# 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 |

Point node table:

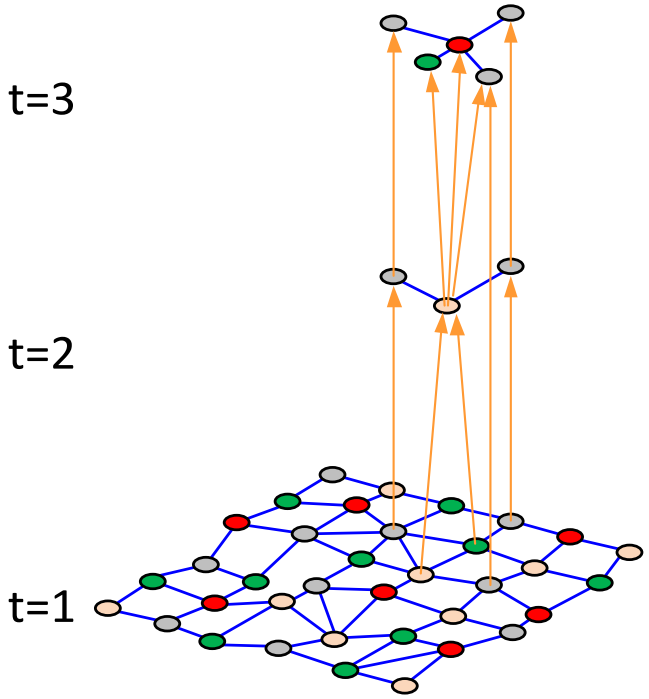
| id | Name | Address | Latitude | Longitude |
|---|---|---|---|---|
| P1 | Consulate of Italy | 150 S. Independent Mall West #1026 | -75.14895 | 39.94884 |
| P2 | Congress Hall | 41 N 6th Street | -75.14920 | 39.94899 |
| P3 | Independence Hall | 520 Chestnut Street | -75.15000 | 39.94889 |
| P4 | Graduate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

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 |
| . | . | . |
| . | . | . |
| . | . | . |

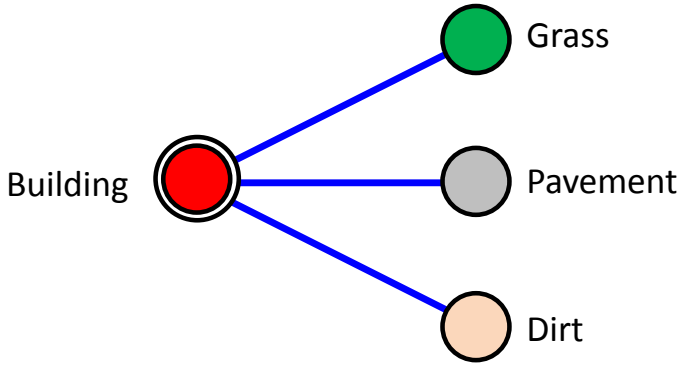
# Representing Change Over Time

- Encode change:
  - Node attributes include duration seen.
  - Only construct new nodes for changes.
  - “Changed-to” arcs encode time evolution.
  - Graph complexity focuses on change areas.

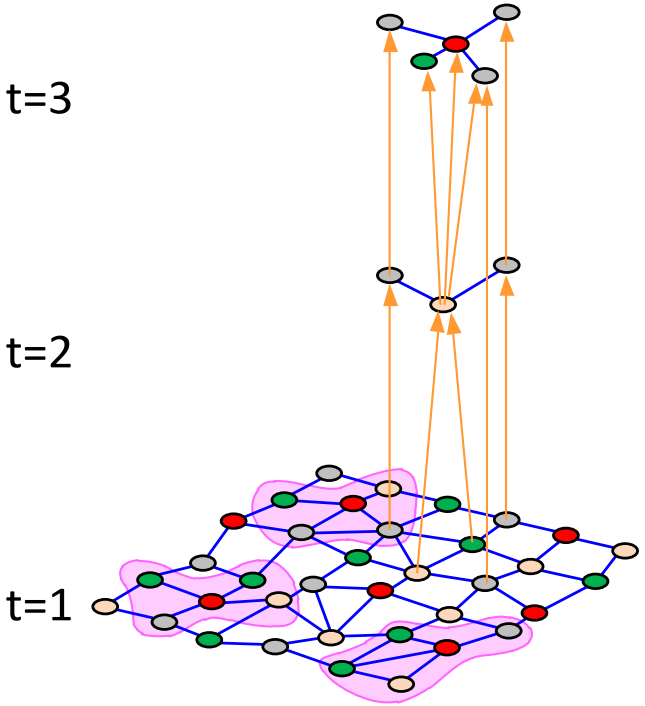


# Signature Search

- A signature encodes a desired question.
- Example: “Where are buildings with nearby grass, pavement, and dirt?

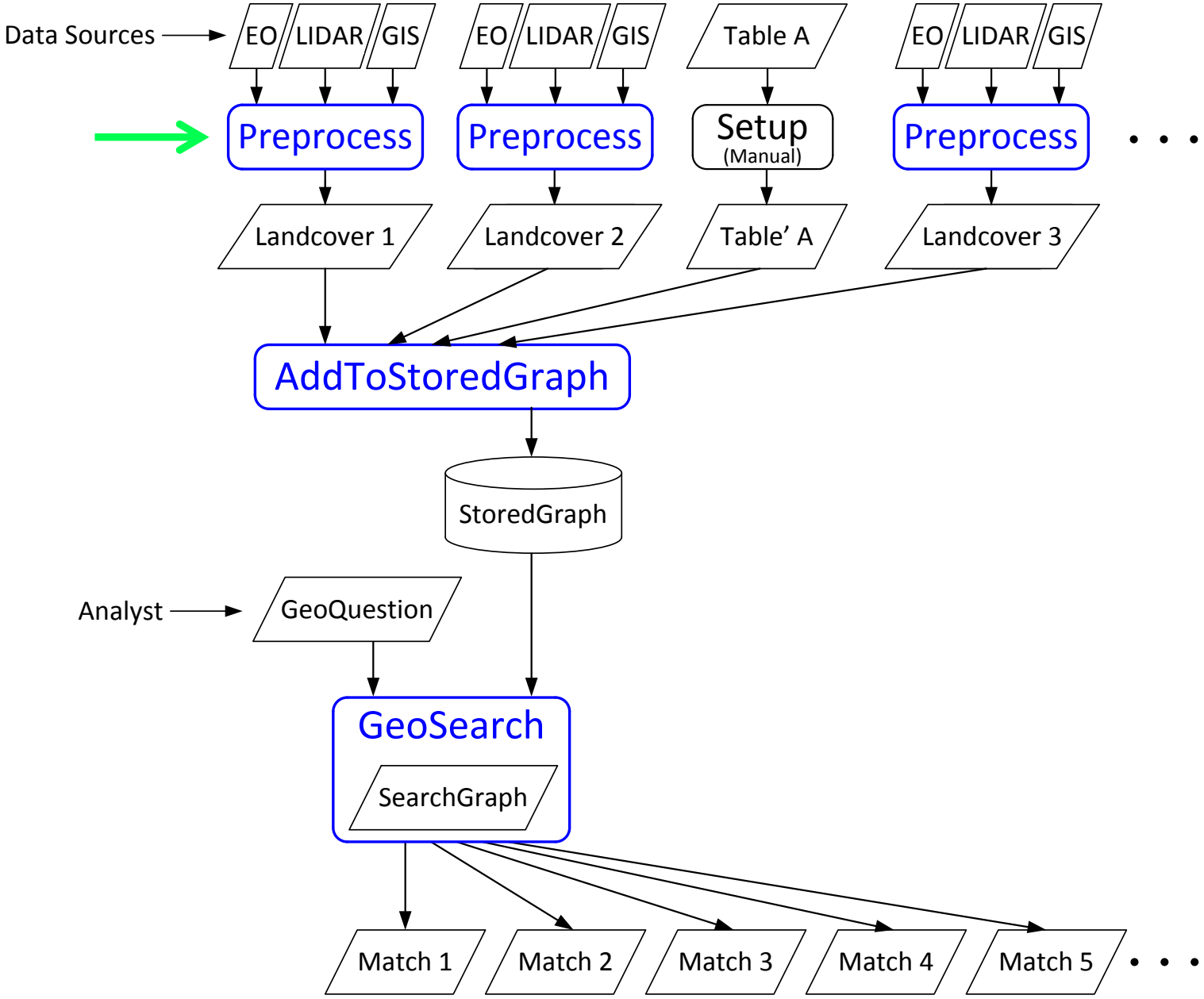



Query Template

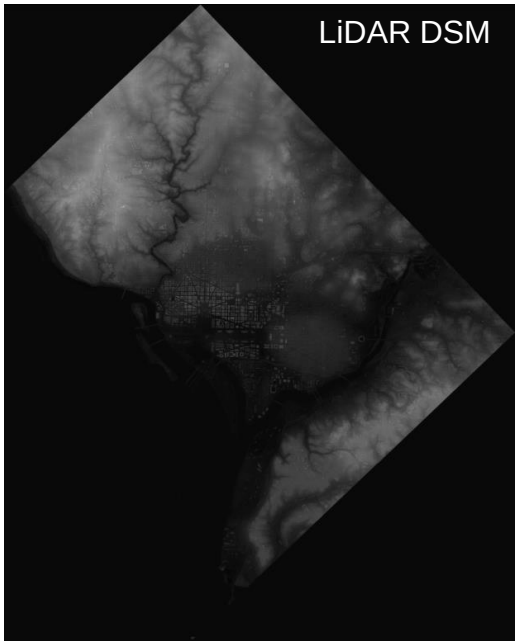



Search Results

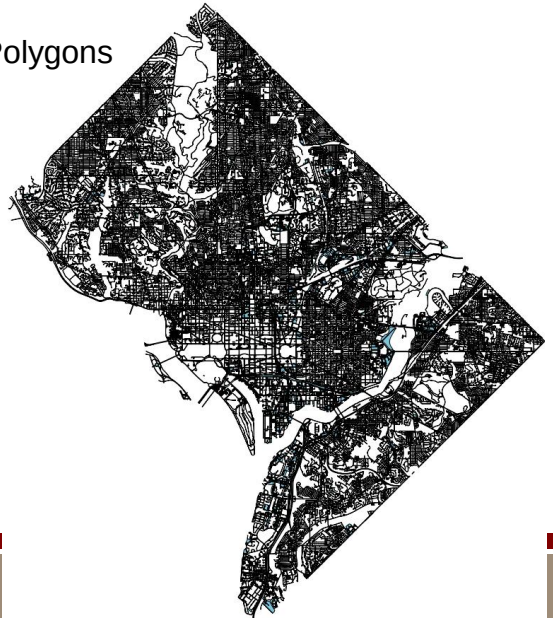
# Primary Data Flow



# Input Data


GIS Road Polygons



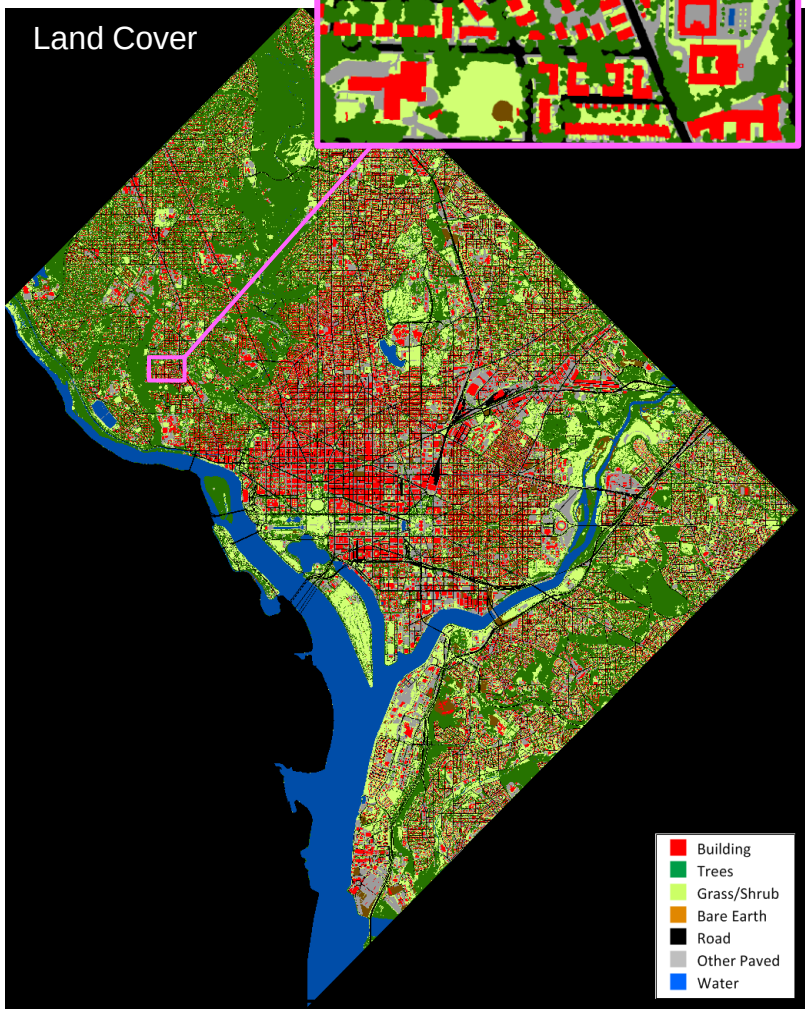
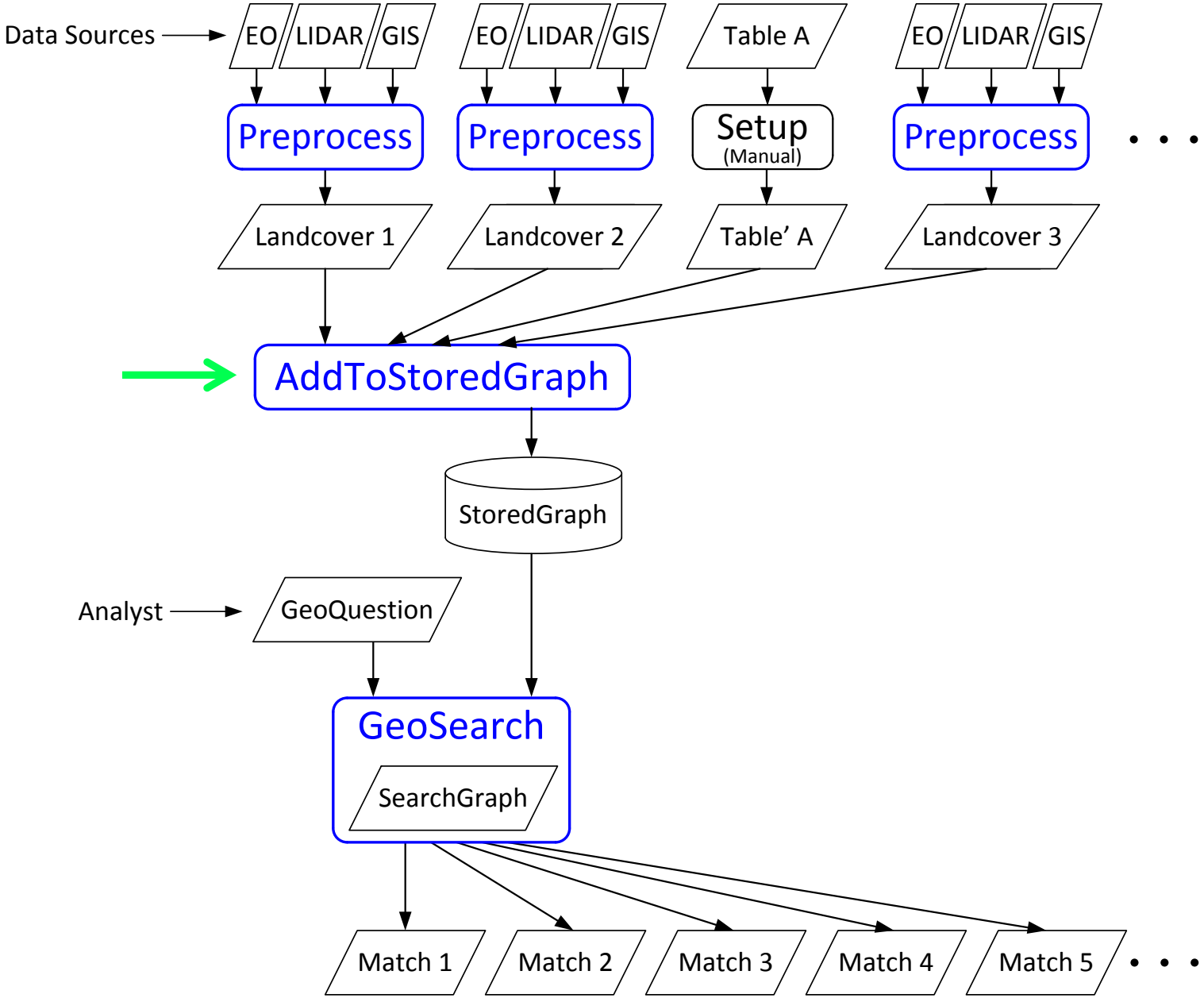



Image data and processing by University of Vermont Spatial Analysis Lab [O'Neil-Dunne 2012].

[O'Neil-Dunne 2012] O'Neil-Dunne, et al, An object-based system for LiDAR data fusion and feature extraction, Geocarto (28), pp. 227–242, 2012.

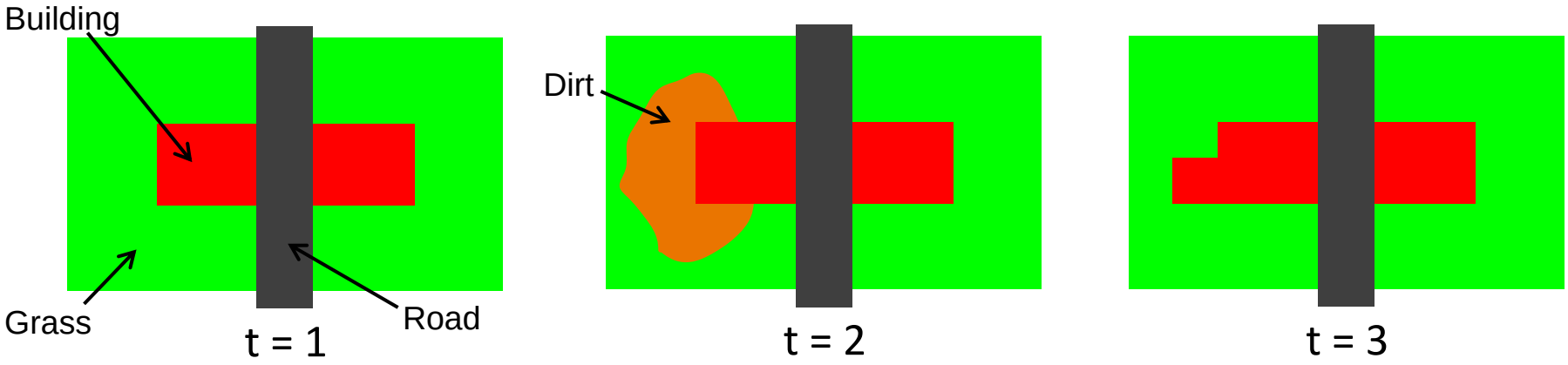
# Primary Data Flow



# Example Scenario

Sandia National Laboratories

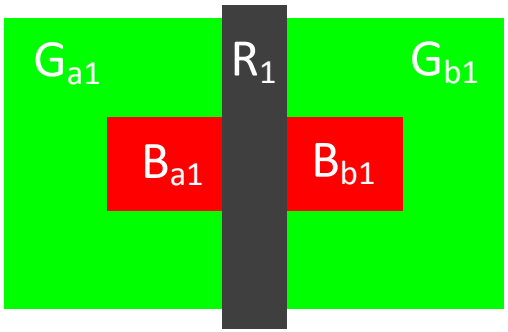
Data sequence:



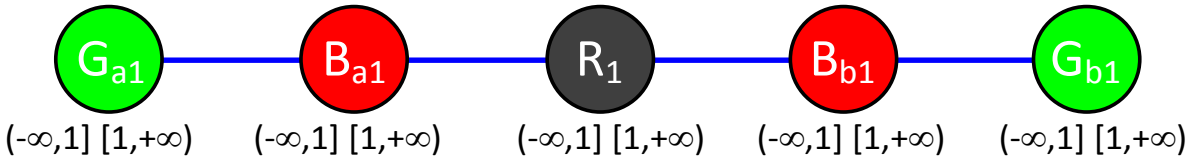
# StoredGraph Construction

Sandia National Laboratories

Start with land cover #1:




t = 1




Each node has several attributes:
area, centroid, moments, eccentricity, orientation, bounding box, etc...

Legend:

G Durable nodes

Adjacency edges

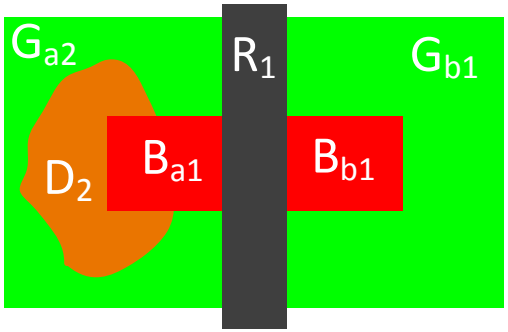
Data semantics:

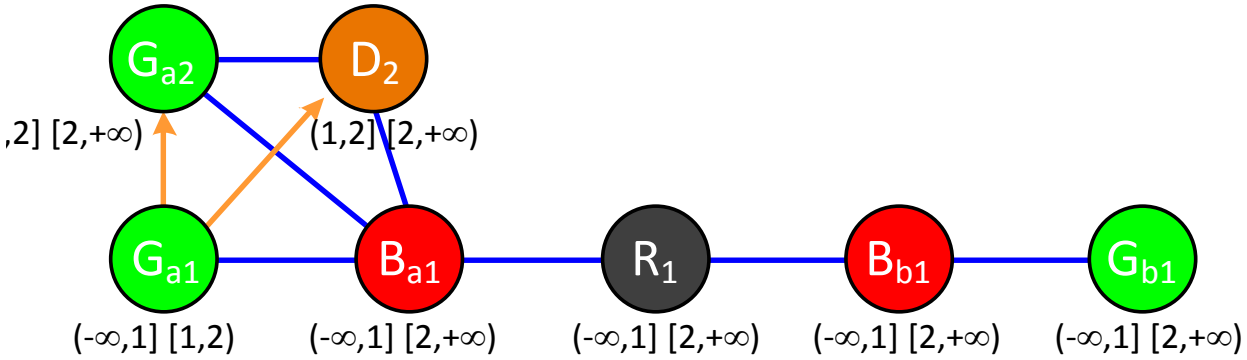
| | |
|---|---|
| Building | B |
| Grass | G |
| Dirt | D |
| Road | R |

* Road↔Grass edges omitted for clarity throughout.

# StoredGraph Construction

Add land cover #2:




Data semantics:

| | |
|---|---|
| Building | B |
| Grass | G |
| Dirt | D |
| Road | R |

Legend:

G Durable nodes

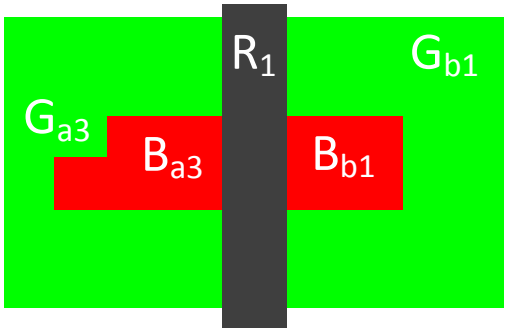
Adjacency edges

Change edges

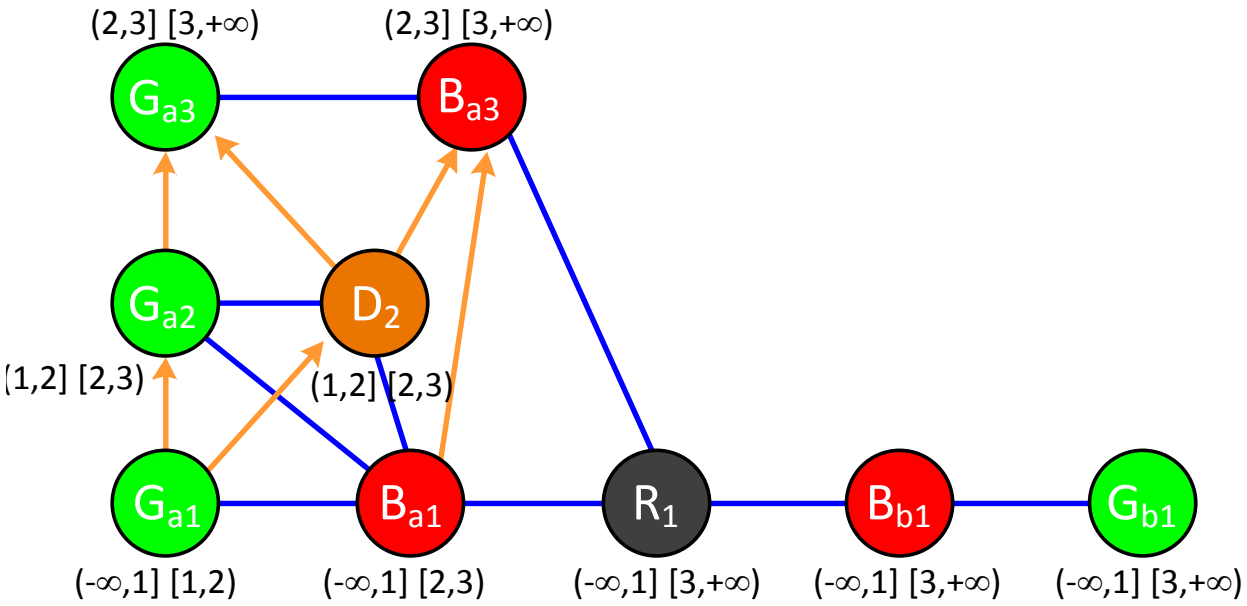
# StoredGraph Construction

Sandia National Laboratories

Add land cover #3:




t = 3




Data semantics:

| | |
|---|---|
| Building | B |
| Grass | G |
| Dirt | D |
| Road | R |

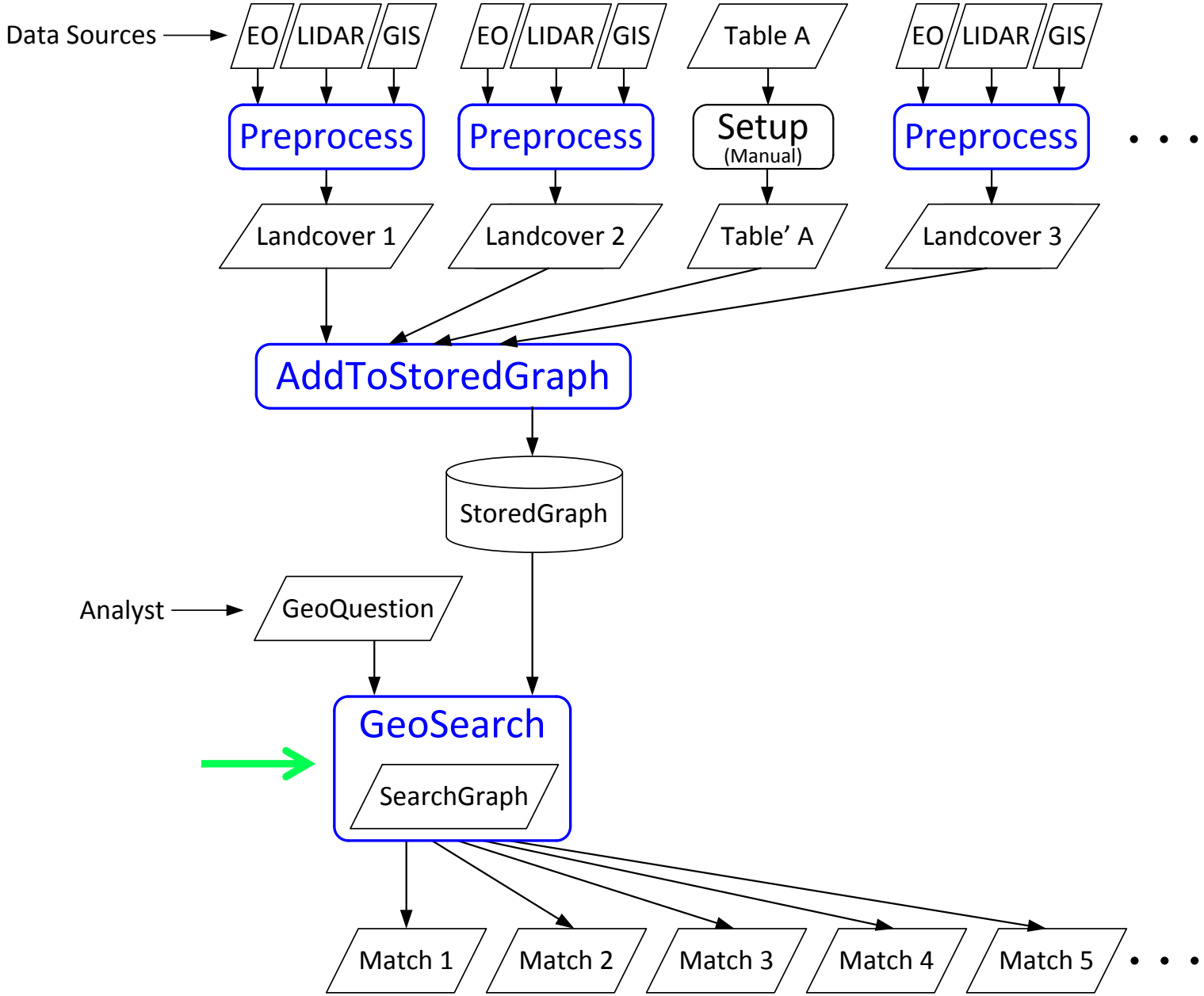
Legend:

G Durable nodes

Adjacency edges

Change edges

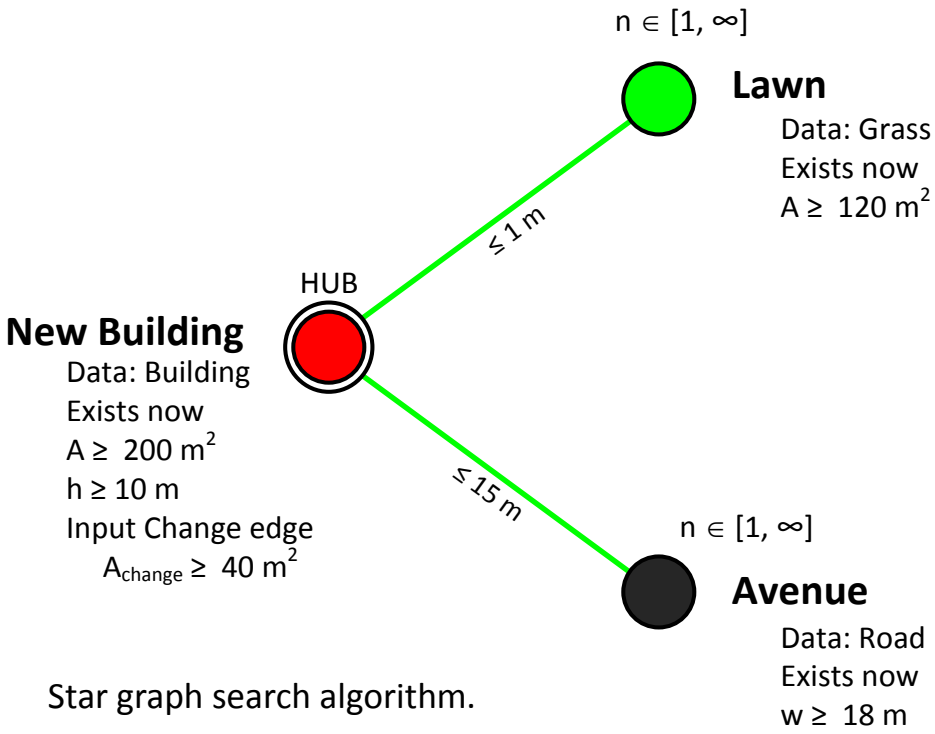
# Primary Data Flow



# Query Definition

Question: Where is a new luxury office building?

Star graph template:




Star graph search algorithm.

Question semantics:

| | |
|---|---|
| New Building | N |
| Lawn | L |
| Avenue | A |

Data semantics:

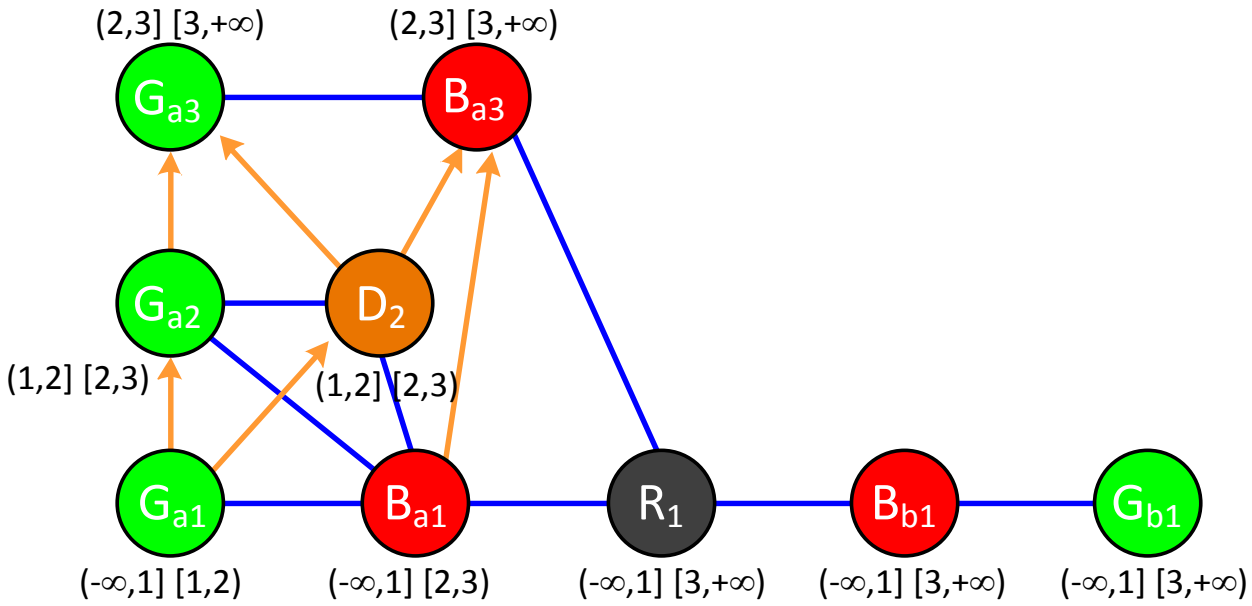
| | |
|---|---|
| Building | B |
| Grass | G |
| Dirt | D |
| Road | R |

This configuration is consistent with a luxury office building.

# Search Graph Source: StoredGraph

Sandia National Laboratories

Available StoredGraph:




Legend:

- G — Durable nodes
- Adjacency edges
- Change edges

Data semantics:

| | |
|---|---|
| Building | B |
| Grass | G |
| Dirt | D |
| Road | R |

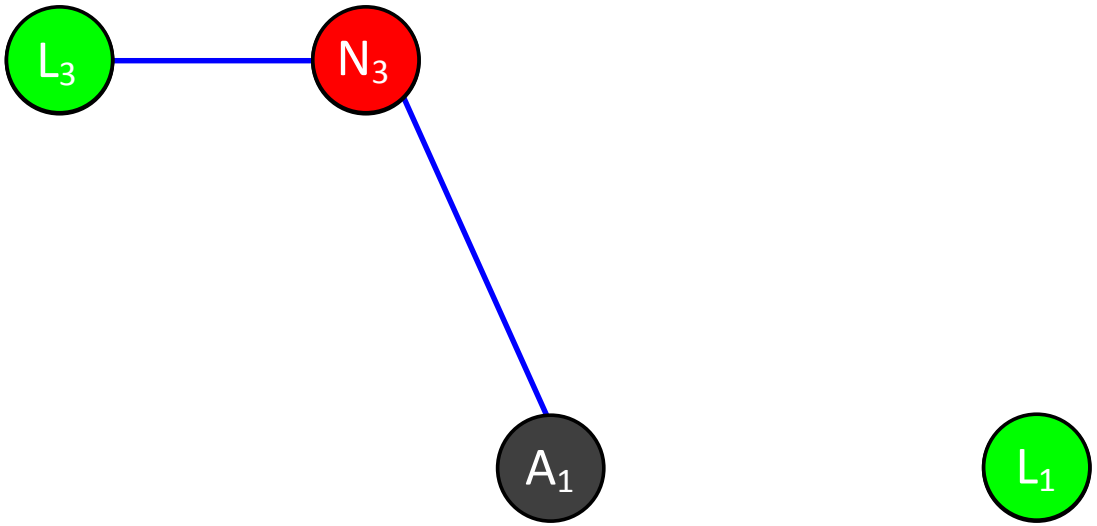
# Constructed SearchGraph

Sandia National Laboratories

<u>After adding edges:</u>

Signature nodes:
- New Building
- Lawn
- Avenue

Signature edges:
- New Building ↔ Lawn
- New Building ↔ Avenue




<u>Question semantics:</u>

| | |
|---|---|
| New Building | N |
| Lawn | L |
| Avenue | A |

SearchGraph only contains elements relevant to the question.

Semantics are now in terms of the question.

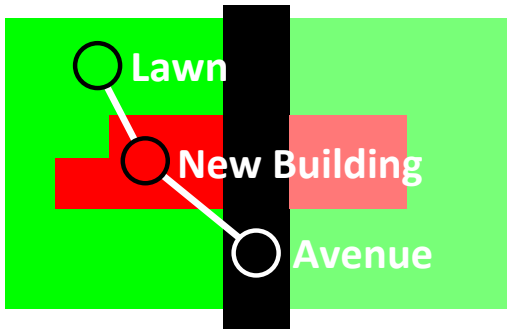
<u>Legend:</u>

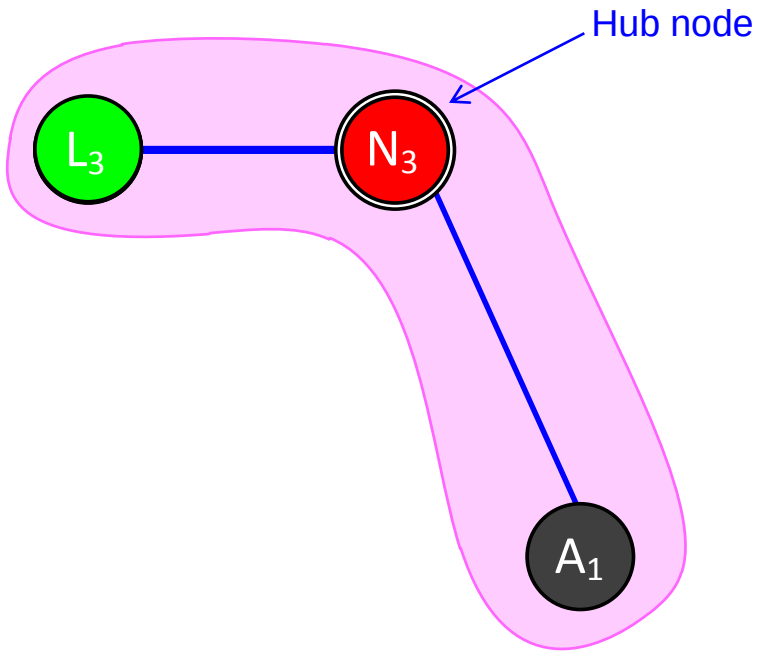
G Nodes

Edges

# Search Result

Matches found:




Remarks:

- This match is consistent with hypothesized signature.
- It does not "prove" this is a luxury office building --- this template might also match a movie star's mansion.

# Graph Search

Sandia National Laboratories

Initial implementation [Watson 2010]:

- Subgraph isomporhism.
- NP-complete! [Cook 1971, Ullmann 1976]

Current approach:

- Identify relevant portion of StoredGraph.
- Lazy constraint, distance edge evaluation, with caching.
- Simple graph search algorithms:
  - Star graph.
  - Connected component.
  - Interrupted star, using transitive closure.
- Postprocessing calculations.

[Cook 1971] S. A. Cook, "The Complexity of Theorem-Proving Procedures," 3rd ACM Symposium on Theory of Computing, pp. 151–158, 1971.

[Ullmann 1976] J. R. Ullmann, "An Algorithm for Subgraph Isomorphism," Journal of the ACM 23(1), pp. 31–42, 1976.

[Watson 2010] J. P. Watson, “Complex Signature Detection Using Geospatial/Temporal Semantic Graphs,” Simulations, Algorithms, and Modeling Program Review Meeting (SAM2010), April 2010.

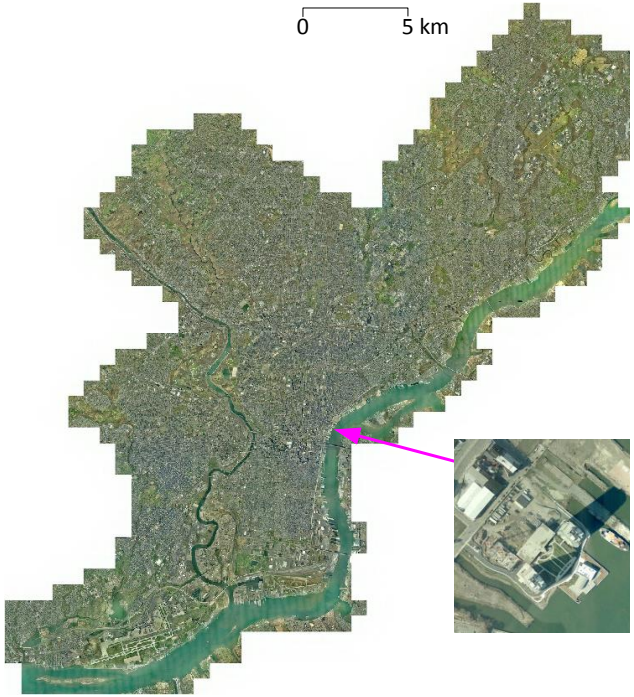
# Overview

Sandia National Laboratories

- Motivation.
- Computation.
- Results.
  - Data.
  - Power Plant Search.
  - Refinery search.
  - Change Analysis.
  - New Building complex search.
- Discussion.

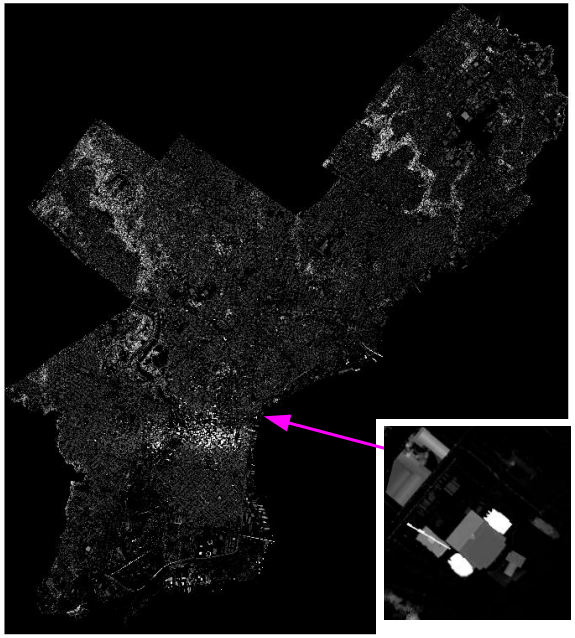
# Philadelphia 2008

Primary input:



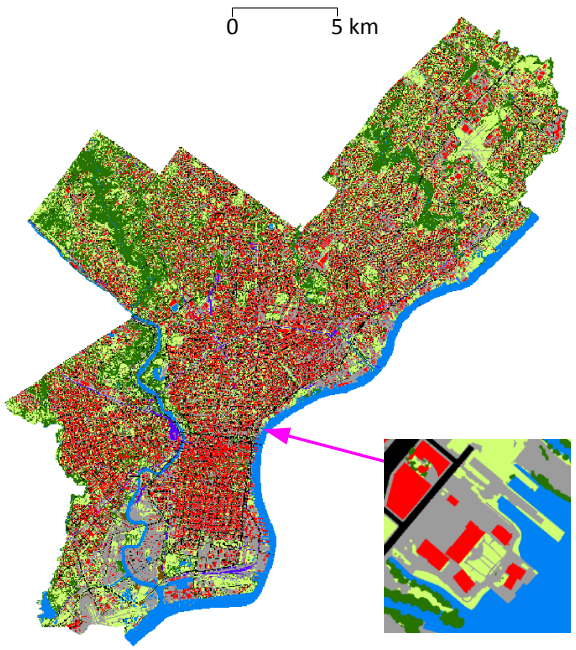

Optical Image

Pixel size 0.1 m
307,531 × 330,033 pixels
(101.5 Gpix)
7,669 MB



LiDAR nDSM

Pixel size 0.3 m
89,540 × 100,294 pixels
(9.0 Gpix)
2,084 MB



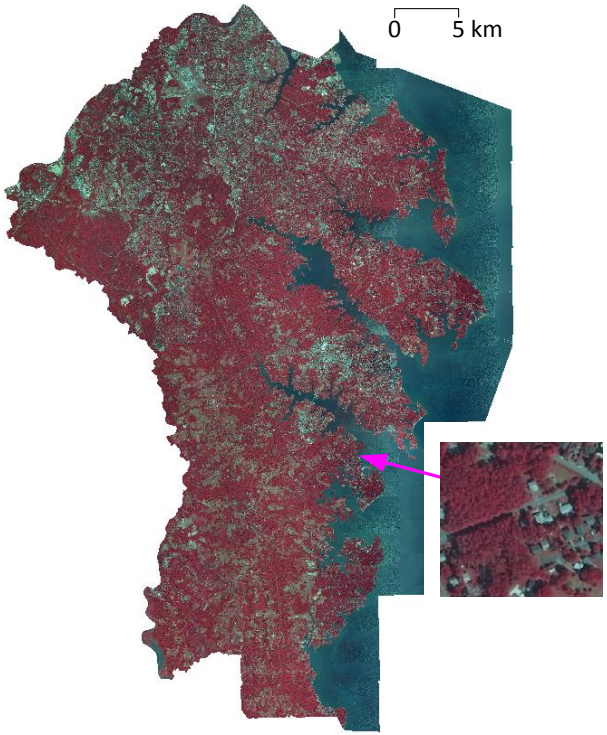

Land Cover

Pixel size 0.3 m
89,548 × 100,303 pixels
(9.0 Gpix)
8,775 MB

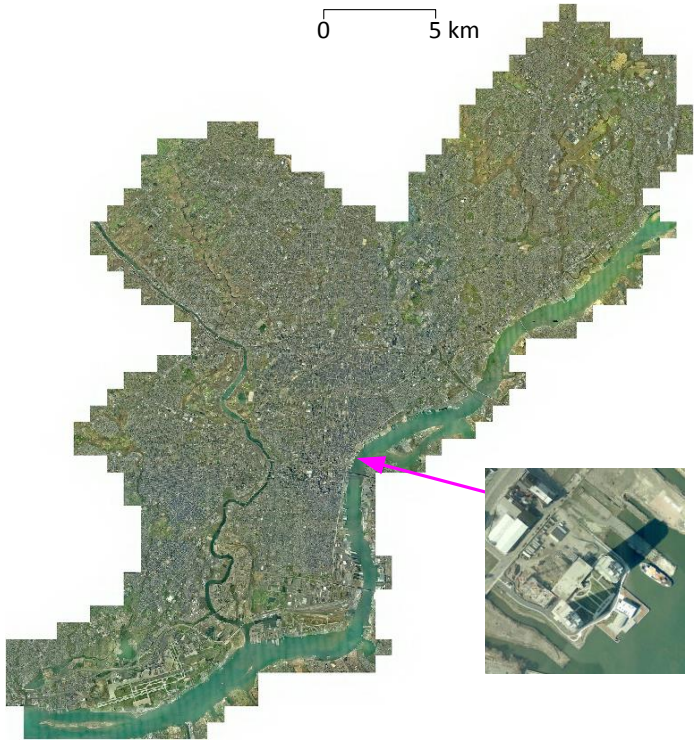
# Three Data Regions

Sandia National Laboratories

Search results:




Anne Arundel County, MD




Philadelphia, PA


Washington, DC

Total:

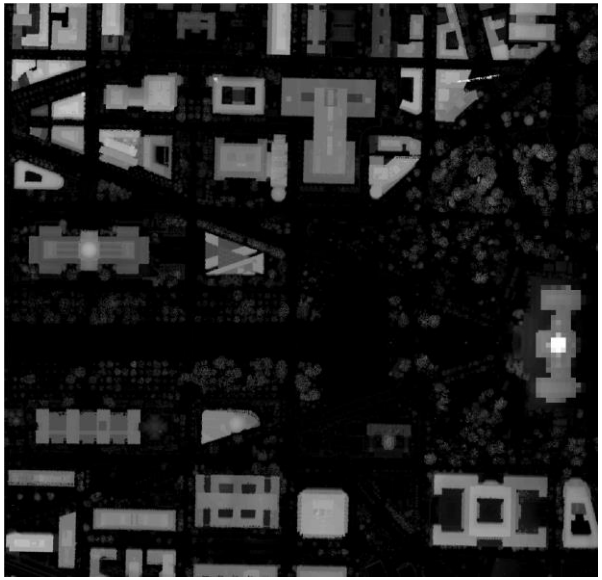
| | |
|---|---|
| 2,067 | $km^2$ total area |
| 135 billion | Pixels |
| 3.6 million | Features |

Total file size was about 88 GB.

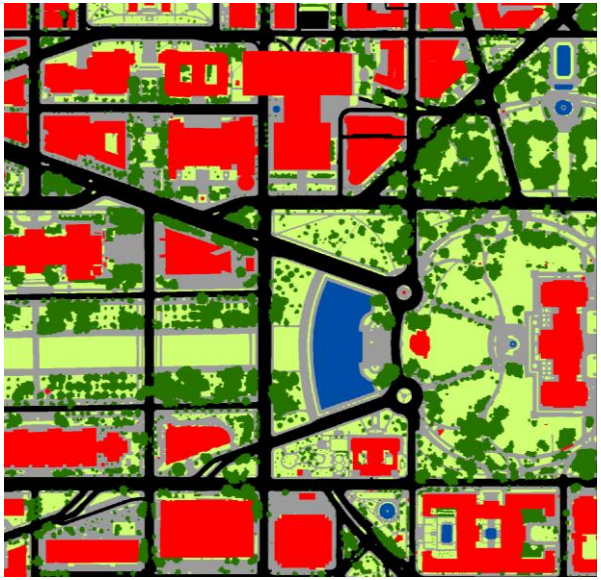
# Washington, DC Data

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).

# Power Plant Search

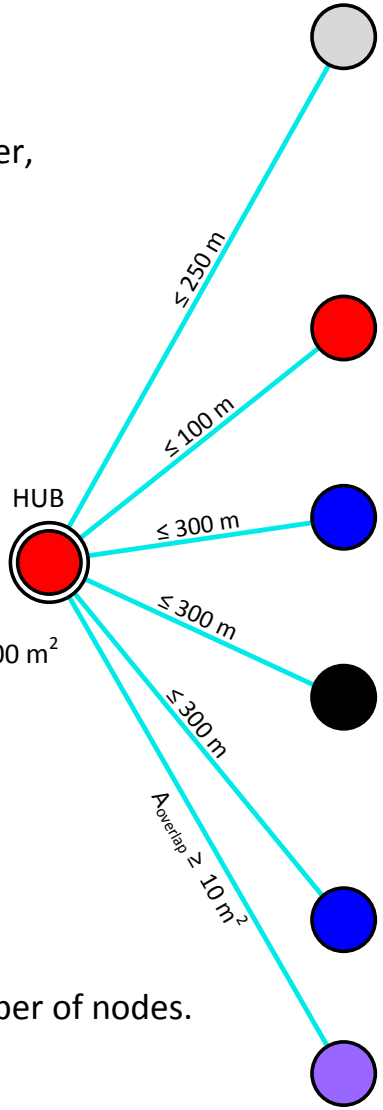
Sandia National Laboratories

Query specification:

A power plant is a heat building with a transformer, and optional storage tank, evaporation pond, coal pile, body of water.

Question semantics:
Heat Building
Transformer
Cooling Tower
Evaporation Pond
Body of Water
Coal Pile
Storage Tanks
Processing Tower
Pipe Network
⋮

Data semantics:
Building
Road
Other Paved
Grass/Shrub
Trees
Dirt
Water




**Transformer** n ∈ [1, ∞] (required)
Data: Other Paved
Chunk (15 × 15)
A ≥ 5,000 m²
Eccentricity ≤ 3.5
12 m ≤ $h_{max}$ ≤ 35 m
At least 5 m² above 10 m

**Tank** n ∈ [0, 6] (optional, limited)
Data: Building
Circular
90 m² ≤ A ≤ 3,000 m²

**Evaporation Pond** n ∈ [0, ∞] (optional)
Data: Water
1,500 m² ≤ A ≤ 19,500 m²
Eccentricity ≤ 6.0

**Coal Pile** n ∈ [0, ∞] (optional)
Data: Dirt
A ≥ 30,000 m²
Eccentricity ≤ 3.0
$[R\ G\ B]_{max}$ ≤ 0.4

**Body of Water** n ∈ [0, ∞] (optional)
Data: Water
A ≥ 20,000 m²

**Constructed** n ∈ [0, 0] (forbidden)
Data: Building
A ≥ 100 m²
New, Extended, Changed

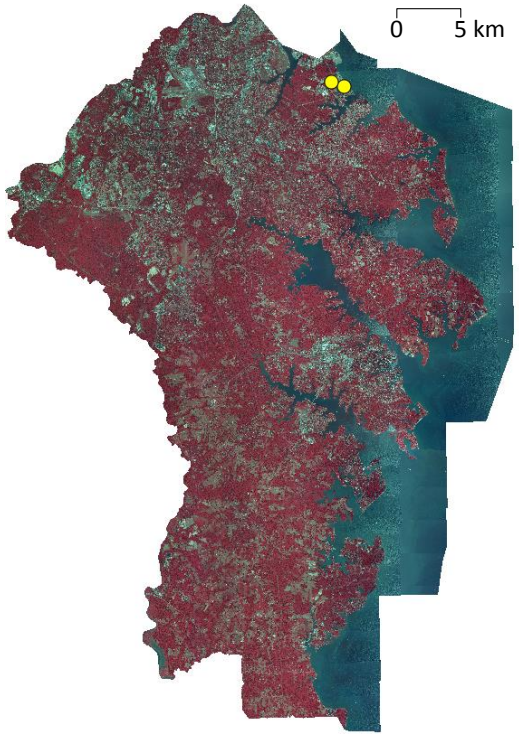
Star graph search.

Sort matches by number of nodes.

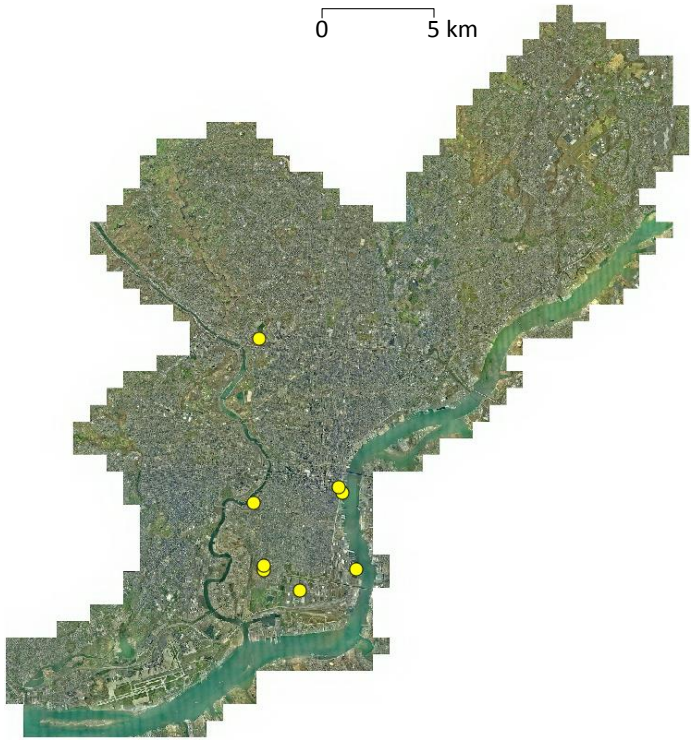
Note: "Exists Now" is omitted for clarity.

# Power Plant Search

Search results:




Anne Arundel County, MD




Philadelphia, PA


Washington, DC

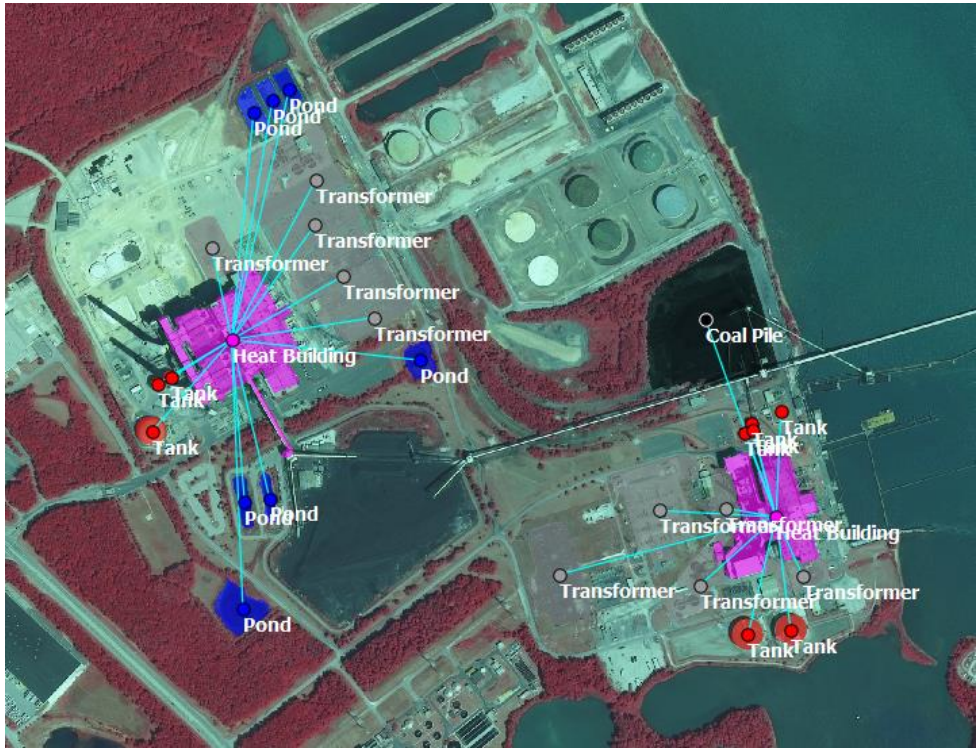
Input:

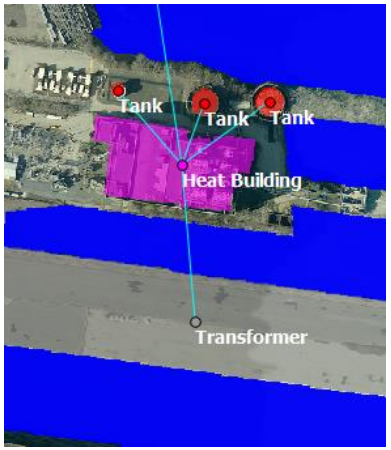
| | |
|---|---|
| 2,067 | $km^2$ total area |
| 135 billion | Pixels |
| 3.6 million | Features |

Output:

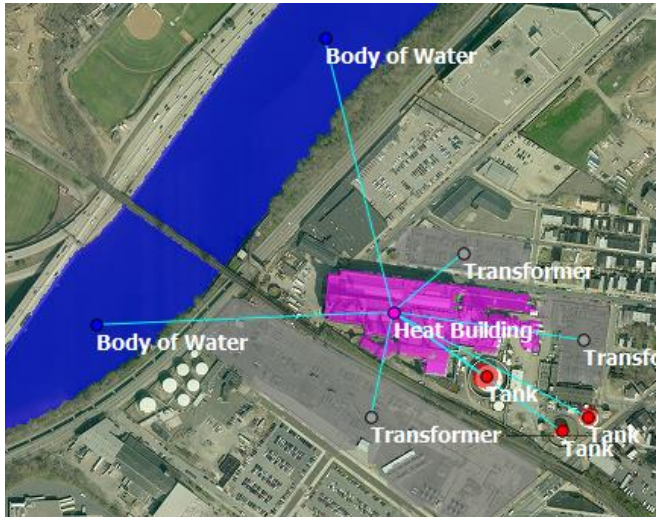
| | |
|---|---|
| 6 | True positives |
| 9 | False positives |
| 2 | False negatives |

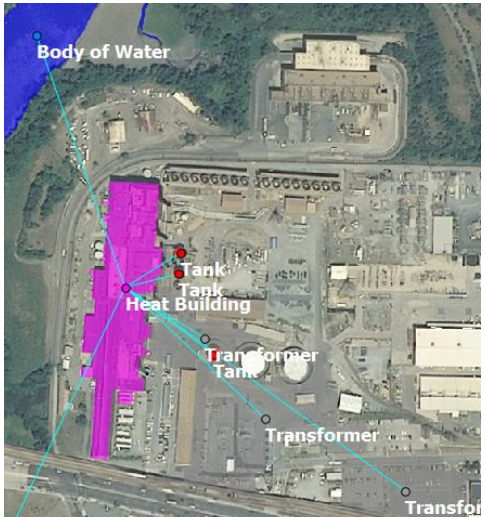
# Power Plant Results: True Positives



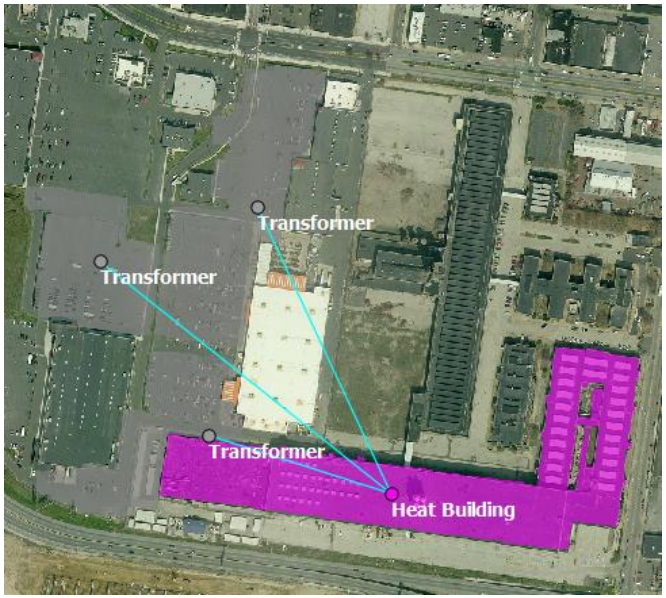



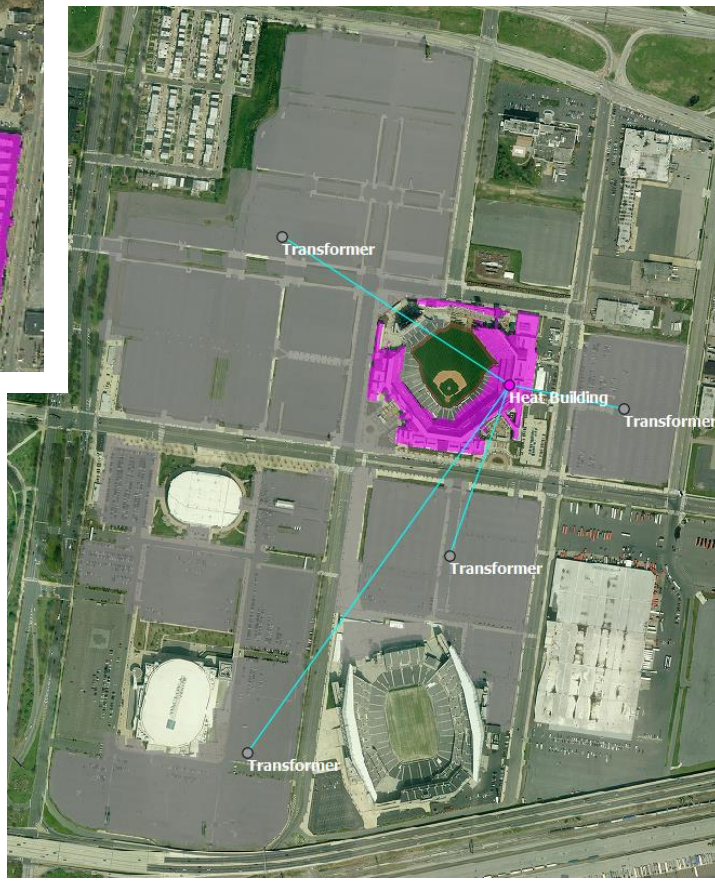

Note that power plants were found despite several land cover classification errors.

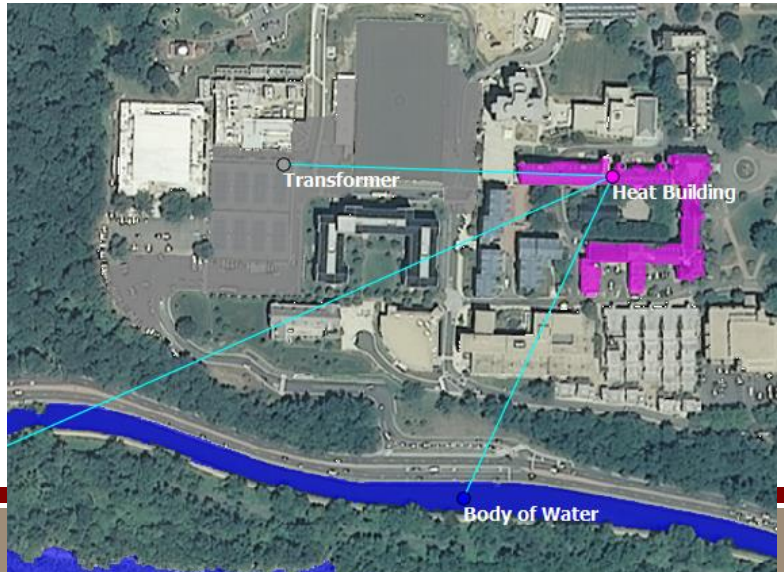



# Power Plant Results: False Positives

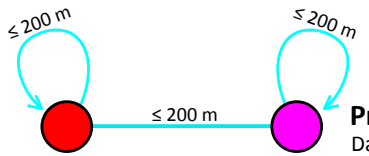



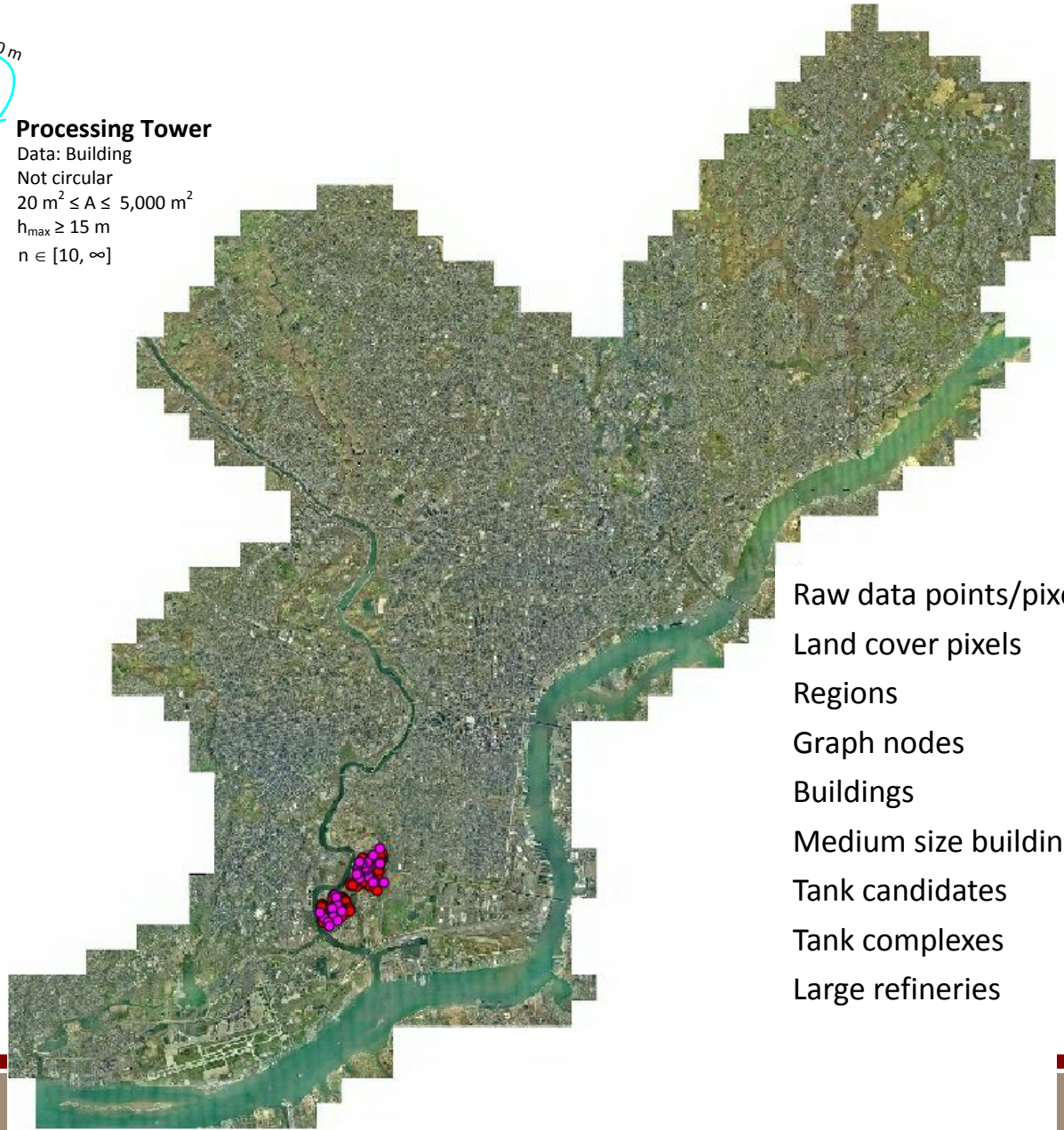

A better transformer filter would eliminate these.

# Large Refinery Search




**Tank**
Data: Building
Circular
150 $m^2 \leq A \leq$ 3,000 $m^2$
$n \in [10, \infty]$

**Processing Tower**
Data: Building
Not circular
20 $m^2 \leq A \leq$ 5,000 $m^2$
$h_{max} \geq 15$ m
$n \in [10, \infty]$



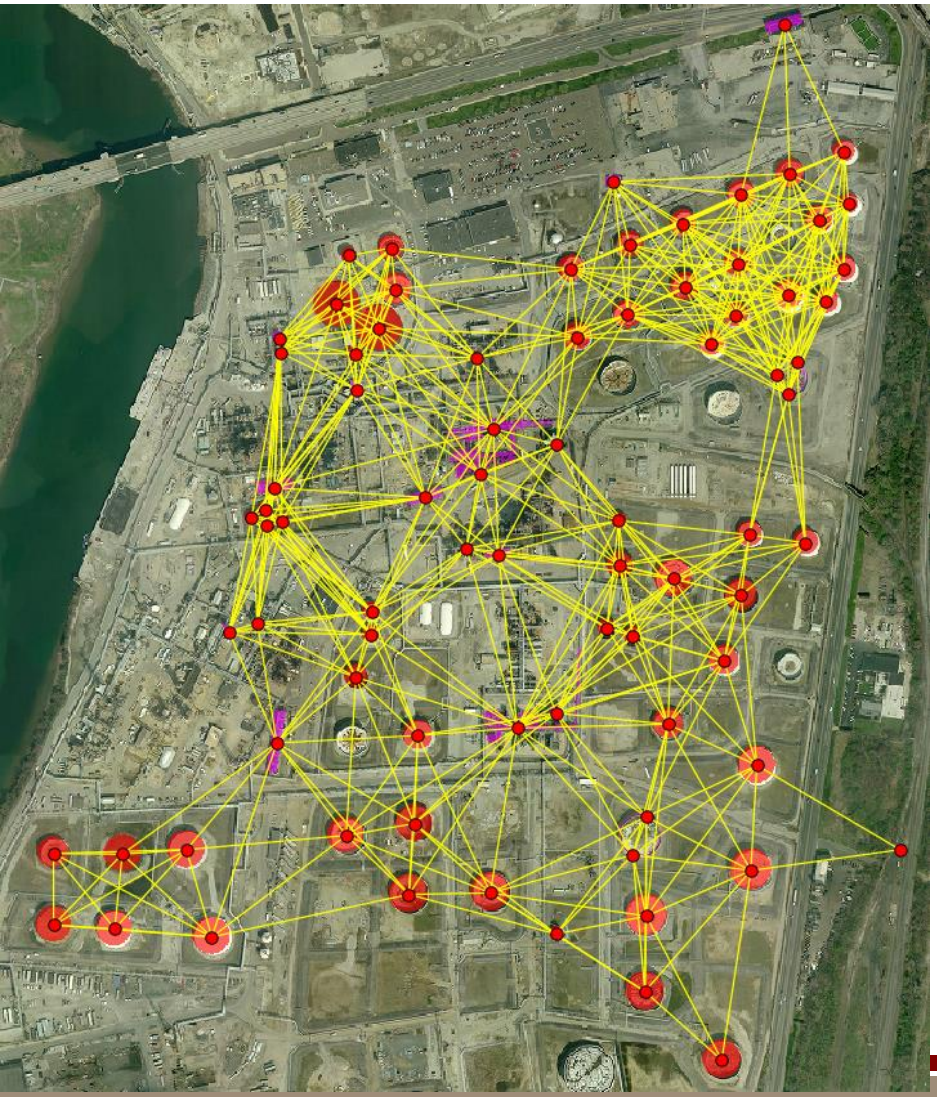
0 False positives
0 False negatives

| | |
|---|---|
| Raw data points/pixels | 101,495,378,523 |
| Land cover pixels | 8,981,933,044 |
| Regions | 1,133,822 |
| Graph nodes | 1,133,822 |
| Buildings | 154,062 |
| Medium size buildings | 87,170 |
| Tank candidates | 371 |
| Tank complexes | 28 |
| Large refineries | 2 |

# Large Refinery Search

Sandia National Laboratories

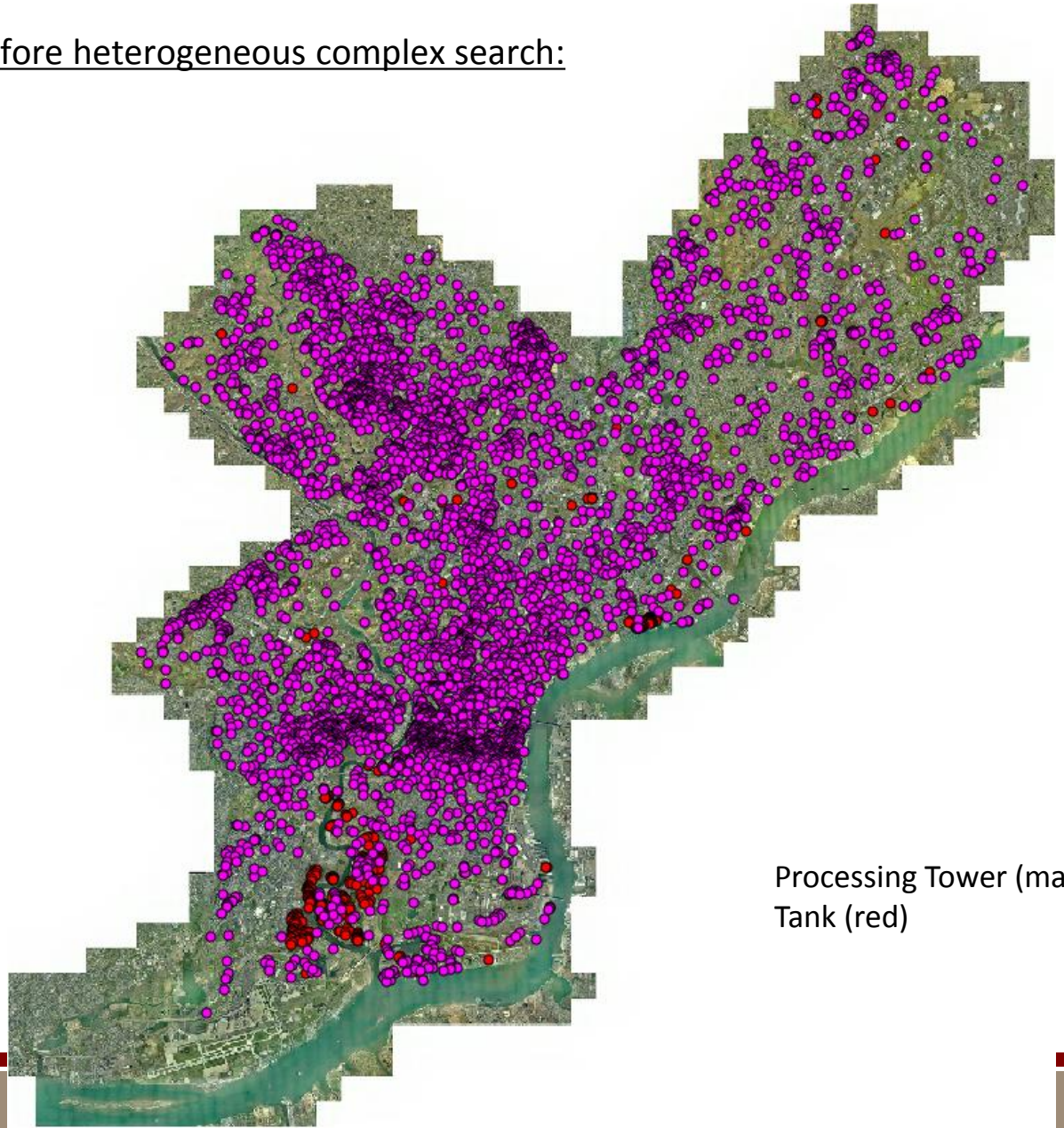
Refineries found:



# Large Refinery Search

Sandia National Laboratories

SearchGraph, before heterogeneous complex search:



| | |
|---|---|
| Processing Tower (magenta) | 4,909 |
| Tank (red) | 371 |

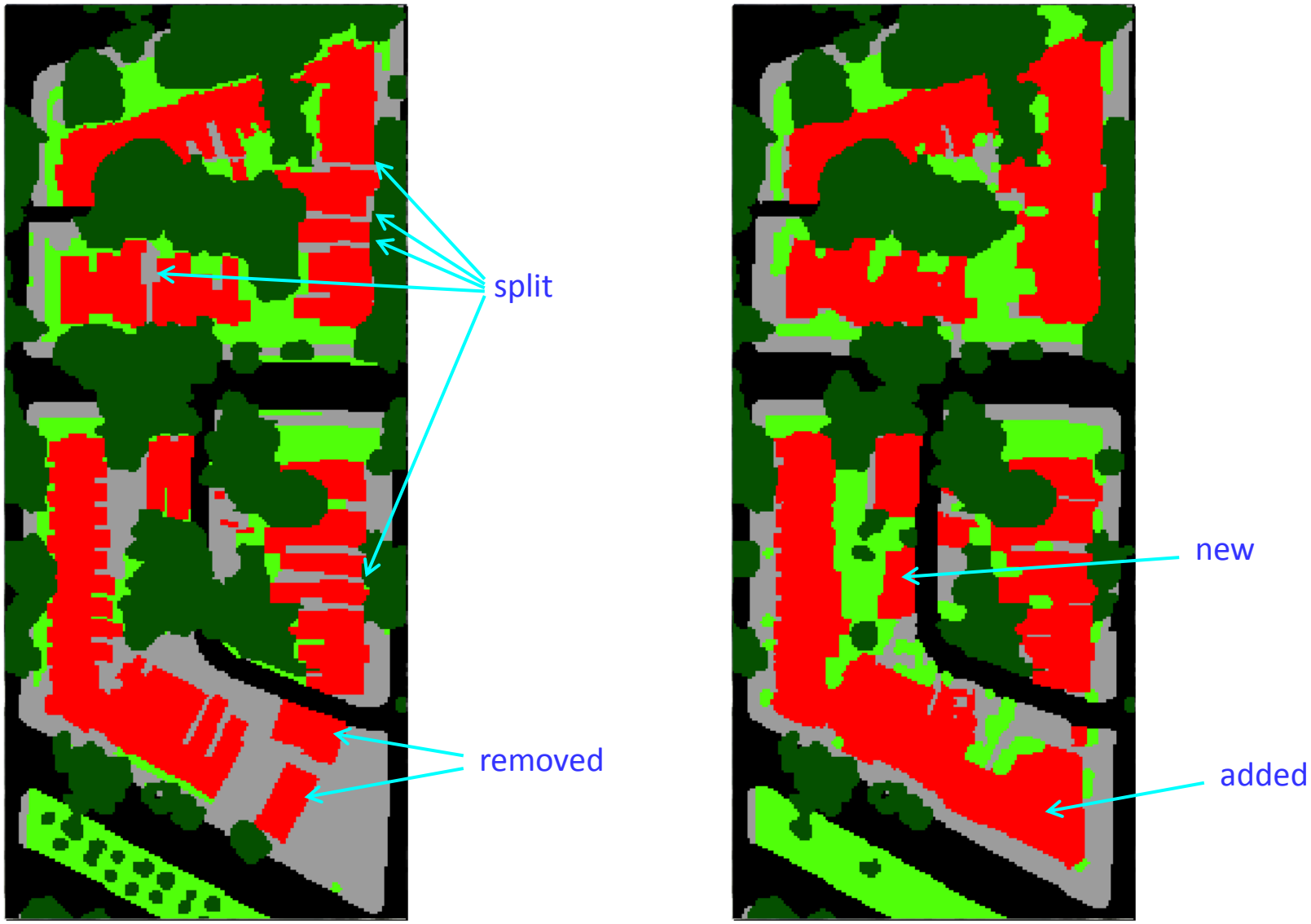
# Simple Change Example

Sandia National Laboratories

Context:

# Simple Change Example

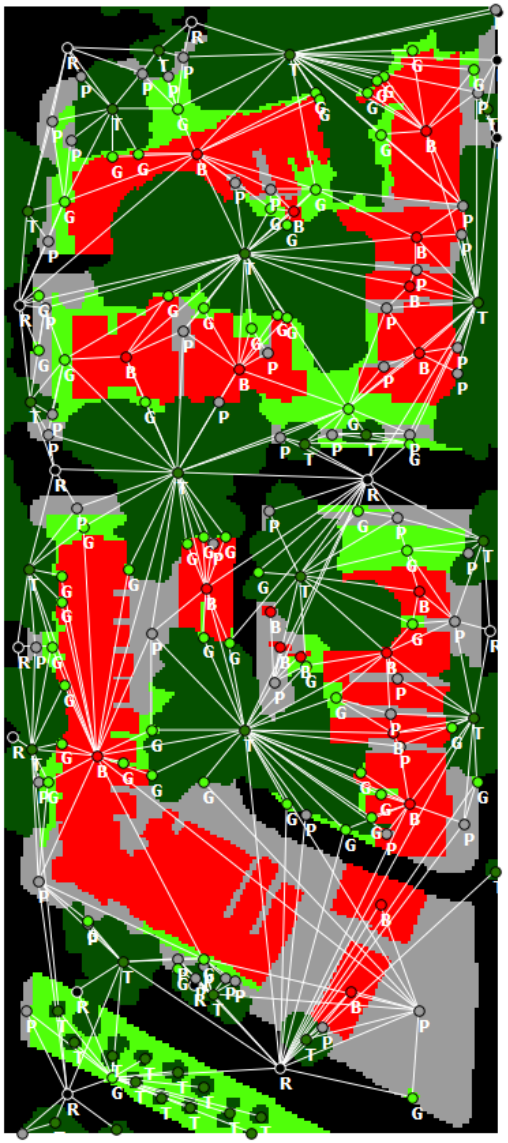
Input land cover:



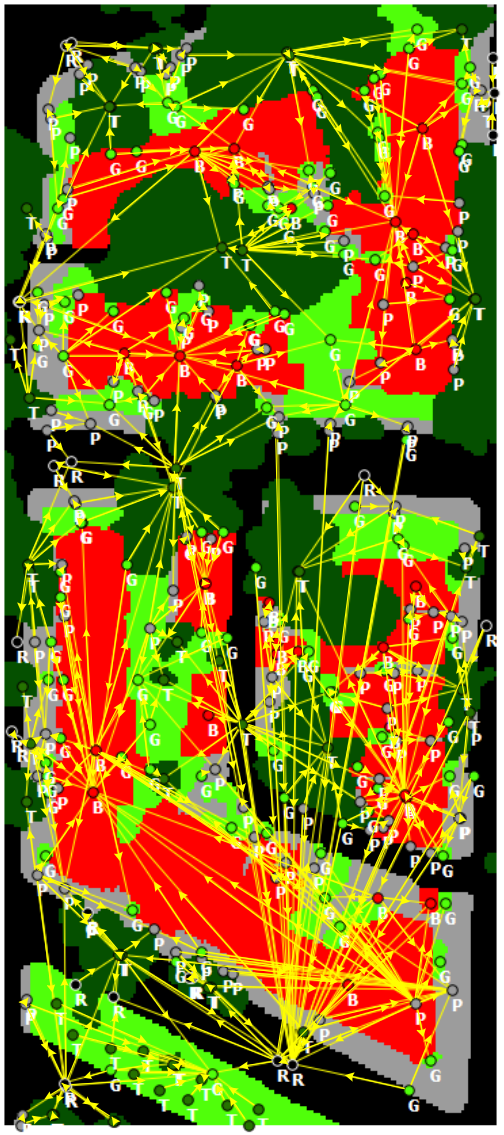
# Graph-Based Analysis

342 nodes
808 adjacency edges
559 change edges
0 distance edges

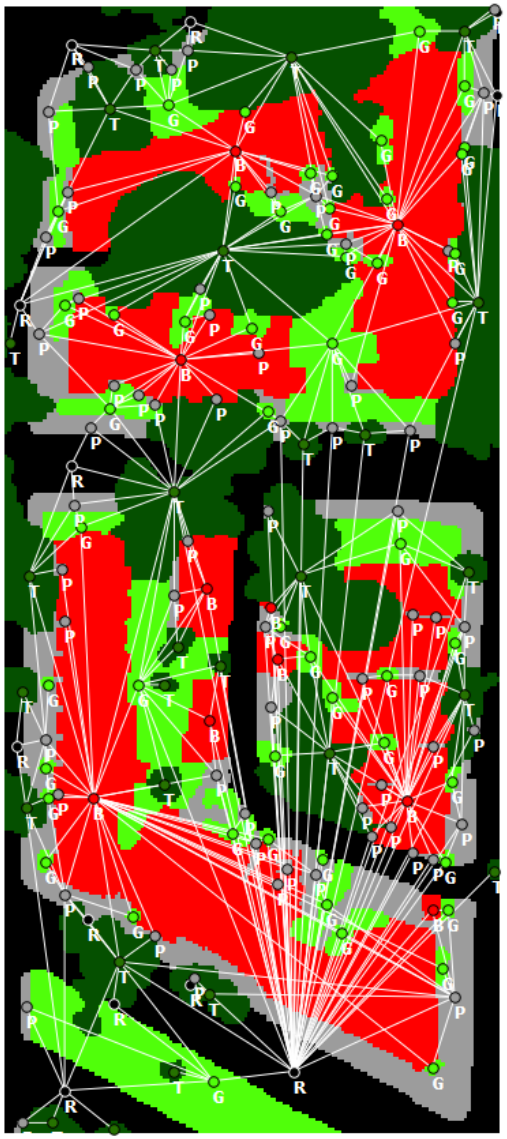
Geospatial-temporal graph:



2006 Adjacency



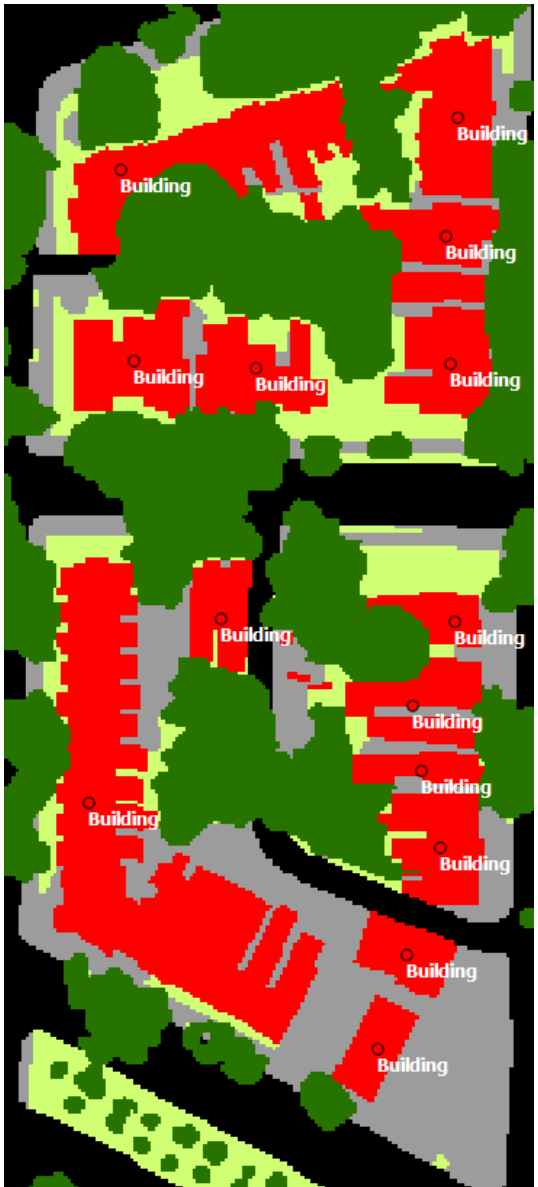
2006 → 2011 Change



2011 Adjacency

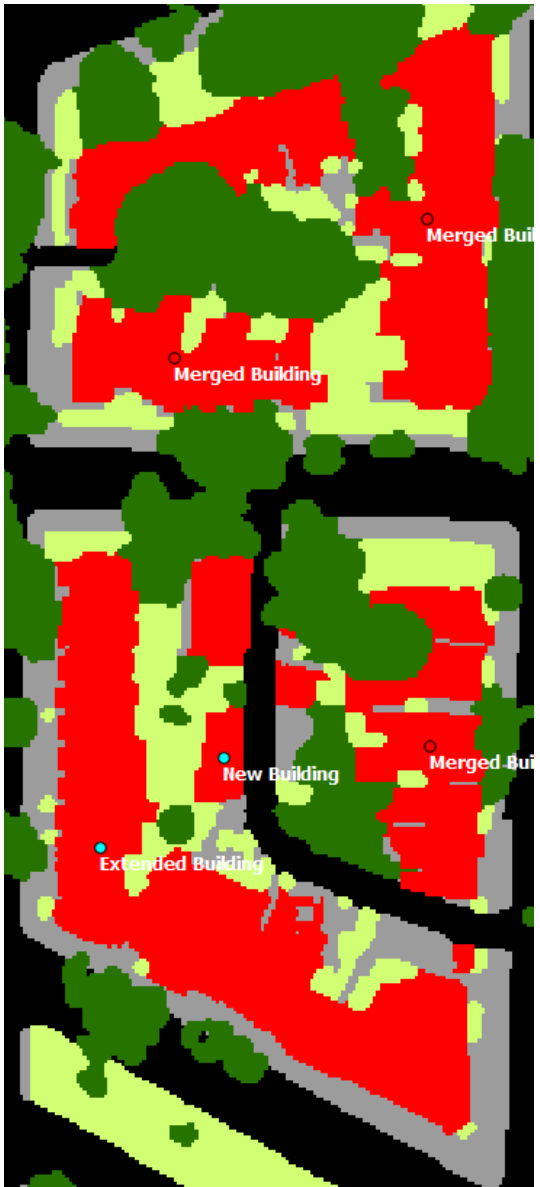
# Simple Change Example – Graph-Based

Before graph-based change analysis:





Graph-based diagnosis:

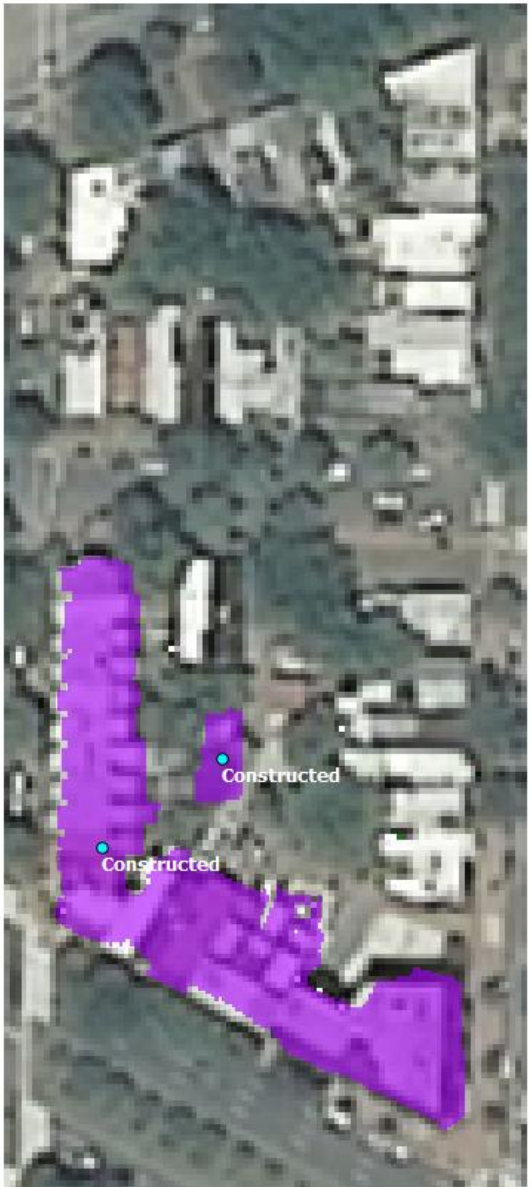


# Differentiating Important Change

Significant change diagnosis:

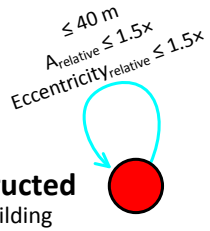


Shown over latest image:



# New Complexes

Seek complexes of new buildings, across the entire city:




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

2006*

2011



2006*



2011

* Image from DigitalGlobe.

# Overview

- Motivation.
- Computation.
- Results.
- Discussion.

# Summary

Sandia National Laboratories

What we have shown:

- Imagery + LiDAR + GIS ⇒ Land cover model [O'Neil-Dunne 2012].
- Sequence of above ⇒ Geospatial-temporal graph.
- Given query, spatial search for power plants, refineries, high schools…
- Given query, spatial-temporal search for change, construction complexes…
- Over a wide area (2,067 $km^2$, 135 billion pixels, 3.6 million graph nodes).

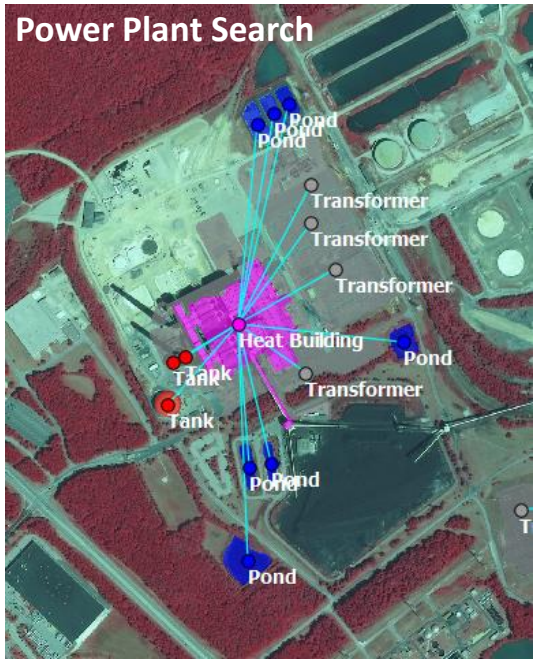
What we have NOT shown:

- Continent-scale robust image pre-processing.
- Recognition scope.
- User-friendly query construction.
- Complex multi-step change analysis.

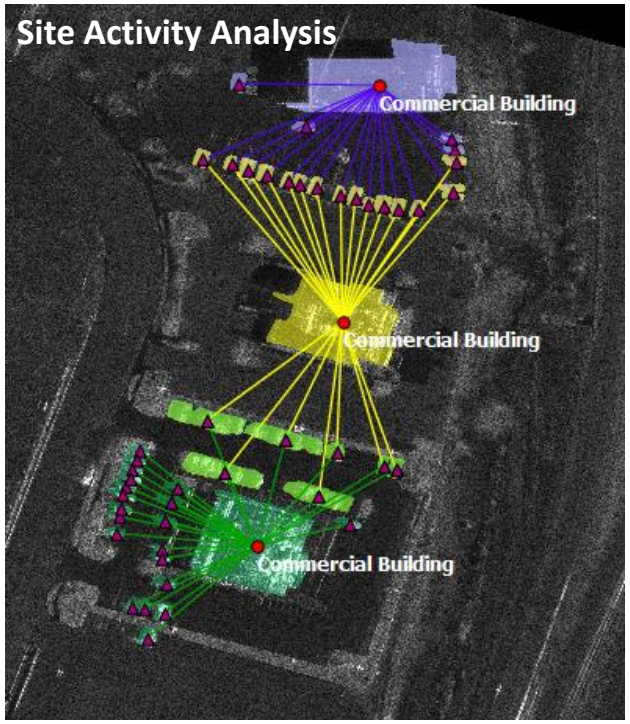
} Areas of on-going work.

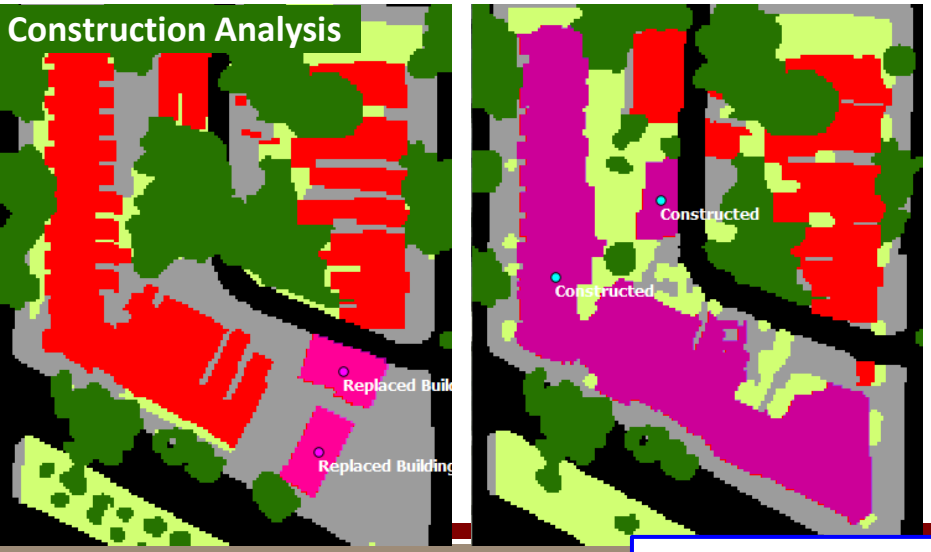
[O'Neil-Dunne 2012] O'Neil-Dunne, et al, An object-based system for LiDAR data fusion and feature extraction, Geocarto (28), pp. 227–242, 2012.

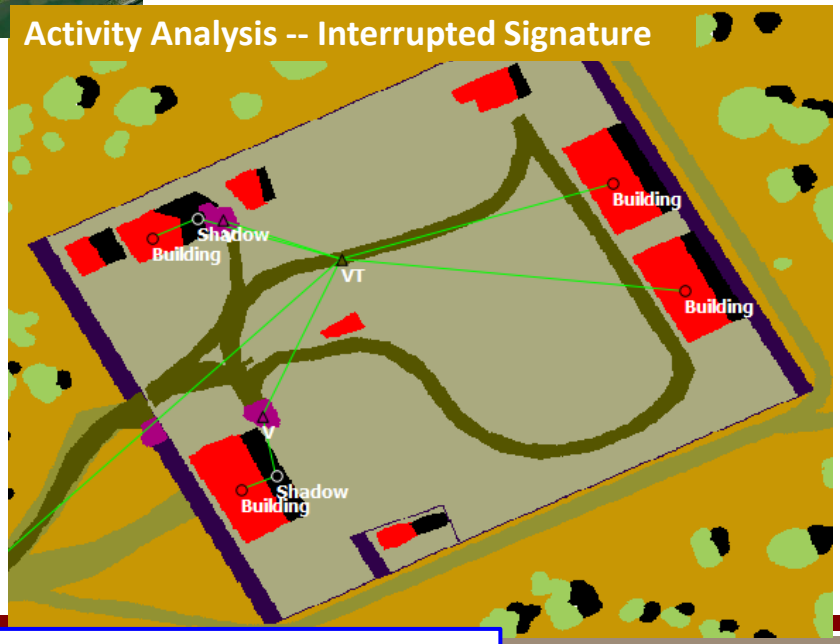
# Diversity of Problems








All of these were solved by the same code.

Sandia National Laboratories

# BACKUP SLIDES

# Discussion

Sandia National Laboratories

- Graph-based search tolerates varying topology.
- Query templates do not need to specify particular shapes, or all features.
- Search can tolerate some pre-processing errors.
- Change detection can be challenging if data methods are not consistent across time.
- Our approach is to assemble simple pieces into an effective sequence.
- We did not attempt to optimize run time.
- Our search approach exploits context, and can be viewed as analogous to context-based image preprocessing (GEOBIA), at a higher level of abstraction.

# Comparison to GEOBIA*

Same as:

- Analyzes multi-modality overhead image data.
- Includes shrink/grow operations.
- *Future: Direct output of graph tables.*

System Difference:

- Rule sets must be programmed for a new query.
- Queries with multi-constraint filtering or variable topology require difficult manual setup.
- Re-usable distance edges.
- *Quality score/sort.*

Mathematical Difference:

- Heterogeneous complex.
- Temporal graph, chronology representation.
- Explicit representation of multi-step change.
- Interrupted star algorithm.
- Infer connected tree.
- *Multi-step sequence analysis.*
- *Forward/backproject, with temporal feasibility.*
- *Path sequence analysis.*

Shortfalls:

- Shape file input.
- Particular functions.

…eCognition might output graph tables directly.

* Geospatial Object-Based Image Analysis, particularly eCognition. See GEOBIA 2014 for recent publications. Preliminary.

# Comparison to Spatial SQL Databases*

Same as:

- Superset of SQL. *(Future: Spatial SQL.)*
- R trees.
- Coarse/fine tests.
- Star-graph equivalent through complex query.

Shortfalls:

- Shape file input.
- Particular functions.

…would be resolved by converting to Spatial SQL.

System Difference:

- Direct scan from raster image.
- Lazy constraints (circularity, height, RGB, *shape scan, transformer*).
- Caching.
- Some variables, such as area, pre-computed.
- Provenance fetch.
- Re-usable distance edges.
- Temporal chronology representation/history.
- Chunking to provide separation of joined spatial regions into semantic components.
- Some queries conceptually simpler (?).
- *Quality score/sort.*

Mathematical Difference:

- Transitive closure.
- Heterogeneous complex.
- Temporal graph, automatic change analysis.
- Interrupted star algorithm.
- Infer connected tree.
- *Multi-step sequence analysis.*
- *Forward/backproject, with temporal feasibility.*
- *Path sequence analysis.*
- Heterogeneous relation types: Distance, change, adjacency , spatiotemporal constraints, *is-a, component-of.*
- *Layers of abstraction.*

* Preliminary. Further, graph databases, semantic graph databases not addressed here.

# Key Data Structures

Sandia National Laboratories

## StoredGraph

- Accumulates all observed data.
- Large, persistent.
- Semantics match data: Building, Trees, Grass_Shrub, Dirt, Road, Other_Paved, Water, …[1]

## GeoQuestion

- Defines semantics of interest.
  Example: "Where are high schools?" ⇒ Classroom_Building, Football_Field, Parking_Lot.
- Defines signature: items and relationships.
- Defines desired search procedure.

## SearchGraph

- Parts of the StoredGraph relevant to the question.
- Smaller, with full topology.
- Semantics match question: Classroom_Building, Football_Field, Parking_Lot.[1]

## Match

- A set of SearchGraph nodes and edges, satisfying the signature.

1. Contrast with:
   - Kolas, Dean, and Hebeler, "Geospatial semantic web: architecture of ontologies," IEEE Aerospace Conference, page 10, 2006.
   - Arpinar, et al, "Geospatial ontology development and semantic analytics," Transactions in GIS, 10(4):551–575, 2006.
   - Ashish and Sheth, eds, Geospatial Semantics and the Semantic Web, Springer, 2011.

# StoredGraph Representation

Sandia National Laboratories

Data semantics.

Node types:

- Region, Durable (building, tree,…).
- Region, Ephemeral (cloud, tire tracks,...).
- Point, Durable (hospital name and address,…).
- Point, Ephemeral (concert date and location,…).
- Ensemble (high school complex,…). [PENDING]

Attributes:

- Region: class, area, eccentricity, orientation,…
- Point: specialized semantics. [FULL-FEATURED PENDING]
- All have back-pointers to original data.

Time attributes:

- Durable: $(t_{last.absent}, t_{first.seen}]$ $[t_{last.seen}, t_{first.absent})$.
- Ephemeral: t
- Also parameter history (durable only).

Edge types:

- Adjacency.
- Distance.
- Change.
- Component-Of. [PENDING]

# GeoQuestion Representation

Sandia National Laboratories

GeoQuestion

- Semantics: Node, edge "roles."
- SearchGraph node specs.
- SearchGraph edge specs.

} These define the SearchGraph.

- Search method type and parameters.
- Postprocessing options.
- Quality spec (not implemented).

The design of an effective user interface to design queries remains an important topic for PANTHER research.

# GeoQuestion Details

Sandia National Laboratories

GeoQuestion

- Bookkeeping: Name, directory.
- Node role map.
- Edge role map.
- List of SearchGraphNodeSpecs
  - List of LazyConstraintSpecs
- List of SearchGraphEdgeSpecs
  - List of InterNodeConstraintSpecs
- GraphSearchMethodEnum
  - Entire graph.
  - Connected components:
    - Min, max components allowed.
  - Star Graph, Interrupted Star:
    - Hub role
    - List of SpokeSpecs (spoke role, min/max counts).
    - List of SpokeConstraintSpecs (spoke role, other role, list of InterNodeConstraintSpecs)
- List of PostprocessMatchSpec (calculation type, roles to consider, min/max limits).
- Quality spec (not implemented).

We do not expect analysts to program this!

The design of an effective query input method is an important PANTHER research topic.

# SearchGraph Representation

Sandia National Laboratories

SearchGraph

- Lightweight nodes.
- Can be viewed as a graph with colored nodes and edges.
- Attributes not included, but points back to original StoredGraph nodes.
- Full topology included in memory.
- Typically much smaller than StoredGraph.
- It is convenient to think of SearchGraph as subset of StoredGraph.
- But in general, StoredGraph nodes map many-to-many to SearchGraph nodes.
- Also has edges that may not exist in the StoredGraph.

# Match Representation

Match

- A set of SearchGraph nodes and edges.
- Associated post-process results (quality score, etc).

# AddToStoredGraph Summary

Sandia National Laboratories

1. Read data file name and auxiliary info.
2. Add provenance lookup pointers to StoredGraph.
3. Scan the image, adding nodes and adjacency edges to StoredGraph.
   - Perform change analysis for each node, creating change edges.

Use streaming algorithms to process very large files without exceeding memory.

# AddToStoredGraph Context

Sandia National Laboratories

Context: For each input data set:

1. Prepare the input.
2. Add auxiliary provenance files.
3. Run AddToStoredGraph program.

…Repeat as new data arrives.

# AddToStoredGraph Algorithm

Sandia National Laboratories

Input:

- Posterized image file.
  *or*
  csv table file.
- Companion file explaining time window, background information.

Algorithm (for image):

Write observation context to StoredGraph.
Scan image file lines, grouping pixels into regions. ← Streaming algorithm allows very large images to be processed without exceeding memory.
Each time a durable region* $R_i$ closes:
  Use $R_i$'s bounding box to look up candidate overlap nodes in the StoredGraph.
  Analyze whether $R_i$ is new, a repeat, or a change to a previous node.
    New $\Rightarrow$ write $R_i$ node to StoredGraph.
    Repeat $\Rightarrow$ update time history of existing StoredGraph node.
    Change $\Rightarrow$ write $R_i$ node to StoredGraph, and
      for each overlapping prior node:
        write change edge to $R_i$ and update prior node's time history.

* Ephemeral region and point nodes are simpler.

# GeoSearch Algorithm

Sandia National Laboratories

Primary steps:

1. Load data, construct GeoQuestion.
2. Construct SearchGraph.
   - Nodes
     - SQL query constraints.
     - Lazy-evaluated node constraints.
   - Edges
     - Generate node pairs.[1]
     - Evaluate inter-node constraints.
     - If passes, construct edge.
3. Graph search.
   - Graph search algorithm → set of Matches.
   - Postprocess each match, filter by result.
   - Quality scoring and sorting.*
4. Save results to database.*

* Not implemented yet.

1. Among other techniques, R-Trees are used to limit average-case complexity. See:
Guttman, "R-Trees: A Dynamic Index Structure for Spatial Searching," ACM SIGMOD 1984, p. 47.
Boost Geometry Library. http://www.boost.org (2013).

# GeoSearch Detail: Graph Search Algorithms

- **Subgraph isomorphism** – General template match (not supported).[1]
- **Connected components** – Find groups of related items.[2]
- **Star graph** – Hub and spokes, with various constraints.
- **Interrupted star** – Allows interruption of hub and/or spokes.
- **Heterogeneous complex** – Useful when there is no obvious hub.

1. Ullmann, "An algorithm for subgraph isomorphism", Journal of the ACM 23 (1): 31–42, 1976.

2. Hopcroft and Tarjan, "Efficient algorithms for graph manipulation," Communications of the ACM 16(6), 372-378, June 1973.

# GeoSearch Detail: Star-Graph Search

Sandia National Laboratories

Input:

- Hub role $r_h$.
- Spoke specs: { <role $r_j$ , $n_{j.\min}$ , $n_{j.\max}$>,… } ← Spoke $[n_{j.\min}, n_{j.\max}]$ interval enables:
  - Variable match topology (set $n_{j.\min} \neq n_{j.\max}$).
  - Allowable and preferred configurations.
  - Optional items (set $n_{j.\min} = 0$).
  - Forbidden items (set $n_{j.\max} = 0$).
- Spoke-to-spoke constraints.

Algorithm:

For each candidate hub node $h_i \in \mathcal{H}$:

Walk neighbors of $h_i$, collect candidate spoke nodes $\{s_j\}$.

Check spoke-to-spoke constraints, mark $s_j$ nodes that pass.

Discard unmarked nodes from $\{s_j\}$.

Check $[n_{j.\min}, n_{j.\max}]$ bounds:

If not met, discard hub $h_i$.

Else assemble star $\mathbf{S}_i \leftarrow h_i \cup \{s_j\}$, add $\mathbf{S}_i$ to match list $\mathcal{M}$.

endfor

Return $\mathcal{M}$.

- This algorithm solves a restricted class of problems.
- More efficient than general subgraph isomorphism.
- More flexible than subgraph isomorphism (see above).
- Narrower class of templates.
- Geospatial problems allow edge replacement.

# GeoSearch Detail: Post-Processing Algorithms

- Count nodes – Count nodes in match
  - Optional node filtering.
  - Min/max thresholds.
- Sum area – Sum node area in match, with optional filtering.
  - Optional node filtering.
  - Min/max thresholds.
- Break cycles – Removes redundant chain edges (preliminary).

# Land Cover Primitive Recognition

Sandia National Laboratories

Steps:

1. Assemble input:
   - RGB+IR imagery.
   - LiDAR data.
   - GIS road polygons.
2. Construct normalized digital surface model (nDSM).
3. Using nDSM, identify buildings and trees.
4. Using RGB+IR, identify grass, bare earth, water, impervious.
5. Using GIS data, distinguish roads from other paved.

Primitive recognition algorithm and data results due to Jarlath O'Neil-Dunne, et al, University of Vermont Spatial Analysis Lab.

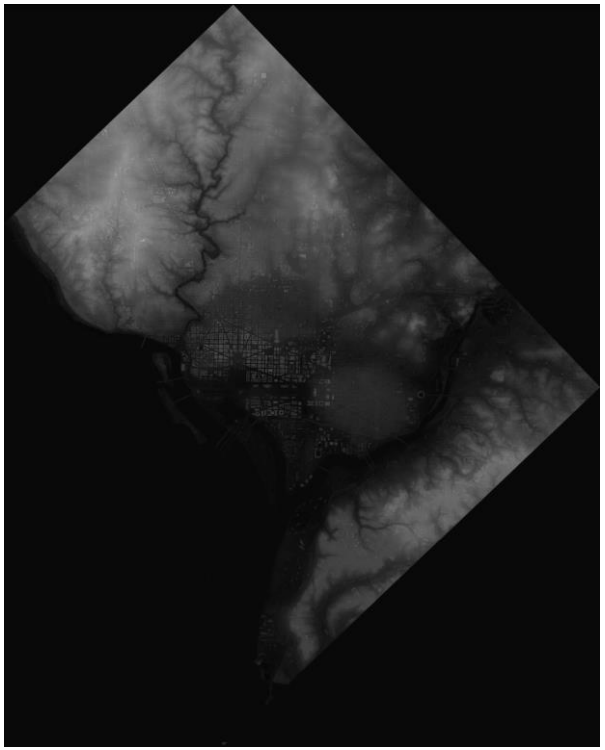
For details, see O'Neil-Dunne, et al, "An Object-Based System for LiDAR Data Fusion and Feature Extraction," Geocarto International, 2012.

# Input

Sandia National Laboratories

Washington, DC example:

4-Band (RGB+IR) Imagery

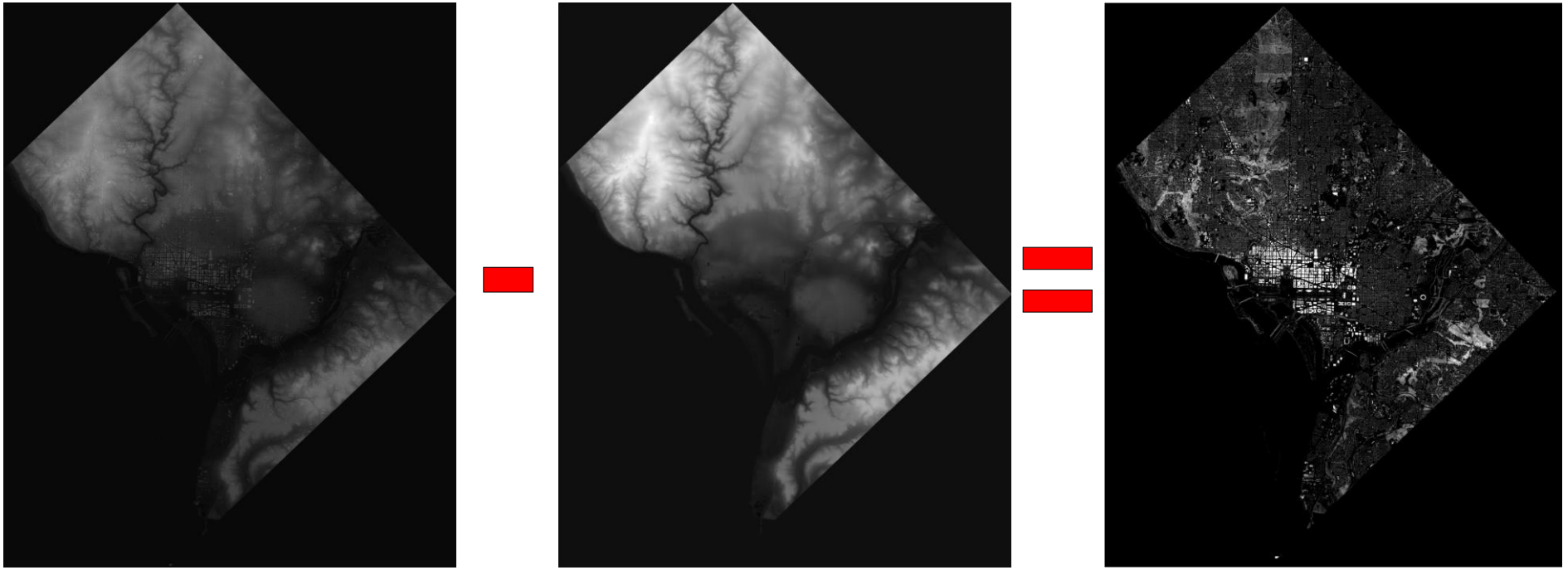


Digital Surface Model, from LiDAR

GIS Road Polygons

Data provided by Jarlath O'Neil-Dunne at the University of Vermont Spatial Analysis Lab.

# nDSM Generation



Digital Surface Model
(DSM)

Digital Elevation Model
(DEM)

Normalized Digital Surface Model
(nDSM)

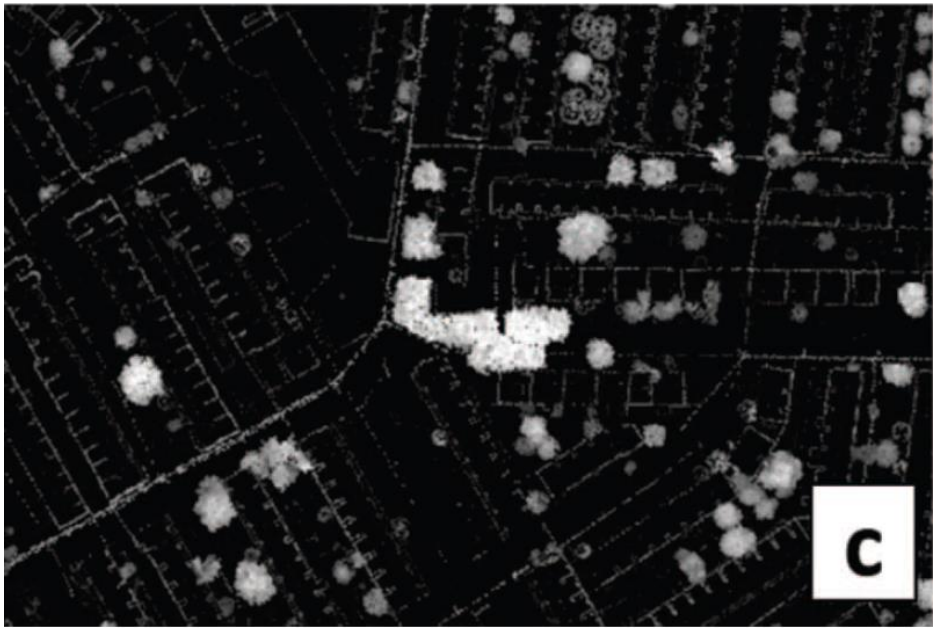
Data provided by Jarlath O'Neil-Dunne at the University of Vermont Spatial Analysis Lab.

# Building and Tree Segmentation

Sandia National Laboratories

Distinguishing buildings and trees:

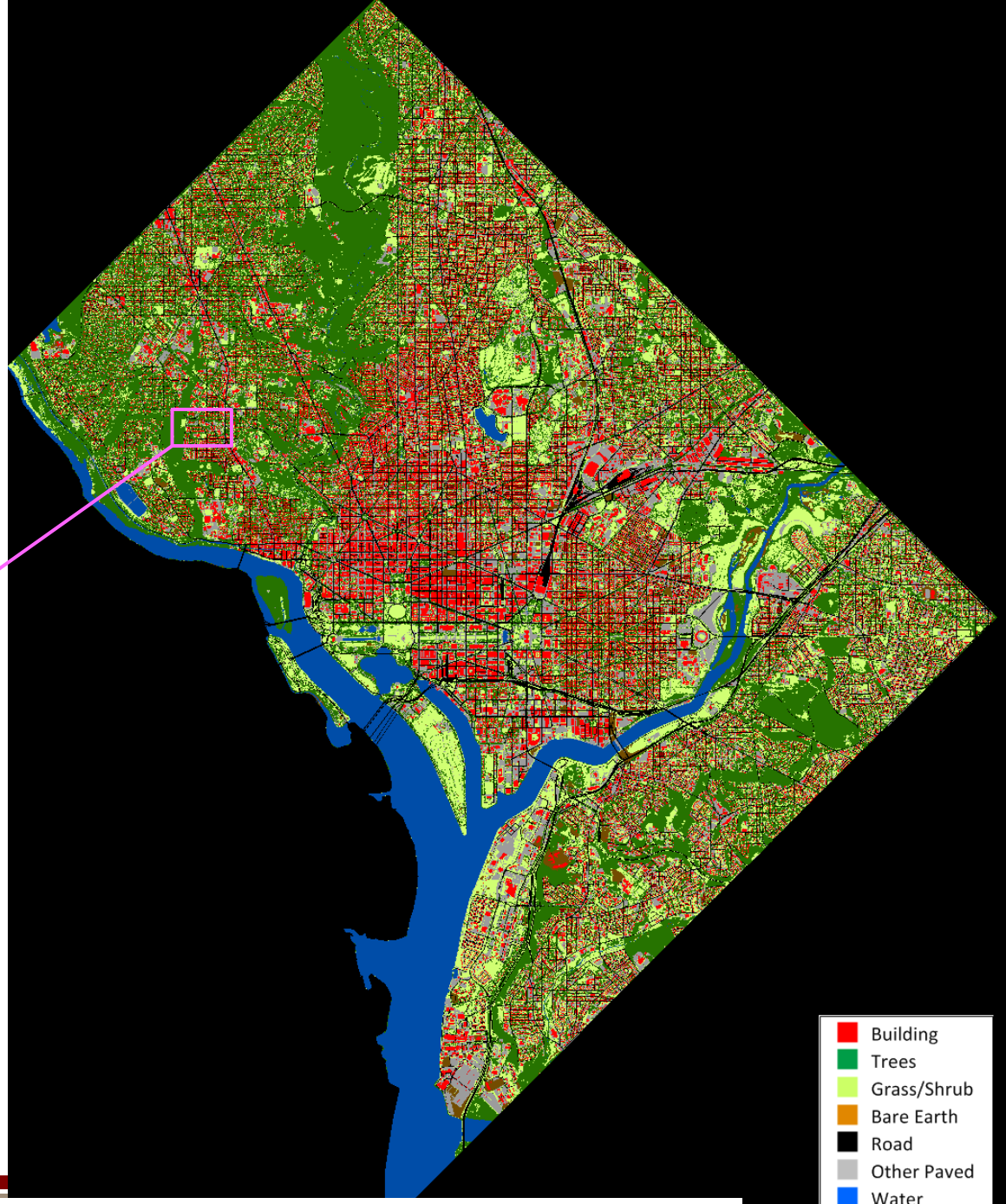

Z Height




Z Deviation

From O'Neil-Dunne, et al, “An Object-Based System for LiDAR Data Fusion and Feature Extraction,” Geocarto International, 2012.

# Output

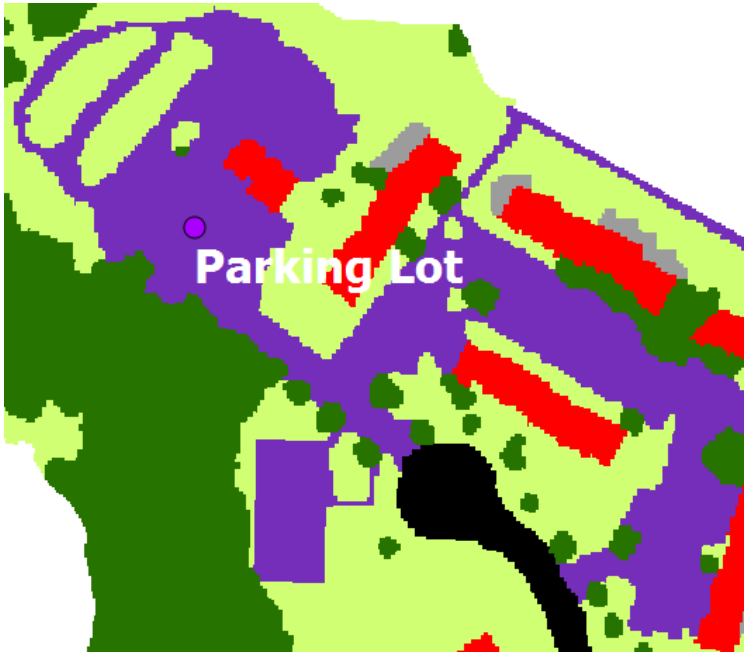



Data provided by Jarlath O'Neil-Dunne at the University of Vermont Spatial Analysis Lab.

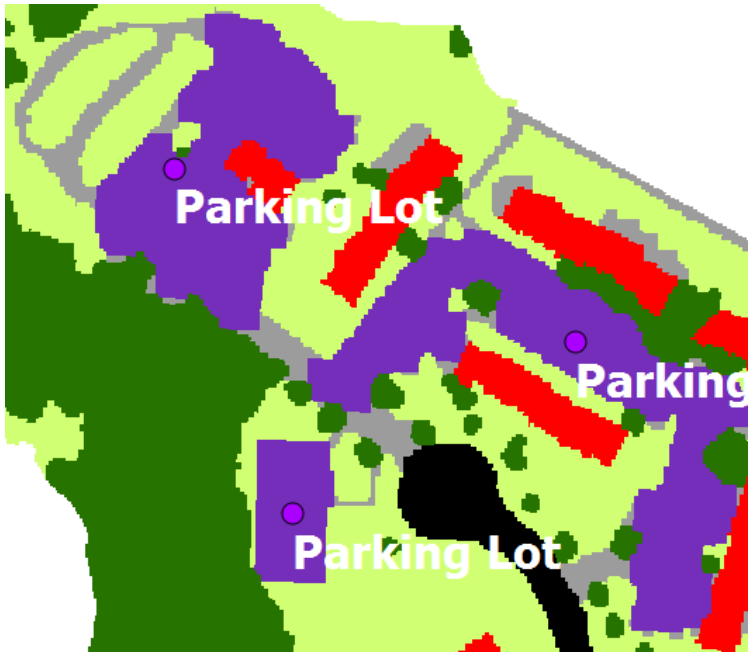
# Chunk Operation

Chunking breaks up linked regions into significant sub-regions:

Before Chunking




After Chunking




Implemented by a geometric shrink/grow operation, with controllable magnitude.

# GeoGraphy Code Implementation

Sandia National Laboratories

Implementation details:

- C++, a few scripts.
- Third-party libraries:
  - Boost (www.boost.org) - Smart pointers, Program options, R-trees, variable data types.
  - SQLite (www.sqlite.org) – Lightweight and efficient SQL database storage.
  - GDAL (www.gdal.org) – Tools for abstracting geospatial data.
- Roughly 53,000 lines of C++ code, 370 files.
- Three major applications, six supporting utility programs.
- Externalized semantics.
- Maturity level:
  - Code runs fully automatically.
  - Single processor, single thread – multi-processor speedups deferred.
  - Input format designed for programmers.
  - Output rendering capabilities support technical communication.
  - No interface built for user community.

# Data Collected

Sandia National Laboratories

GeoGraphy

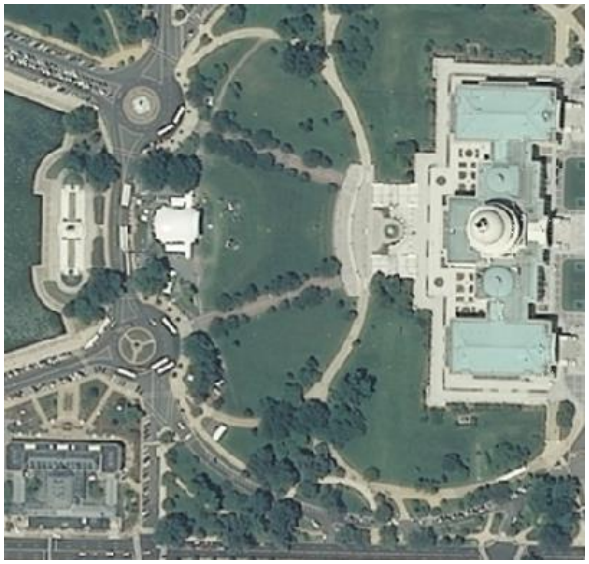
Data sets in hand:

- Single time:
  - Anne Arundel County.
  - Greater Baltimore
    (Baltimore City, Baltimore County, Howard County, Anne Arundel County).
- Two-time:
  - Washington, DC.
  - Philadelphia (2nd-generation pre-processing):
    - Improved feature recognition.
    - Railroad separated from road.
    - Shapes based on prior landcover, vary only if significant change.
- The above include provenance data (imagery, LiDAR, GIS).
- Also prepared many sample tiles of varying size, for development.

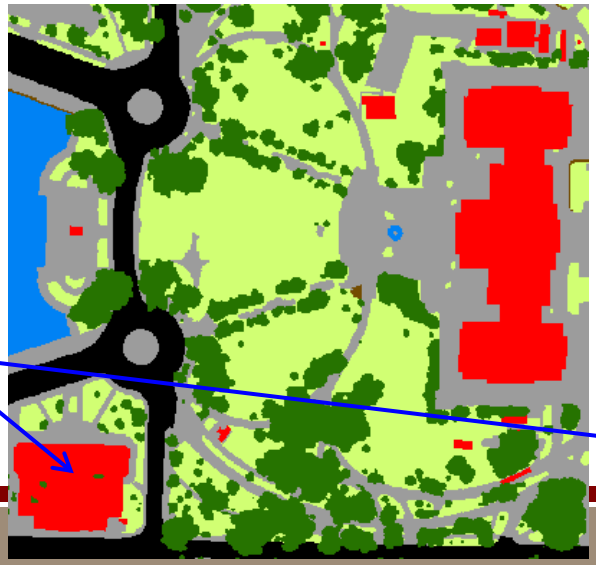
# Washington, DC Data

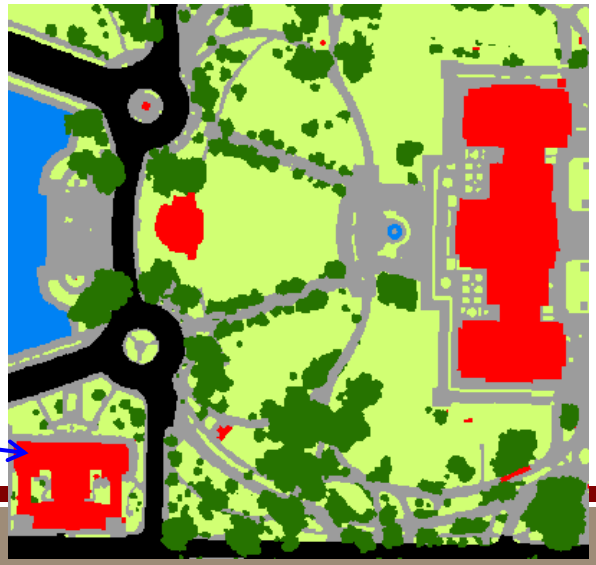
Comparing 2006 and 2011:

2006



2011





Difference in precision causes pseudo-change.

# Graph Construction

Large graphs constructed so far:

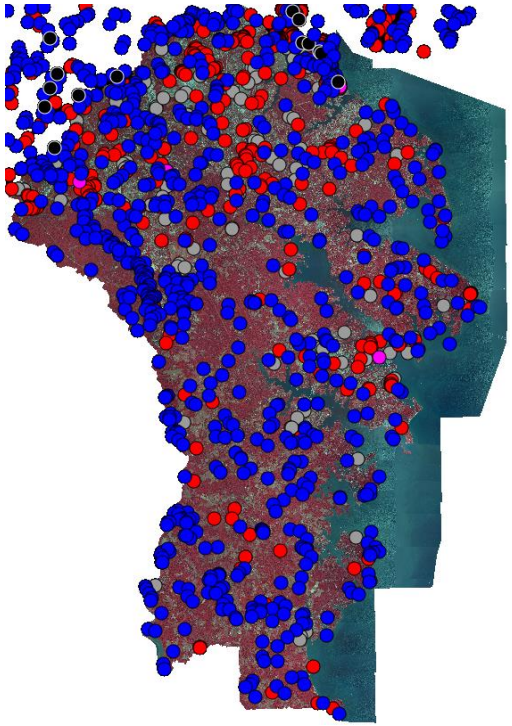
| | Landcover Pixels* | Nodes | Graph File |
|---|---|---|---|
| ▪ Single-time: | | | |
| ▪ Anne Arundel County. | 4.1 GPix | 1,204,087 | 1.6 GB |
| ▪ Greater Baltimore (sliced). | 11.3 GPix | 3,488,712 | 4.5 GB |
| ▪ Two-time: | | | |
| ▪ Washington, DC. | 1.0 GPix | 1,326,212 | 1.7 GB |
| ▪ Philadelphia. | 3.8 GPix | 1,540,909 | 2.0 GB |

* Only includes land-area pixels, not blank filler space.

# Power Plant SearchGraph

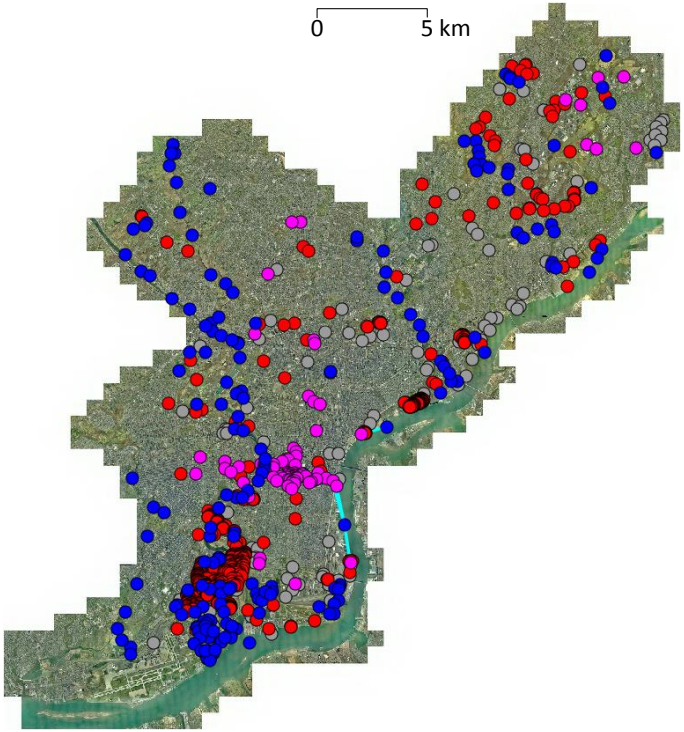
Sandia National Laboratories

SearchGraph, before star graph search algorithm:



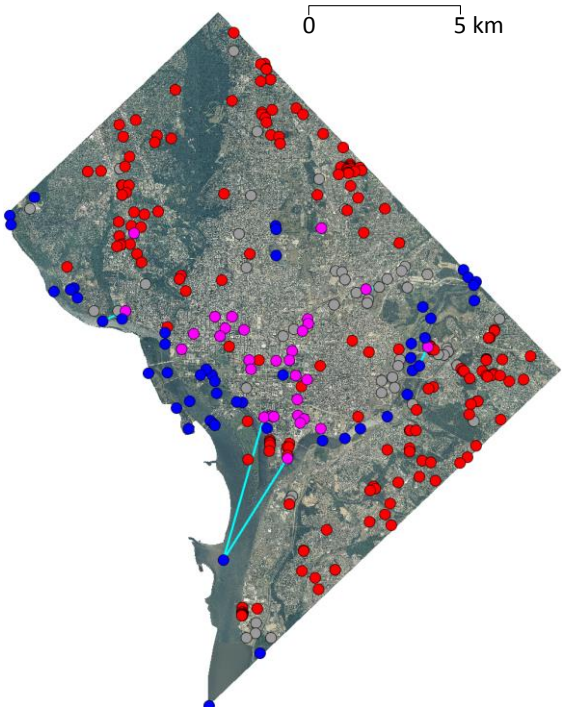
Anne Arundel County, MD

| | |
|---|---|
| Heat Building (magenta) | 5 |
| Transformer (grey) | 128 |
| Tank (red) | 500 |
| Evaporation Pond (blue) | 774 |
| Coal Pile (black) | 14 * |
| Body of Water (blue) | 62 |




Philadelphia, PA

| | |
|---|---|
| Heat Building (magenta) | 77 |
| Transformer (grey) | 104 |
| Tank (red) | 520 |
| Evaporation Pond (blue) | 138 |
| Coal Pile (black) | 0 |
| Body of Water (blue) | 53 |




Washington, DC

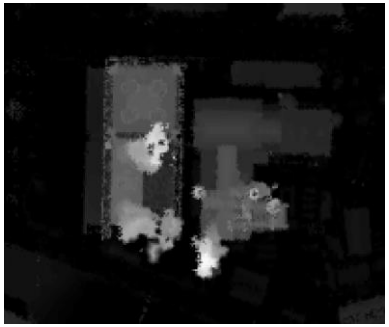
| | |
|---|---|
| Heat Building (magenta) | 31 |
| Transformer (grey) | 52 |
| Tank (red) | 174 |
| Evaporation Pond (blue) | 20 |
| Coal Pile (black) | 0 |
| Body of Water (blue) | 31 |

(Rejected construction nodes not shown.)

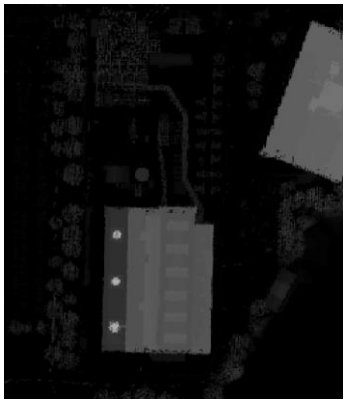
# Individual Heat Buildings – Correct

Sandia National Laboratories

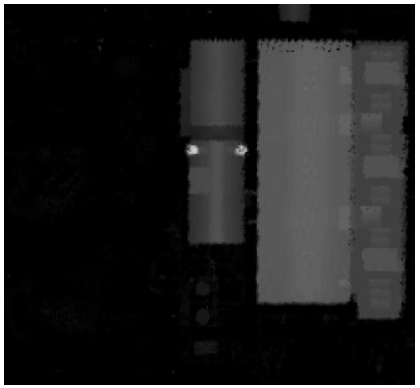
A: Capitol Power Plant




B: Buzzard Point Power Plant

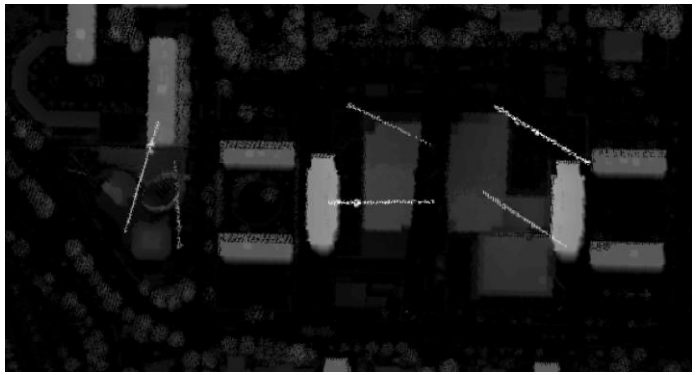



C: Surprise Power Plant

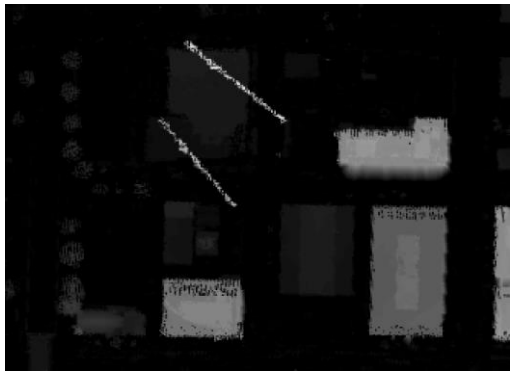


# Individual Heat Buildings – False Positive

D



E



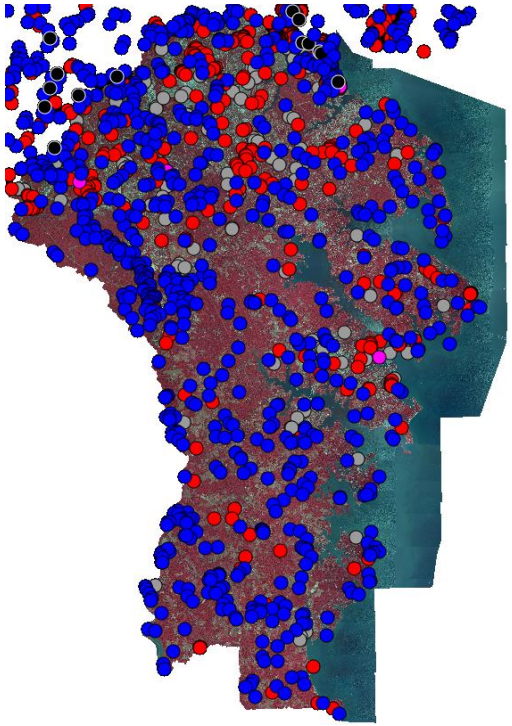
Construction cranes are interpreted as thin tall building components (like exhaust stacks).

Note: This reinforces the importance of providing provenance back-pointers to the user.

# Power Plant SearchGraph

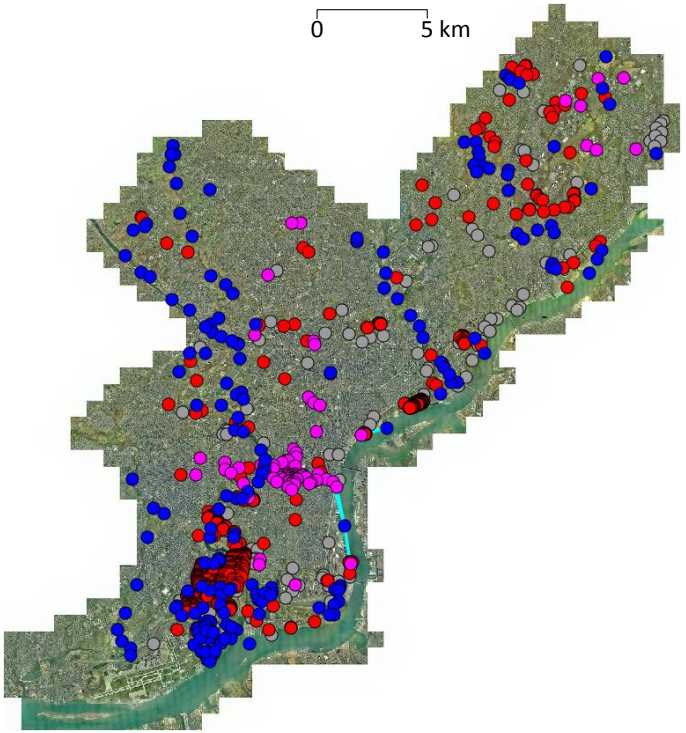
Sandia National Laboratories

SearchGraph, before star graph search algorithm:



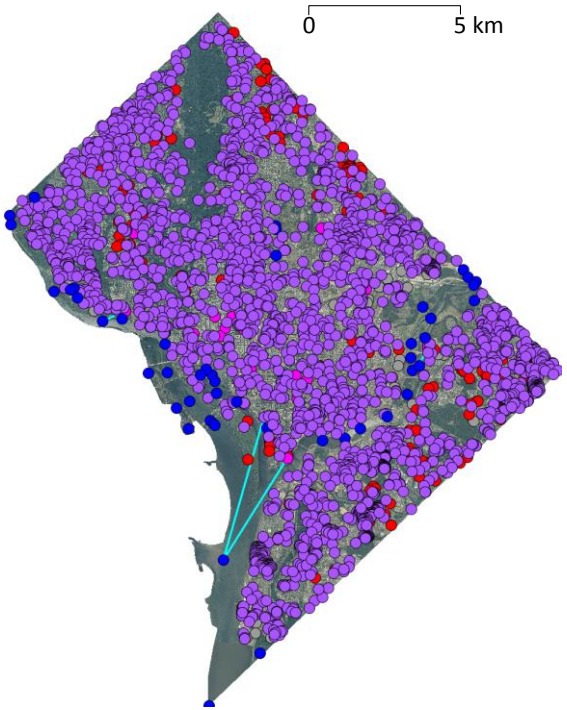
Anne Arundel County, MD

| | |
|---|---|
| Heat Building (magenta) | 5 |
| Transformer (grey) | 128 |
| Tank (red) | 500 |
| Evaporation Pond (blue) | 774 |
| Coal Pile (black) | 14 * |
| Body of Water (blue) | 62 |
| Constructed (purple) | N/A |




Philadelphia, PA

| | |
|---|---|
| Heat Building (magenta) | 77 |
| Transformer (grey) | 104 |
| Tank (red) | 520 |
| Evaporation Pond (blue) | 138 |
| Coal Pile (black) | 0 |
| Body of Water (blue) | 53 |
| Constructed (purple) | N/A |




Washington, DC

| | |
|---|---|
| Heat Building (magenta) | 31 |
| Transformer (grey) | 52 |
| Tank (red) | 174 |
| Evaporation Pond (blue) | 20 |
| Coal Pile (black) | 0 |
| Body of Water (blue) | 31 |
| Constructed (purple) | 5,979 |

* Some coal piles due to image clipping.

# A Predicted False Negative

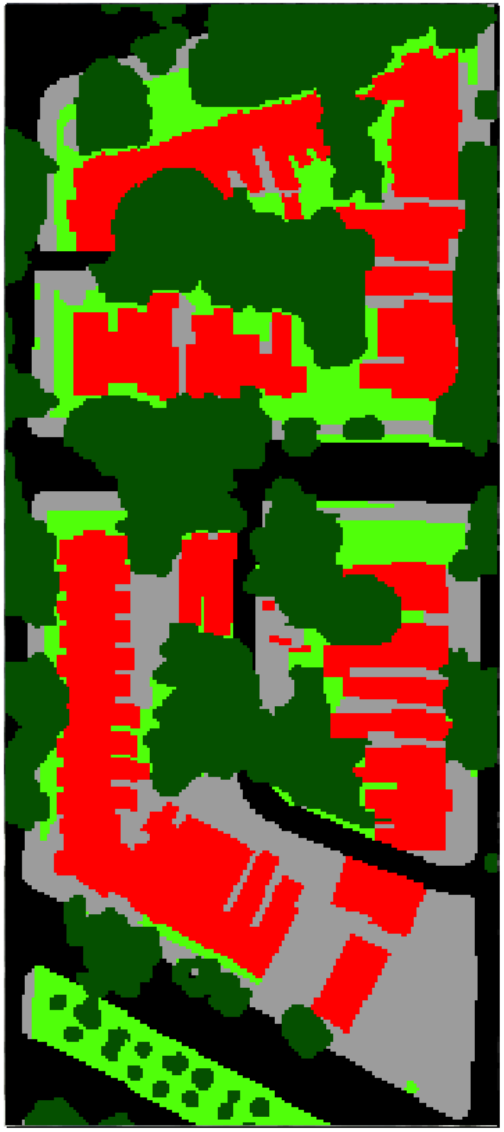
Sandia National Laboratories

If we searched for medium-to-small refineries, this would be missed:

The pipe network caused tanks and processing towers to be linked into one large region. Redefining the query to utilize chunking is likely to solve this.

# Simple Change Example

Resulting land cover:

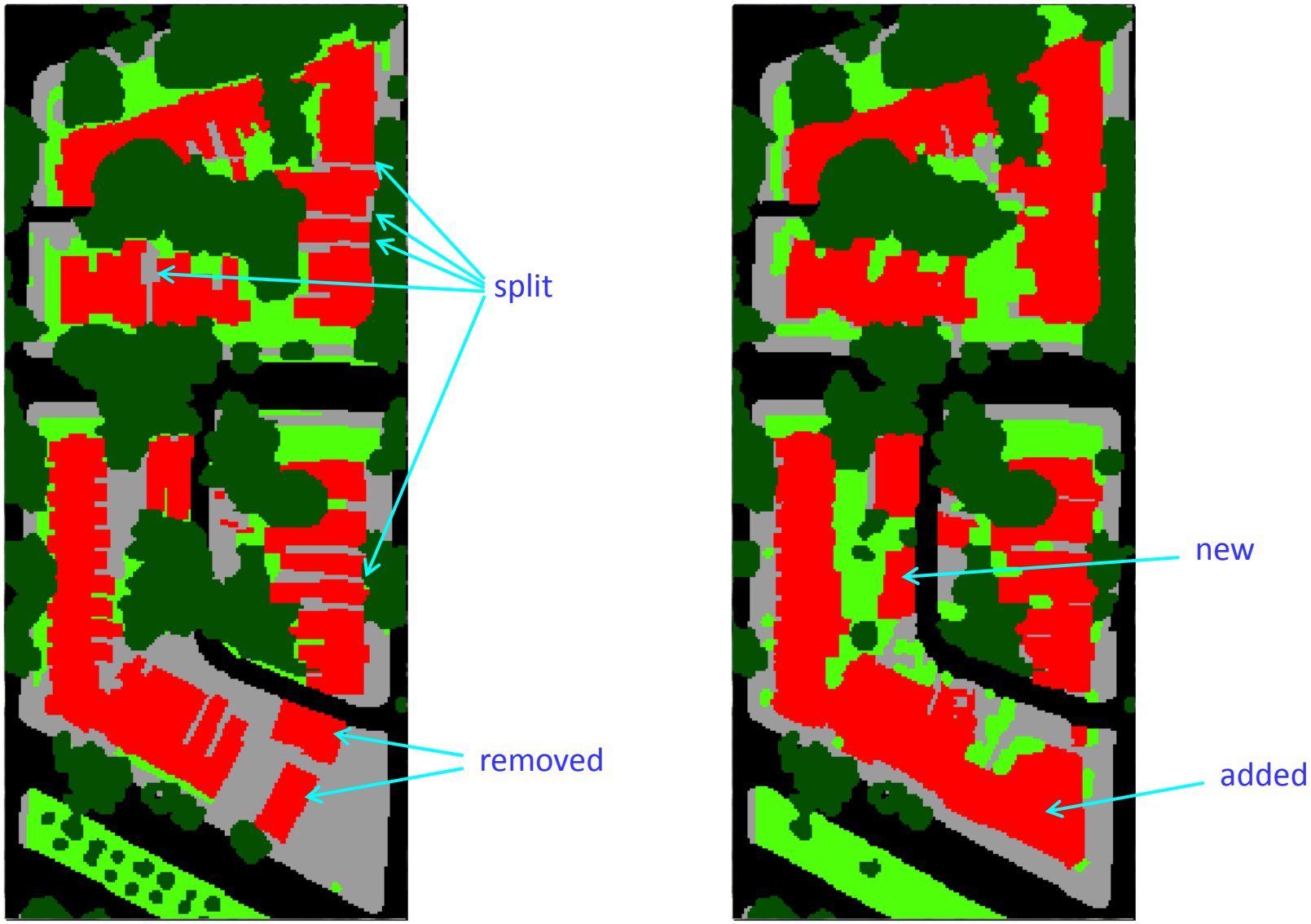


2006

2011

# Geometric Change Analysis

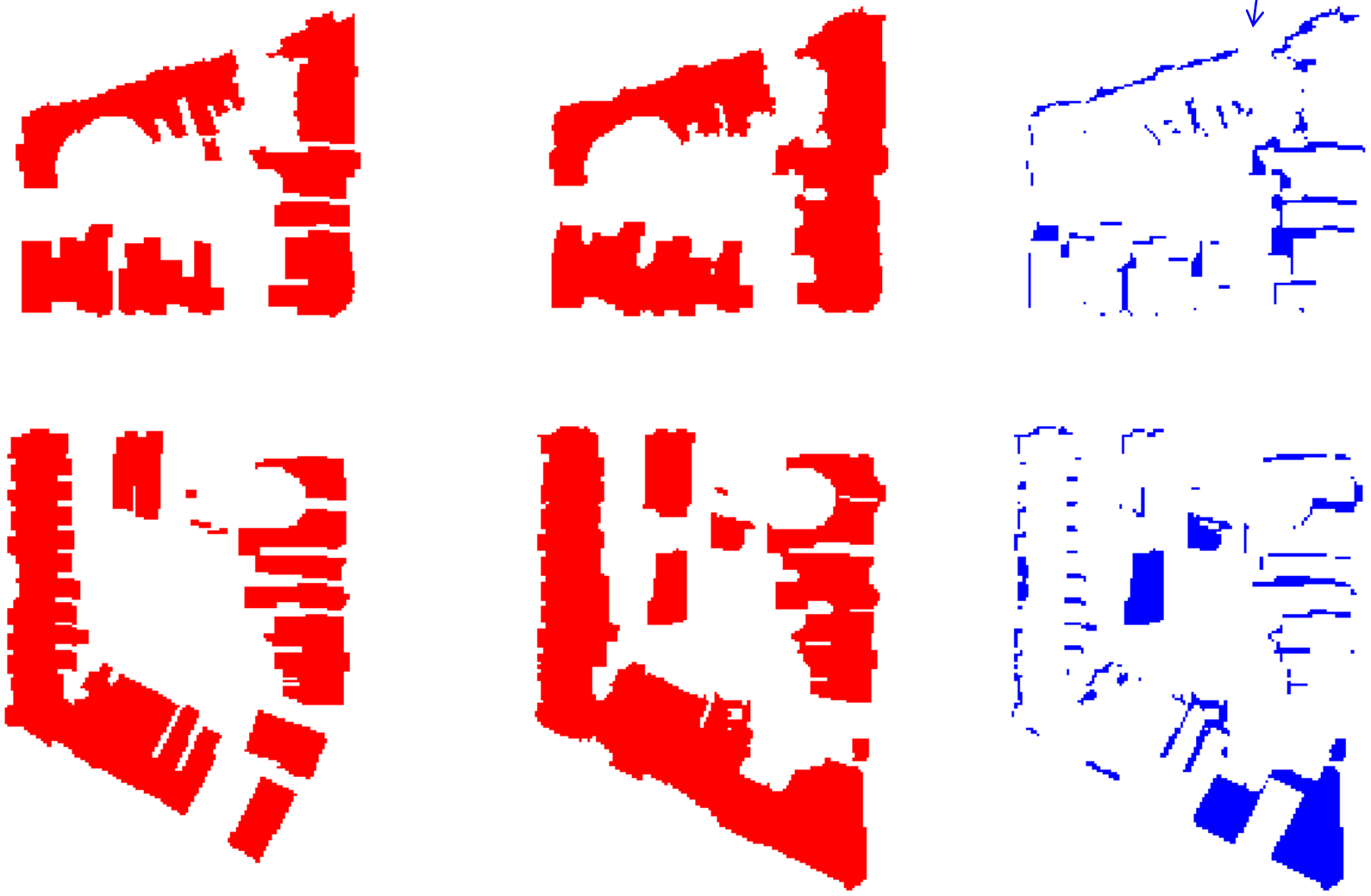
Input land cover:



# Geometric Change Analysis

Ghost boundaries result from sensor noise and registration error.

Geometric subtraction:



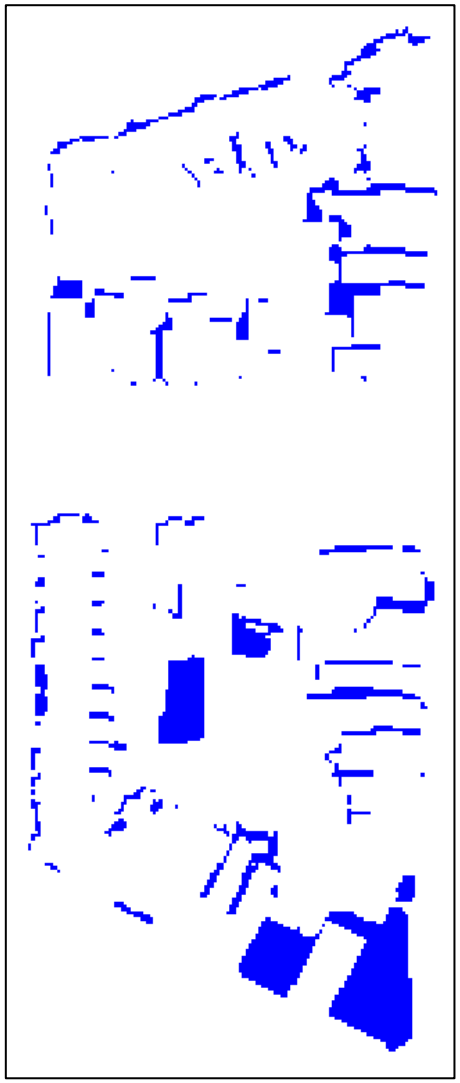
2006 Buildings

2011 Buildings

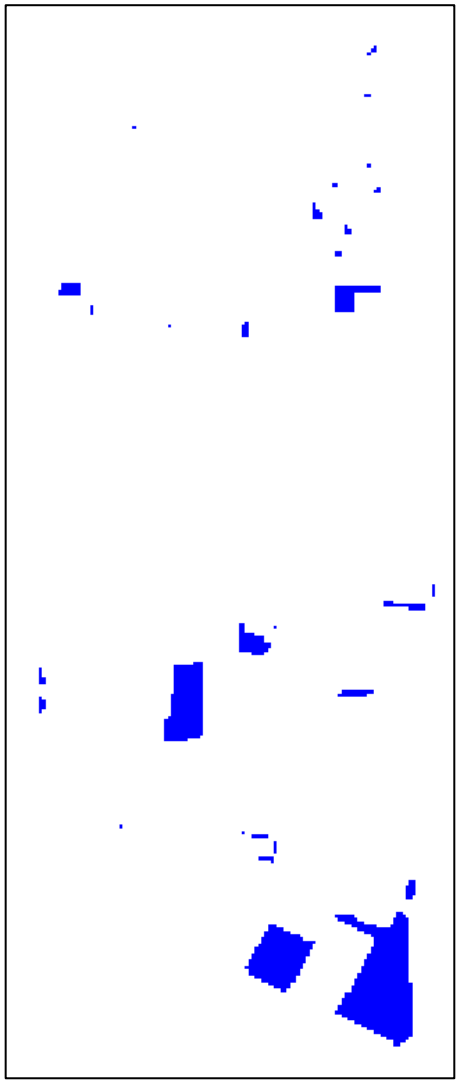
2011 minus 2006

# Geometric Change Analysis

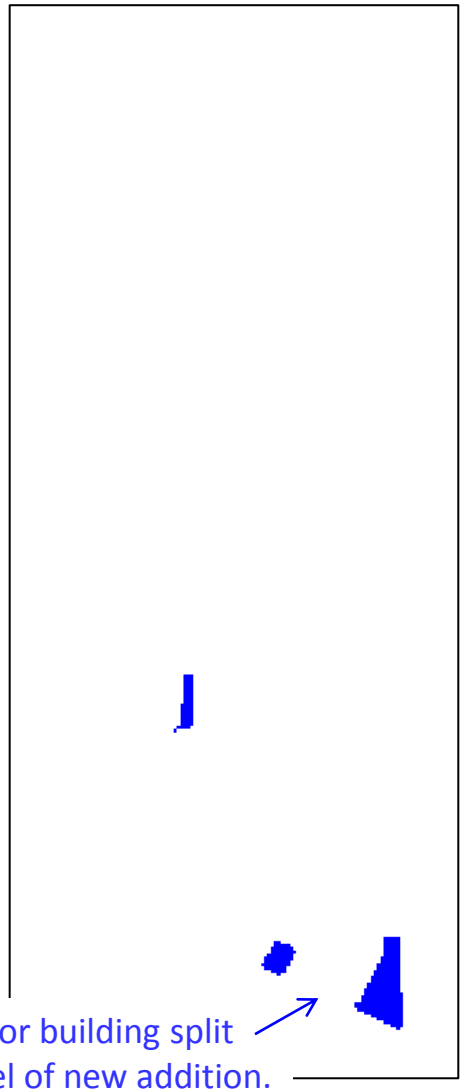
Shrinking to remove ghost boundaries:



2011 minus 2006



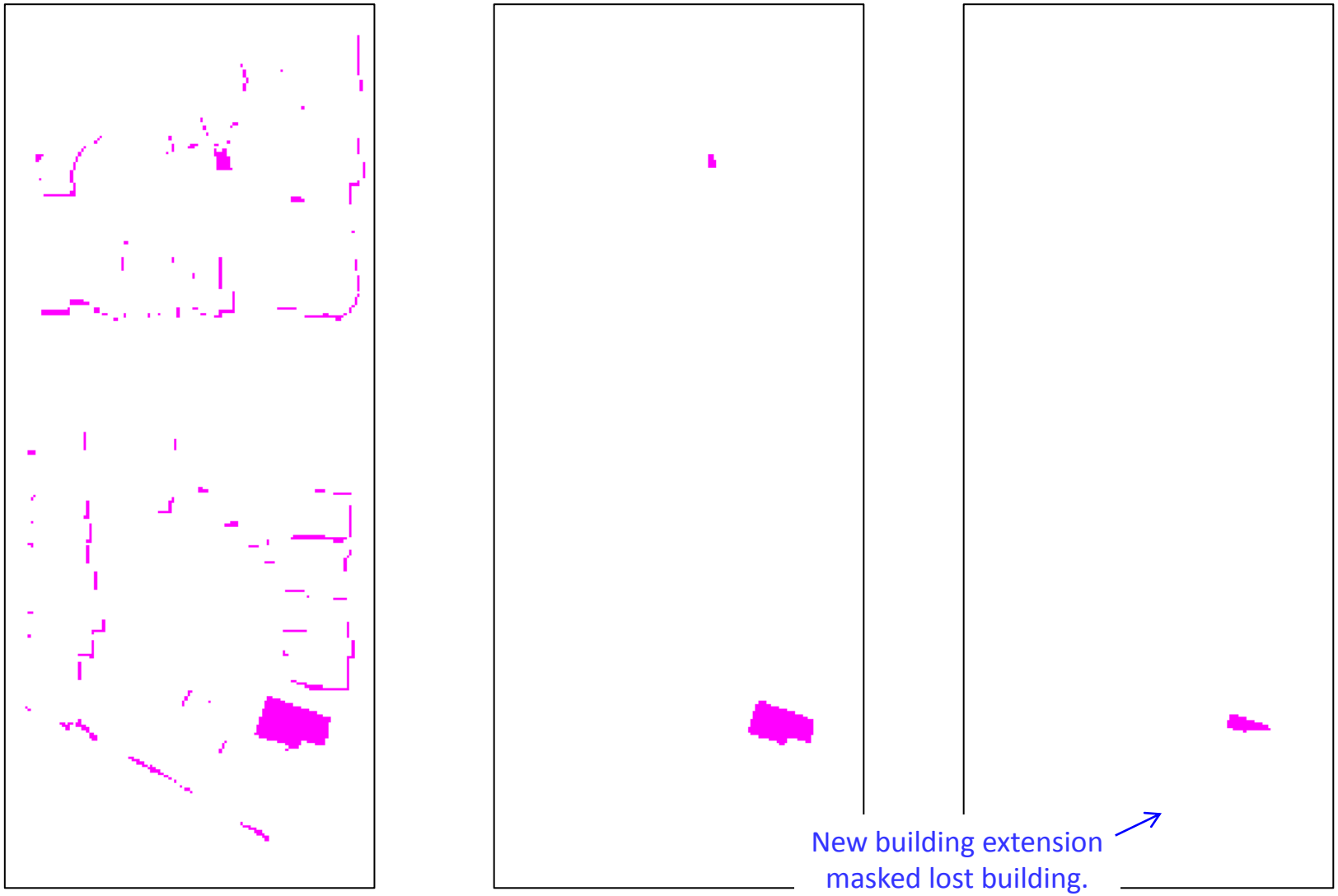
Shrink 1 Pixel




Shrink 4 Pixels

# Geometric Change Analysis

Finding lost buildings (old minus new):




2006 minus 2011 | Shrink 1 Pixel | Shrink 4 Pixels

# Geometric Change Analysis Results

Input land cover:

removed

missed this one

2006

new

added split in two; misses interior change section.

2011

Change results do not match full shape.

# Review of Change Analysis Methods

Approaches to change detection:

1. Pure pixel-based change analysis.
   - Simple, but image changes can result from illumination, color balance, moving shadows, etc., which are not real changes in the world.

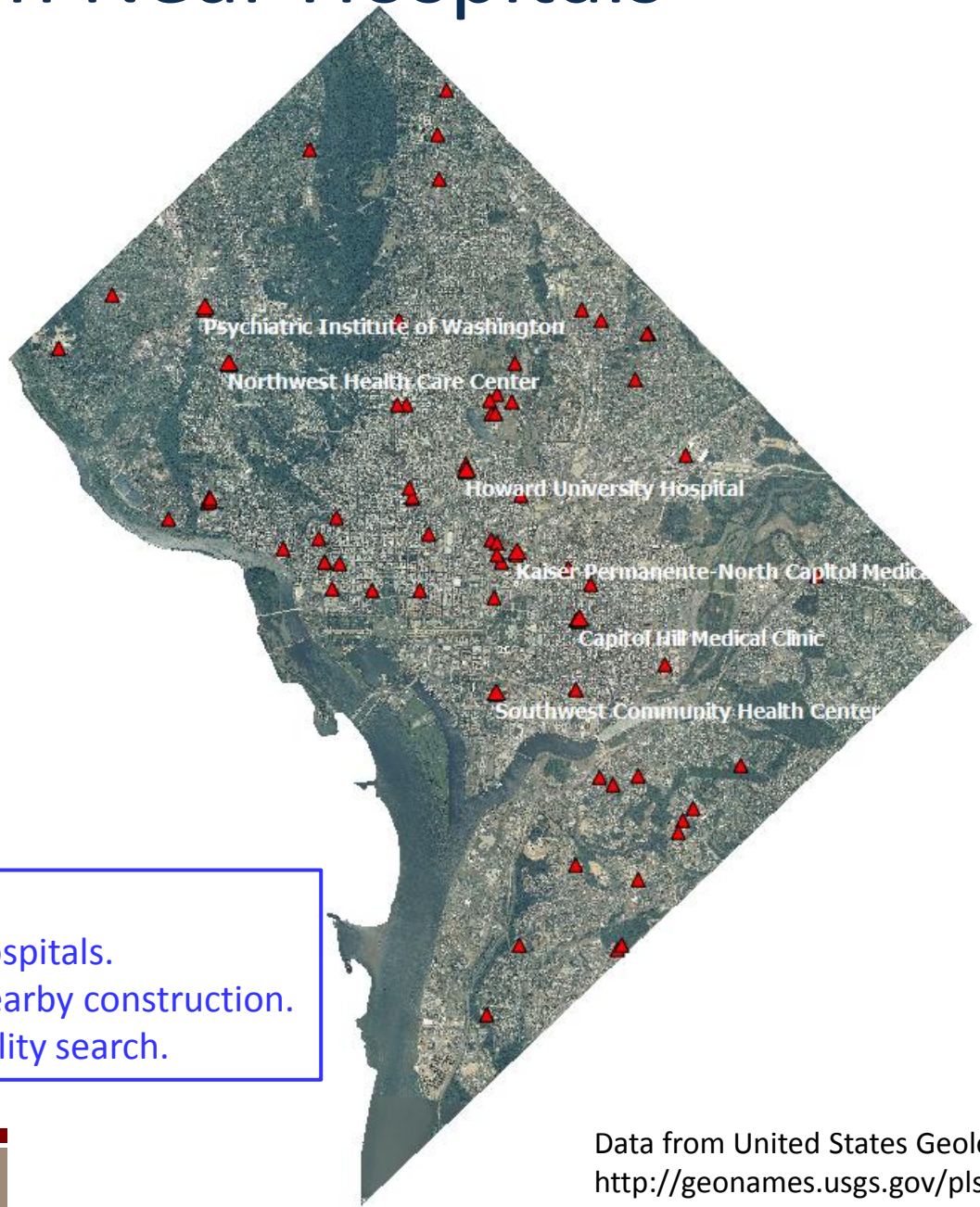
2. Geometric feature-based change analysis.
   - Better, but noise and geometric interference can lead to confusing or misleading results.

3. Graph-based change analysis.
   - Can reject false change, clarify change relationships, and present correct before/after representations.

# Construction Near Hospitals

Sandia National Laboratories

Hospitals in the DC area:




Goals:

1. Find construction near hospitals.
2. Estimate magnitude of nearby construction.
3. Demonstrate multi-modality search.

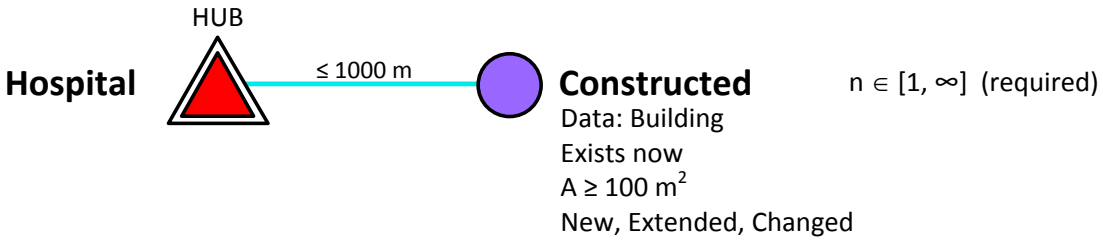
Data from United States Geological Survey.
http://geonames.usgs.gov/pls/gnispublic

# Construction Near Hospitals

Query specification:

How much construction has occurred within 1 km of each hospital?

Question semantics:
Building
Changed Building
Extended Building
Merged Building
New Building
Replaced Building
Constructed
Hospital
⋮




Data semantics:
Building
Road
Other Paved
Grass/Shrub
Trees
Dirt
Water
Hospital

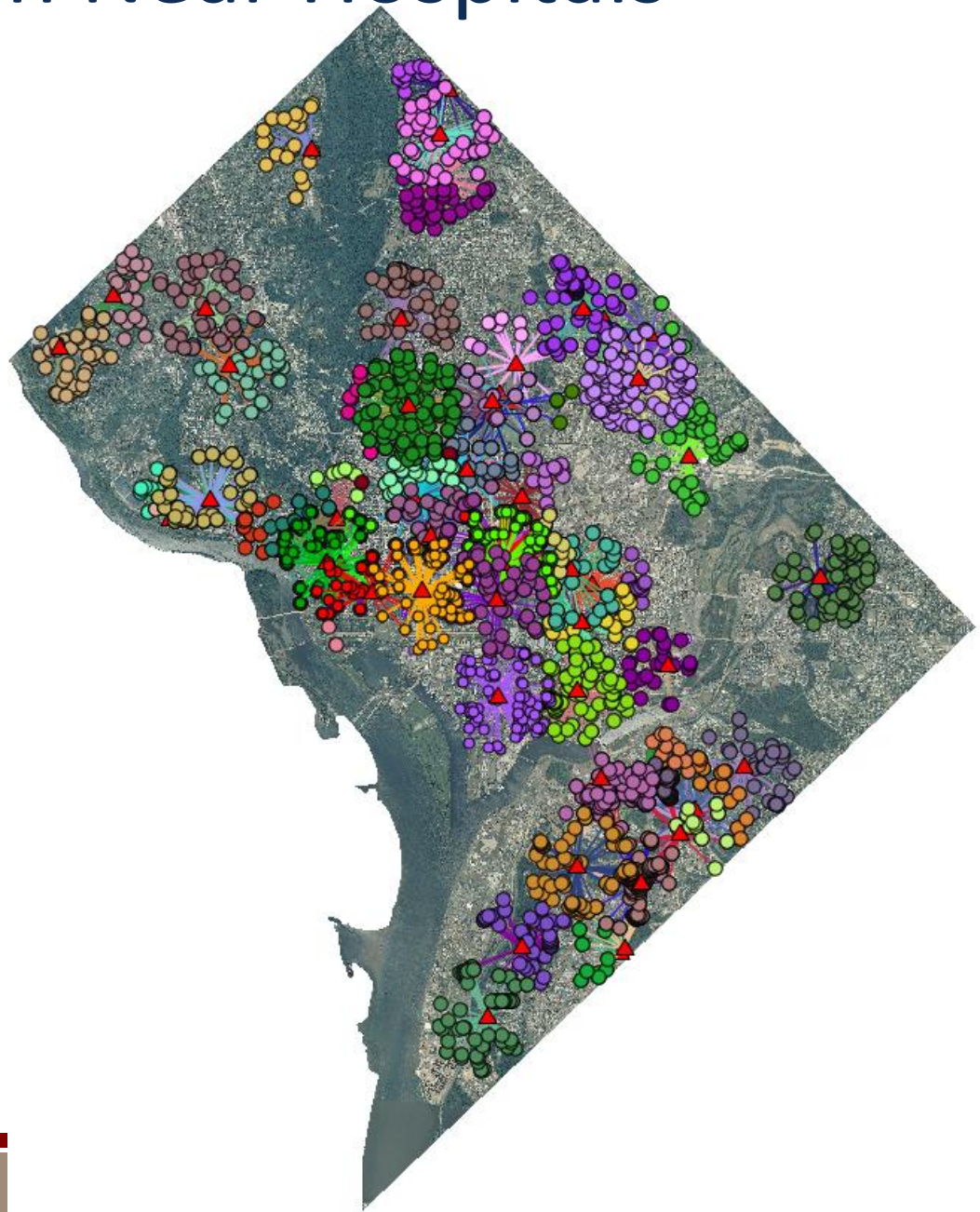
Star graph search.

Rank in order of most construction first.

# Construction Near Hospitals

Sandia National Laboratories

All matches:



# Construction Near Hospitals

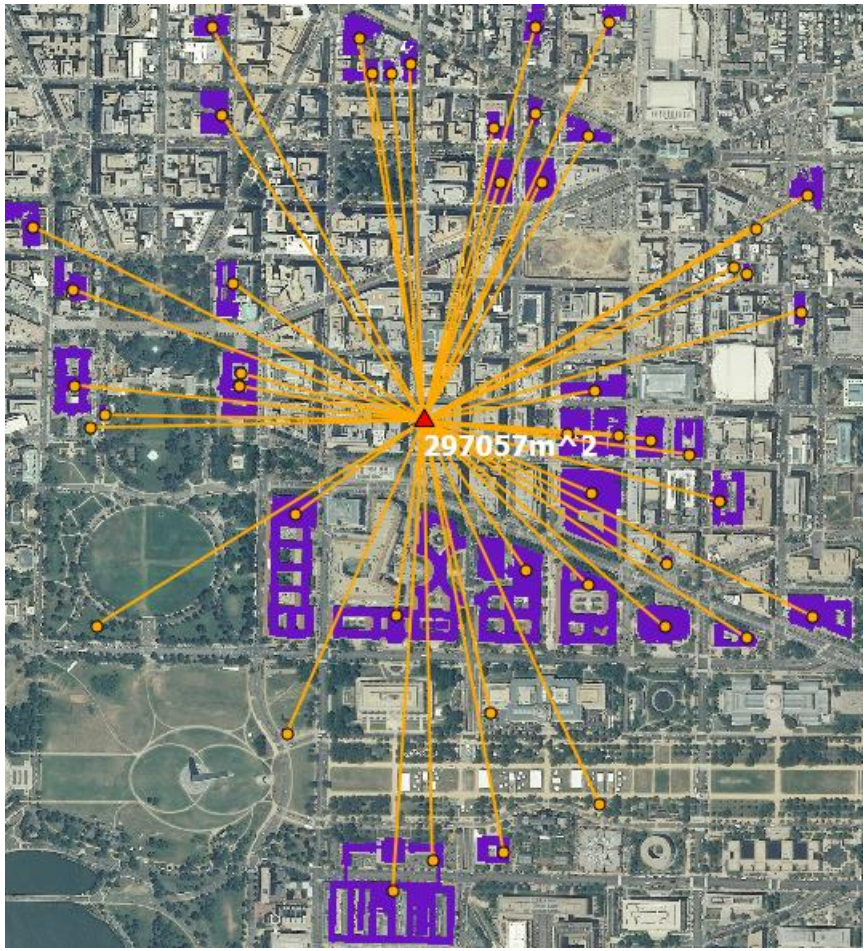
Sandia National Laboratories

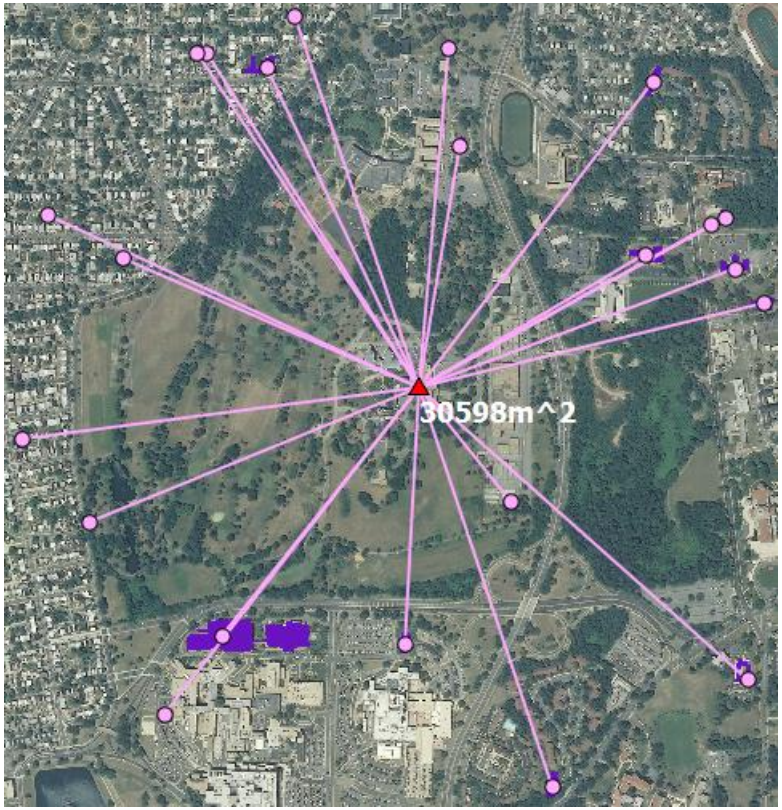
Select two for close examination:

# Construction Near Hospitals

Sandia National Laboratories

Hospital with different amounts of nearby construction:*




One computation includes:

- Temporal change detection.
- Construction diagnosis.
- Image and text data.
- Post-processing numerical sum.

* Includes pseudo-construction. See above.