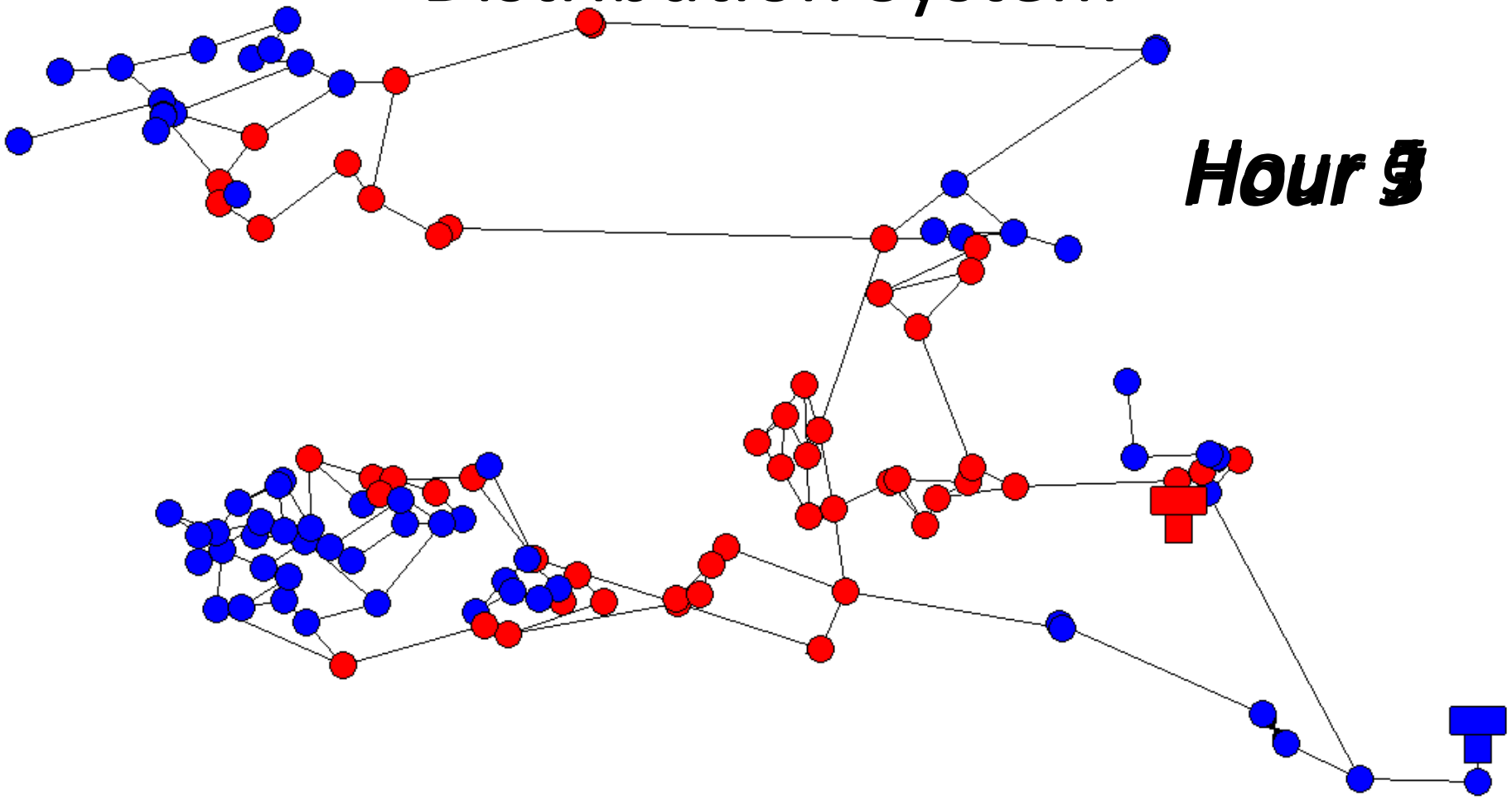
Tracking Dynamic Water-borne Outbreaks with Temporal Consistency Constraints

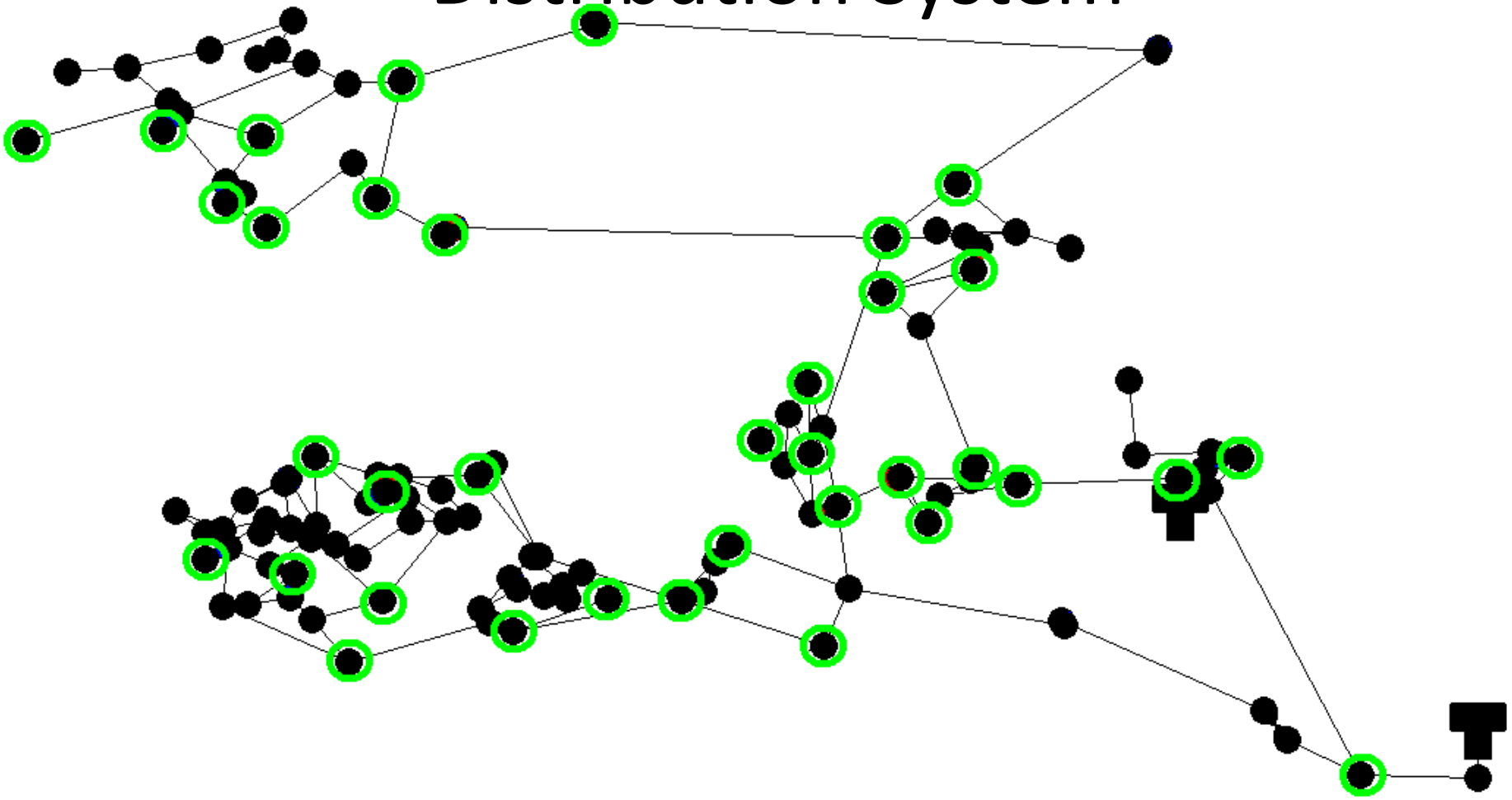
Skyler Speakman, Yating Zhang, & Daniel B. Neill
Event and Pattern Detection Lab
Heinz College, Carnegie Mellon University

This work was partially supported by NSF grants:
IIS-0916345, IIS-0911032, and IIS-0953330

Spreading Contaminants in a Water Distribution System



Spreading Contaminants in a Water Distribution System



Problem Definition

Given:

Network structure

-Water Pipe System

Noisy sensor readings at each node over time

-False positive triggers (0.1)

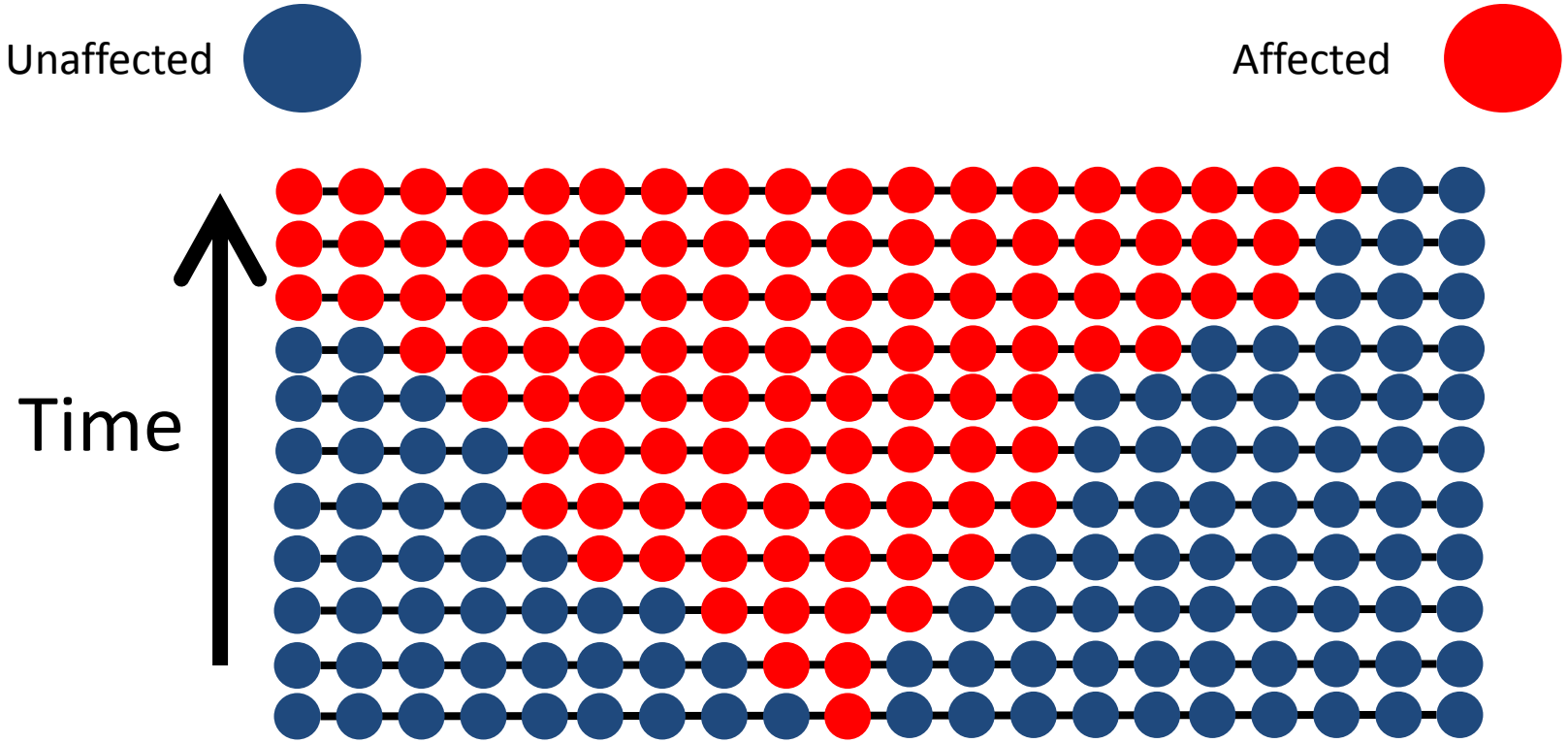
-True positive triggers (0.9)

Identify:

Temporally consistent subset of nodes that tracks a *dynamic pattern* spreading through the network.

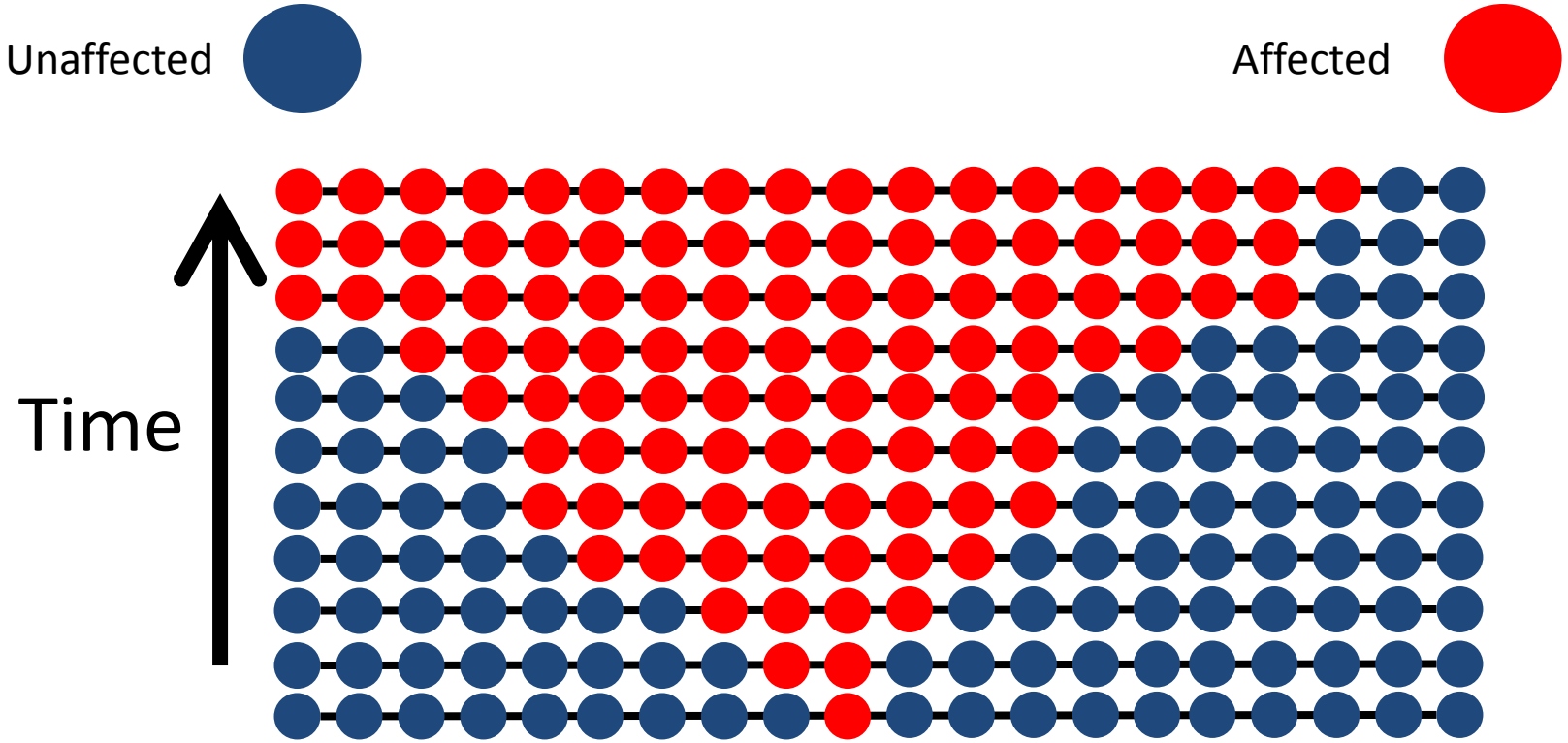
Space-time scan statistics, Pattern detection, Sensor fusion

1-Dimensional Chain Example



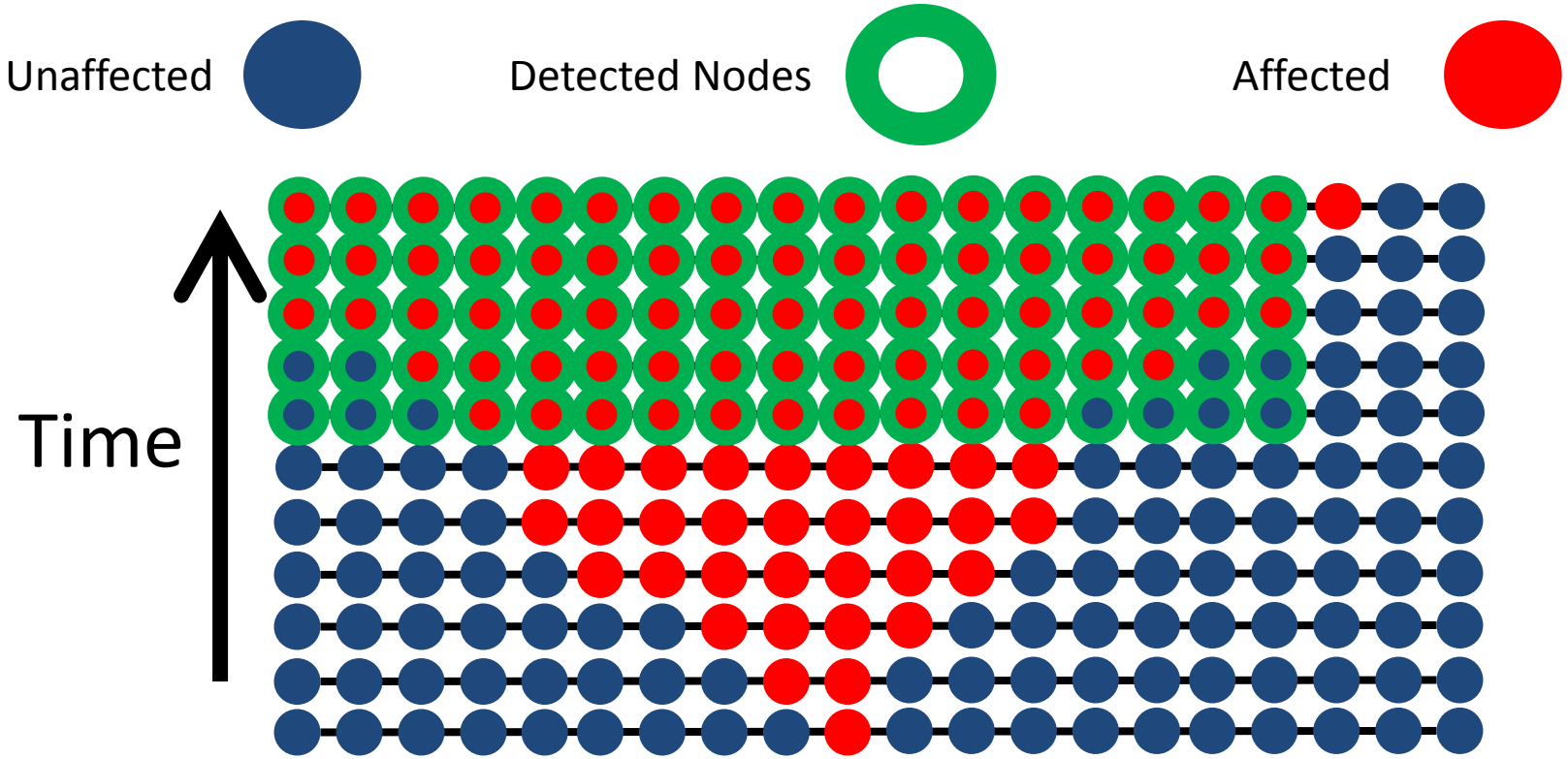
“Static” Method

Assumes the affected subset of nodes does not change over time.
Constrained to search over spatial-temporal “cylinders”.



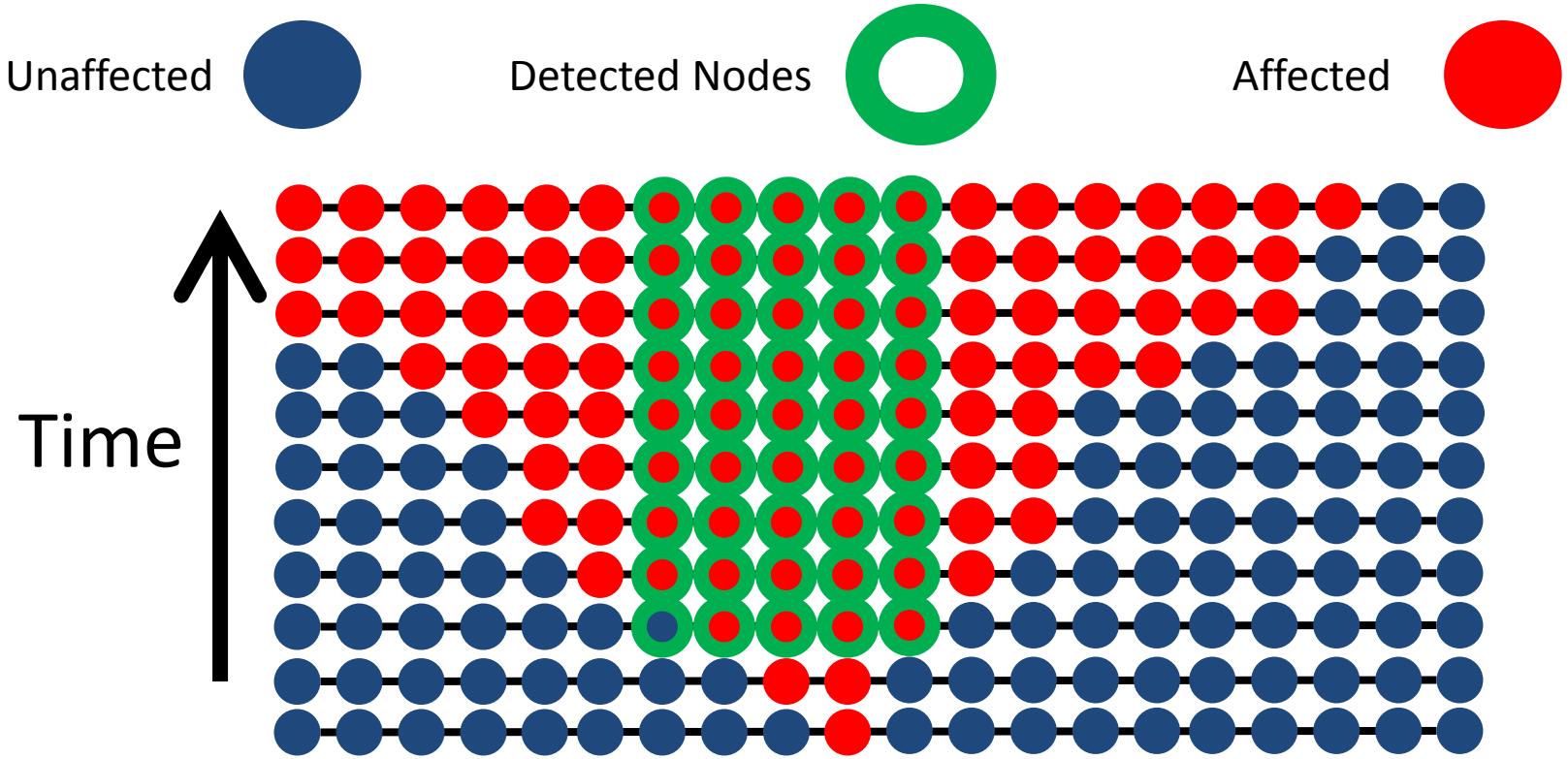
“Static” Method

Assumes the affected subset of nodes does not change over time.
Constrained to search over spatial-temporal “cylinders”.



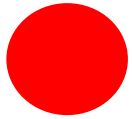
“Static” Method

Assumes the affected subset of nodes does not change over time.
Constrained to search over spatial-temporal “cylinders”.



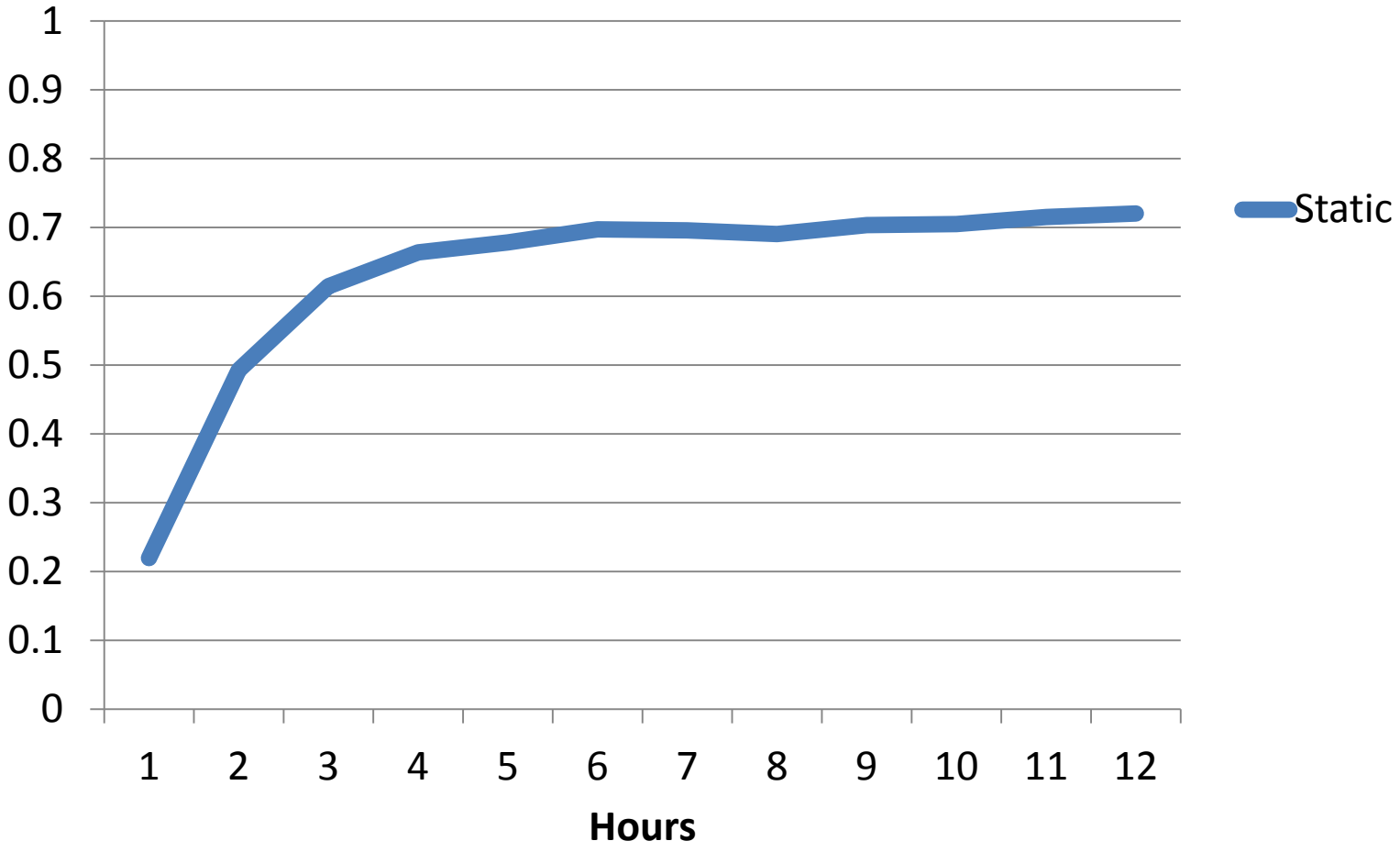
Spatial-Temporal Overlap

Measures how well the detected subset covers the affected subset.
It is a combination of precision and recall.

Detected Nodes  Affected Nodes 

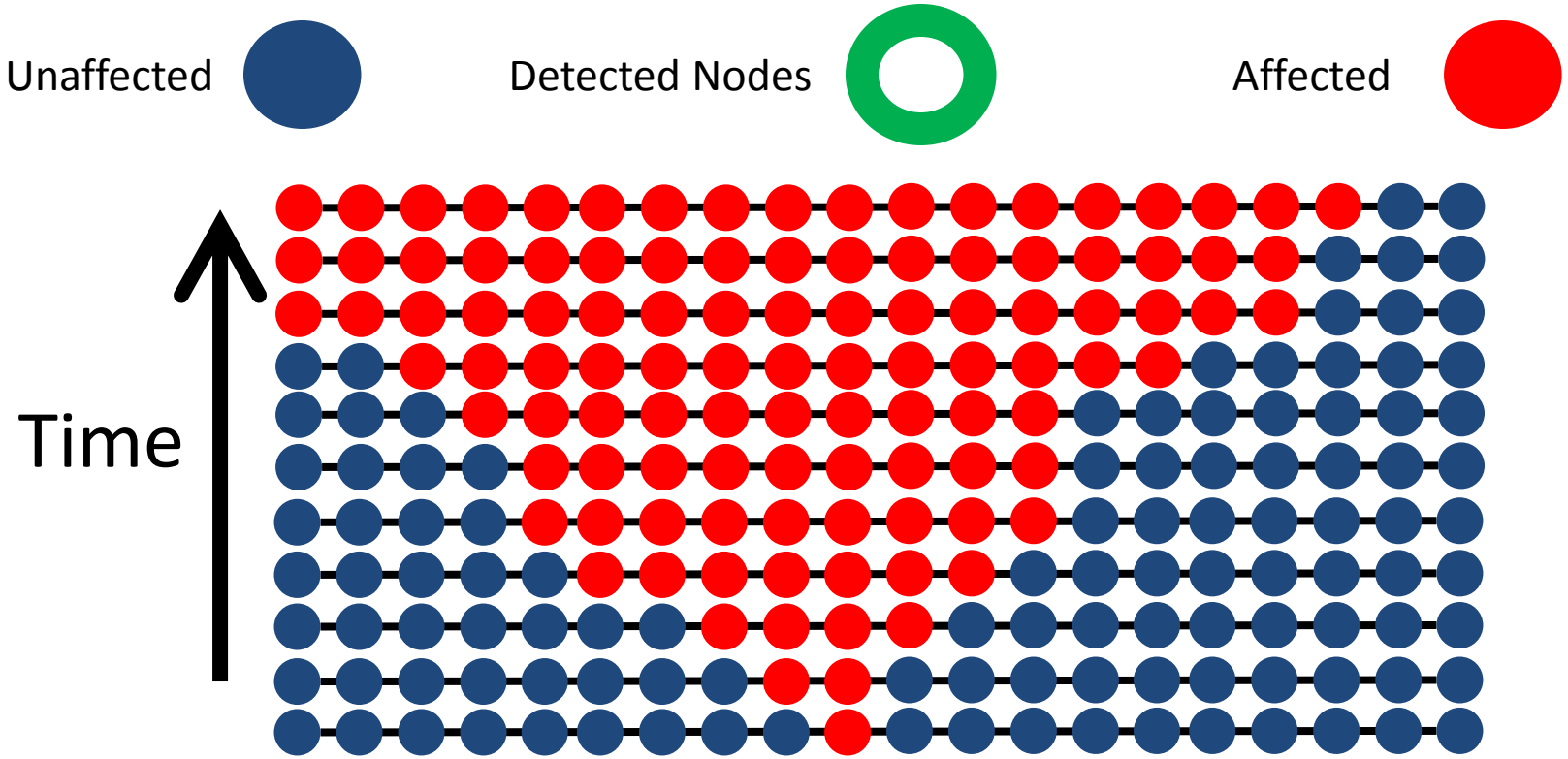
$$\text{Overlap} = \frac{\text{Number of } \bigcirc\!\bullet}{\text{Number of } \bigcirc + \text{Number of } \bullet}$$

Spatial Temporal Overlap



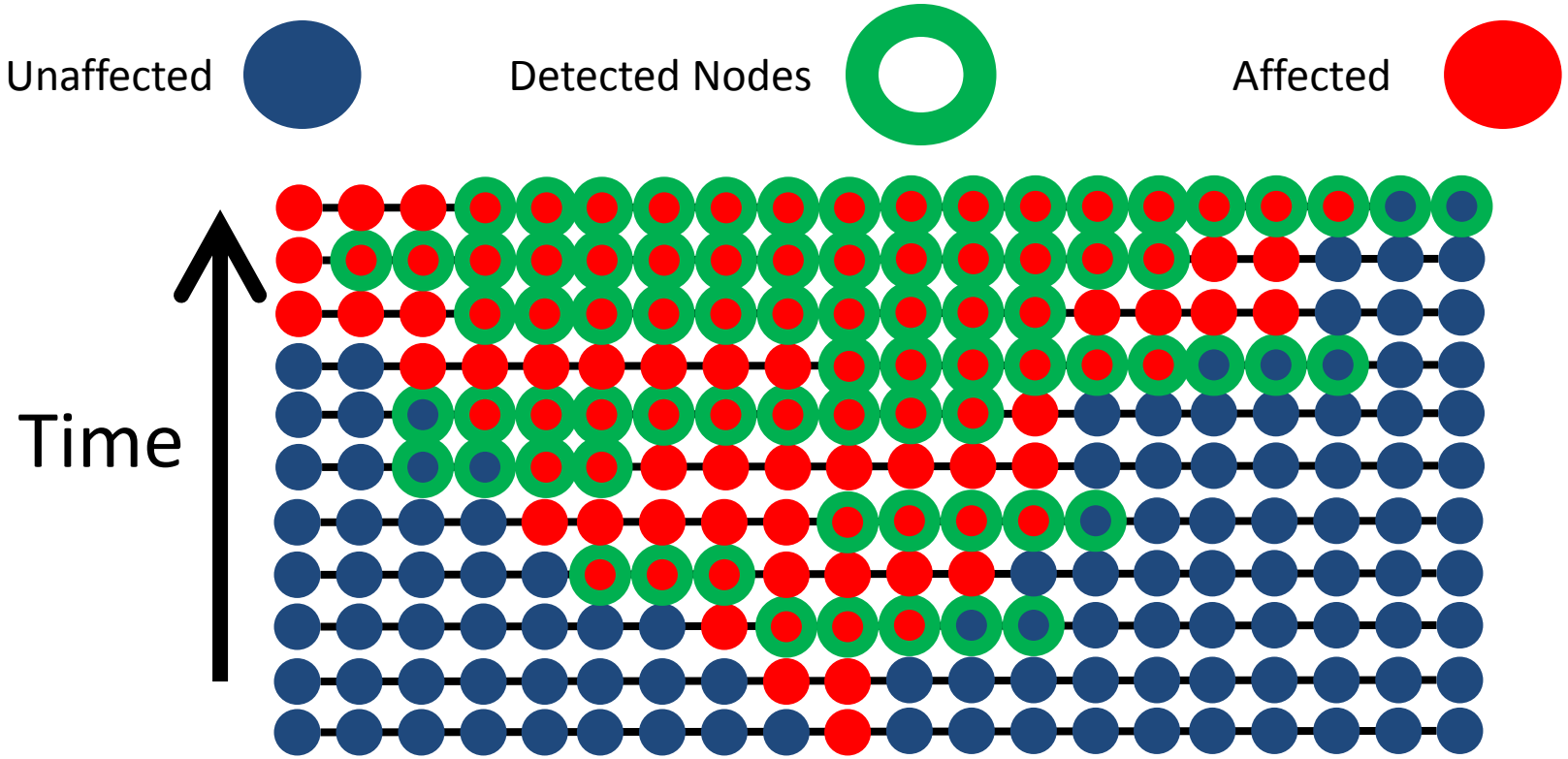
“Independent” Method

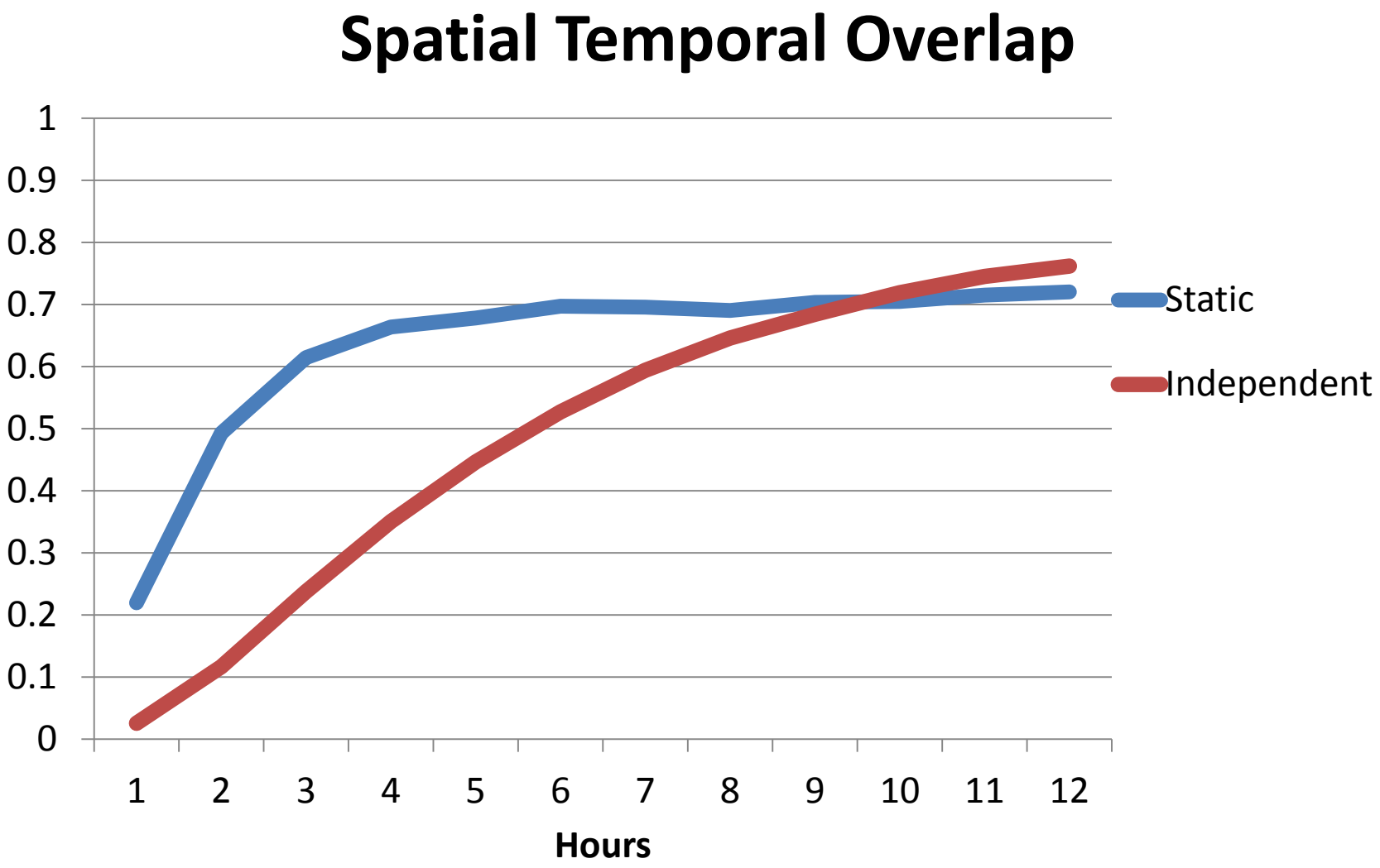
Allows the affected subset of nodes to change over time.
Identifies an independent spatial subset for each time step.



“Independent” Method

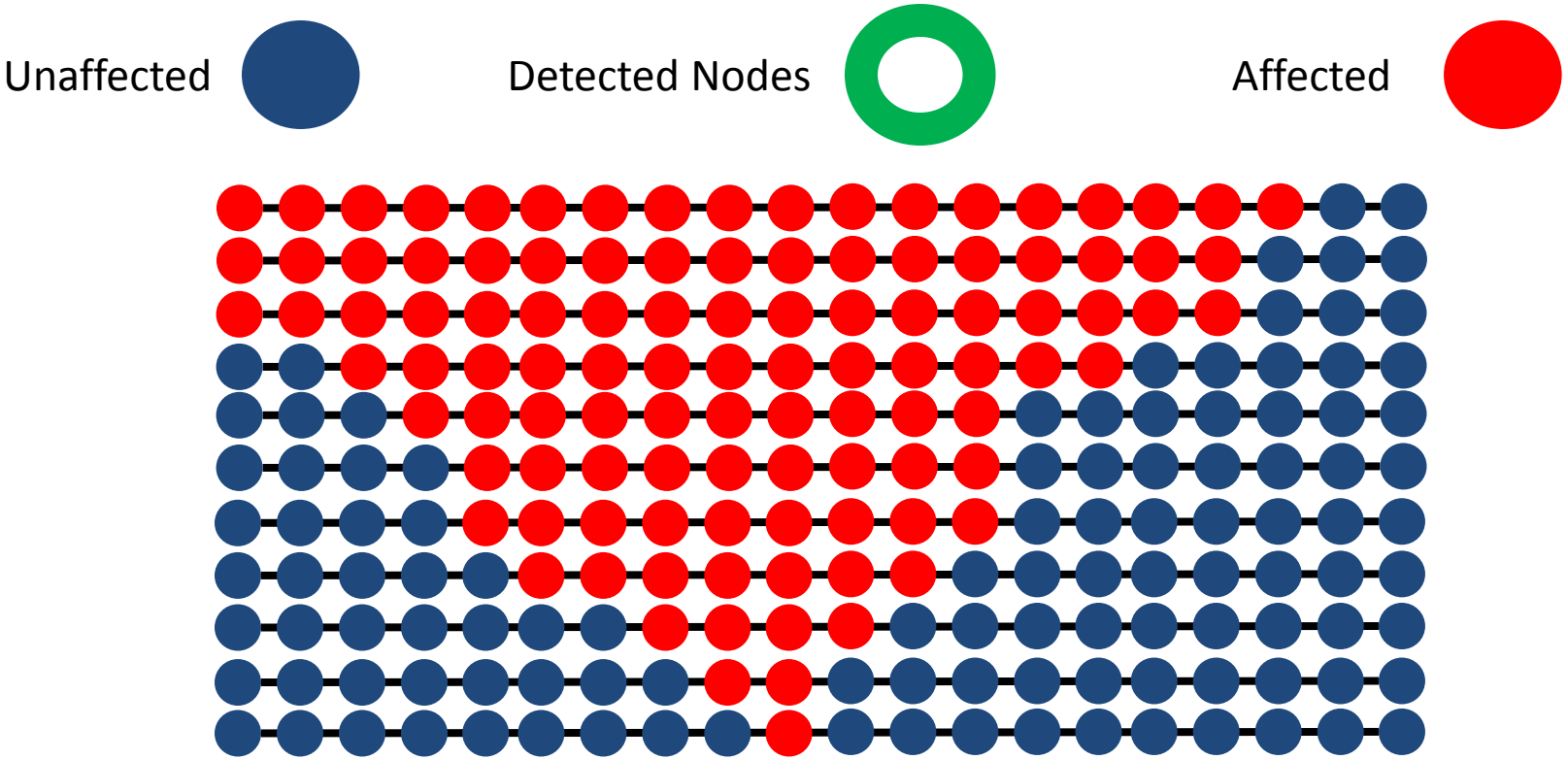
Allows the affected subset of nodes to change over time.
Identifies an independent spatial subset for each time step.




Spatial Temporal Overlap
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
12
Hours
Static
Independent

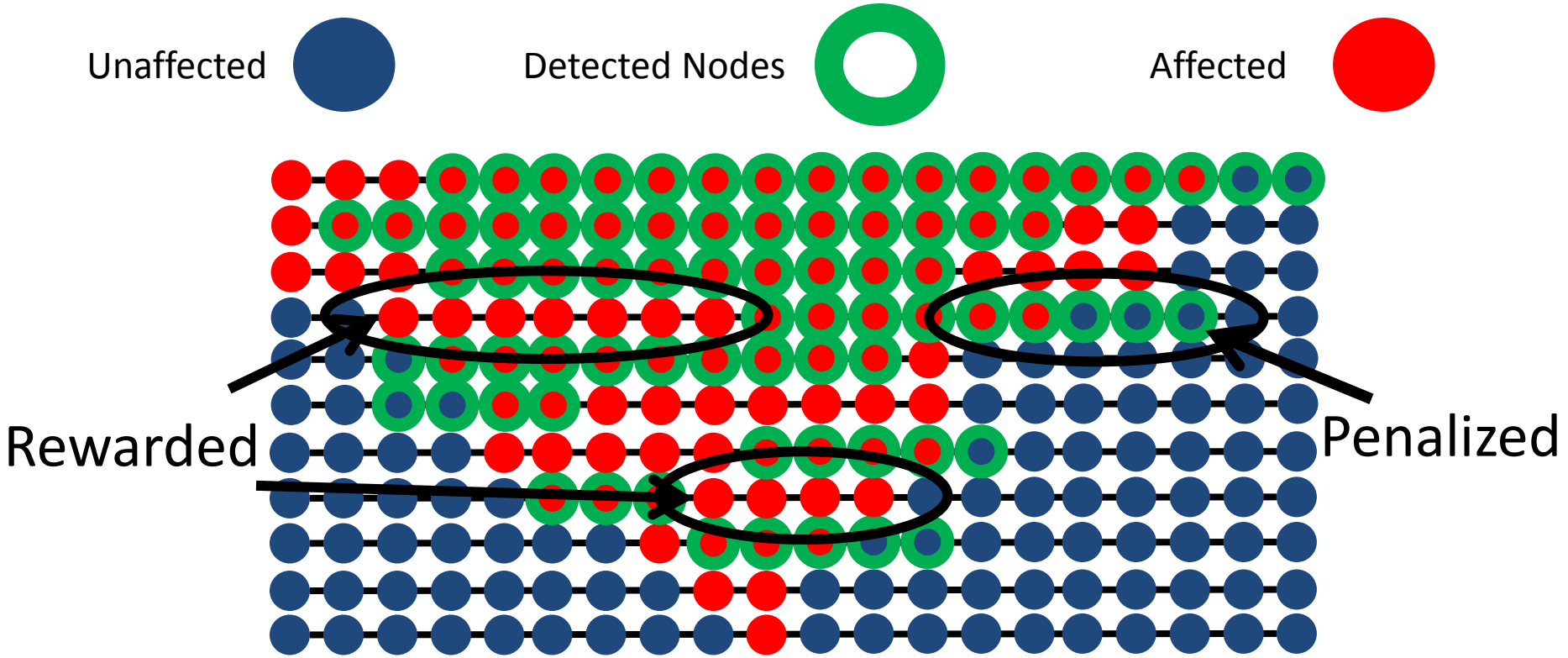
“Dynamic” Method

Allows the affected subset of nodes to change over time.
Rewards spatial subsets that are *temporally consistent.*



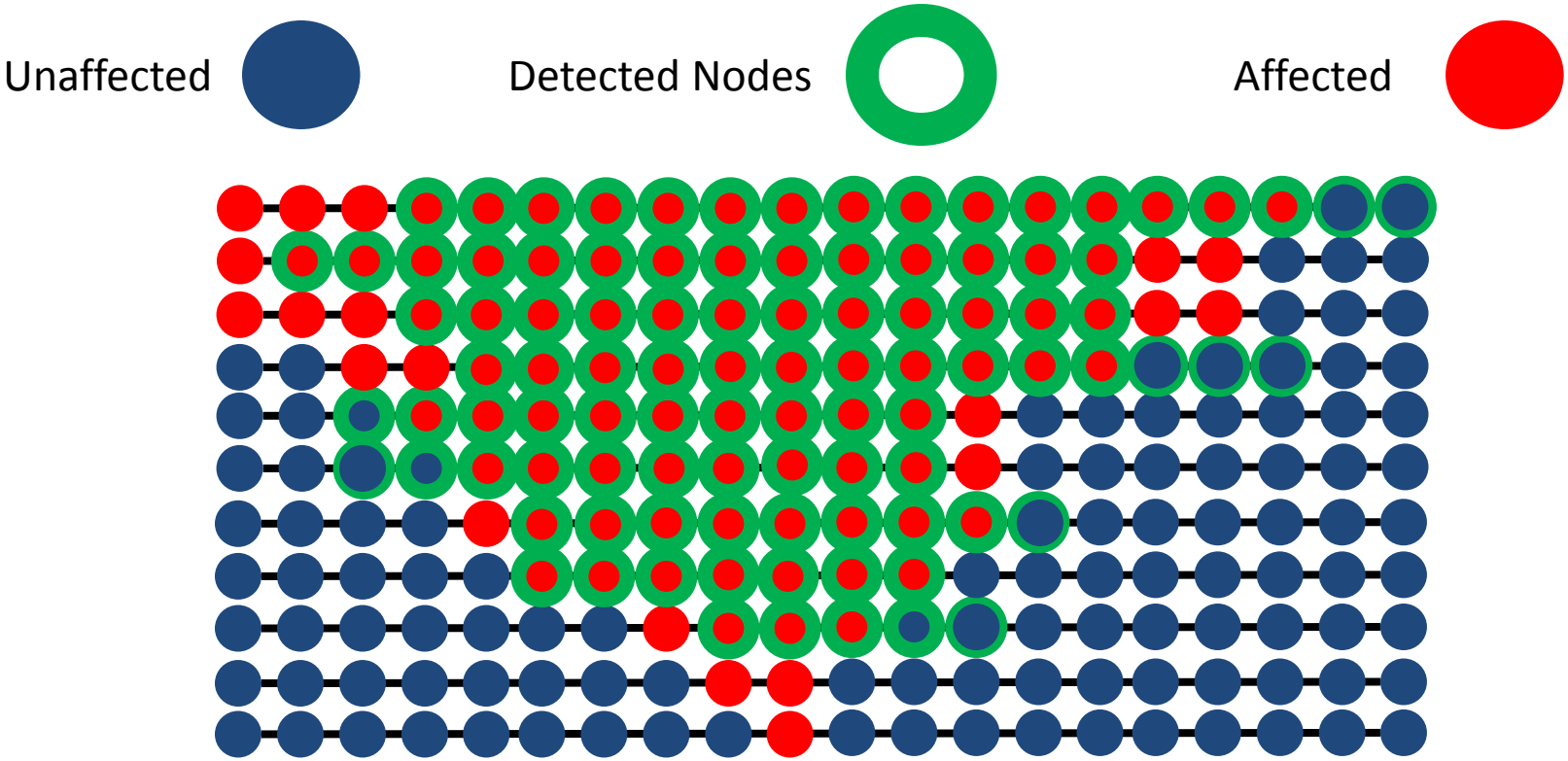
“Dynamic” Method

Allows the affected subset of nodes to change over time.
Rewards spatial subsets that are *temporally consistent.*



“Dynamic” Method

Allows the affected subset of nodes to change over time.
Rewards spatial subsets that are *temporally consistent.*



Temporal Consistency Constraints

Let p_i^t be the prior probability that node *i* is included in the optimal subset at timestep *t.*

Let X_i^{t-1} be 1 if node *i* is included in the optimal subset at timestep *t-1* and 0 otherwise.

We model the prior log-odds as:

$$\log\left(\frac{p_i^t}{1-p_i^t}\right) = \beta_0 + \beta_1 X_i^{t-1}$$

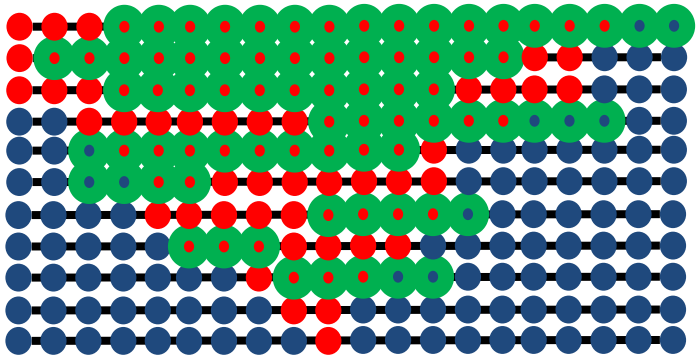
Temporal Consistency Constraints

$$\log\left(\frac{p_i^t}{1-p_i^t}\right) = \beta_0 + \beta_1 X_i^{t-1}$$

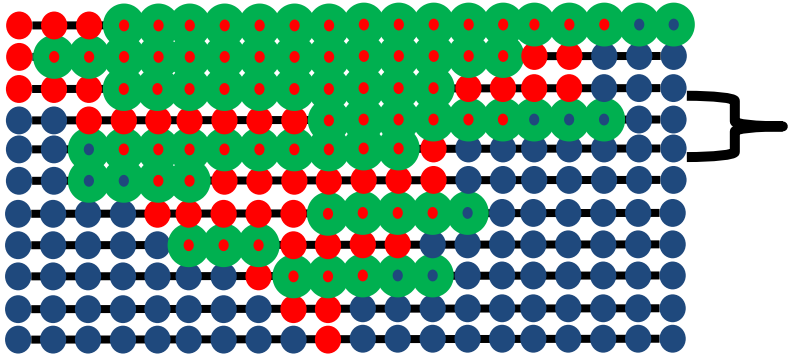
For example, if $\beta_0 = -1.5$ and $\beta_1 = 5$ then a node included in an optimal subset at timestep *t-1* has a **97**% prior probability of being included in the optimal subset at time *t.*

A node that was *not* included at timestep *t-1* has only a **18**% prior probability of being included in the optimal subset at time *t*.

From Independent to Dynamic

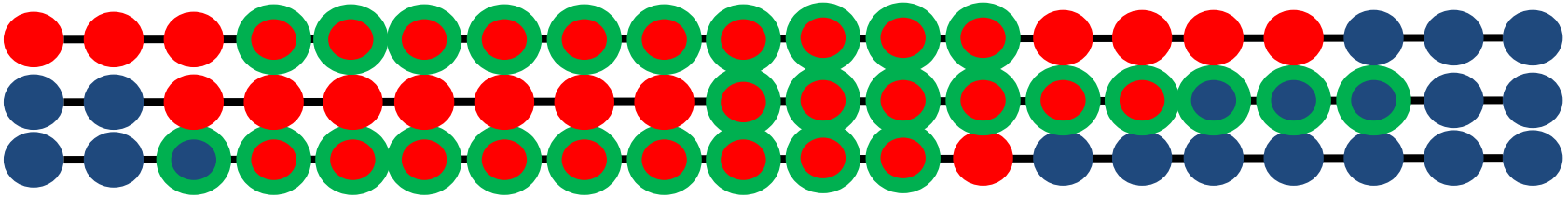


Initialize the spatial subsets independently

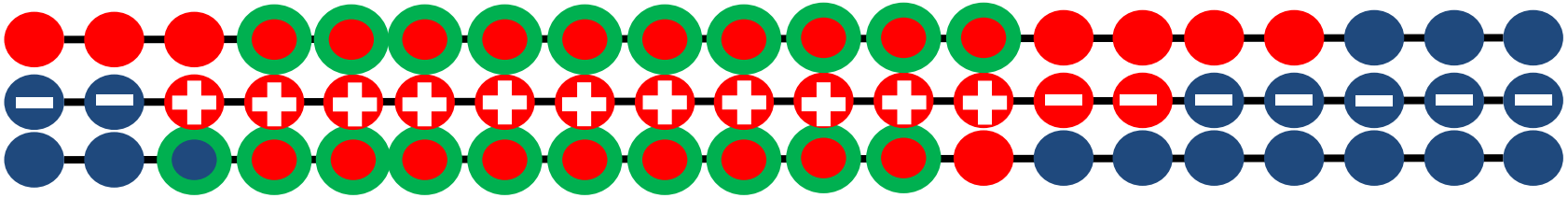


Randomly select a timestep *t* and its temporal neighbors *t-1* and *t+1*

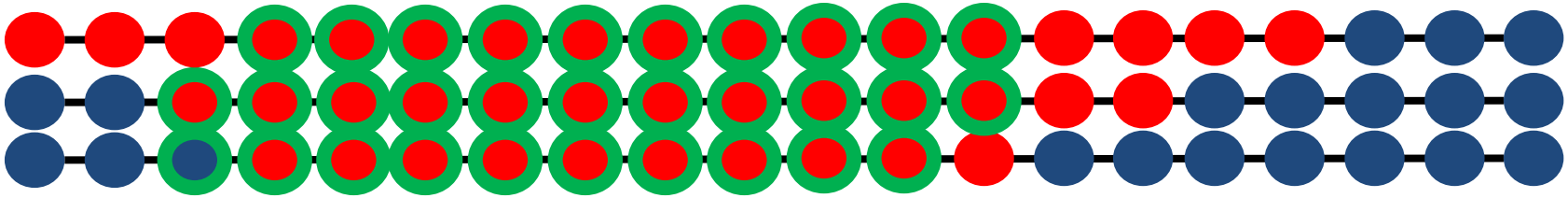
From Independent to Dynamic



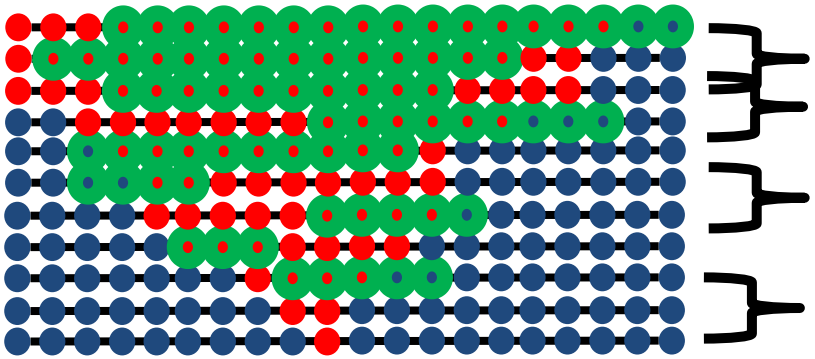
Assign prior probabilities based on subsets *t-1* and *t+1*



Recalculate optimal spatial subset incorporating temporal consistency



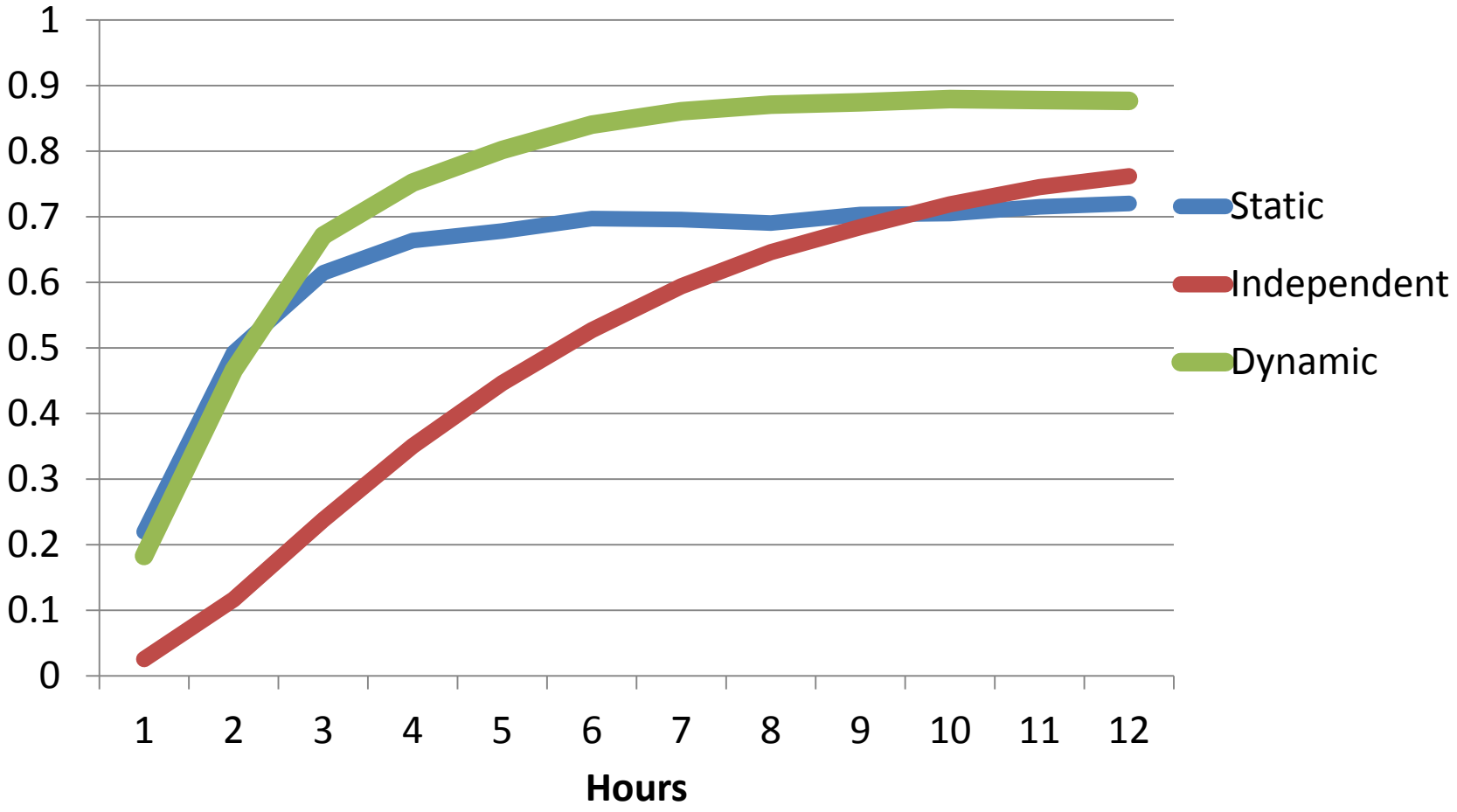
From Independent to Dynamic



Repeat this process until no changes improve the posterior probability of the resulting spatial temporal subset.
(convergence to a local maximum)

The resulting spatial-temporal subset of nodes and timesteps is returned as the detected subset for the “Dynamic” method.

Spatial Temporal Overlap



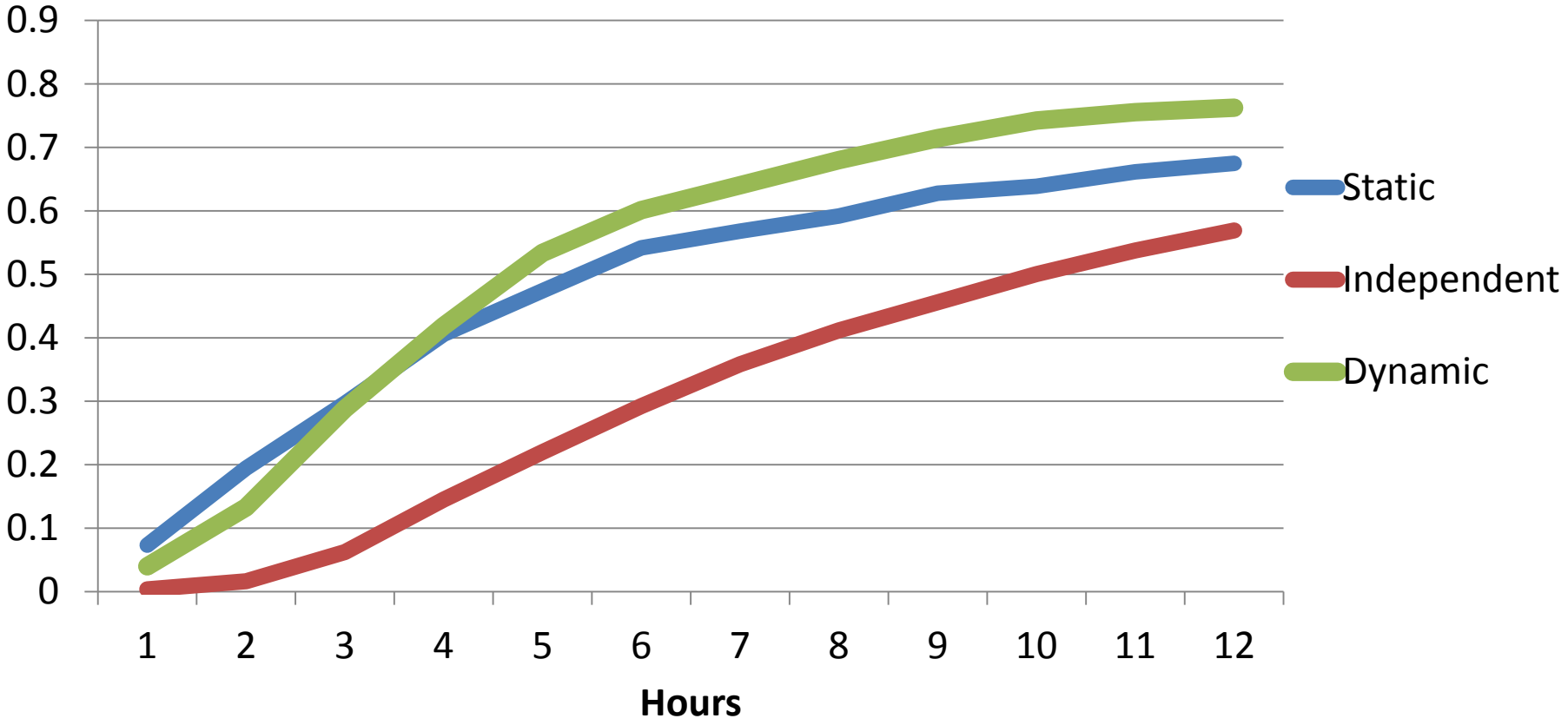
Temporal Consistency Constraints and Neighbors

Let n_i^{t-1} be the number of neighbors of node *i* included in the optimal subset at timestep *t-1* and k_i be the degree of node *i.*

Then neighboring nodes can influence the prior probability as well.

$$\log\left(\frac{p_i^t}{1-p_i^t}\right) = \beta_0 + \beta_1 X_i^{t-1} + \beta_2 \frac{n_i^{t-1}}{k_i}$$

Spatial Temporal Overlap for FPR = 0.2 and TPR = 0.8



The “Dynamic” method detects plumes approximately ½ hour earlier than the “Static” method with one false alarm per month.

Conclusions

Relaxing constraints on spatial-temporal region shape must be done carefully:

Hard constraints on the spatial-temporal region fails to capture the dynamics of the pattern.

Independent selection of spatial subsets loses important temporal information.

Incorporating temporal consistency constraints addresses these issues and results in higher spatial-temporal accuracy and detection power.

Thank You

References:
Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods, 1997, 26(6): 1481-1496.

Neill DB. Fast subset scan for spatial pattern detection. Journal of the Royal Statistical Society (Series B), 2012, 74(2):337–360.

Speakman S and Neill DB. Fast graph scan for scalable detection of arbitrary connected clusters. Advances in Disease Surveillance, 2010.