Data, Predictions, and Decisions in Support of People and Society

Eric Horvitz

Microsoft

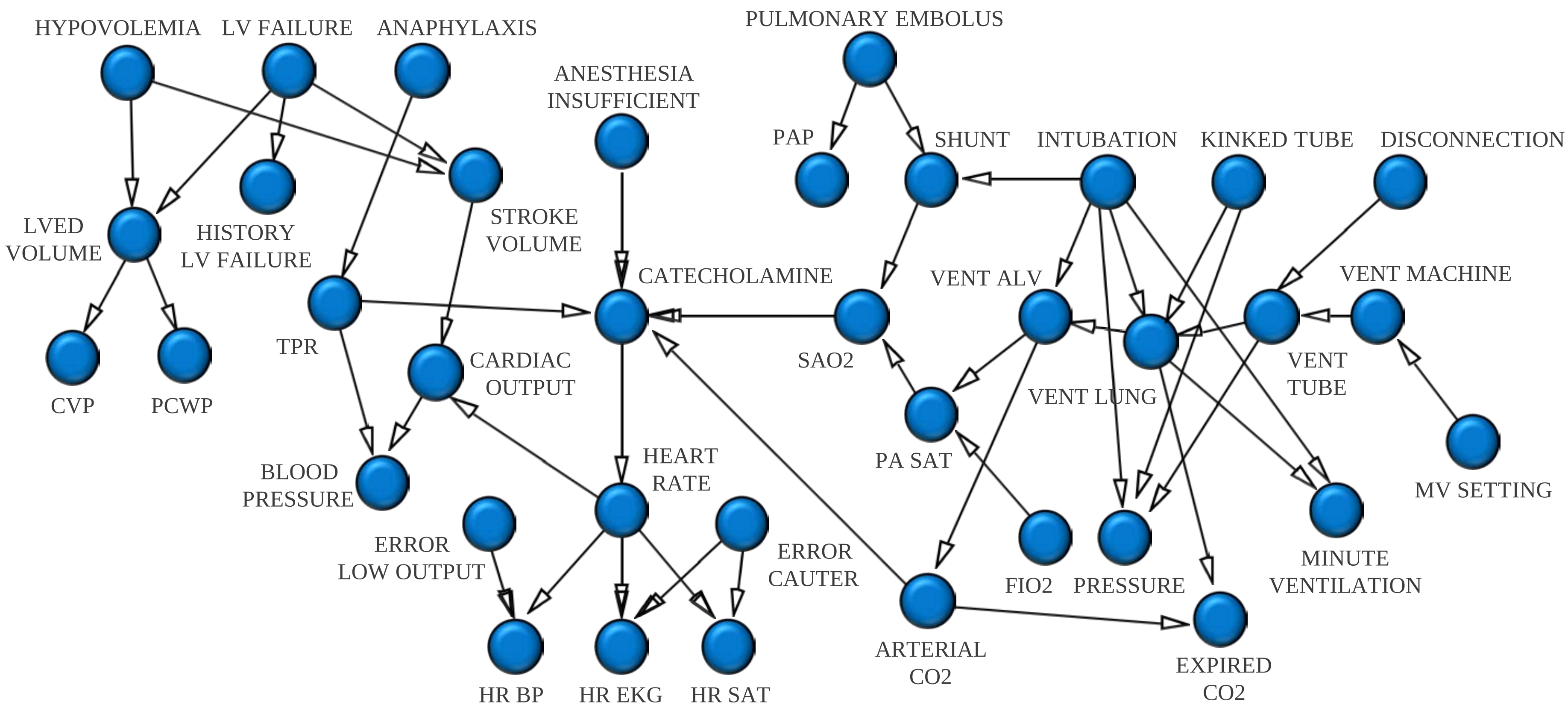
#KDD2014

Data Science for Social Good

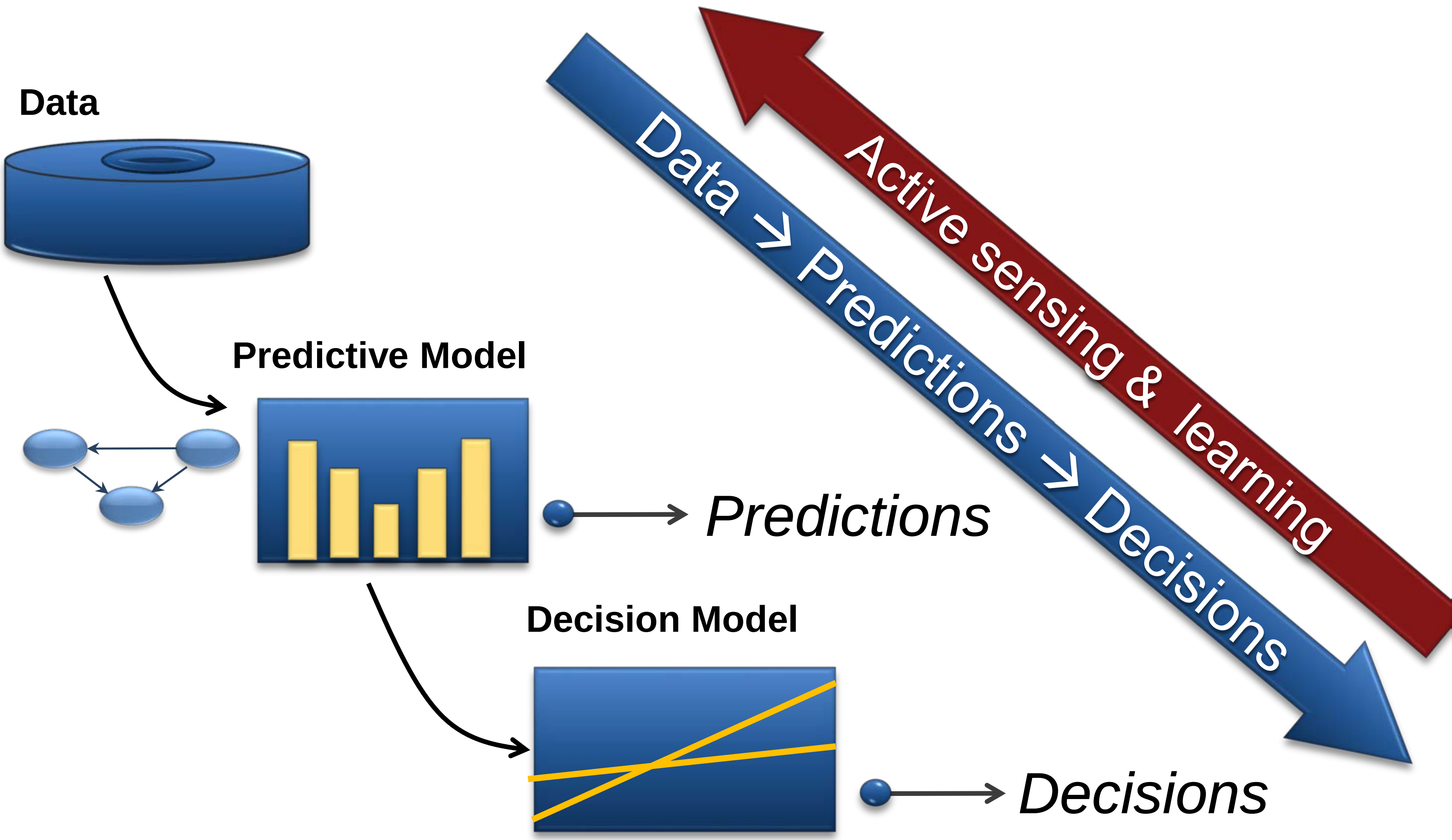
Critical contributions to humanity

Learning, inference, and decision making

Inference for high-stakes challenges

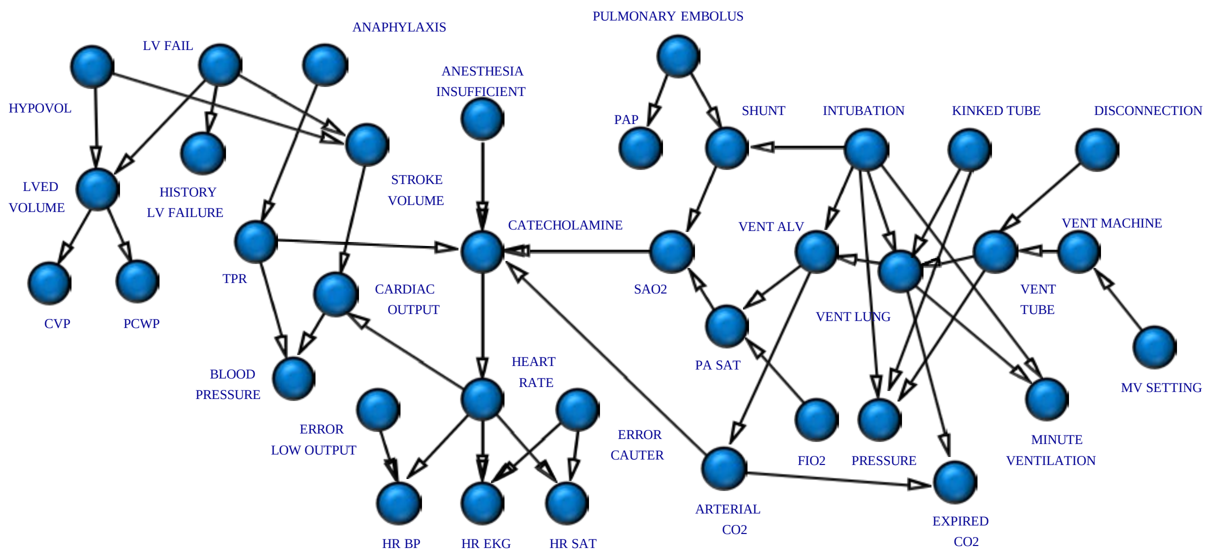


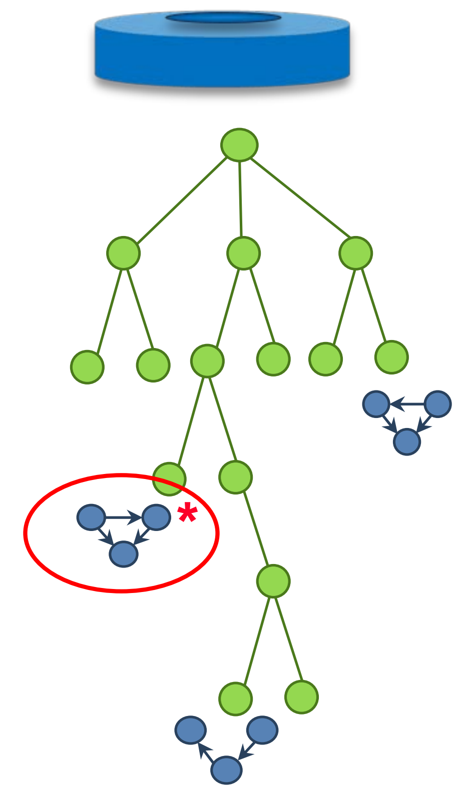
Predictions to Decisions



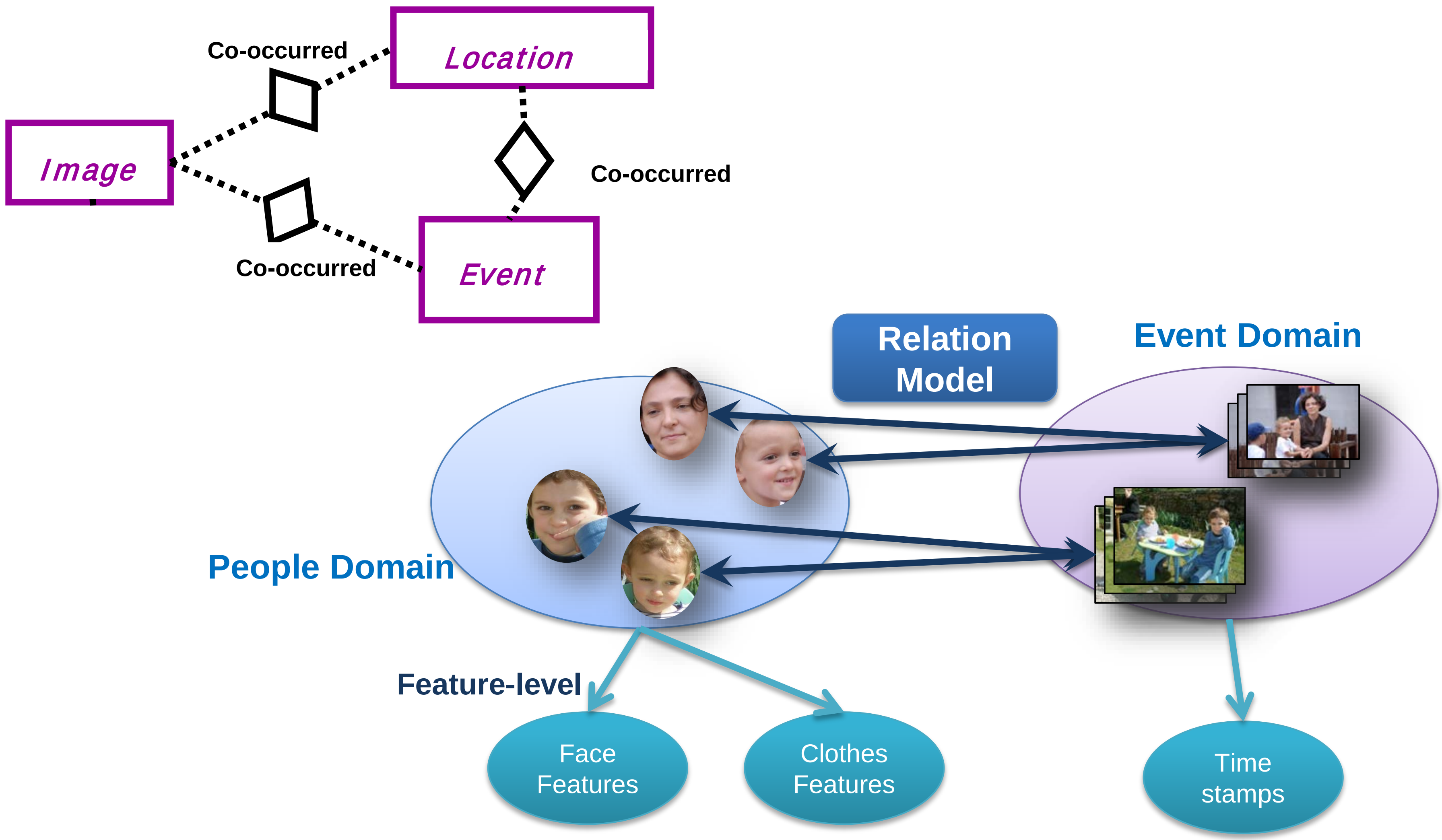
Exciting Times

Learning procedures keeping pace with data





Rise of Rich Representations



Rise of Rich Representations

J. Shotton, J. Winn, C. Rother, A. Criminisi

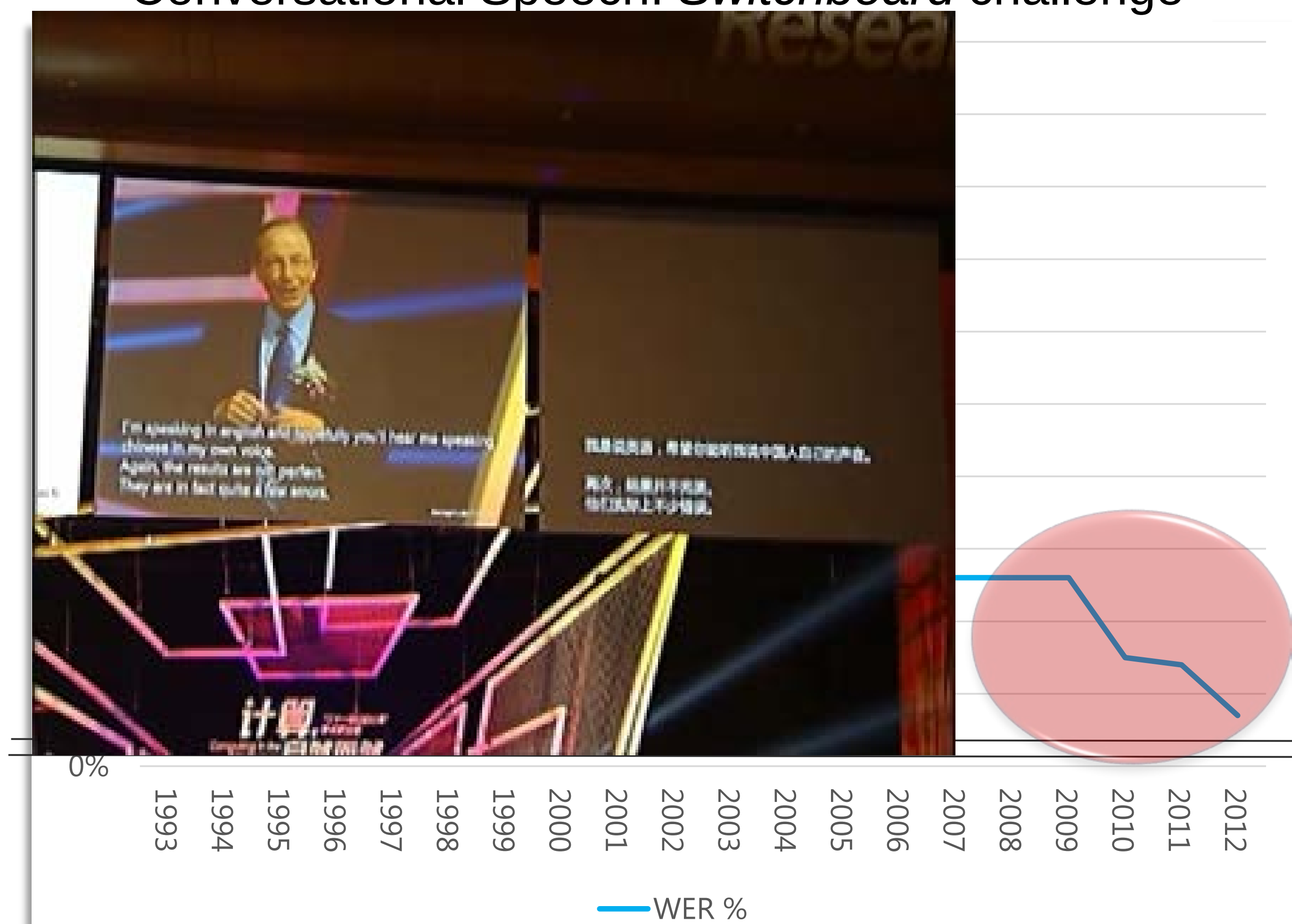
Rise of Rich Representations

KINECT
for XBOX 360

Renaissance of Familiar Methods

Pursuit of speech, vision with stacked representations

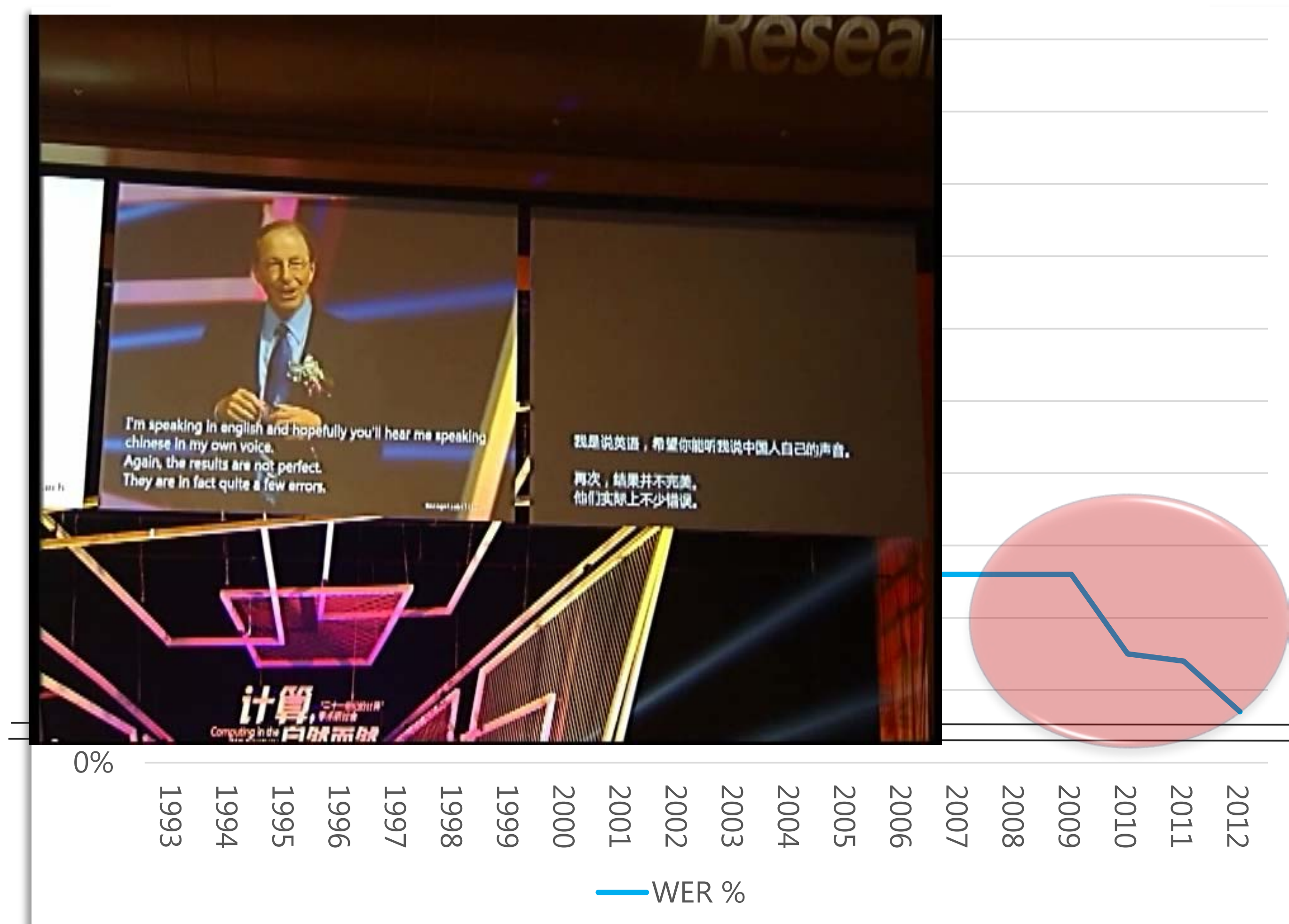
Conversational Speech: *Switchboard* challenge



Renaissance of Familiar Methods

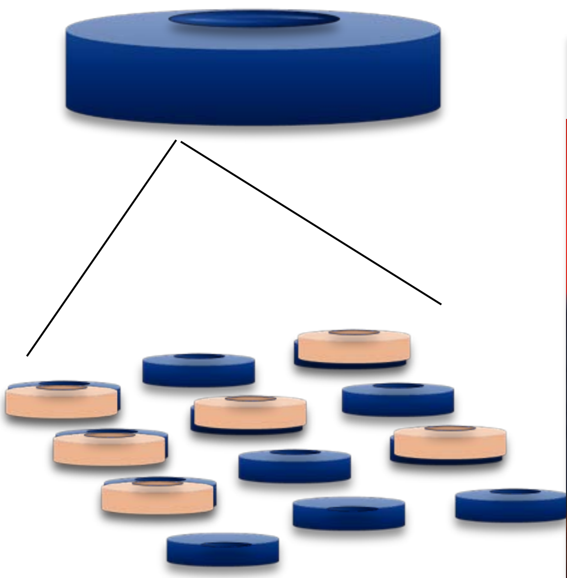
Pursuit of speech, vision with stacked representations

Conversational Speech: *Switchboard* challenge

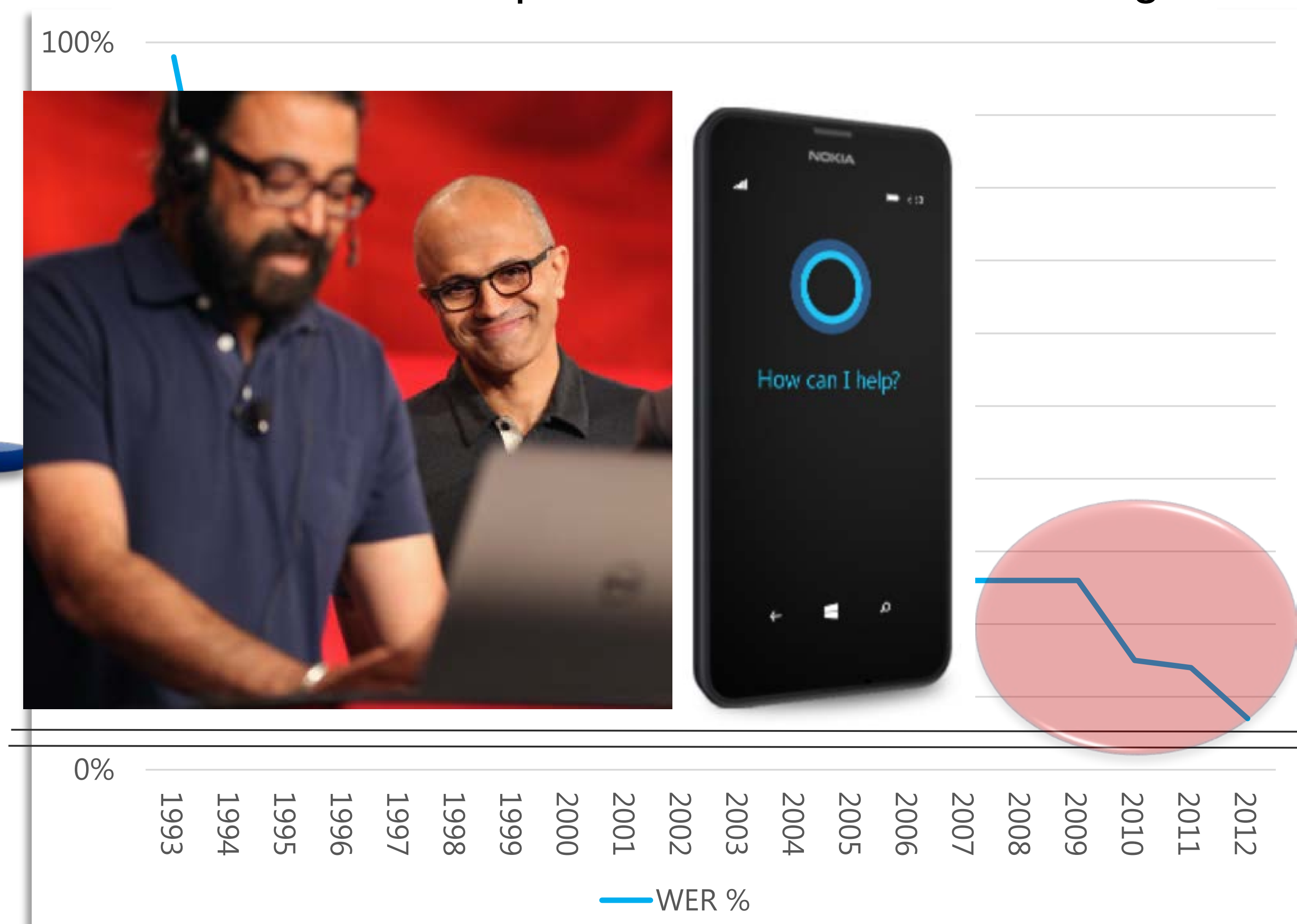


Renaissance of Familiar Methods

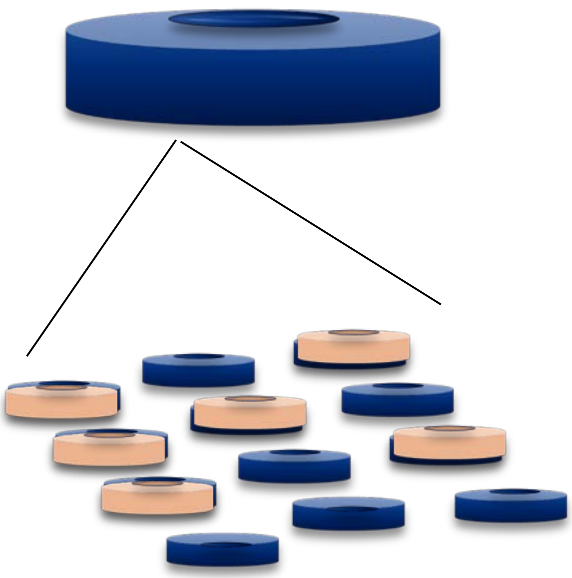
Pursuit of speech, vision with stacked representations

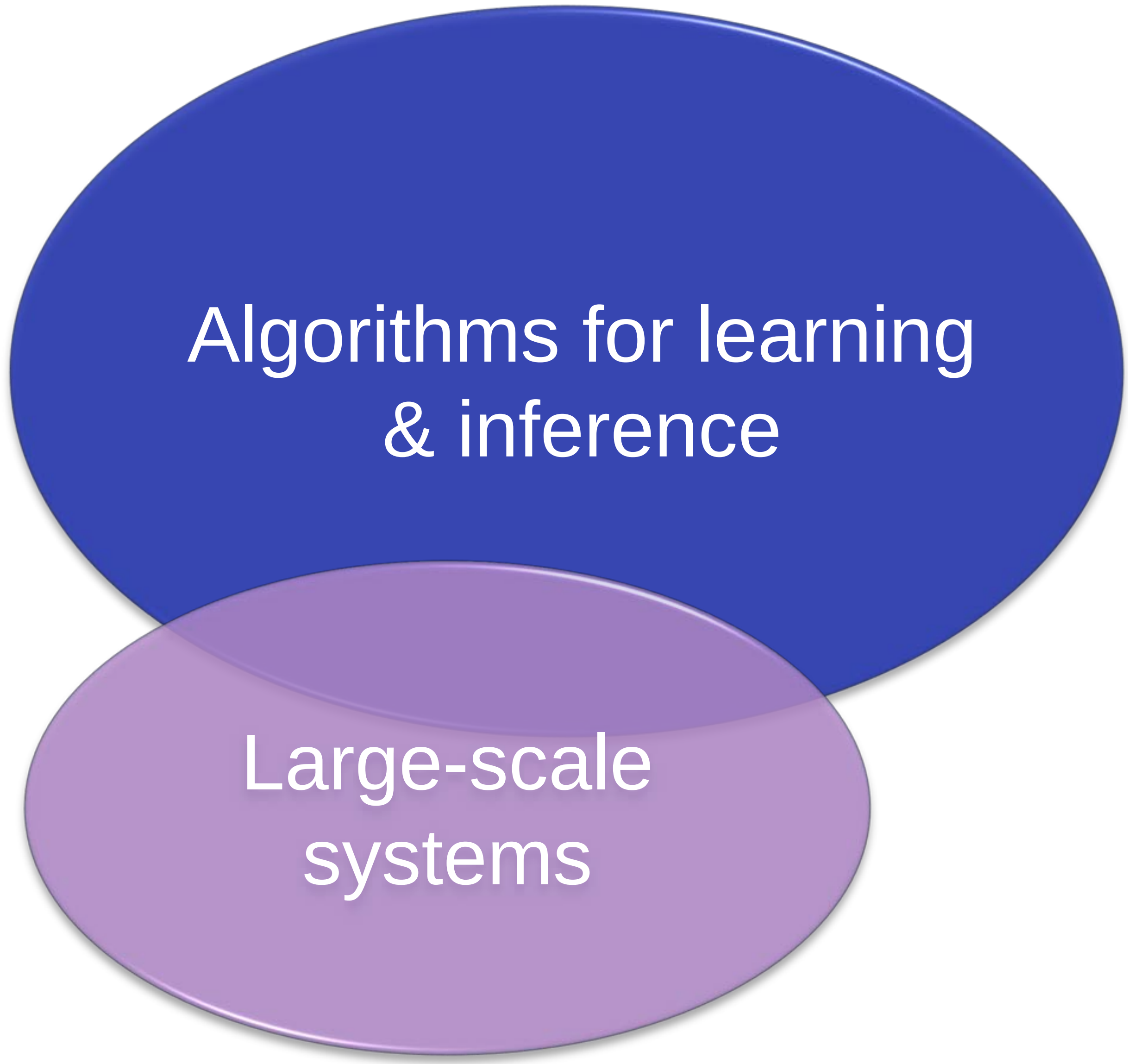


Conversational Speech: *Switchboard* challenge



Data, Learning, and Systems





Beauty and the Bottleneck

Hekaton: Database service

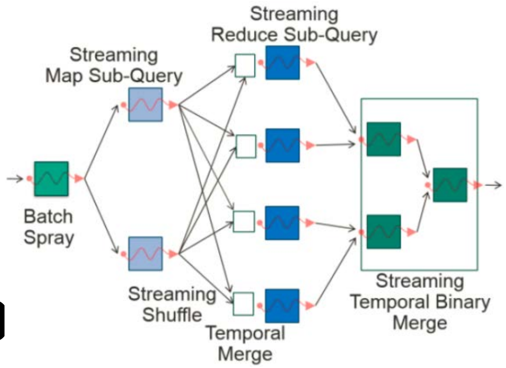
In-memory, manycore, latch-free:

30x speed-up

Trill: Streaming analytics

Column-oriented batches, P3 sort:

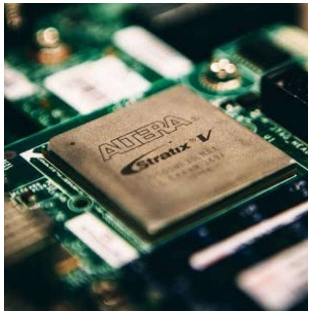
2-4 orders of magnitude speed-up



Catapult: Data center search perf.

Speed-ups via FPGA

40x speed-up



Data Science for Social Good

Transportation

Clinical medicine

Public health

Inference about Traffic

Smartflow, UAI 2005

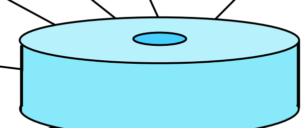
Multiple views on traffic

Weather

Major events

Incident reports

Operator ID: Nick
Heading: INCIDENT
Message: INCIDENT INFORMATION
Cleared 1637: I-405 SB JS I-90 ACC BLK RL CCTV 1623 – WSP, FIR ON SCENE

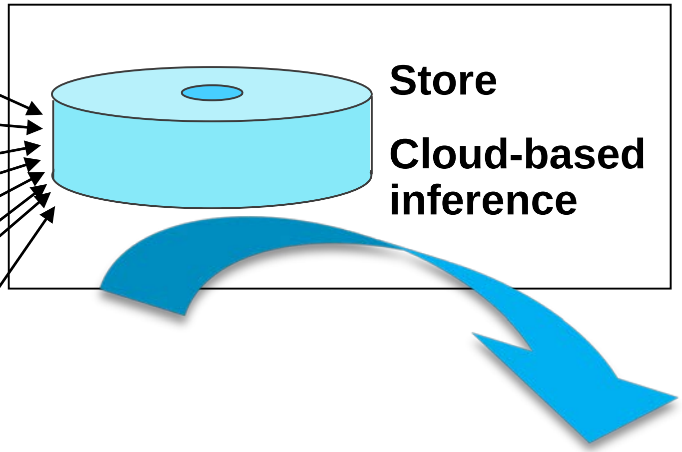


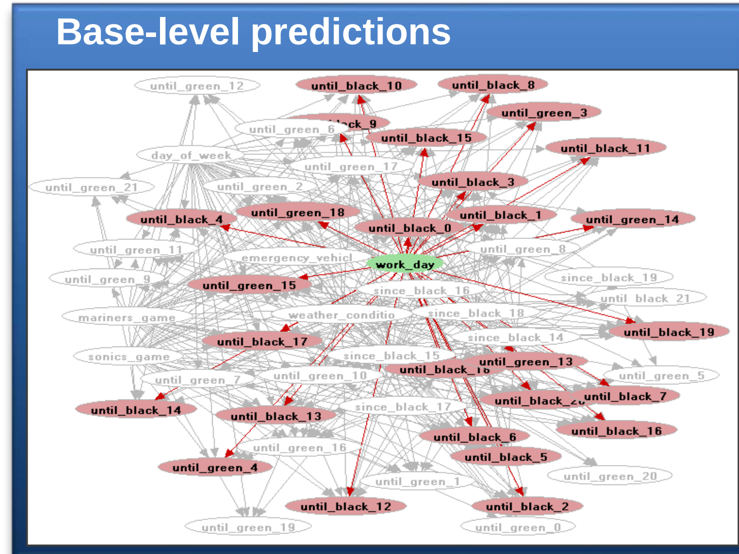
- **Event store**
- **Learning**
- **Reasoning**

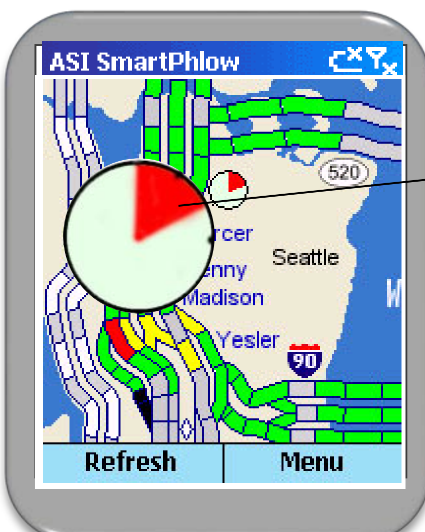
Forecasting Future Traffic

- System-wide status & dynamics
- Incident reports
- Sporting events
- Weather
- Time of day
- Day of week
- Season
- Holiday status

Traffic forecasting service





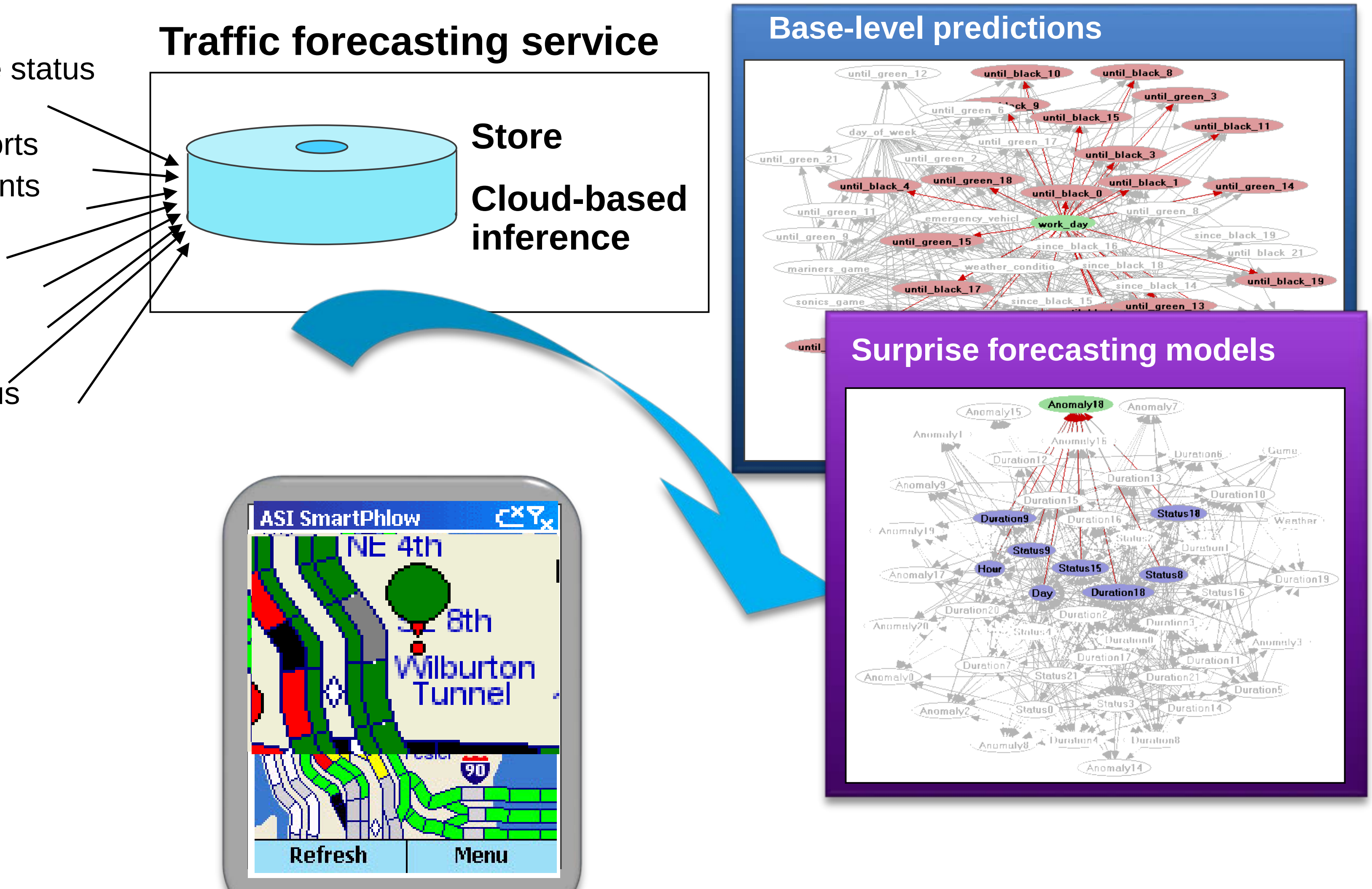


Max likely duration

With J. Apacible, P. Koch, M. Subramani

Forecasting Future Traffic

- System-wide status & dynamics
- Incident reports
- Sporting events
- Weather
- Time of day
- Day of week
- Season
- Holiday status



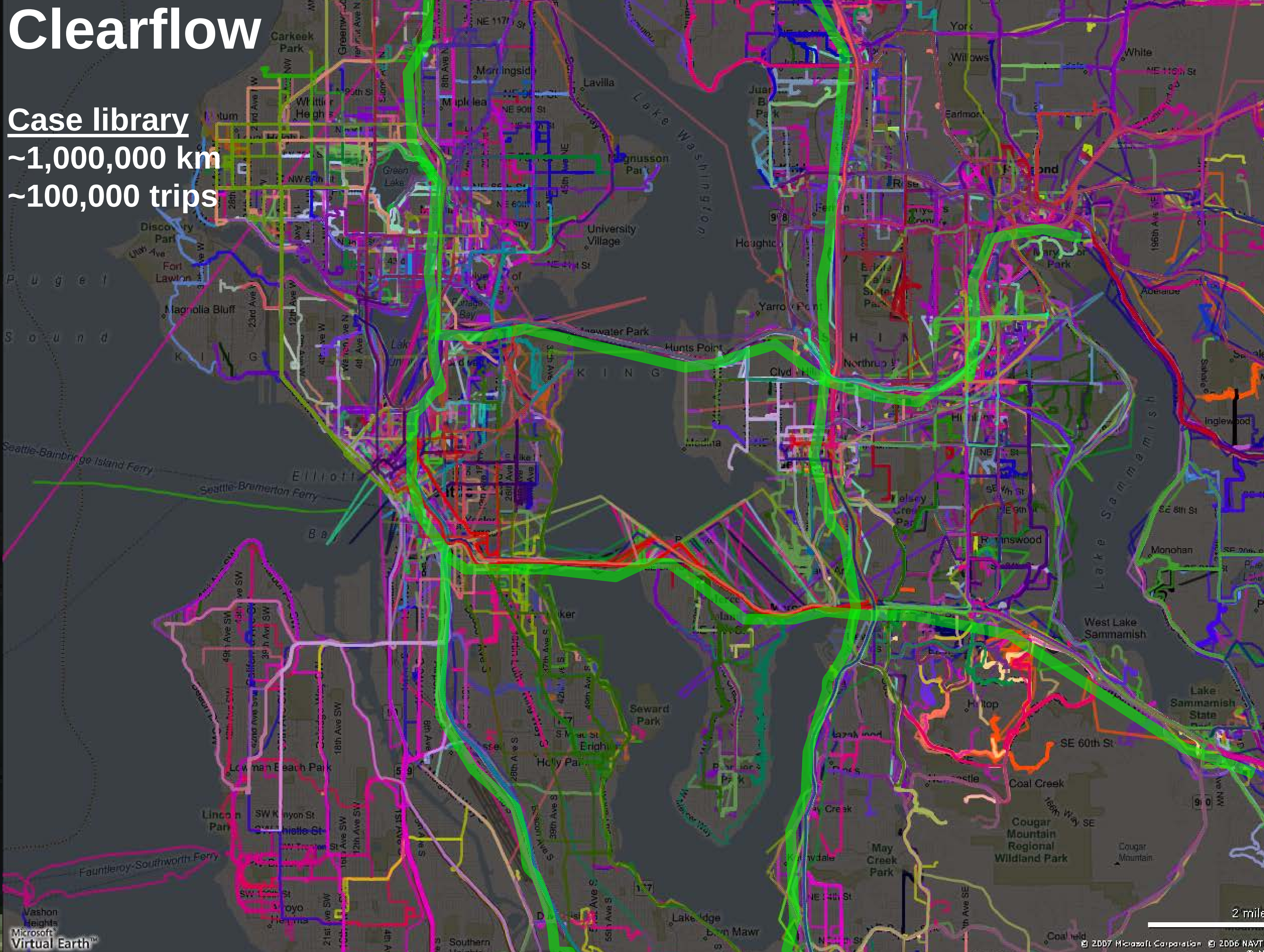
With J. Apacible, P. Koch, M. Subramani

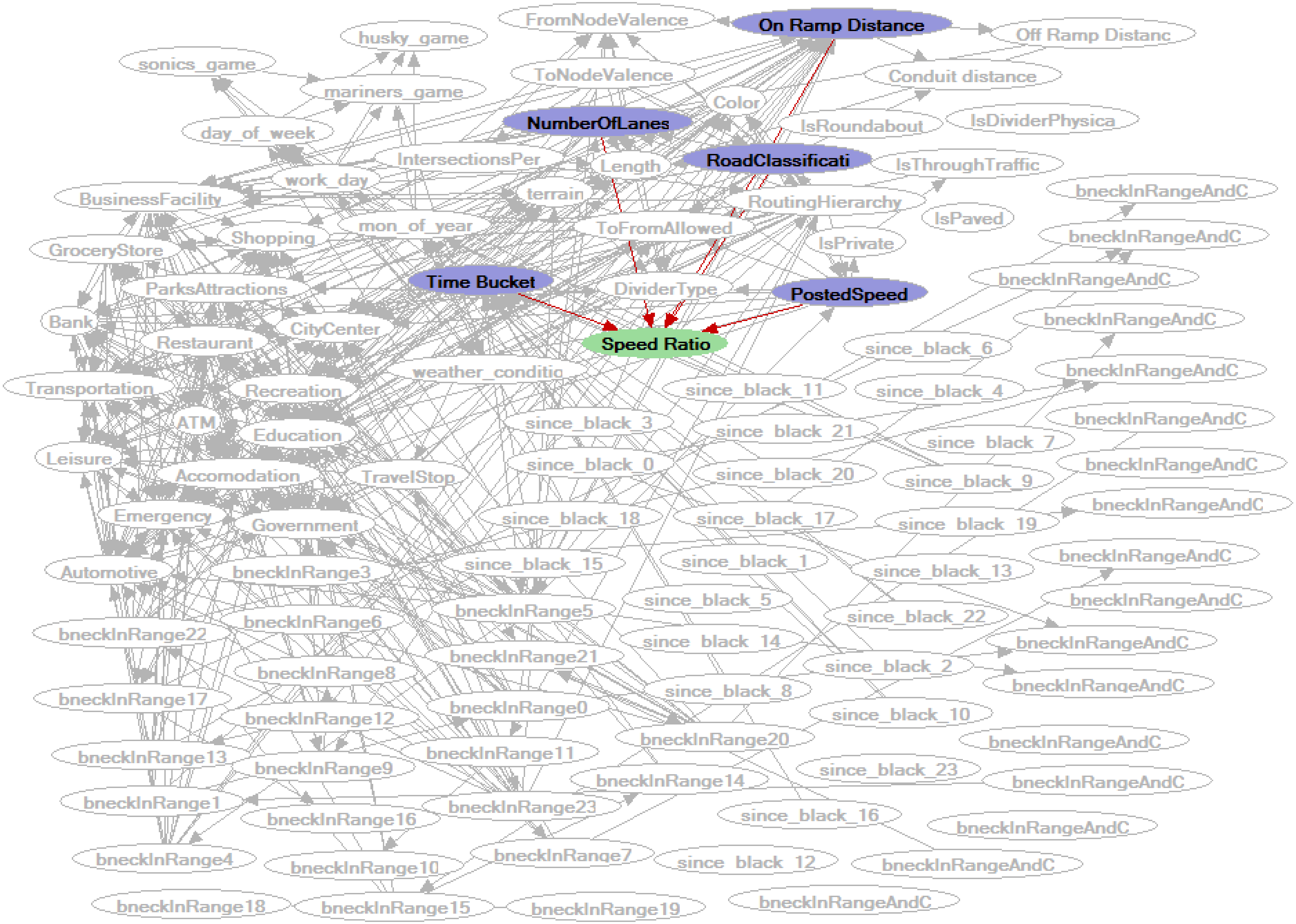
Clearflow

Case library

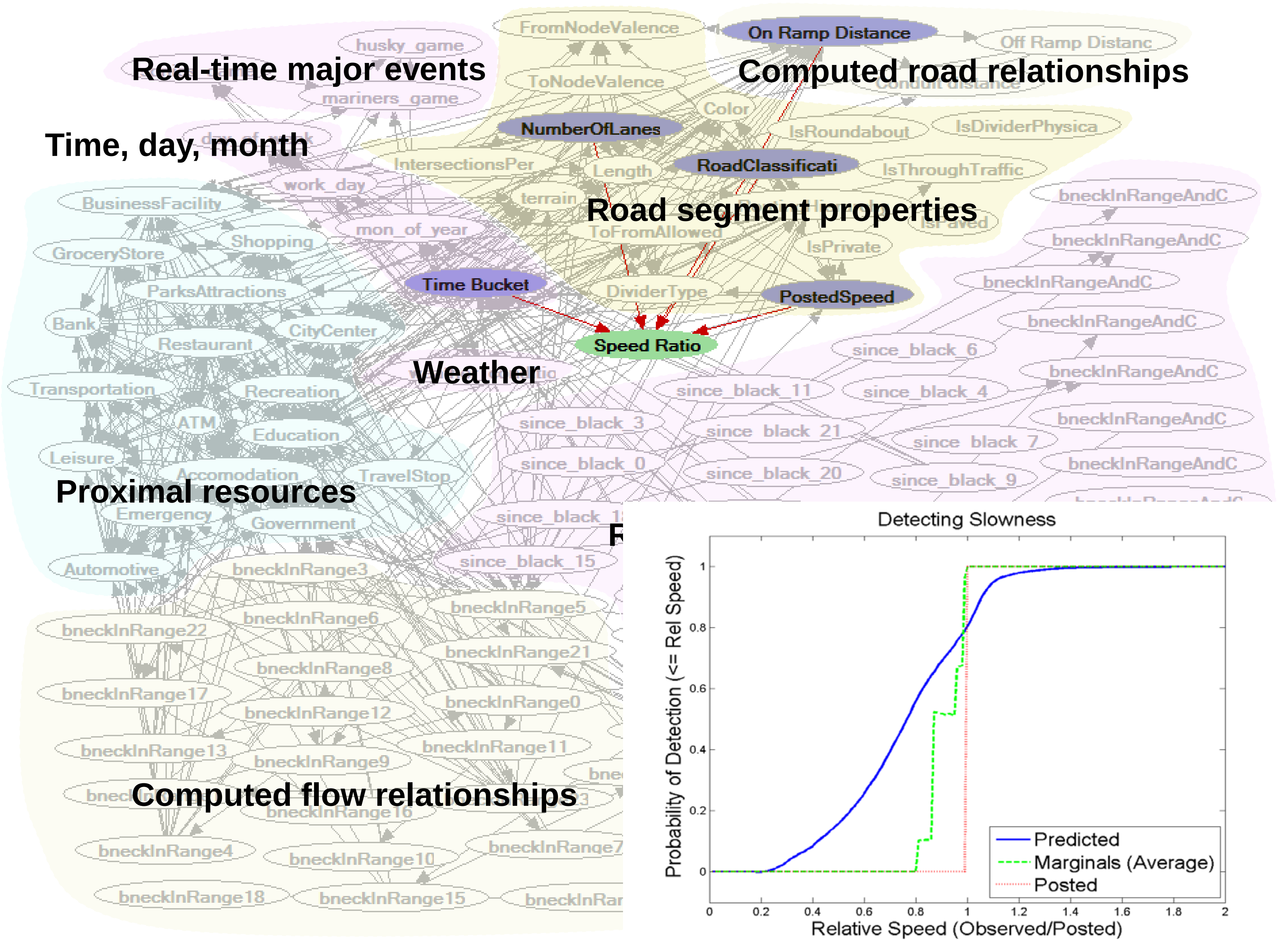
~1,000,000 km

~100,000 trips





FromNodeValence
On Ramp Distance
Off Ramp Distanc
husky_game
sonics_game
ToNodeValence
Conduit distance
mariners_game
Color
day_of_week
NumberOfLanes
IsRoundabout
IsDividerPhysica
IntersectionsPer
Length
RoadClassificati
IsThroughTraffic
work_day
BusinessFacility
terrain
RoutingHierarchy
bneckInRangeAndC
IsPaved
mon_of_year
ToFromAllowed
Shopping
bneckInRangeAndC
GroceryStore
IsPrivate
Time Bucket
bneckInRangeAndC
ParksAttractions
DividerType
PostedSpeed
Bank
bneckInRangeAndC
CityCenter
Restaurant
Speed Ratio
since_black_6
weather_conditio
bneckInRangeAndC
Transportation
since_black_11
Recreation
since_black_4
ATM
since_black_3
bneckInRangeAndC
Education
since_black_21
since_black_7
Leisure
since_black_0
bneckInRangeAndC
Accomodation
TravelStop
since_black_20
since_black_9
bneckInRangeAndC
Emergency
since_black_18
since_black_17
since_black_19
Government
since_black_1
bneckInRangeAndC
Automotive
bneckInRange3
since_black_15
since_black_13
since_black_5
bneckInRangeAndC
bneckInRange5
bneckInRange6
since_black_22
bneckInRange22
since_black_14
bneckInRangeAndC
bneckInRange21
since_black_2
bneckInRange8
bneckInRangeAndC
since_black_8
bneckInRange17
bneckInRange0
since_black_10
bneckInRange12
bneckInRange20
bneckInRangeAndC
bneckInRange13
bneckInRange11
since_black_23
bneckInRange9
bneckInRange14
bneckInRangeAndC
bneckInRange1
bneckInRange23
bneckInRange16
since_black_16
bneckInRangeAndC
bneckInRange4
bneckInRange7
since_black_12
bneckInRangeAndC
bneckInRange10
bneckInRange18
bneckInRange15
bneckInRange19
bneckInRangeAndC


Real-time major events
Computed road relationships
Time, day, month
Road segment properties
Weather
Proximal resources
Computed flow relationships
On Ramp Distance
NumberOfLanes
RoadClassificati
Time Bucket
PostedSpeed
Speed Ratio
Detecting Slowness
Probability of Detection (<= Rel Speed)
Relative Speed (Observed/Posted)
Predicted
Marginals (Average)
Posted
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Microsoft Introduces Tool for Avoiding Traffic Jams

By JOHN MARKOFF
Published: April 10, 2008

SAN FRANCISCO — Microsoft on Thursday plans to introduce a Web-based service for driving directions that incorporates complex software models to help users avoid traffic jams.

Related

Times Topics: Microsoft Corporation

The new service's software technology, called Clearflow, was developed over the last five years by a group of artificial-intelligence researchers at the company's Microsoft Research laboratories. It is an ambitious attempt to apply machine-learning techniques to the problem of traffic congestion. The system is intended to reflect the complex traffic interactions that occu traffic backs up on freeways and spills over onto city streets.

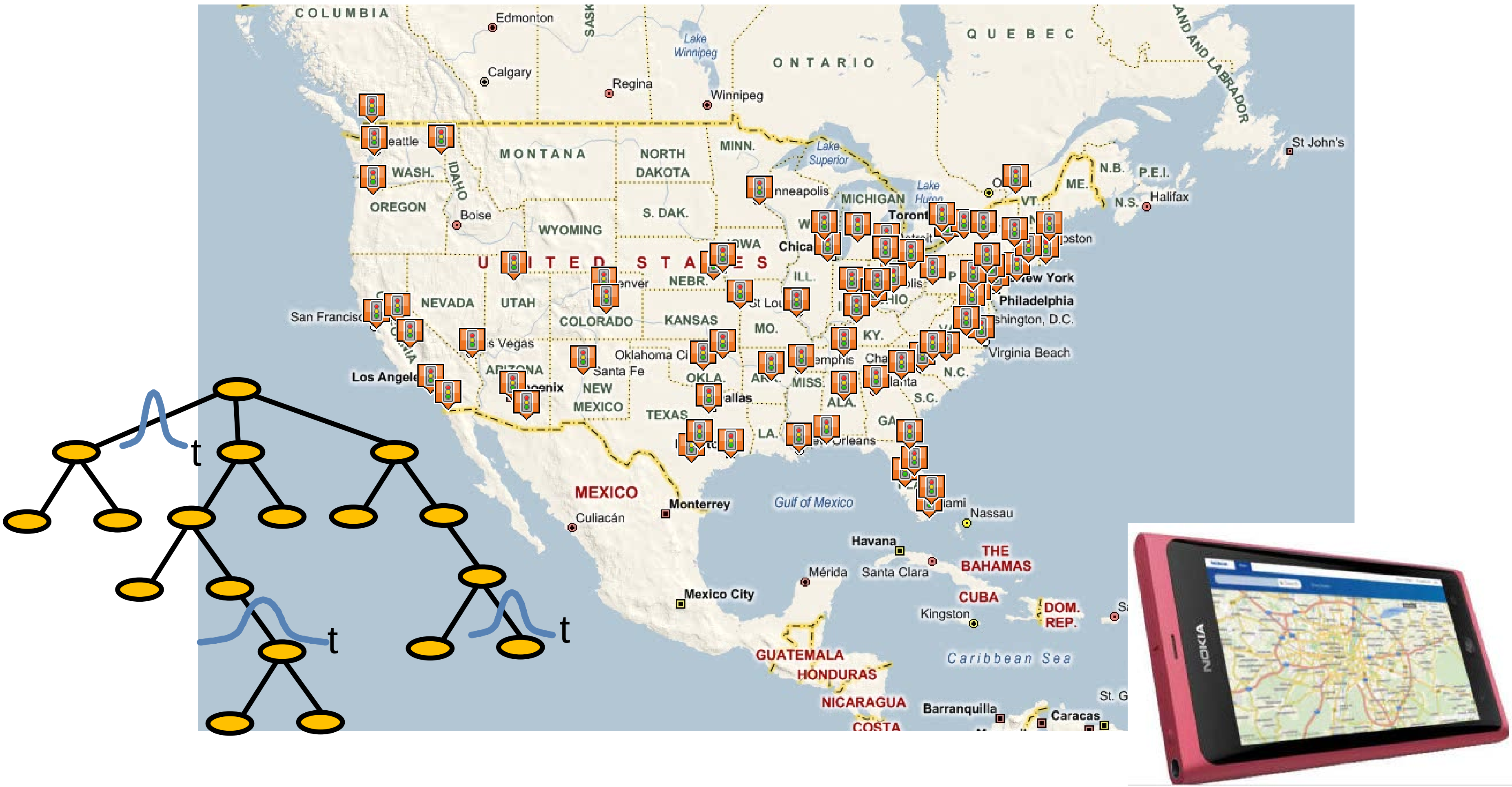
The Clearflow system will be freely available as part of the company's Live.com site (maps.live.com) for 72 cities in the United States. Microsoft says it will give drivers alternative route information that is more accurate and attuned to current traffic pa on both freeways and side streets.

Microsoft now considers surface street traffic as well as freeway speeds in its routing.

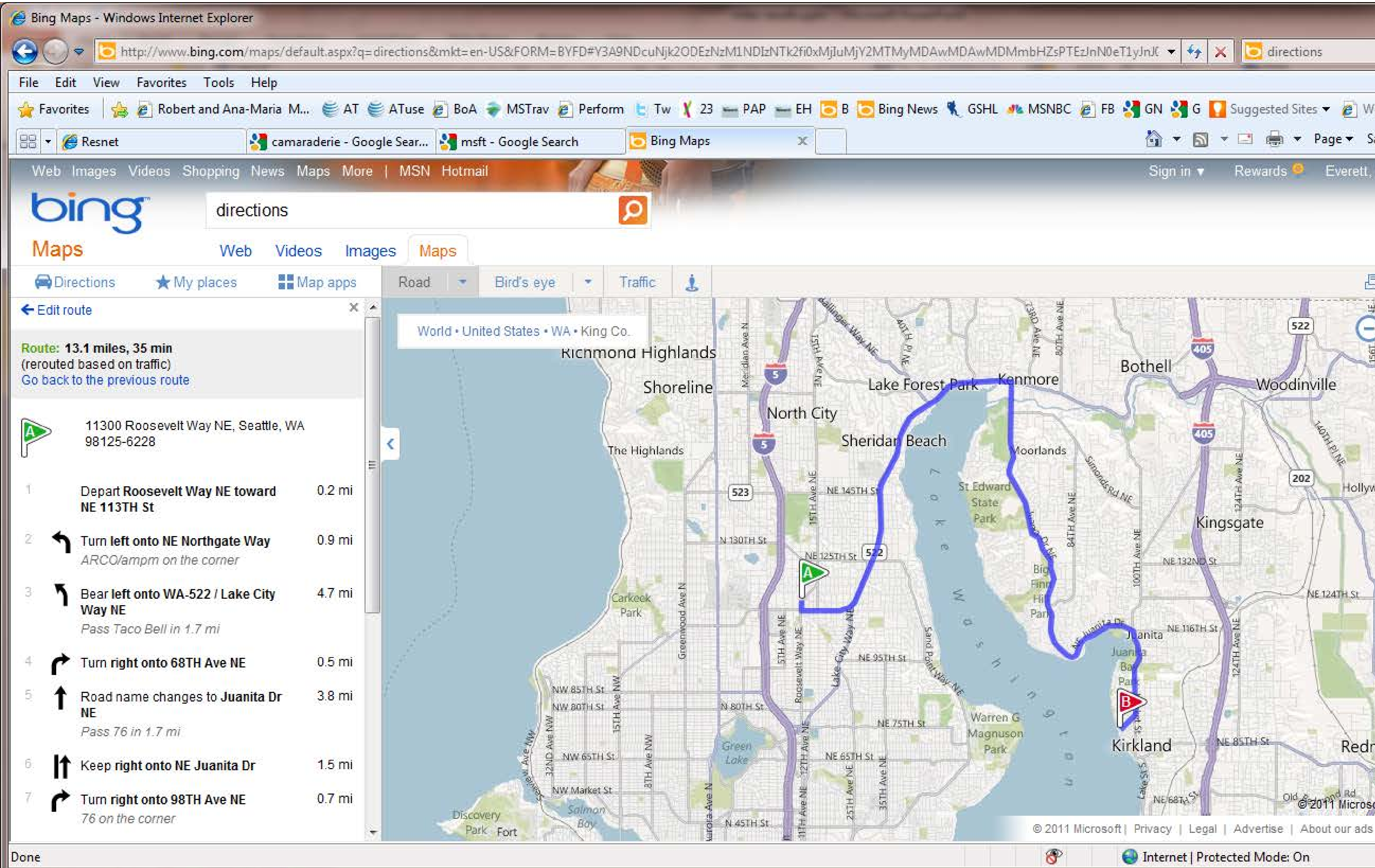
Traffic-Sensitive Routing

72 cities across North America

Flows assigned to ~60 million streets *every few minutes*

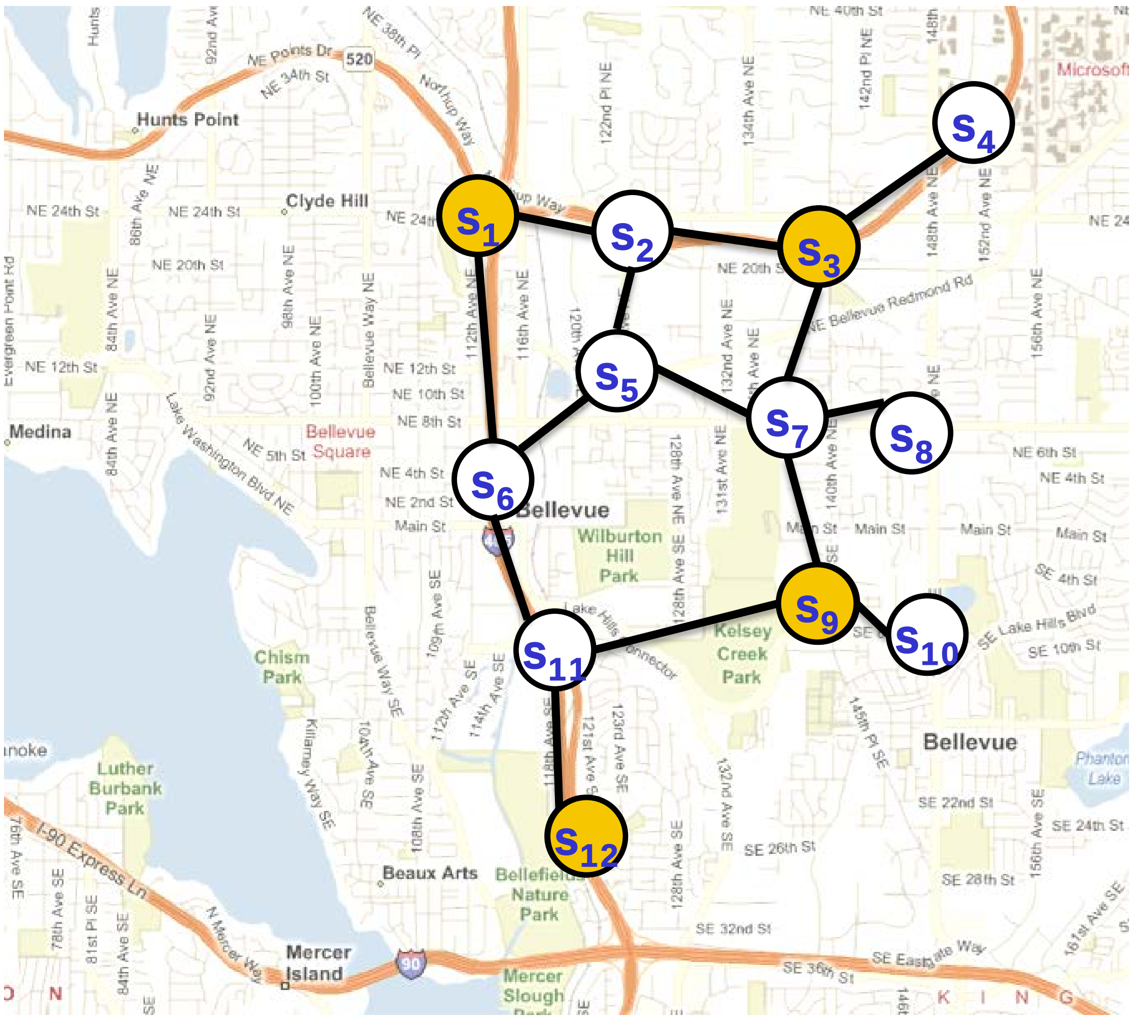


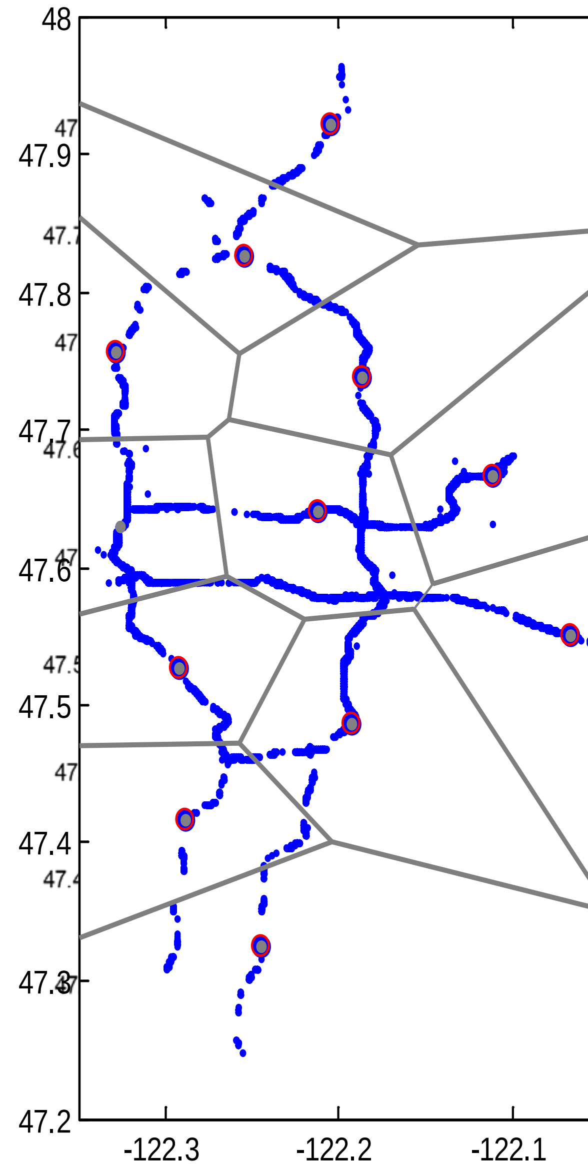
Traffic-Sensitive Routing



Community Sensing

Utilitarian: Demand-weighted value





Krause, H., et al.

Community Sensing

Utilitarian: Demand-weighted value

Phenomenon

Variables of spatiotemporal process

$$\mathrm{Var}(\mathcal{X}_s \mid \mathcal{X}_\mathcal{A} = \mathbf{x}_\mathcal{A}) \quad \mathrm{Var}(\mathcal{X}_s) - \mathrm{Var}(\mathcal{X}_s \mid \mathcal{X}_\mathcal{A} = \mathbf{x}_\mathcal{A})$$

Demand Model

Population needs

$$R(\mathcal{A}) = \sum_{s \in \mathcal{V}} \mathbb{E}\left[\mathcal{D}_s(\mathrm{Var}(\mathcal{X}_s) - \mathrm{Var}(\mathcal{X}_s \mid \mathcal{X}_\mathcal{A}))\right]$$

Sensor Availability

Sharing Preferences

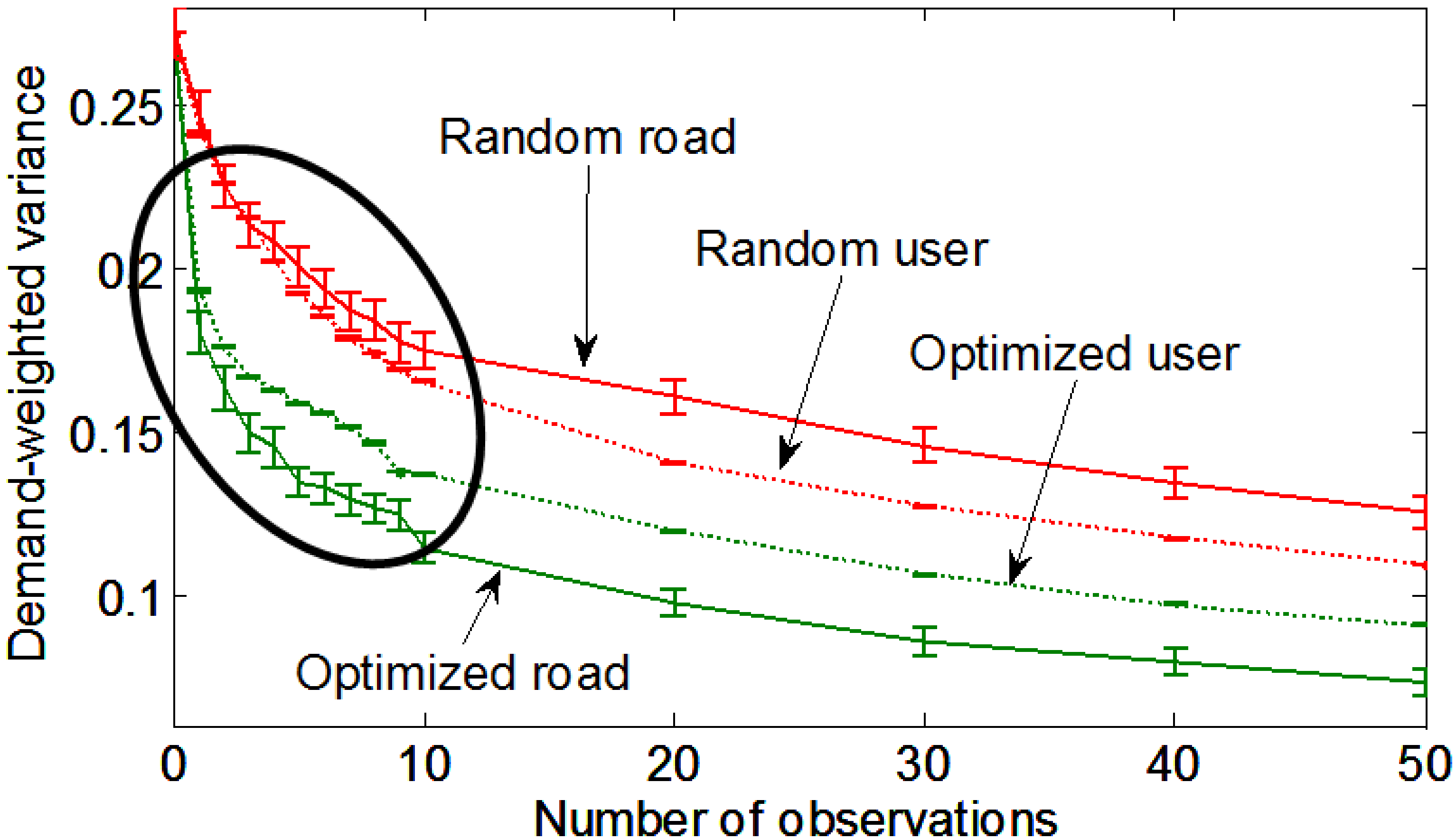
Avail. of observations B at locations A

$$P(\mathcal{A} \mid \mathcal{B})$$

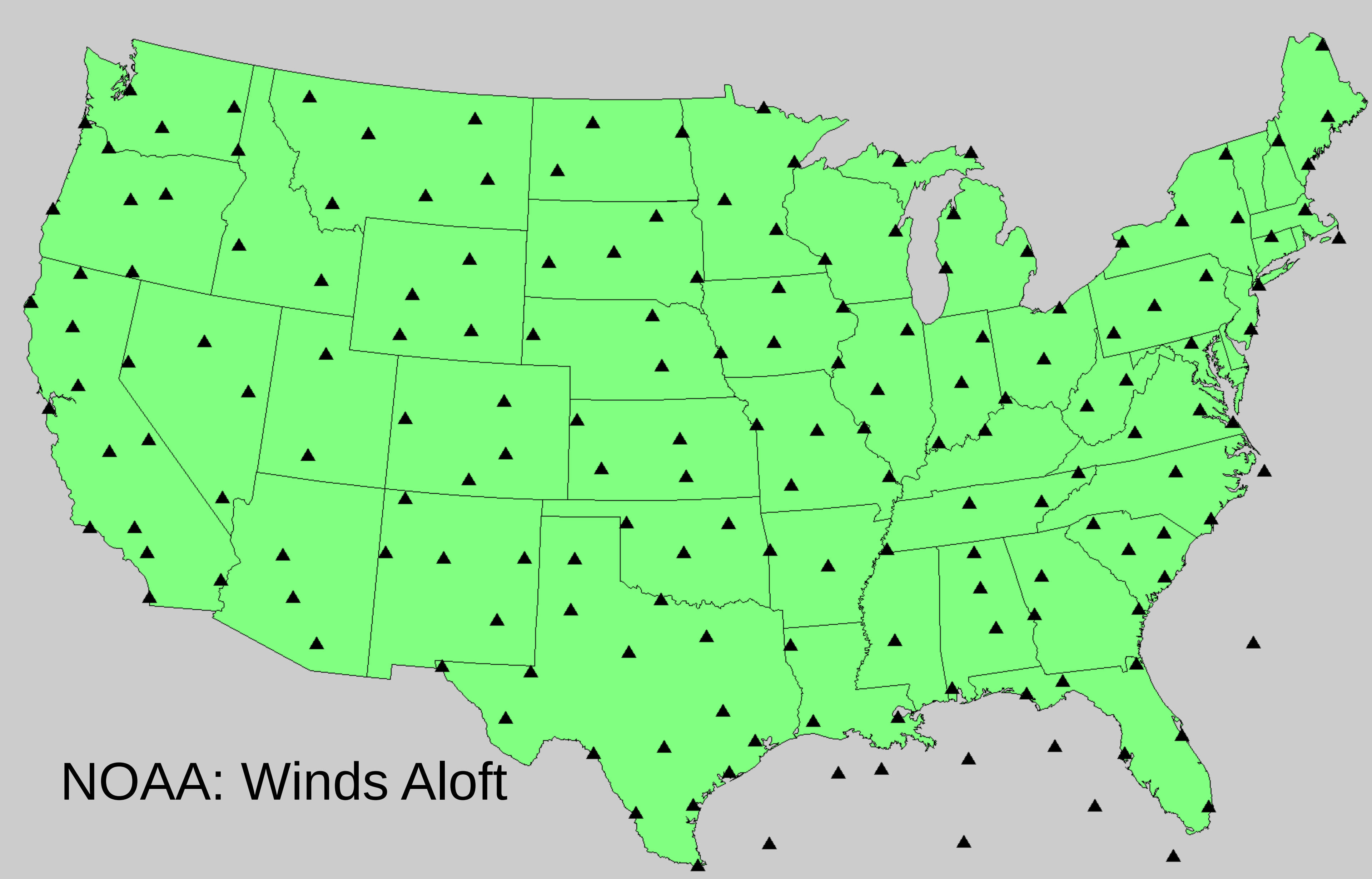
$$F(\mathcal{B}) = \mathbb{E}_{\mathcal{A}|\mathcal{B}}[R(\mathcal{A})] = \sum_{\mathcal{A}} P(\mathcal{A} \mid \mathcal{B}) R(\mathcal{A})$$

Community Sensing

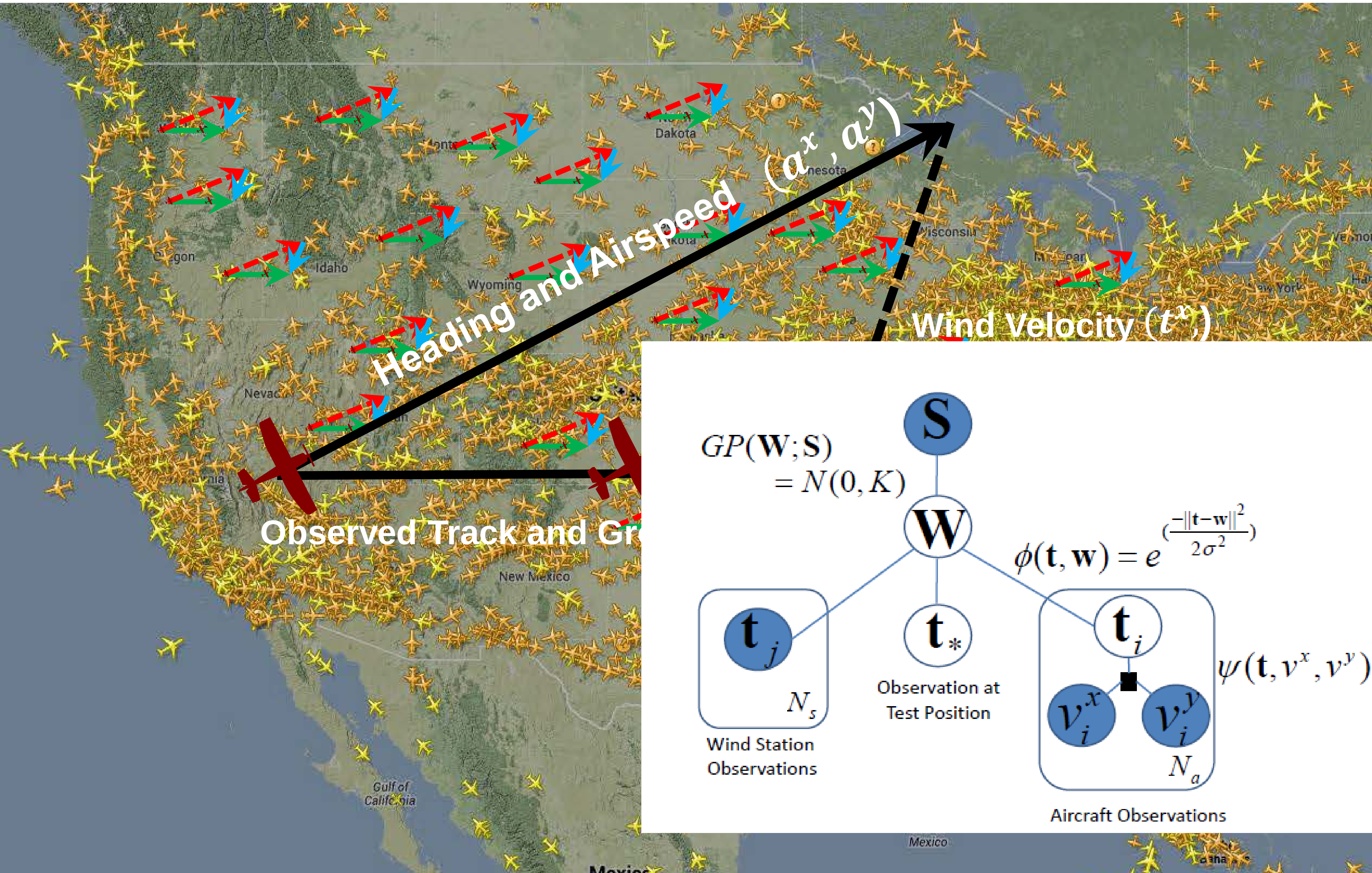
Utilitarian: Demand-weighted value



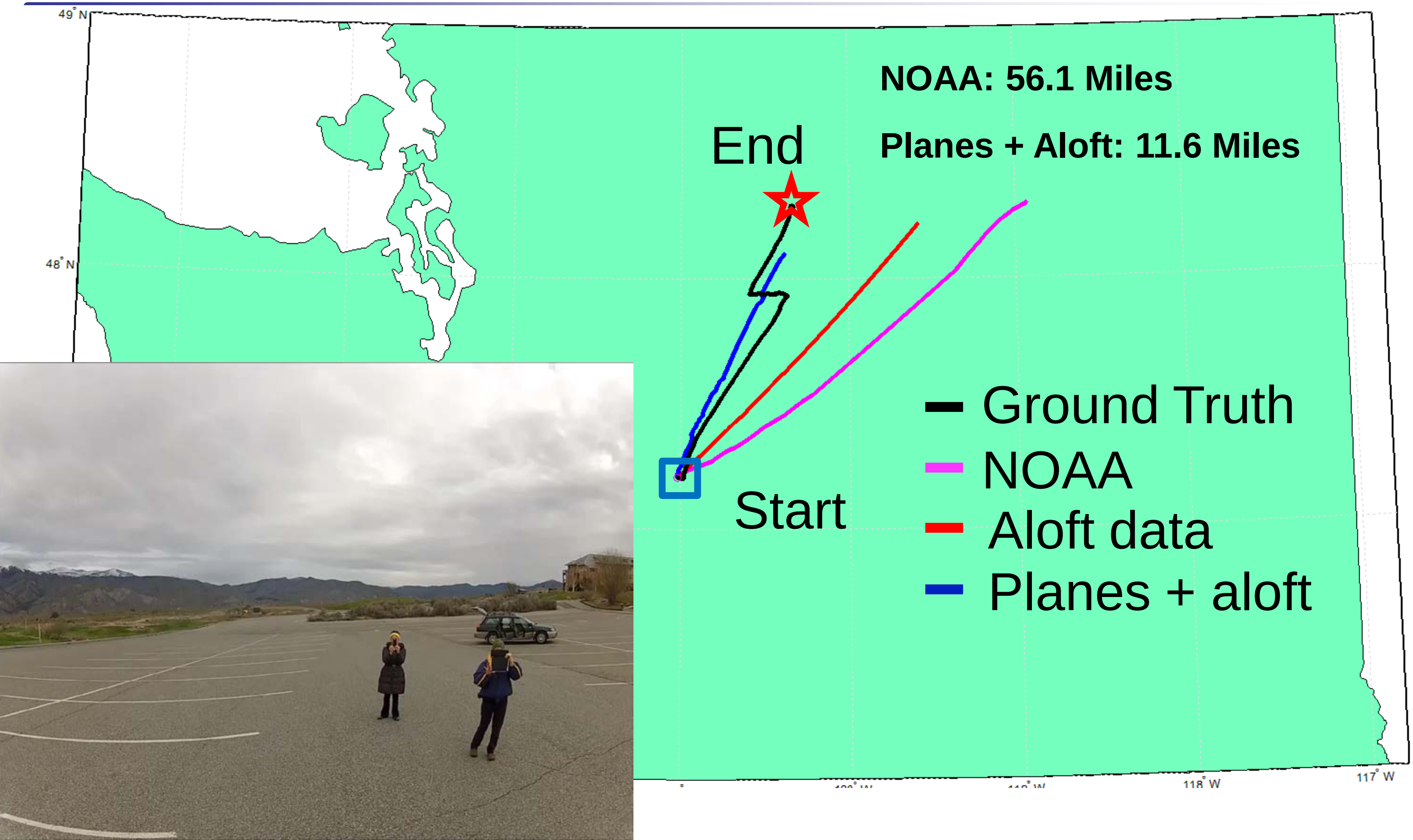
Aiming for the Sky: Aviation



Thousands of Wind Sensors



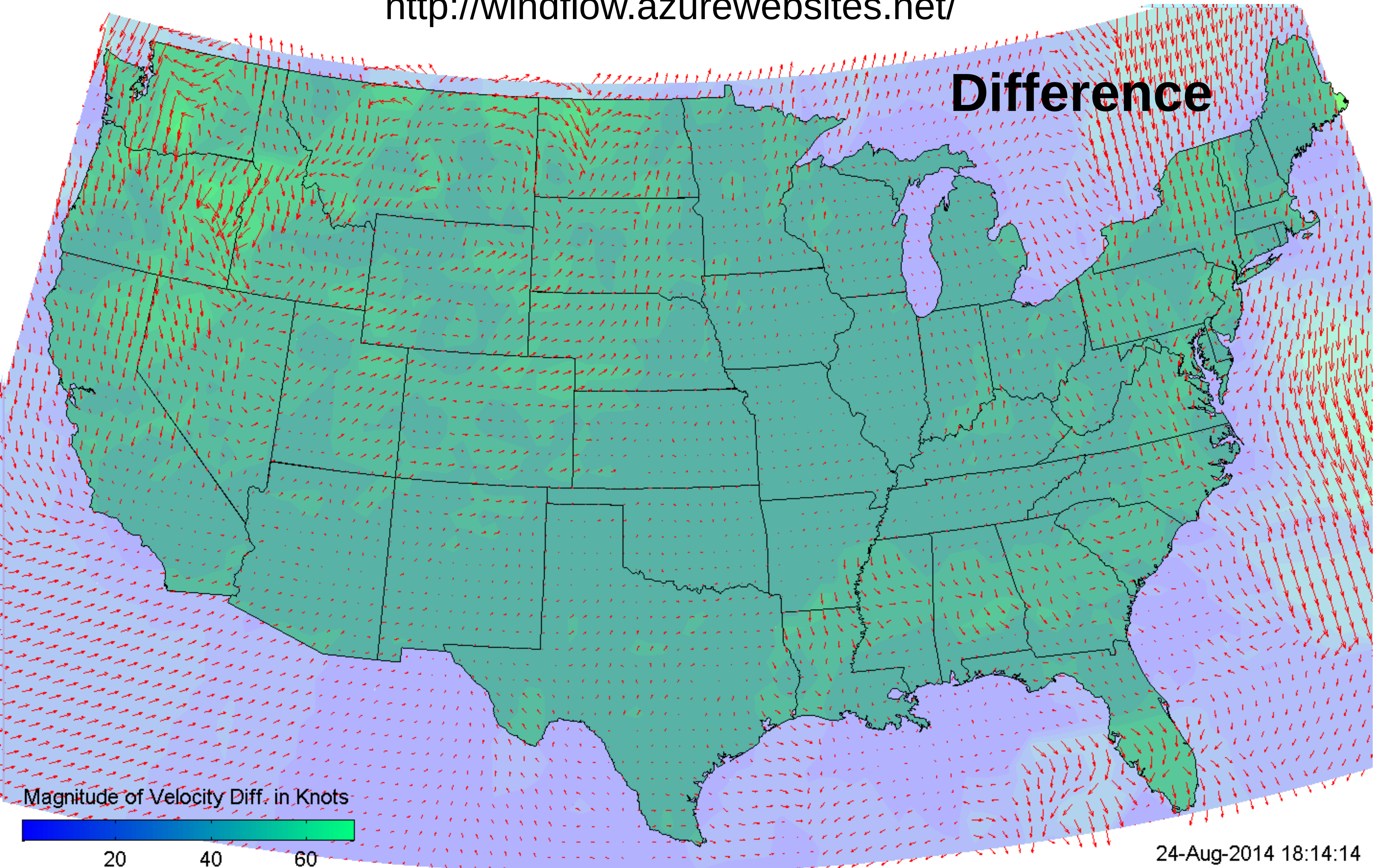
Studies


NOAA: 56.1 Miles
Planes + Aloft: 11.6 Miles
End
Start
Ground Truth
NOAA
Aloft data
Planes + aloft
49° N
48° N
118° W
117° W

Windflow

Azure Cloud Service:

http://windflow.azurewebsites.net/



Clinical Medicine

Rich dataset: All visits,15 years of data

- Admissions, discharge, transfer (ADT)
- Chief complaint in free text
- Age, gender, demographics
- Diagnosis codes (ICD-9)
- Lab results and studies
- Medications
- Vital signs
- Procedures
- Locations in hospital
- Admitting and attending MD codes
- Fees and billing

~30,000 variables available in dataset

Readmissions Challenge

The NEW ENGLAND JOURNAL of MEDICINE

Sign in | Get NEJM's E-Mail Table of Contents — Free | Subscribe

SPECIAL ARTICLE

◄ Previous Volume 360:1418-1428 April 2, 2009 Number 14 Next ►

Rehospitalizations among Patients in the Medicare Fee-for-Service Program

Stephen F. Jencks, M.D., M.P.H., Mark V. Williams, M.D., and Eric A. Coleman, M.D., M.P.H.

ABSTRACT

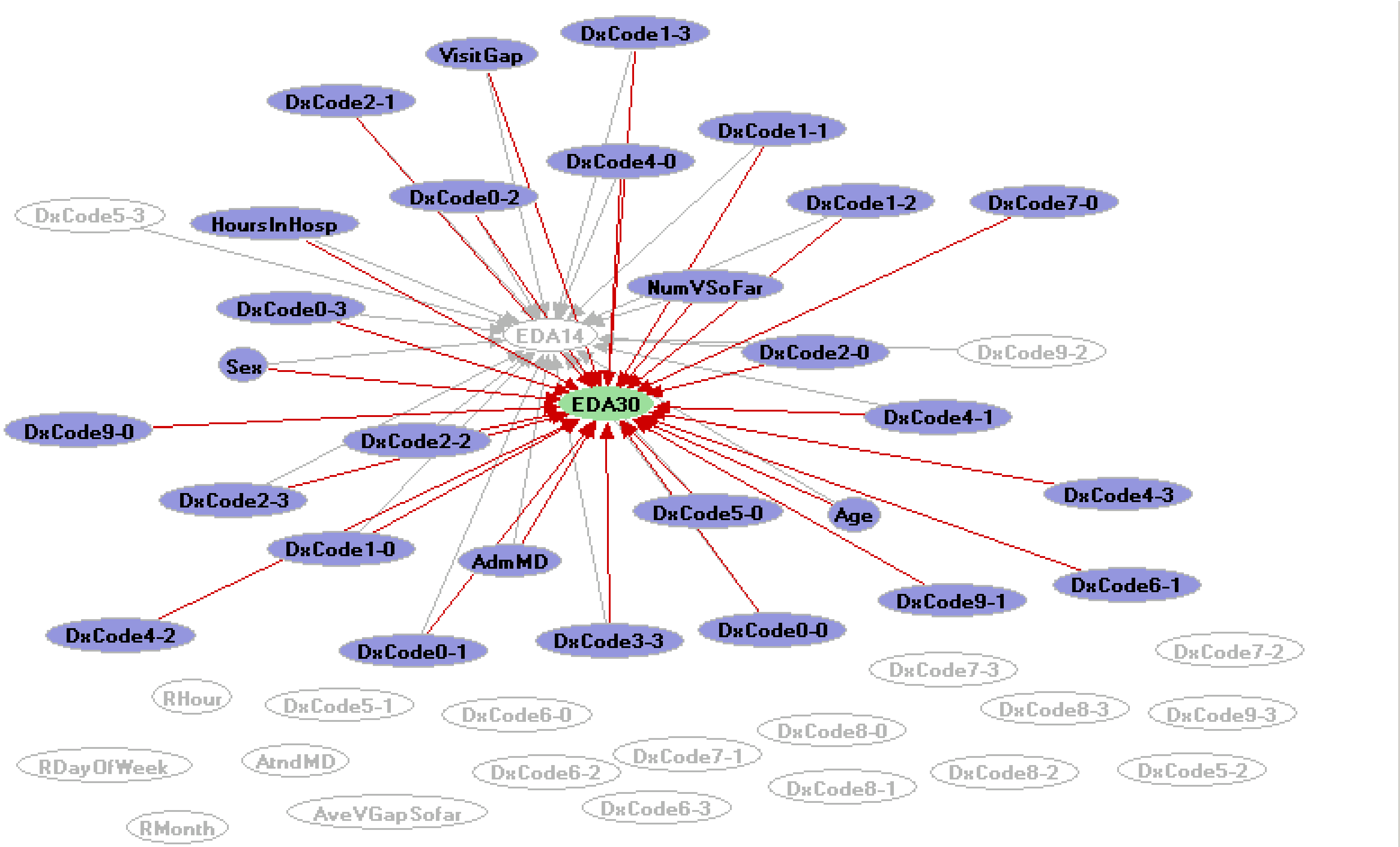
Background Reducing rates of rehospitalization has attracted attention from policymakers as a way to improve quality of care and reduce costs. However, we have limited information on the frequency and patterns of rehospitalization in the United States to aid in planning the necessary changes.

Methods We analyzed Medicare claims data from 2003–2004 to describe the patterns of

- **~20% within 30 days**
- **~35% in 90 days**

➢ ***Estimated cost to Medicare (2004): $17.4 billion***

Predictive Model for Readmission



with P. Koch

Going Live

Readmissions Manager

Reducing Hospital Readmissions is an Impending Priority

Overview

One in five Medicare inpatients is readmitted within 30 days. The Centers for Medicare and Medicaid Services (CMS) considers 40%-75% of these readmissions to be preventable.

In October 2012, CMS will begin to track readmission and impose financial penalties on hospitals with higher–than–expected readmission rates for certain conditions. Other payers will certainly follow.

It is clear that hospital admissions and readmissions are becoming a critical parameter for tracking care delivery from both a financial and quality perspective.

Readmissions Manager for Microsoft Amalga is an innovative solution to help organizations address this very important business need.

At hospitals around the world…

ACCOUNT	ADMITDTTM	DISCHARGEDTTM	AGE	SEX	PROB_NUM_%	FACTOR
	12/03/2010 14:57	12/08/2010 18:03	62	F	37.9	Num past 6m visits = 6 to 10 / P
	12/08/2010 18:45	12/08/2010 18:45	74	M	32.72	stayed <1 day in the hospital / Pa
	11/16/2010 16:14	12/08/2010 18:50	48	M	30.83	Patient had dx = Chronic renal fa
	12/02/2010 13:49	12/08/2010 18:14	68	M	29.05	Patient had dx = Disorders of flui
	12/01/2010 05:26	12/08/2010 18:55	44	M	28.54	
	12/01/2010 19:08	12/08/2010 18:13	61	M	27.36	Patient had dx = Acute renal failu
	11/30/2010 21:50	12/08/2010 18:52	70	M	18.05	Patient had dx = Other personal
	12/08/2010 08:51	12/08/2010 18:45	68	M	16.57	stayed <1 day in the hospital
	12/03/2010 20:32	12/08/2010 17:50	80	M	16.18	Patient had dx = Disorders of flui
	12/01/2010 01:13	12/08/2010 18:06	79	M	15.52	
	12/08/2010 18:39	12/08/2010 18:39	22	F	14.53	stayed <1 day in the hospital / Av
	12/08/2010 19:01	12/08/2010 19:01	25	F	14.42	stayed <1 day in the hospital / Pa
	12/08/2010 18:05	12/08/2010 18:05	24	M	14.39	stayed <1 day in the hospital
	12/08/2010 18:26	12/08/2010 18:26	53	F	13.59	stayed <1 day in the hospital / 44

Challenge: Interpretability

DISCHARGEDTTM	AGE	SEX	PROB_NUM_%	FACTORS_PRO_READMISSION
12/08/2010 18:03	62	F	37.9	Num past 6m visits = 6 to 10 / Patient had dx = Disorders of fluid, electrolyte, an
12/08/2010 18:45	74	M	32.72	stayed <1 day in the hospital / Patient had dx = Disorders of fluid, electrolyte, and
12/08/2010 18:50	48	M	30.83	Patient had dx = Chronic renal failure / 44 < Age < 60
12/08/2010 18:14	68	M	29.05	Patient had dx = Disorders of fluid, electrolyte, and acid-base balance / Patient ha
12/08/2010 18:55	44	M	28.54	
12/08/2010 18:13	61	M	27.36	Patient had dx = Acute renal failure / Patient had dx = Chronic renal failure
12/08/2010 18:52	70	M	18.05	Patient had dx = Other personal history presenting hazards to health / Patient ha
12/08/2010 18:45	68	M	16.57	stayed <1 day in the hospital
12/08/2010 17:50	80	M	16.18	Patient had dx = Disorders of fluid, electrolyte, and acid-base balance / Patient ha
12/08/2010 18:06	79	M	15.52	
12/08/2010 18:39	22	F	14.53	stayed <1 day in the hospital / Ave gap of past yr visits = between 15 and 30 days
12/08/2010 19:01	25	F	14.42	stayed <1 day in the hospital / Patient had dx = Other personal history presentin
12/08/2010 18:05	24	M	14.39	stayed <1 day in the hospital
12/08/2010 18:26	53	F	13.59	stayed <1 day in the hospital / 44 < Age < 60

Interpretability

Considering human interpretability

Procedures that allow end users to understand contribution of individual features

What influence does changing observations x have if other values are not changed?

Interpretability--Power Tradeoff

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

$$y = f_1(x_1) + \ldots + f_n(x_n)$$

$$y = f(x_1, \ldots, x_n)$$

Y. Lou, R. Caruana, J. Gehrke, and G. Hooker. Accurate Intelligible Models with Pairwise Interactions. In KDD, 2013.

Interpretability--Power Tradeoff

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$$

$$y = f_1(x_1) + ... + f_n(x_n)$$

$$y = \sum_i f_i(x_i) + \underline{\sum_{ij} f_{ij}(x_i, x_j)}$$

$$y = \sum_i f_i(x_i) + \underline{\sum_{ij} f_{ij}(x_i, x_j)} + \underline{\sum_{ijk} f_{ijk}(x_i, x_j, x_k)}$$

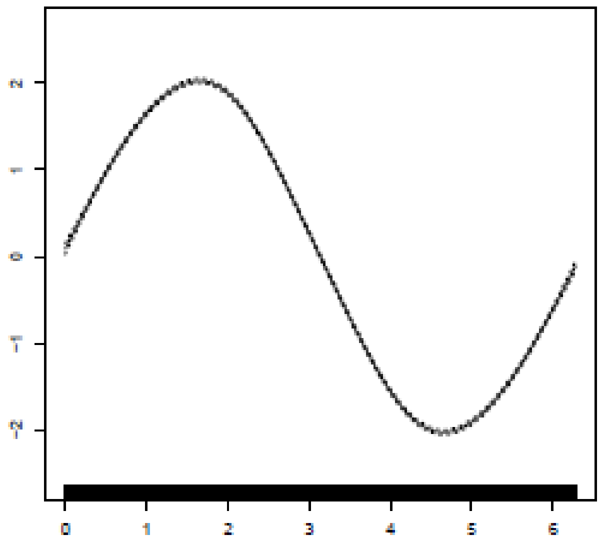
$$y = f(x_1, ..., x_n)$$

Y. Lou, R. Caruana, J. Gehrke, and G. Hooker. Accurate Intelligible Models with Pairwise Interactions. In KDD, 2013.

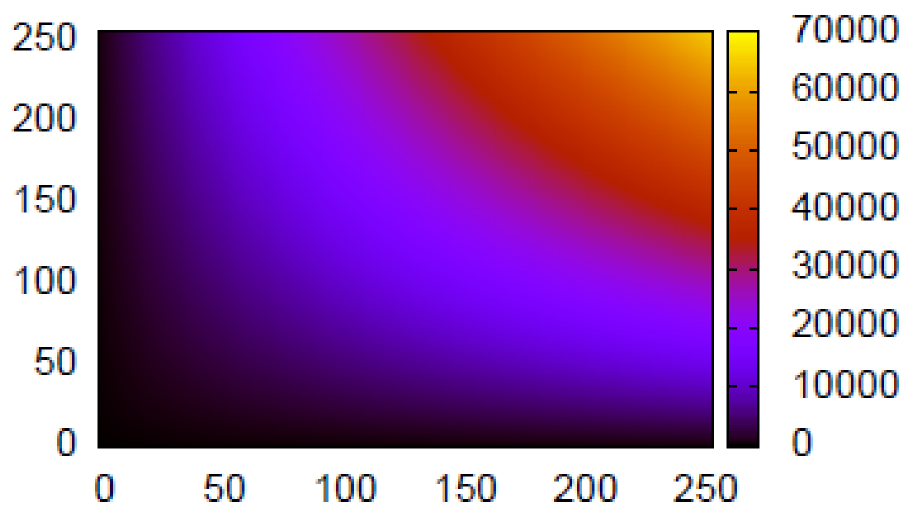
Capturing Key Interactions

Efficient means to identify pairwise interactions

$$y = \sum_i f_i(x_i) + \sum_{ij} f_{ij}(x_i, x_j)$$



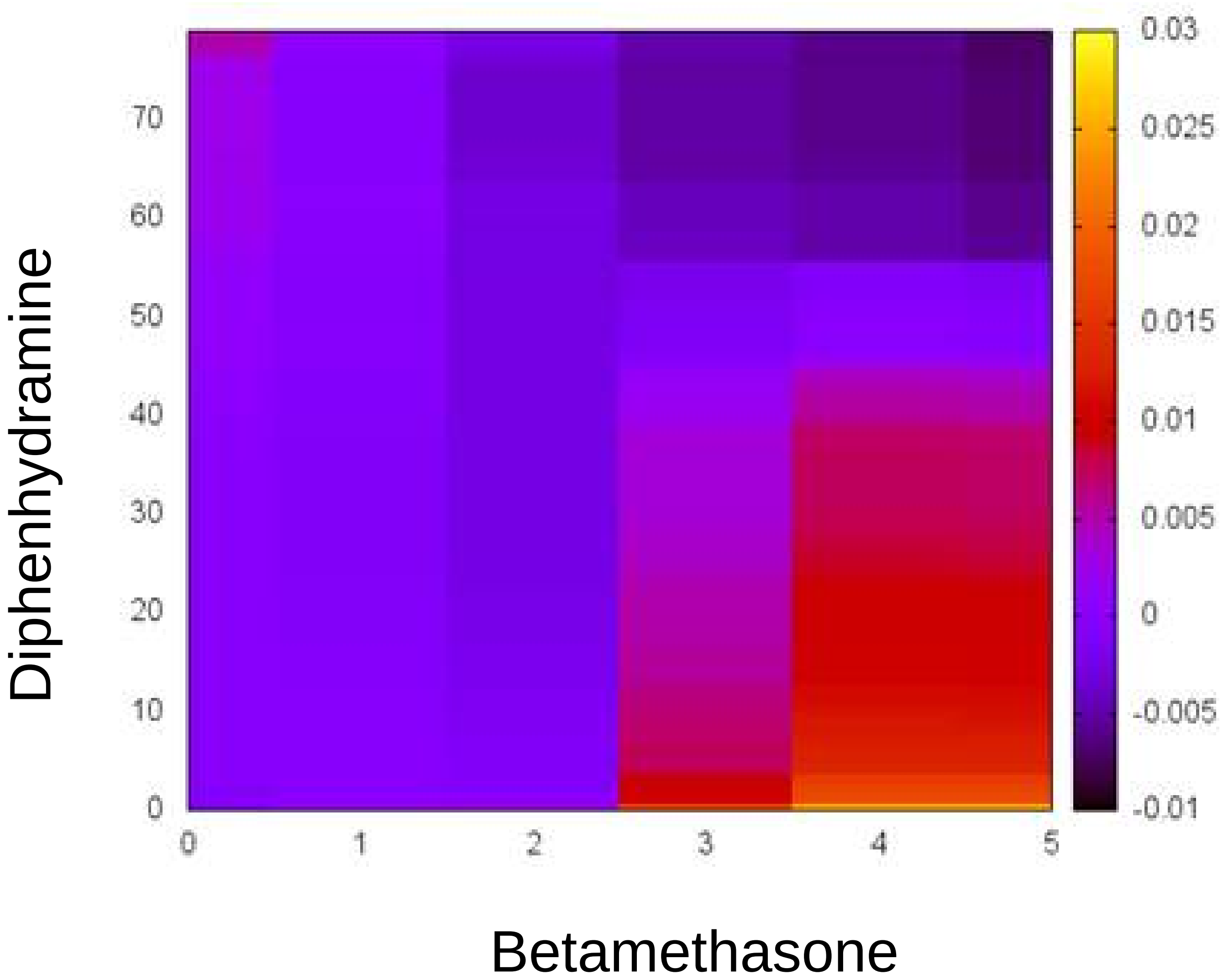
$f_i(x_i)$



$f_{ij}(x_i, x_j)$

Y. Lou, R. Caruana, J. Gehrke, and G. Hooker. Accurate Intelligible Models with Pairwise Interactions. In KDD, 2013.

Insights about Interactions



Decisions

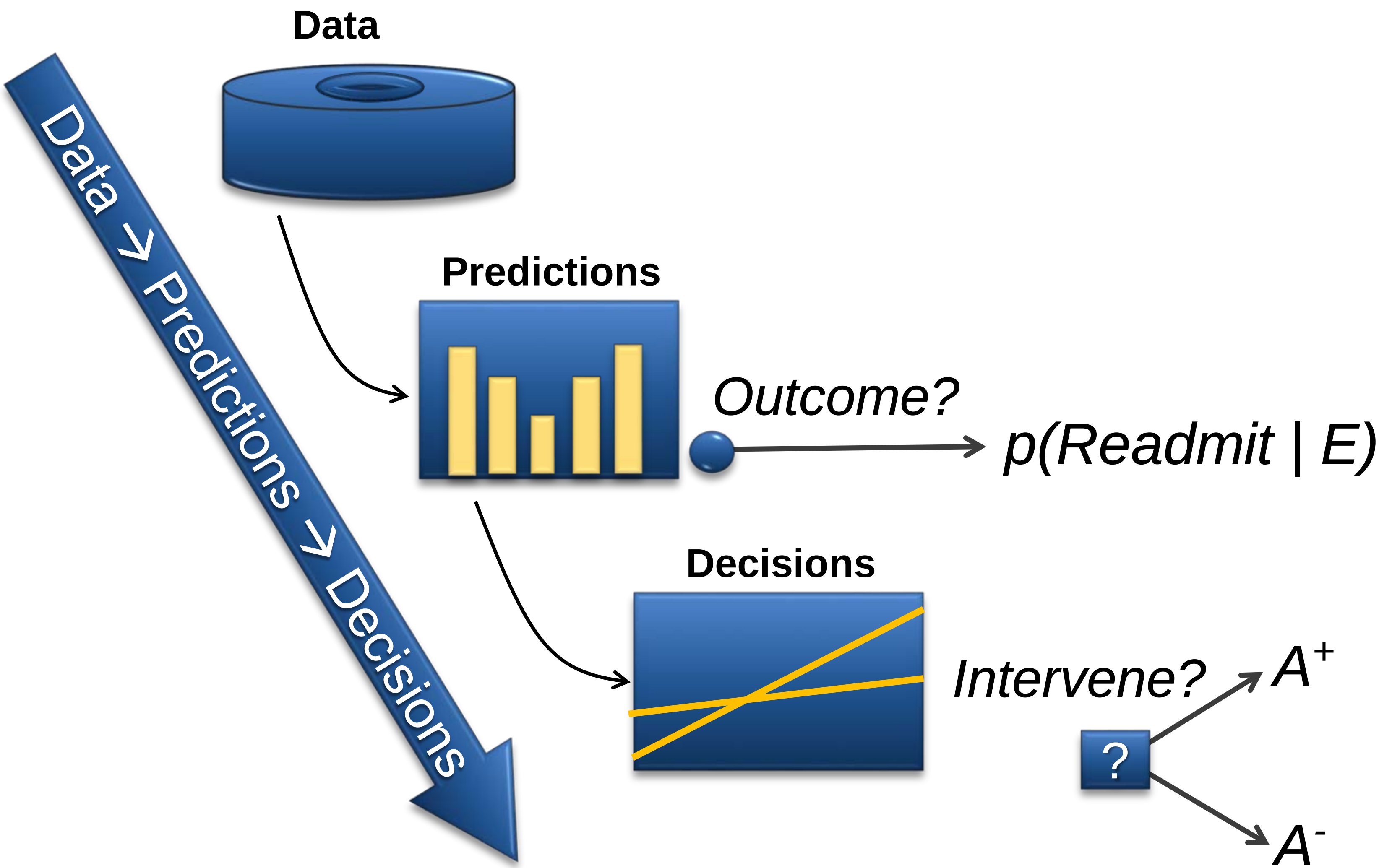
Units 5E/501/8E/9W/8ITCU

Baseline:
Discharges to home/ home health between 10/15/2011 – 4/29/2012
Readmissions Rate (all cases): **13%**
Score ≥ 25: **27%**
Average direct cost/readmission: **$10,888**

	Initial Pilot 4/30/2012 – 7/30/2012	1 Month Post engagement 9/01/2012 – 9/30/2012
Readmissions Rate	12%	10%
Score ≥ 25	23%	20%
# of Admissions Avoided	9	11
Follow up call completion	52%	61%
Follow up call not Completed	32%	21%
Total Annualized savings	$391,968	$1,448,104

↓ Total Readmission Rate by 3% and +$1.4M Savings

Decisions



Data
Data → Predictions → Decisions
Predictions
Outcome?
p(Readmit | E)
Decisions
Intervene?
?
A+
A-

Example: Heart Failure

Most frequent dx for hosp. Medicare patients

6–10% of folks over 65

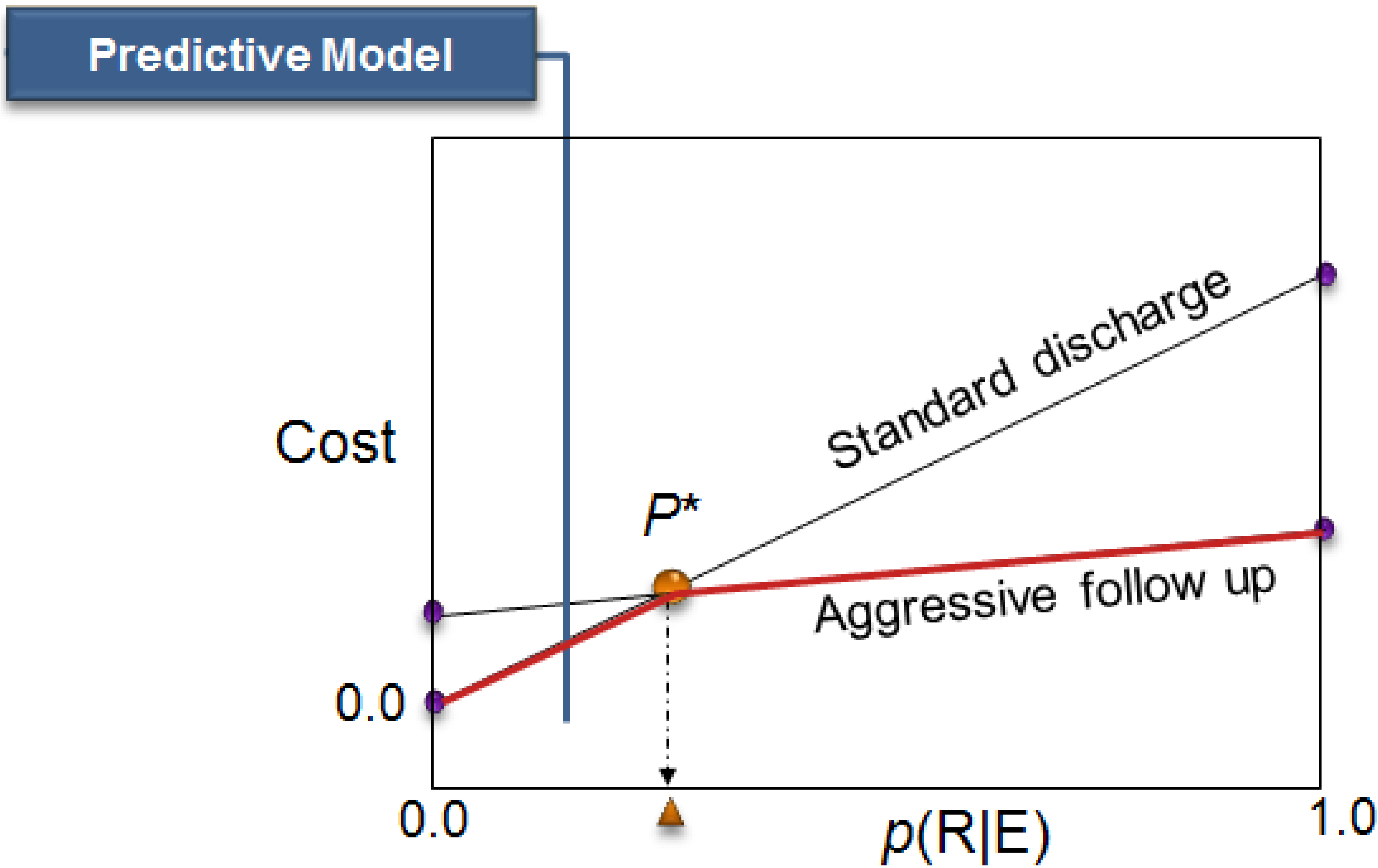
$35 billion/yr US

<u>Decision:</u>

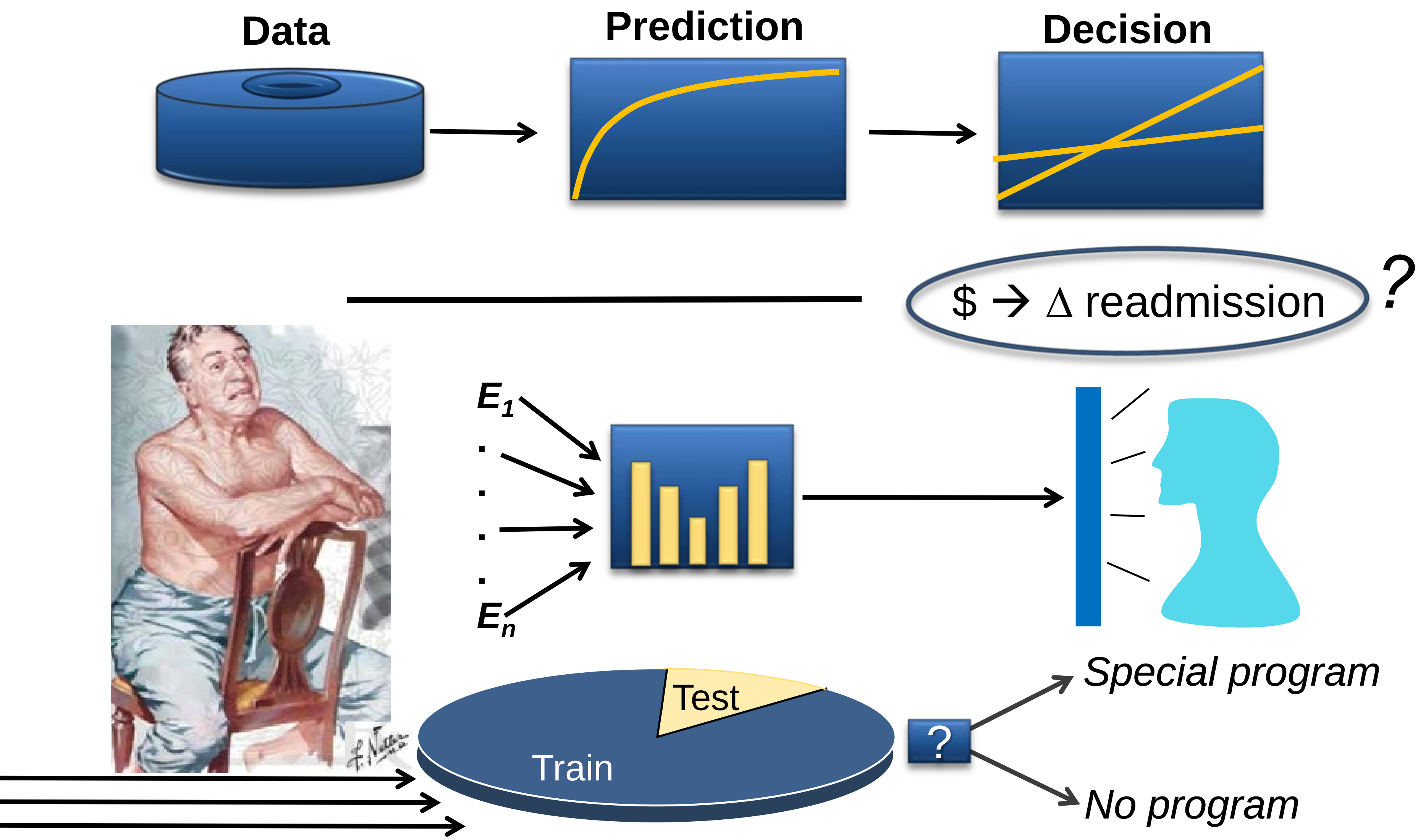
Invest in post-discharge program for patient?

With M. Bayati, M. Braverman, P. Koch, K. Mack, G. Ruiz, M. Smith

Utility Model



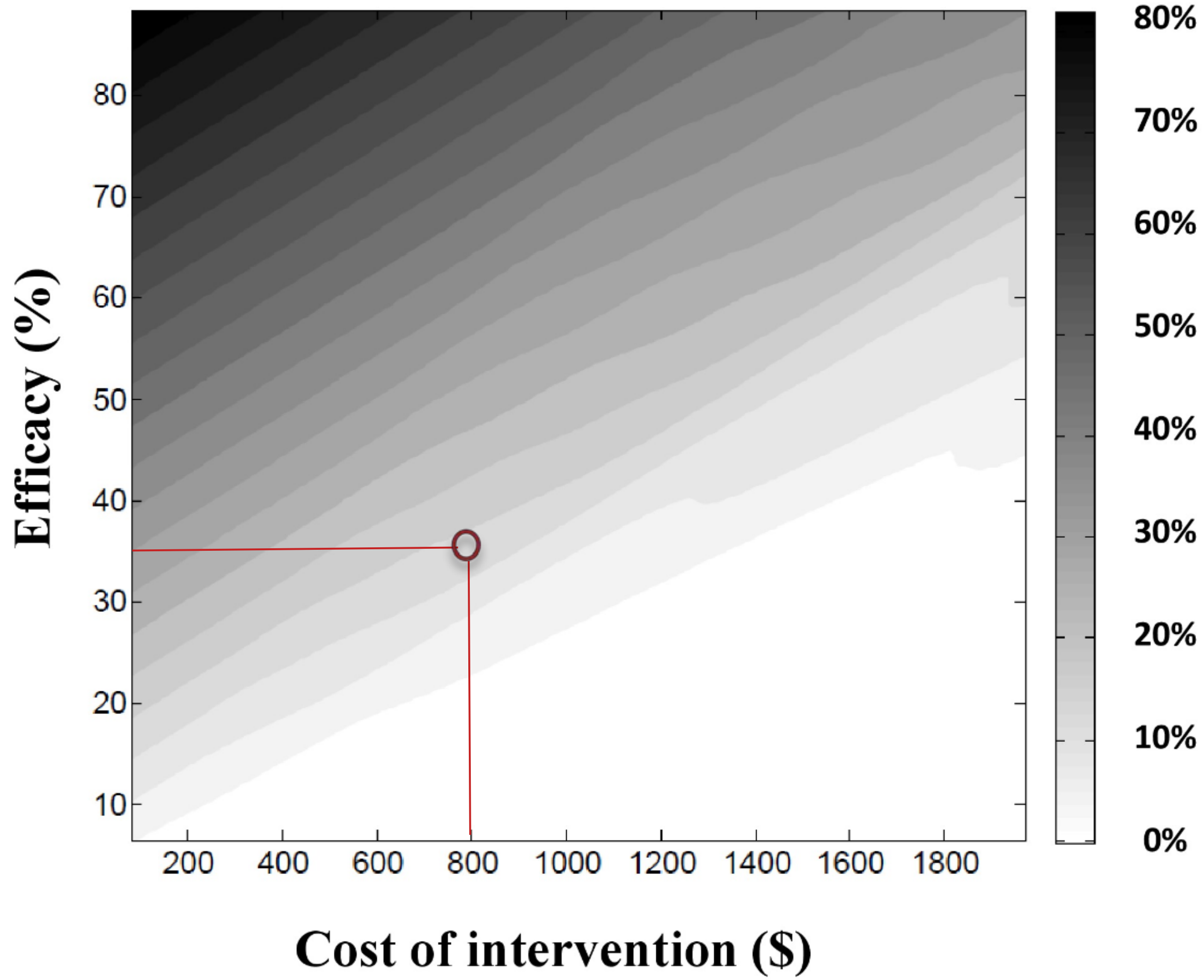
Exploration with Decision Pipeline



Decision Pipeline → Visualization

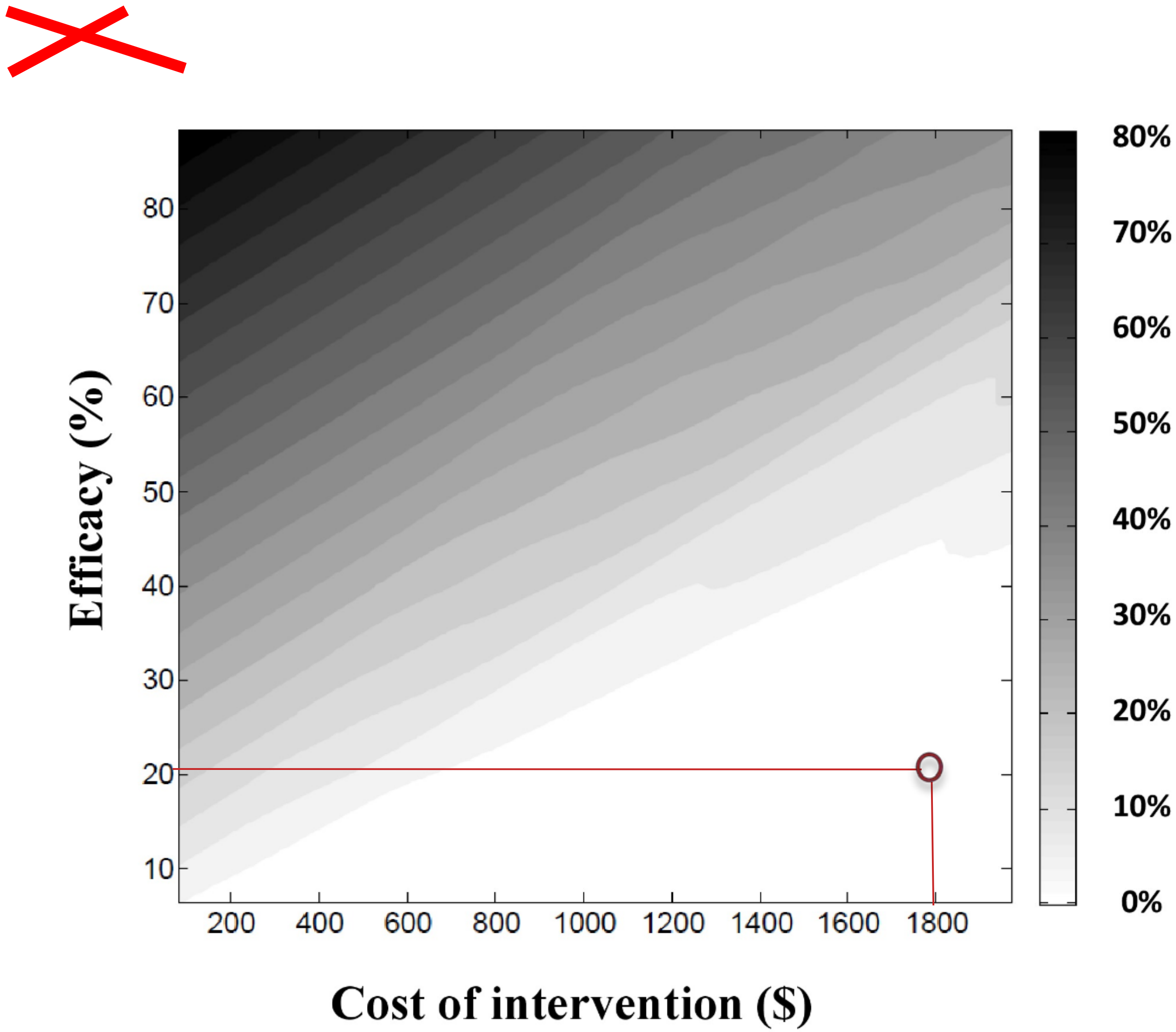
$800 intervention @ 35% efficacy?

↓31.4% readmissions ↓$13.2%.



Decision Pipeline → Visualization

$1800 intervention @ 20% efficacy?



Errors, Adverse Events, and Deaths

Deaths:

44,000 - 98,000 preventable deaths per year

"*To Err is Human*," *Inst. of Medicine, 2000*

Adverse events:

44% preventable.

Levinson, 2010

Costs:

$17 to $29 billion per year in U.S.

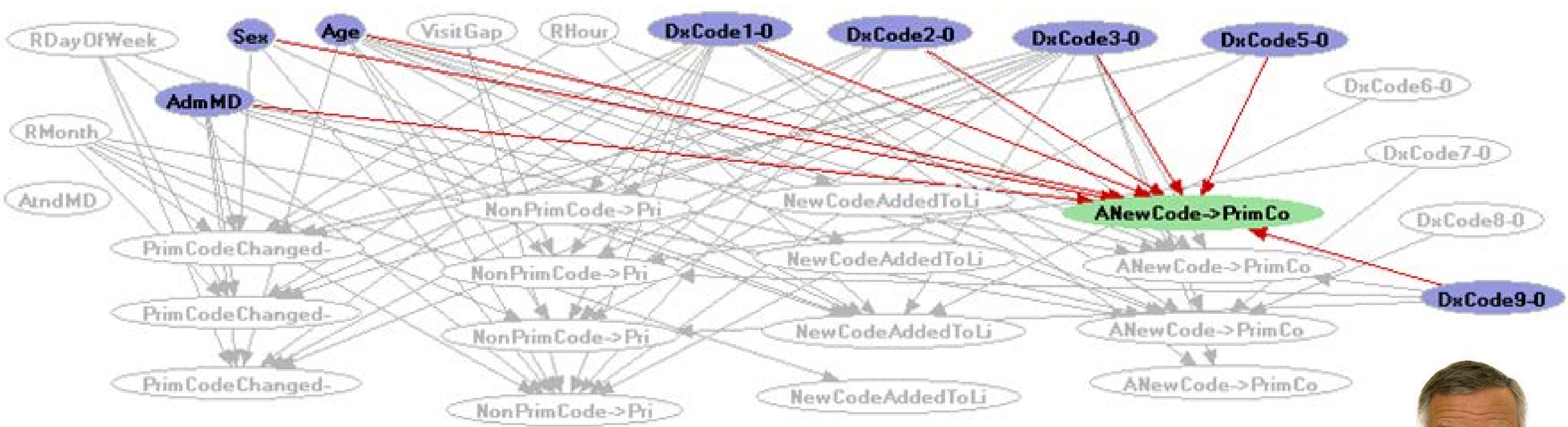
Thomas, et al., 1999

Detecting Errors

e.g., Predict surprise at emergency dept.

At discharge time:

→ p(readmit < *72 hrs.*|*E*) with *<u>new primary diagnosis</u>.*



With M. Bayati, M. Braverman, and J. Gatewood

Hospital-Associated Infection

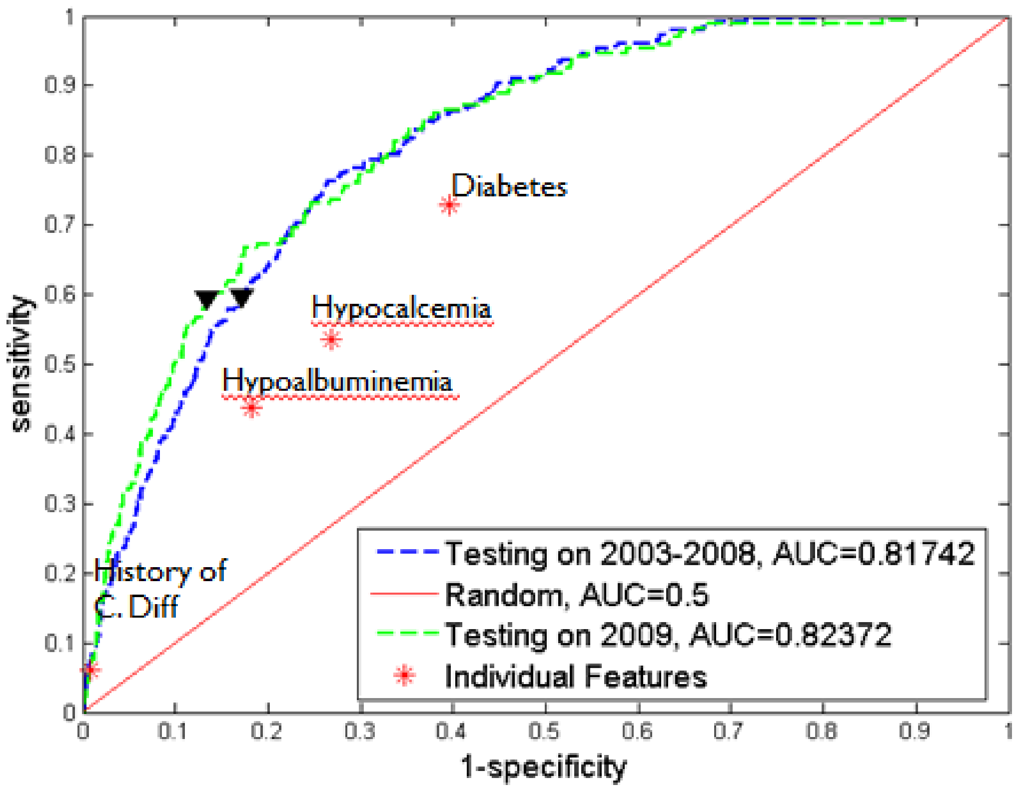
1 in 20 hospitalizations, ~$20 billion/yr.

5% death: top 10 contributor of death in US

Predicting C.Difficile < 48 hrs

SCIENTIFIC AMERICAN™

“Hospitals Fail tc
to Thwart Deadl

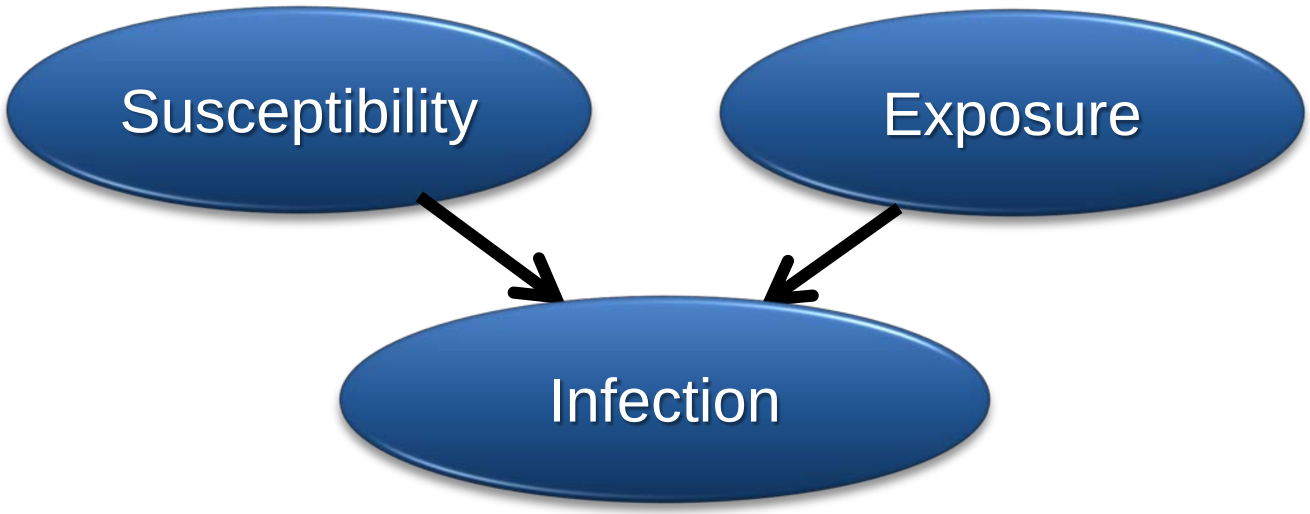


ırk Times

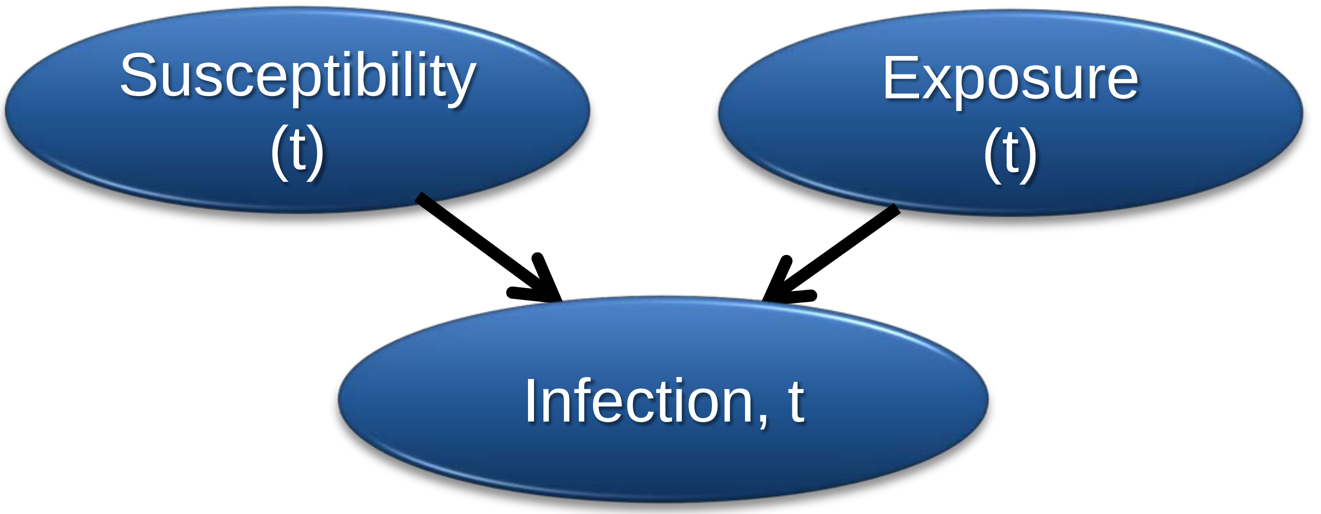
ive Action Urged
ıl Infections”
-May 29th, 2013

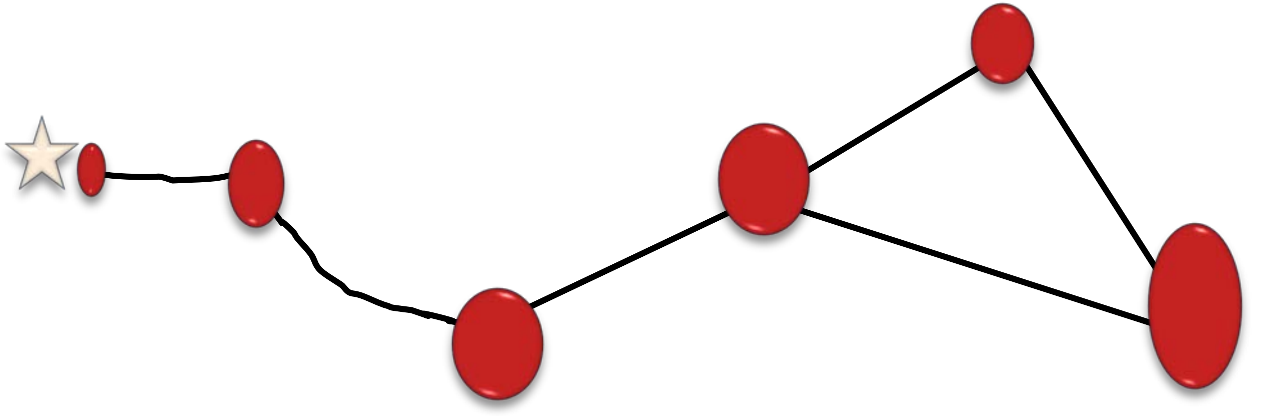
With Wayne Campbell, Ella Franklin, John Guttag, Jenna Wiens

Data on Time and Space

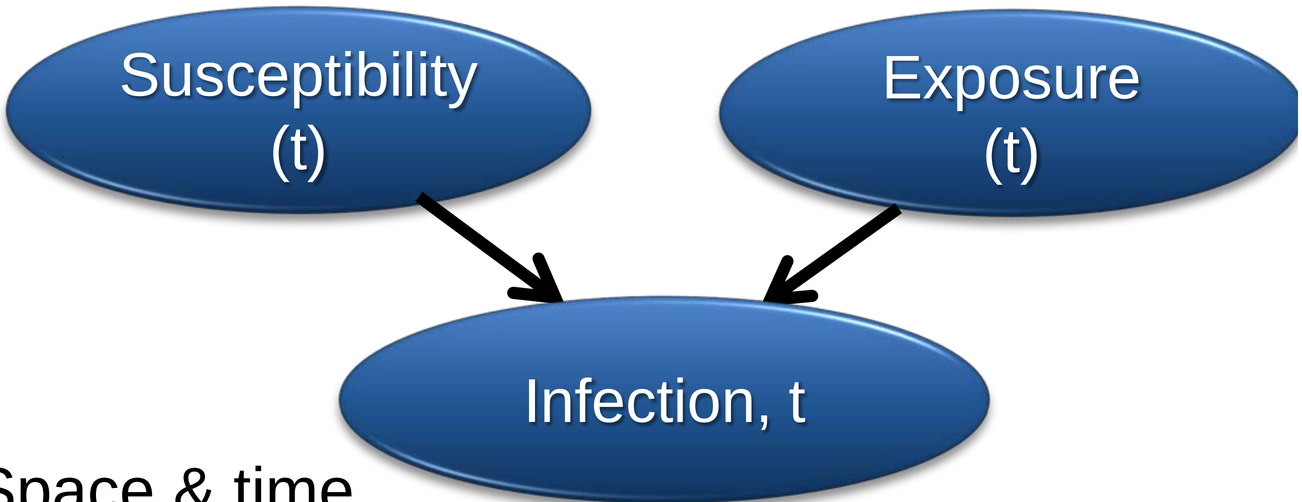


Data on Time and Space

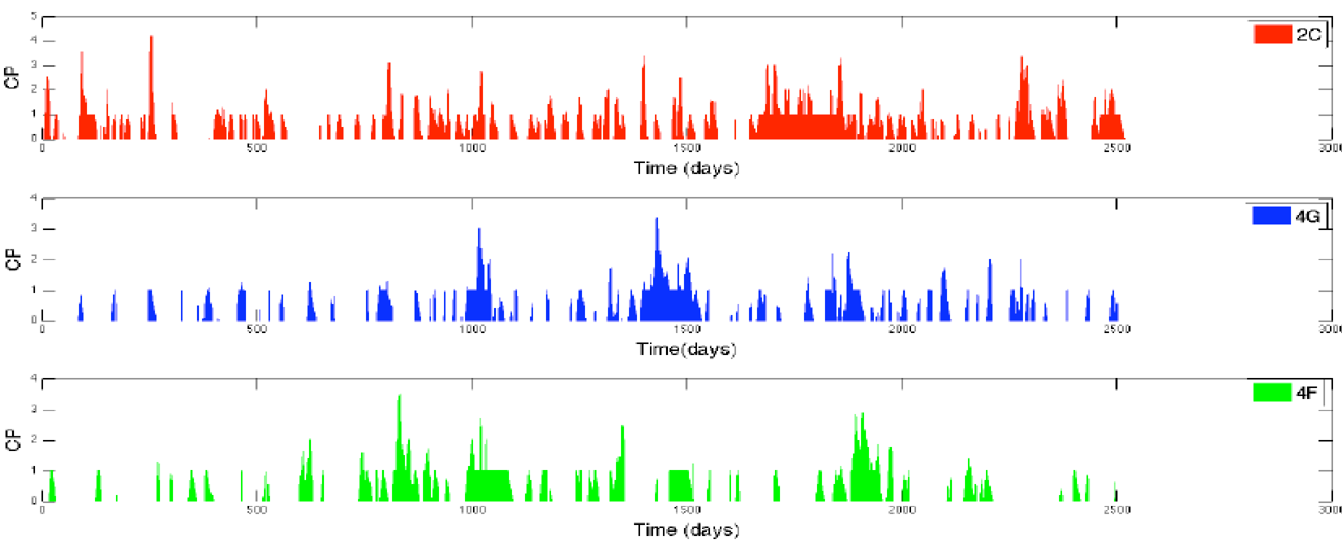




Data on Time and Space



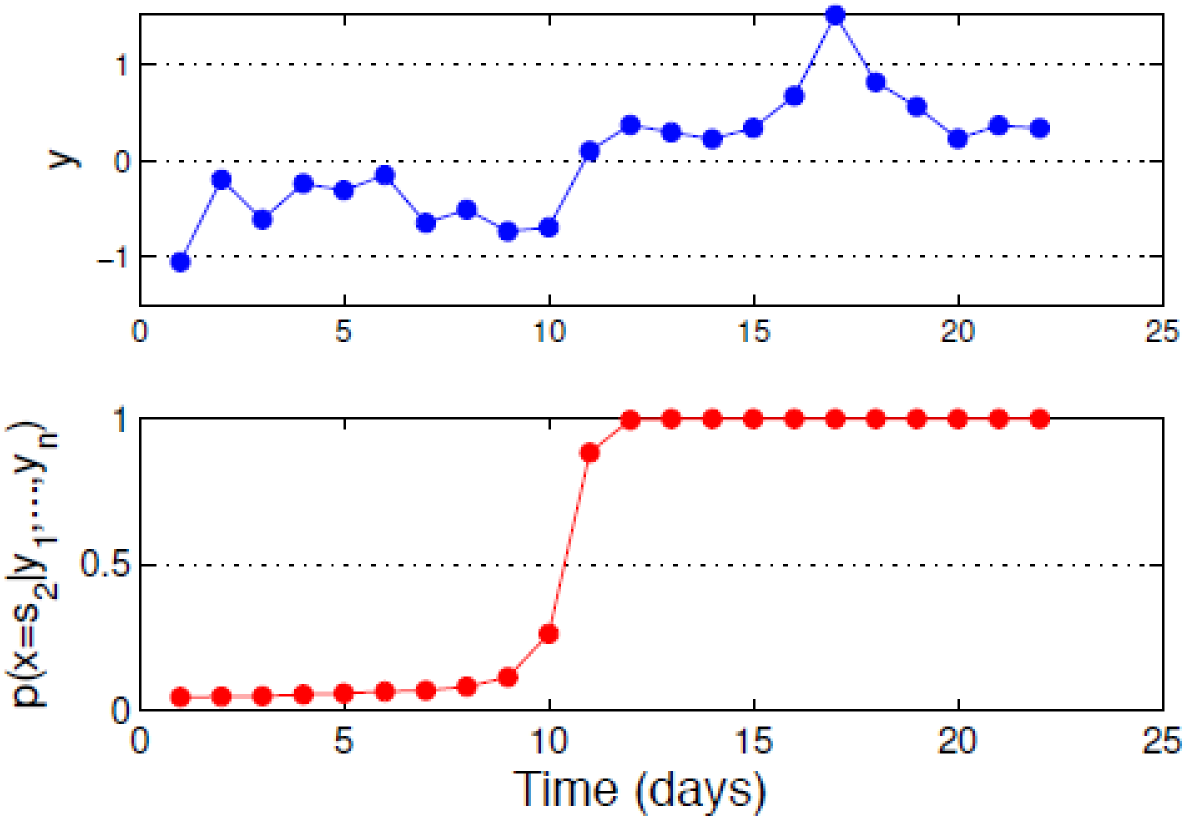
Space & time



Location	Description
1C	Patient Care Unit
1E	Patient Care Unit
1G	MedSTAR ICU
1H	Patient Care Unit
2C	Patient Care Unit
2E	Patient Care Unit
2G	Intensive Care Unit (ICU)
2H	Patient Care Unit
2NE	Patient Care Unit
2NW	Patient Care Unit
3C	Patient Care Unit
3D	Patient Care Unit
3E	Patient Care Unit
3F	Patient Care Unit
3G	Intensive Care Unit (ICU)
3NE	Patient Care Unit
4C	Patient Care Unit
4D	Patient Care Unit
4E	Patient Care Unit
4F	Patient Care Unit
4G	Intensive Care Unit (ICU)
4H	Intensive Care Unit (ICU)
4NW	Patient Care Unit
5C	Patient Care Unit
5D	Patient Care Unit
5E	Patient Care Unit
5F	Patient Care Unit
5NE	Patient Care Unit
5NW	Patient Care Unit

J. Wiens, et al.

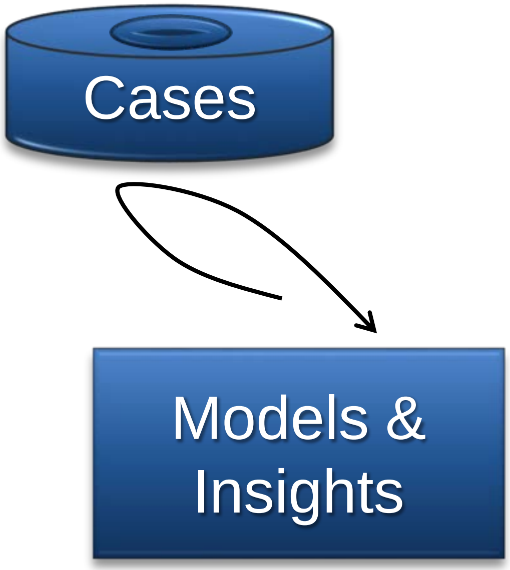
Temporal Models and Prediction



NIPS 2012: AUC: 0.69 → 0.79

J. Wiens, et al.,

Causal Discovery



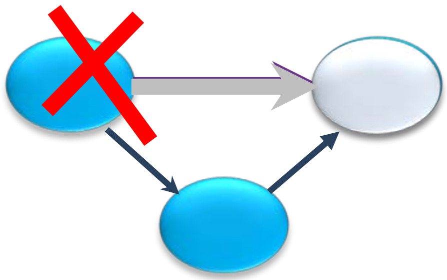
Pt. acquires C. Difficile?

- diabetes = TRUE
- history of C. Diffi = TRUE
- hospital service = gsg (general surgery)
- meds= acetylcysteine (n-acetylcys)
- meds = lidocaine hcl
- meds = clindamycin phosphate
- platelet count = C (thrombocytosis)
- unit = 2g
- albumin = L (hypoalbuminemia)
- admission source = transfer
- attending MD= XXXXXX
- unit = 2d
- CO_2 = L (hypocapnea)
- city = XXXXXX
- employer name = Not Employed
- monocyte percent = H
- 70<=age<80
- wbc = H (white blood cell count)
- admission procedure = catheterization
- admission complaint =gastrointestinal
- last visit meds = fentanyl citrate
- meds = hydromorphone hcl

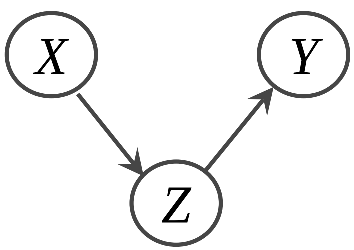
Studies in causality

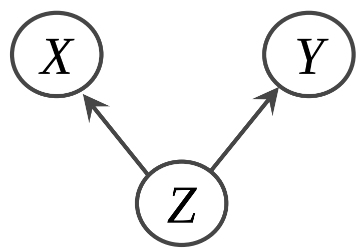
J. Wiens, et al.

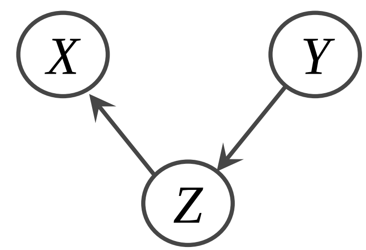
Causal Discovery



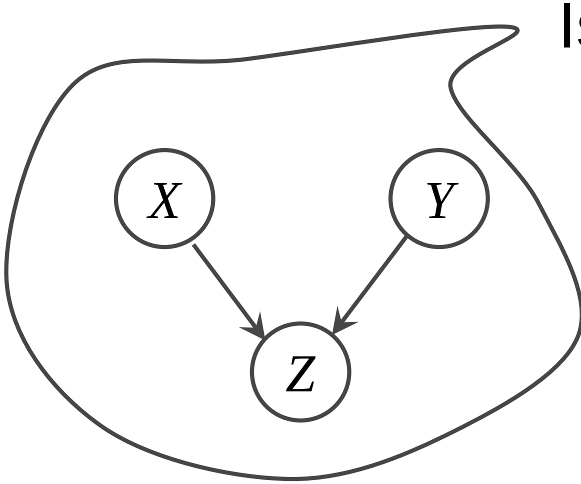
Given $X \perp Y$ *and* $\neg(X \perp Y \mid Z)$,

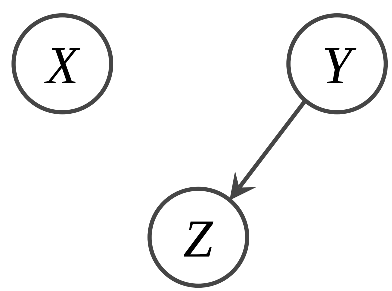




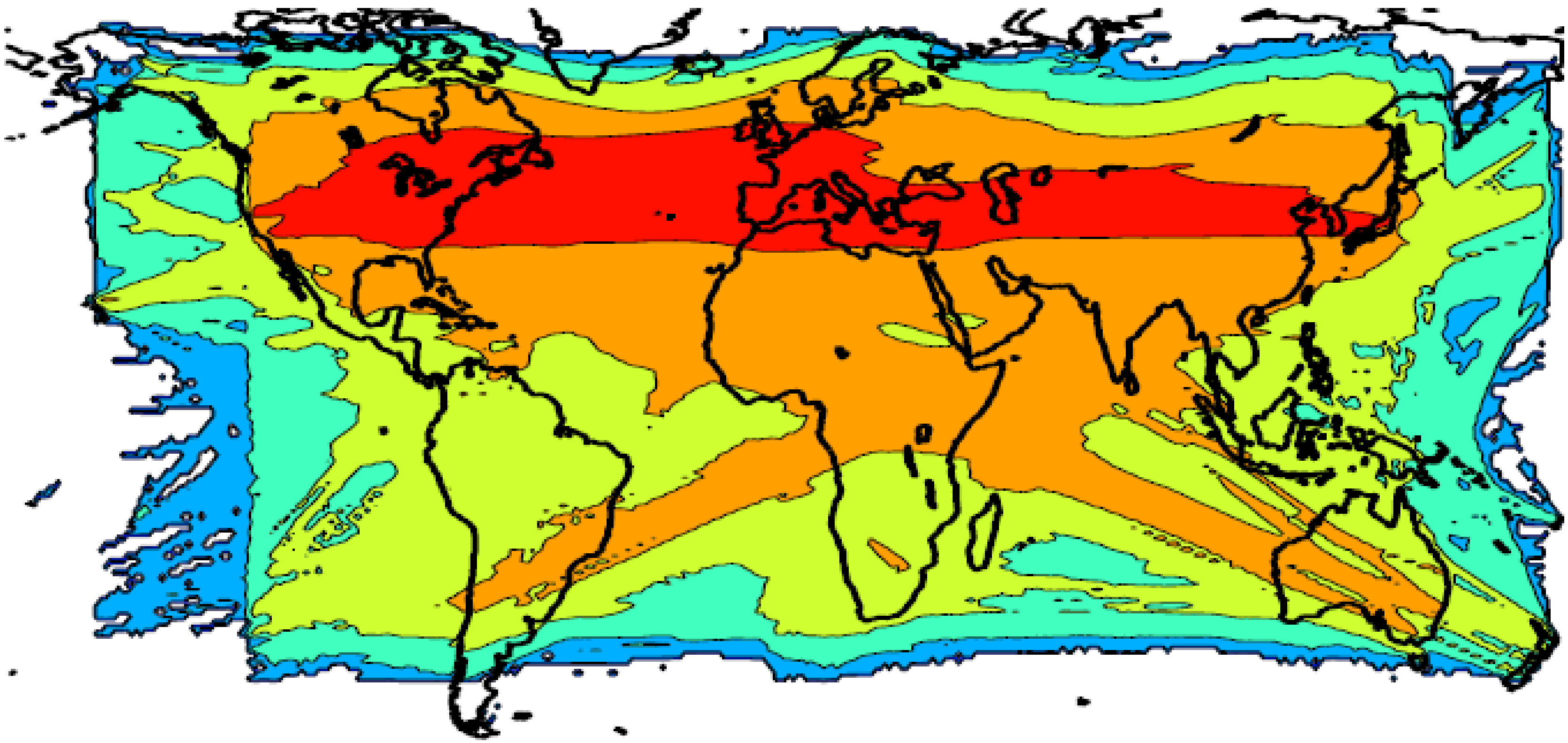


Is the only possible causal model





Web for Planetary-Scale Sensing

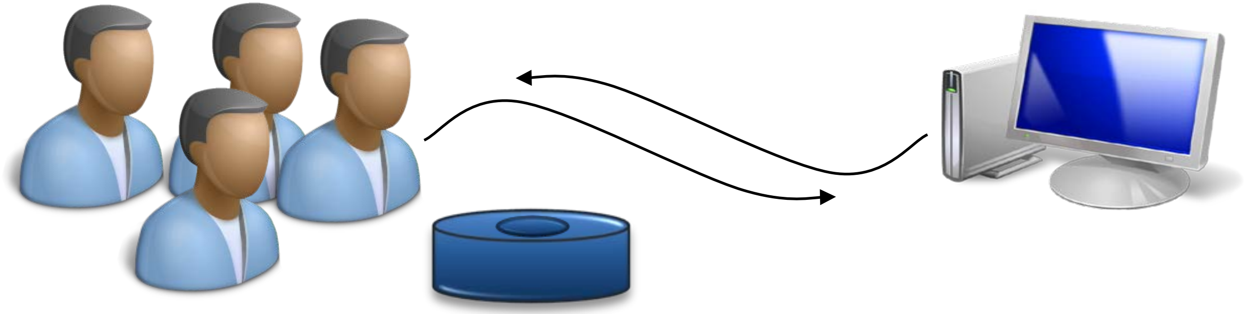


Leskovec & Horvitz (2008)

Signals on Medication Adverse Effects

→ Web search as sensor for side effects?

1 in 250 of people query on top-100 drugs.



Signals on Medication Adverse Effects

Pharmacovigilence: spontaneous reports

FDA *Adverse Event Reporting System* (AERS)

<u>2011 finding (Tatonnetti, et al.)</u>:

Paxil + Pravachol → ⬆ *Hyperglycemia*

Pravachol → ✕ *Hyperglycemia*

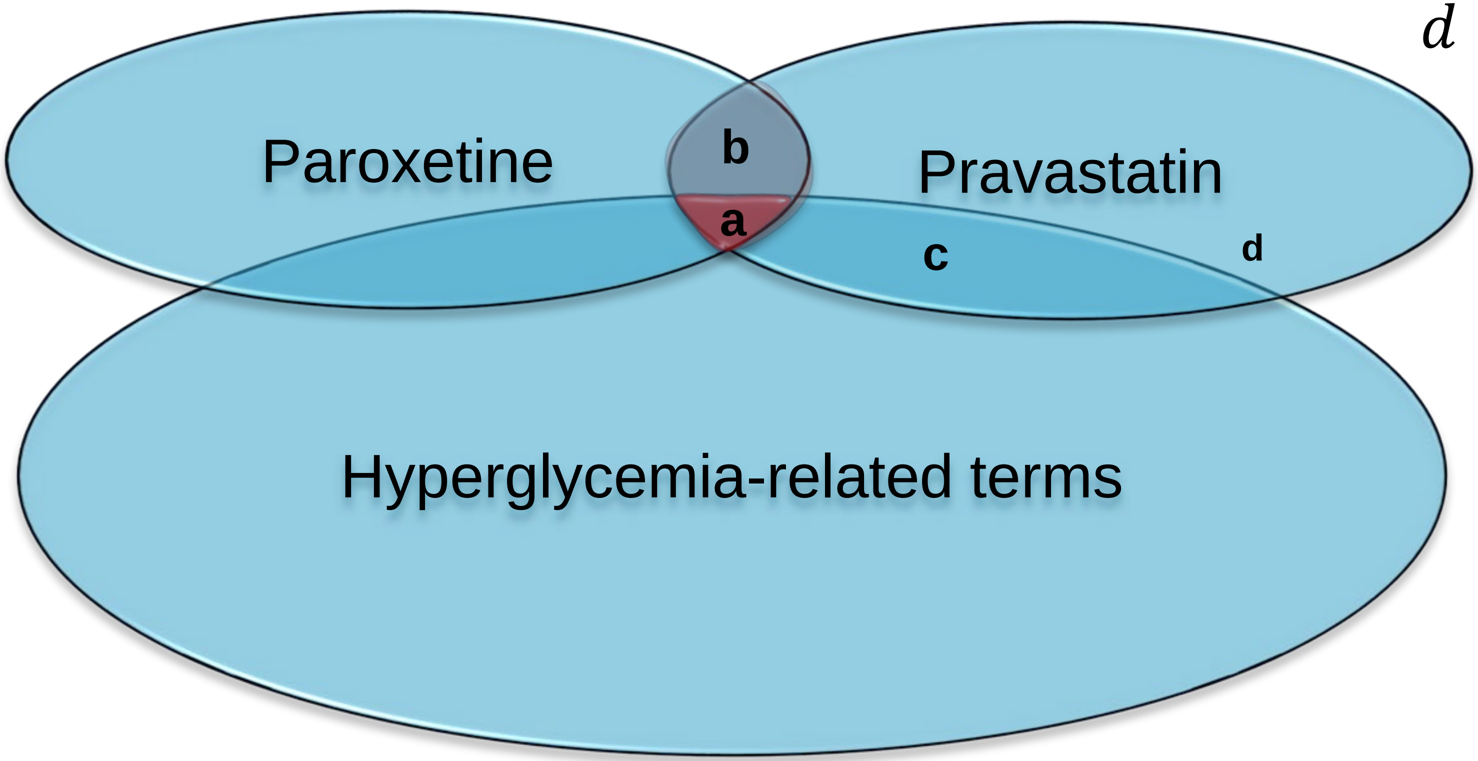
Paxil → ✕ *Hyperglycemia*

R. White, N. Tatonetti, N. Shah, R. Altman, H.

Web-Scale Pharmacovigilance

Disproportionality analysis

- Reporting ratios (RR)--obs. vs. expected: $RR = \frac{\frac{a}{b}}{\frac{c}{d}}$

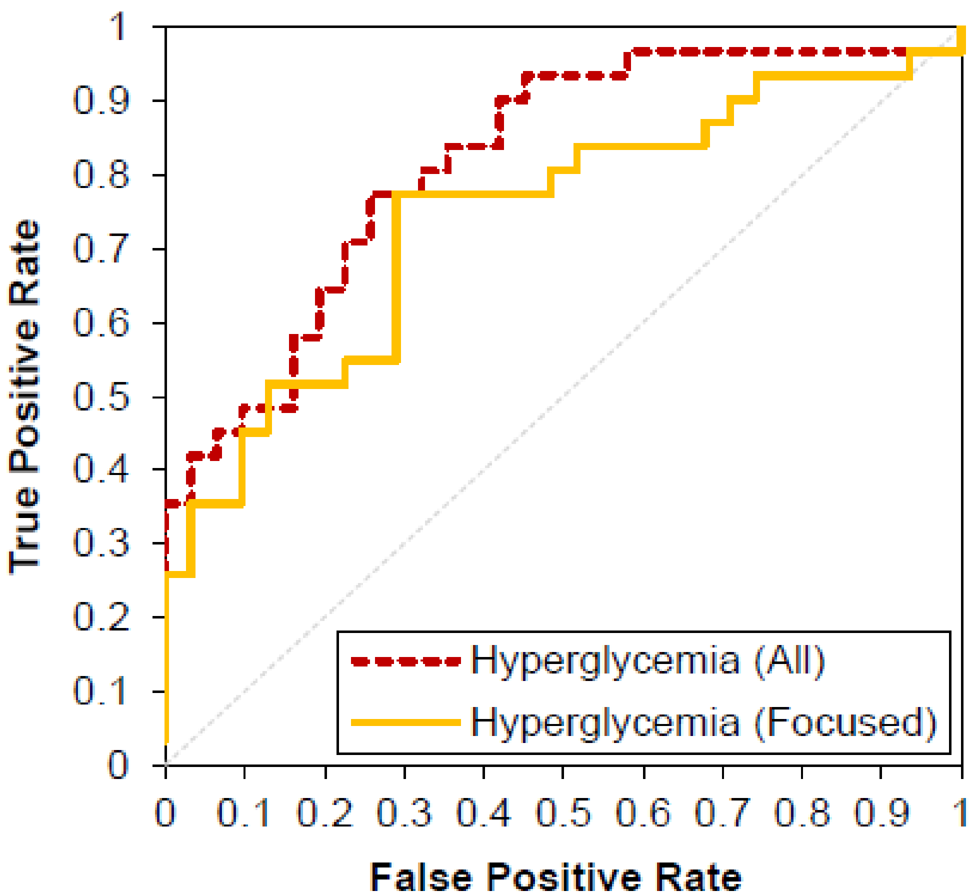


	a	*b*	*c*	*d*	*RR*	*95% CI (Lower, Upper)*	*p-value (one-tailed)*
Expected (pravastatin)	342	2716	2581	56302	2.747	2.438, 3.094	< 0.0001
Expected (paroxetine)	342	2716	3645	71243	2.461	2.189, 2.767	< 0.0001

Characterizing Sensor Error

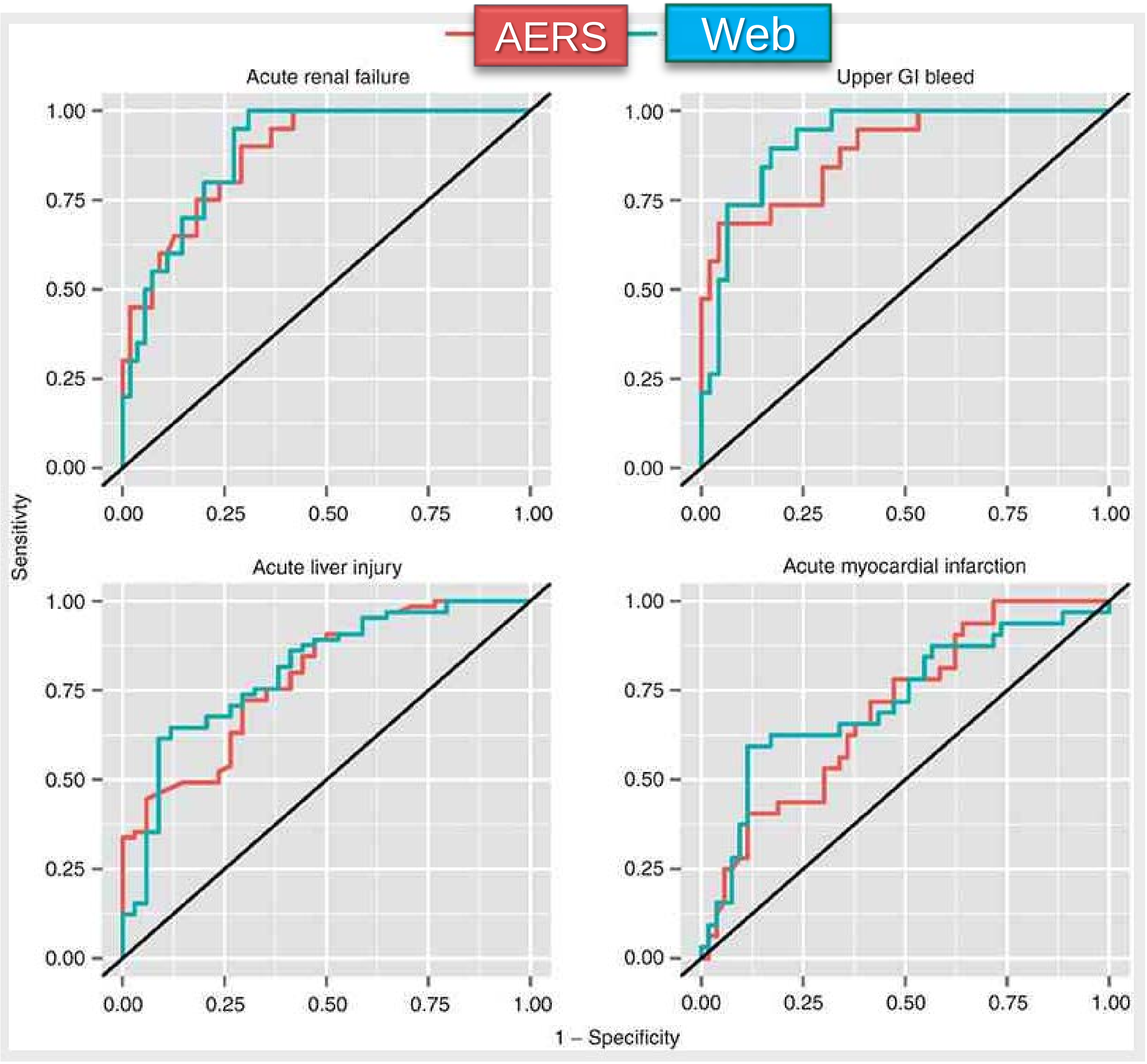
Test on known interactions

- 31 true positives for hyperglycemia
- 31 true negatives for hyperglycemia



Label	*Drug 1*	*Drug 2*
TP	dobutamine	hydrocortisone
TP	dobutamine	triamcinolone
TP	dobutamine	prednisolone
TP	betamethasone	dobutamine
TP	glipizide	phenytoin
TP	dobutamine	methylprednisolone
TP	prednisolone	salmeterol
TP	salmeterol	triamcinolone
TP	betamethasone	terbutaline
TP	dexamethasone	dobutamine
TP	budesonide	salmeterol
TN	hydrochlorothiazide	tazobactam
TN	clindamycin	montelukast
TN	lamotrigine	nystatin
TN	methylprednisolone	rosuvastatin
TP	budesonide	formoterol
TN	loratadine	nystatin
TN	hydroxychloroquine	prochlorperazine
TN	labetalol	sertraline
TN	ciprofloxacin	vecuronium

Rare, Serious Adverse Effects



OMOP

Multi-item Gamma Poisson shrinker algorithm (DuMouchel and Pregibon, KDD)

R. White, R. Harpaz, et al.

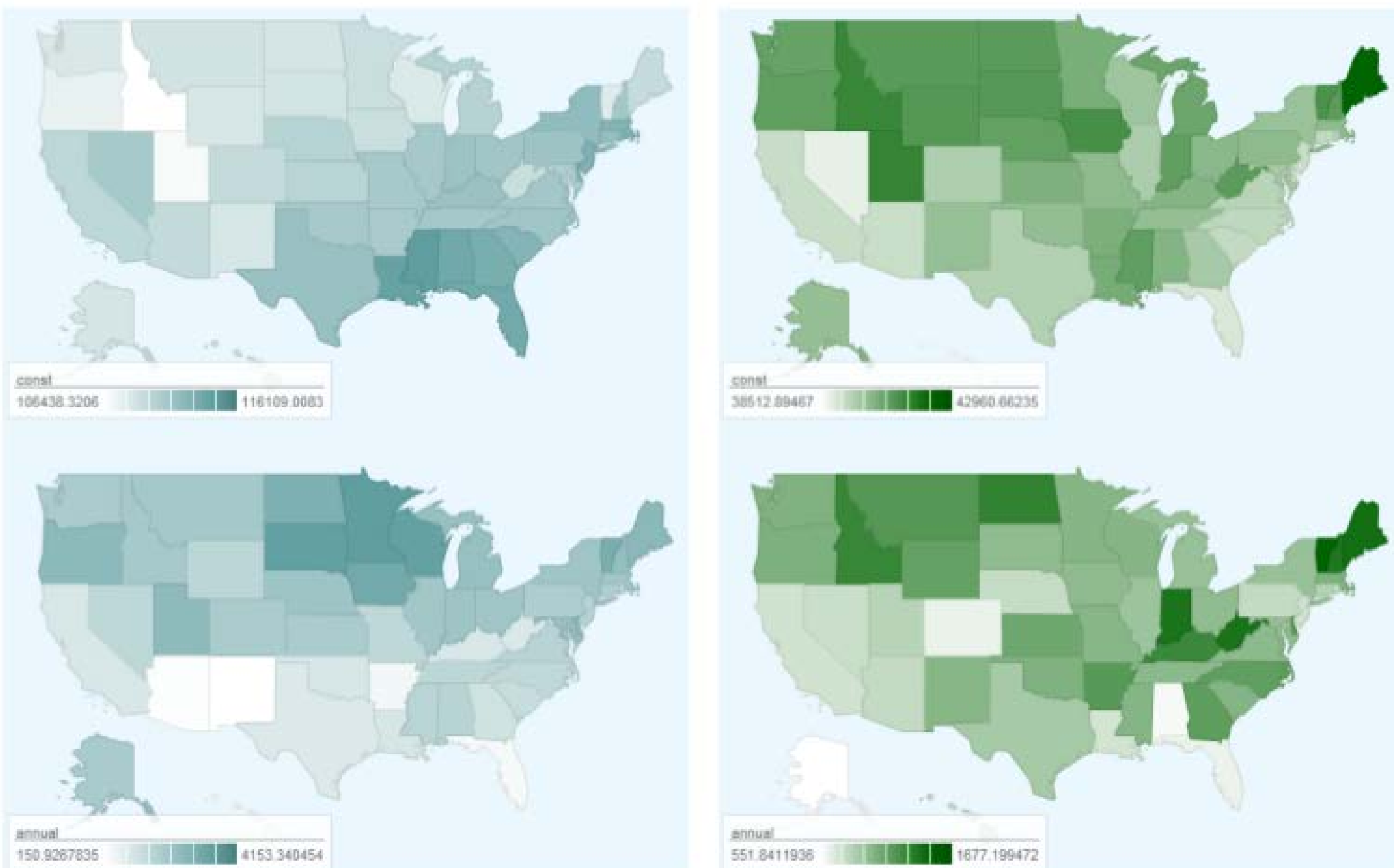
Complementarity of Signals

	AERS	Search	Together
Acute Renal Failure	0.88	0.88	0.93
Upper GI Bleed	0.89	0.92	0.92
Acute Liver Injury	0.79	0.81	0.86
Acute Myocardial Infarction	0.70	0.73	0.75
Average	0.81	0.83	0.86

AUC improvements statistically significant ($p<0.05$)

Wide Range of Studies

e.g., Nutritional content of downloaded recipes



Mean

Annual fluctuation

Total calories / serving

Calories from carbohydrates

R. West, R. White, E. Horvitz

Diet & Illness: Heart Failure

Na+ content in downloaded recipes & admissions (DC metro area)

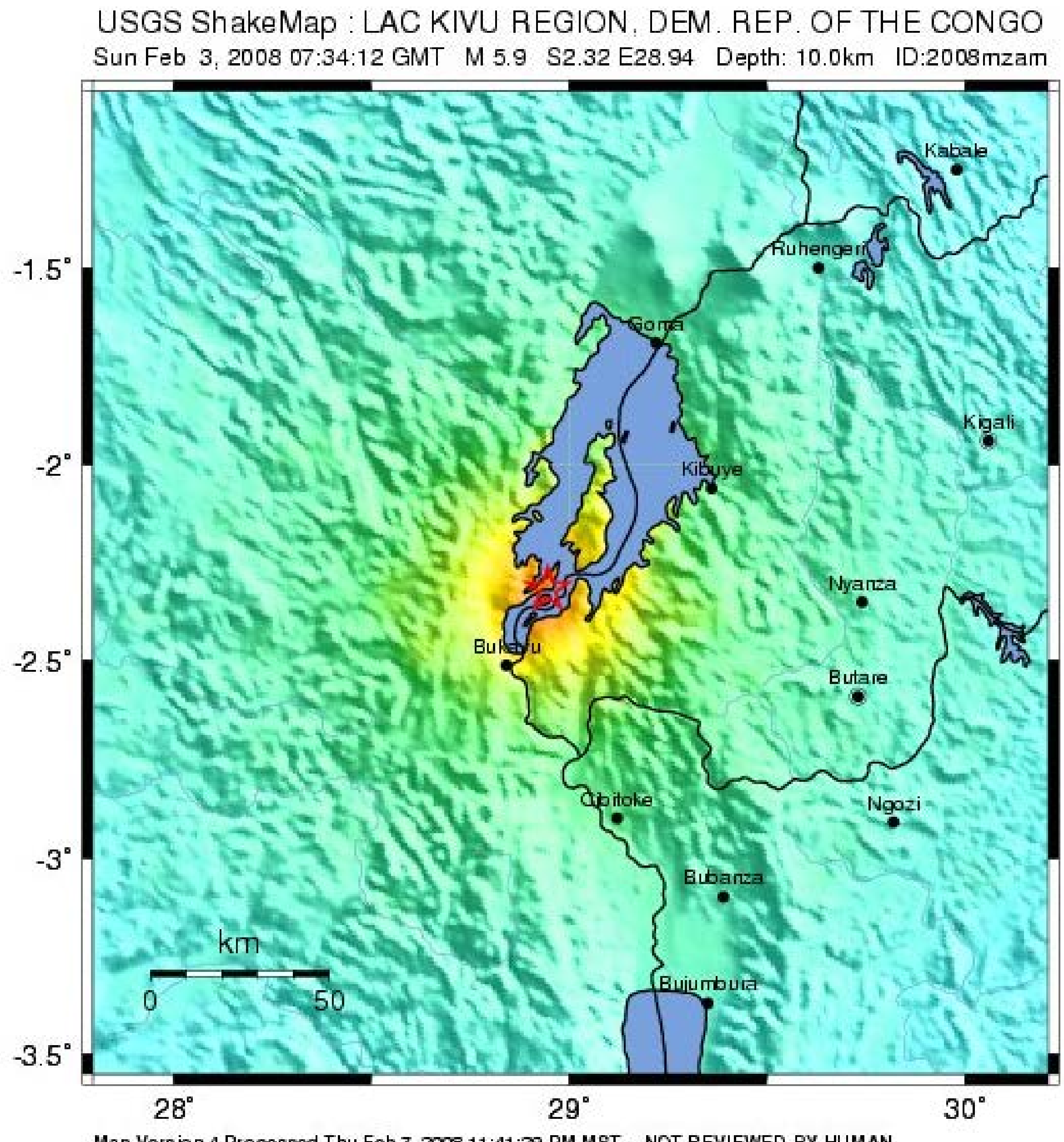
Disruption and Recovery

Disruption and recovery

Lac Kivu quake

Feb 3, 2008

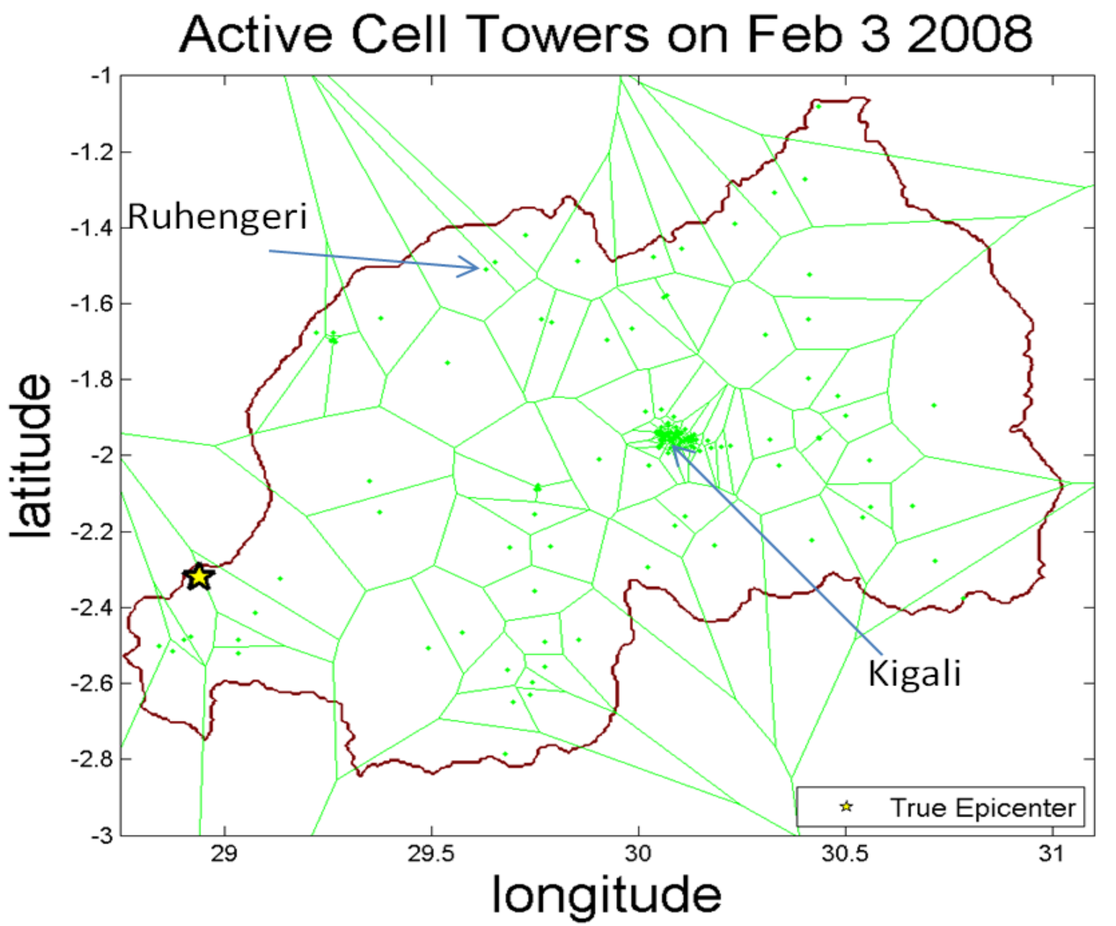
5.9



Cell Tower Call Densities in Rwanda

3 years of logs of ins and outs of comms.

140 cell towers, 6 days: 10,527,799 calls

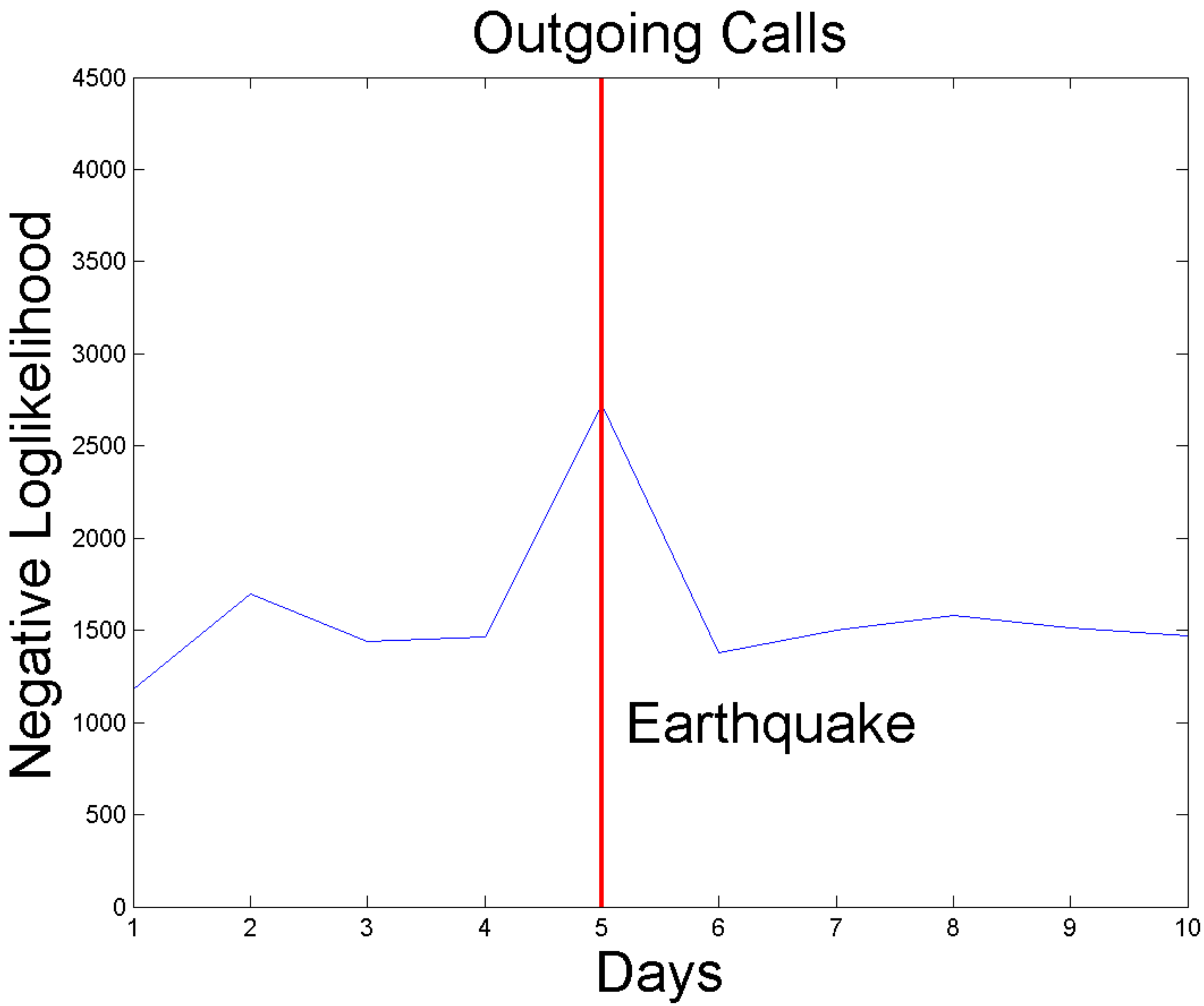


Kapoor, Eagle, Horvitz, 2010

Assumptions

1. Cell traffic deviates from normal in case of unusual events
2. Deviations inversely proportional to distance from event center
3. Larger disruptions have deviations that persist longer

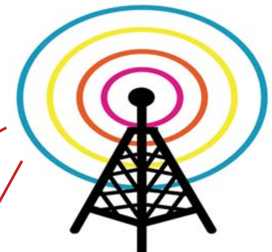
Detecting the Earthquake



Inferring the Epicenter

Modeling deviations from the trend

$$p(a_i \mid Event) \sim N(m_i(1+\Delta_i), \Sigma_i)$$



x_i, y_i

$$\Delta_i = \frac{\alpha}{\beta + \left[(e_x - x_i)^2 + (e_y - y_i)^2\right]^\gamma}$$

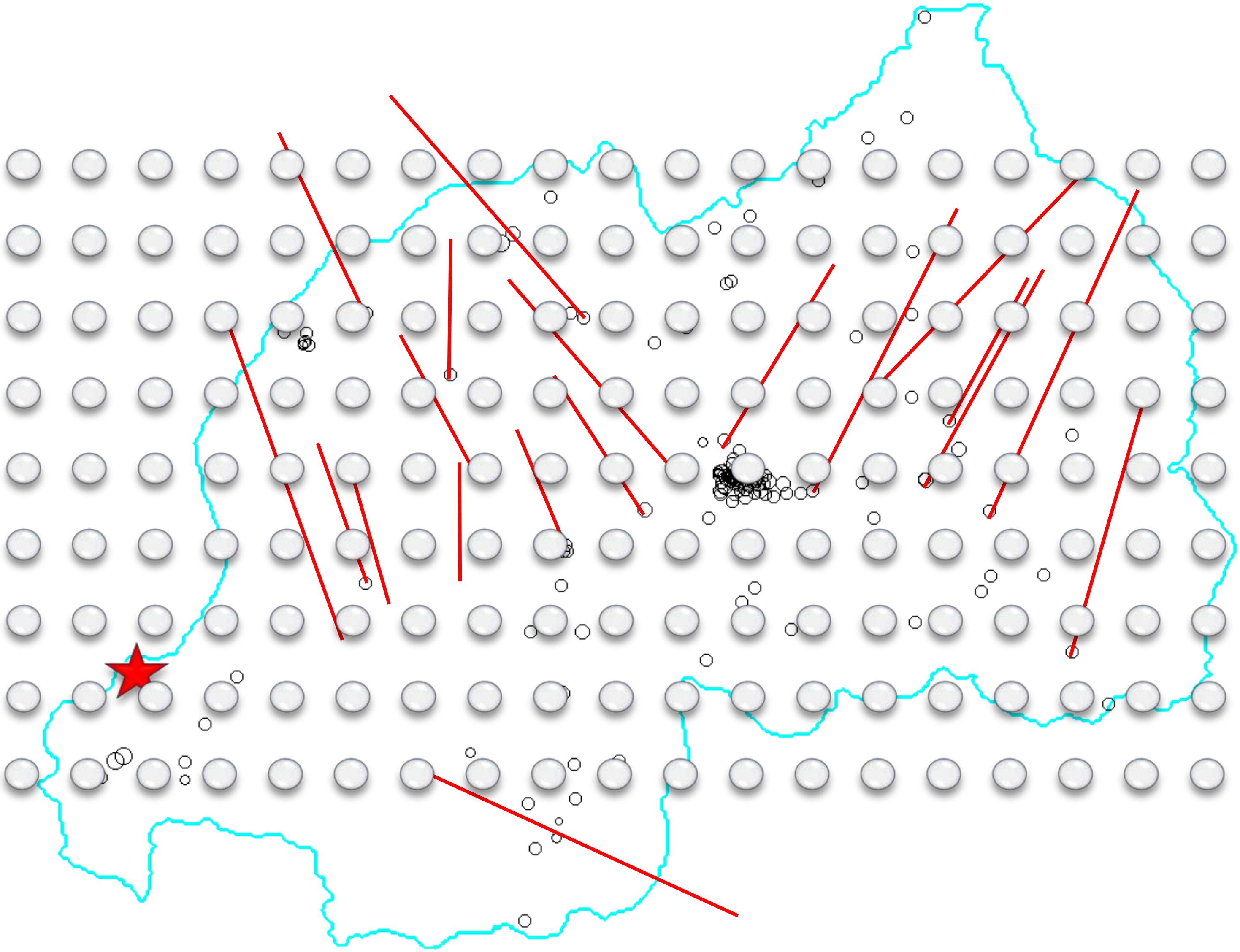
Unknown parameters: $\theta = \left(\alpha, \beta, \gamma, e_x, e_y\right)$

epicenter

$$\theta = \arg\max_\theta \sum_{i=1}^{T} \log p_\theta(a_i \mid Event)$$

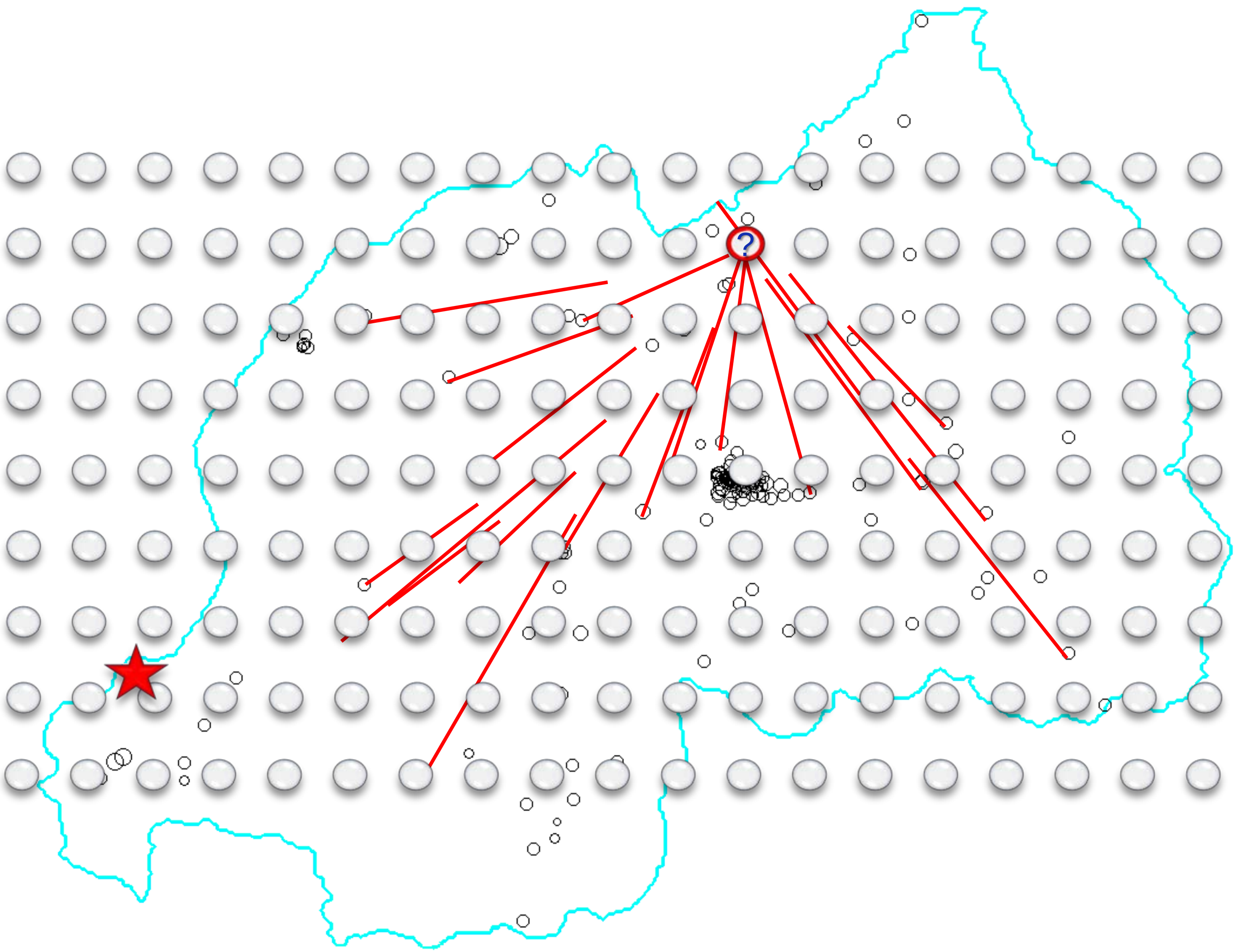
Determining the Epicenter

- Radius of towers = % increase in calls



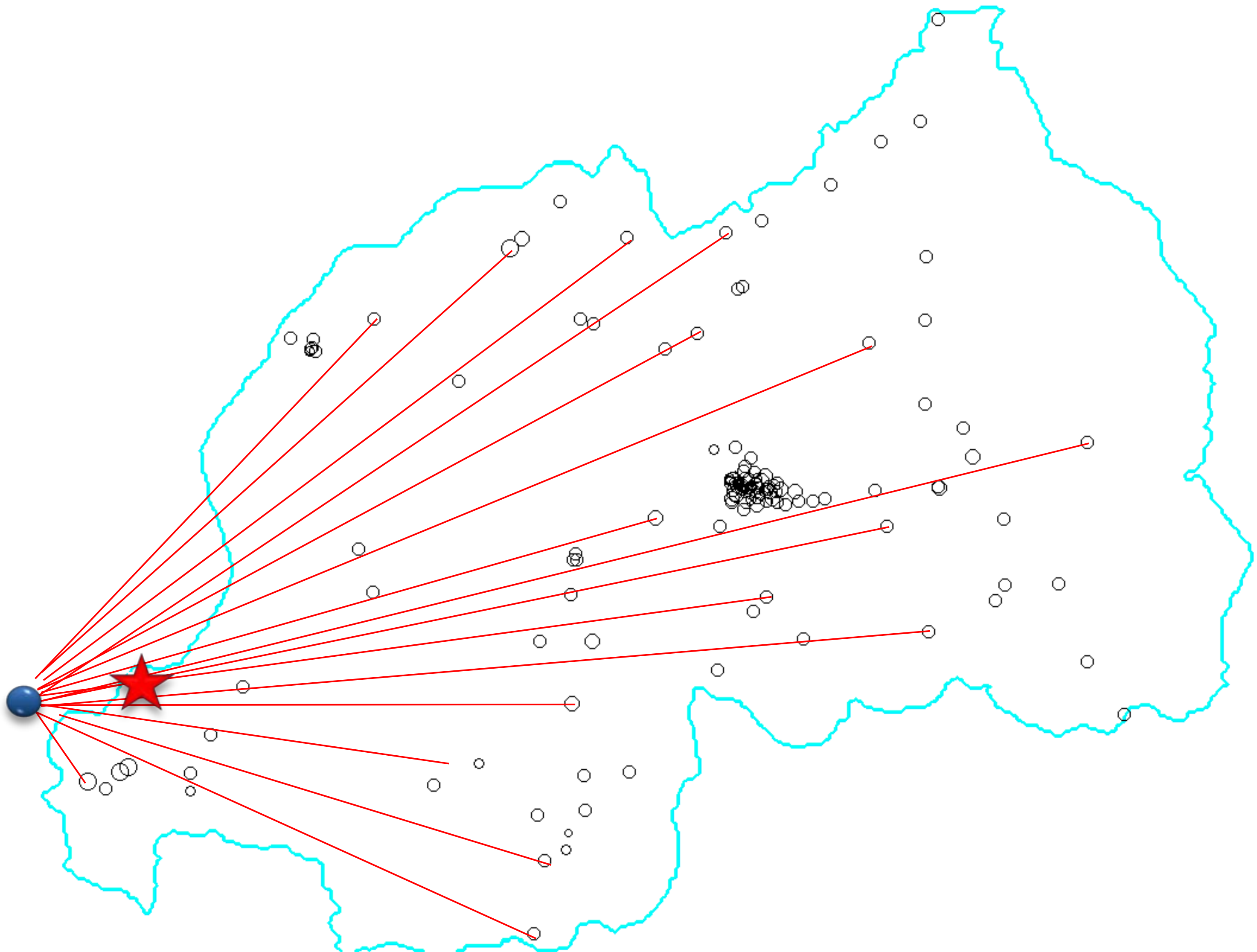
Determining the Epicenter

- Radius of towers = % increase in calls

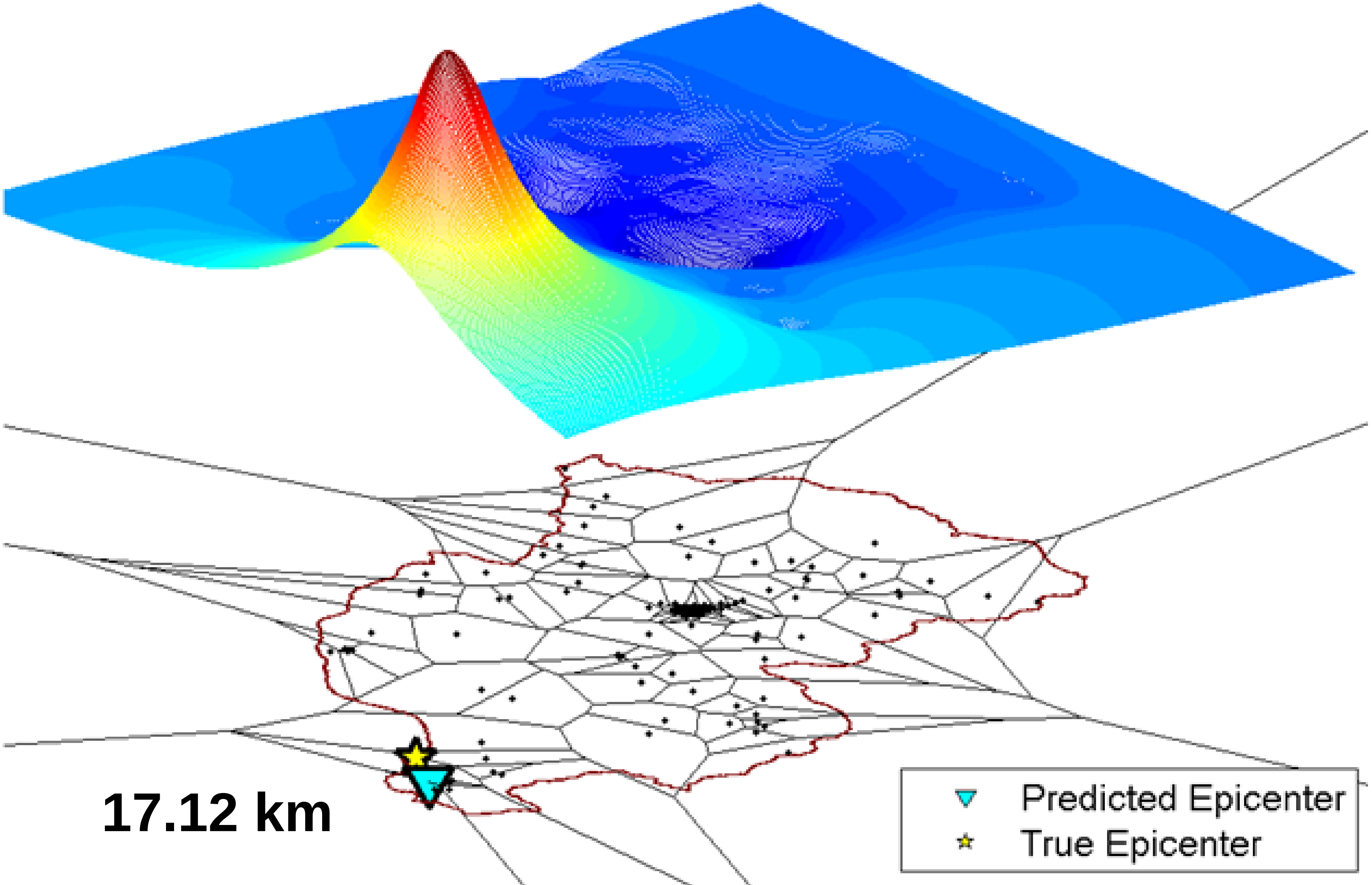


Determining the Epicenter

- Radius of towers = % increase in calls

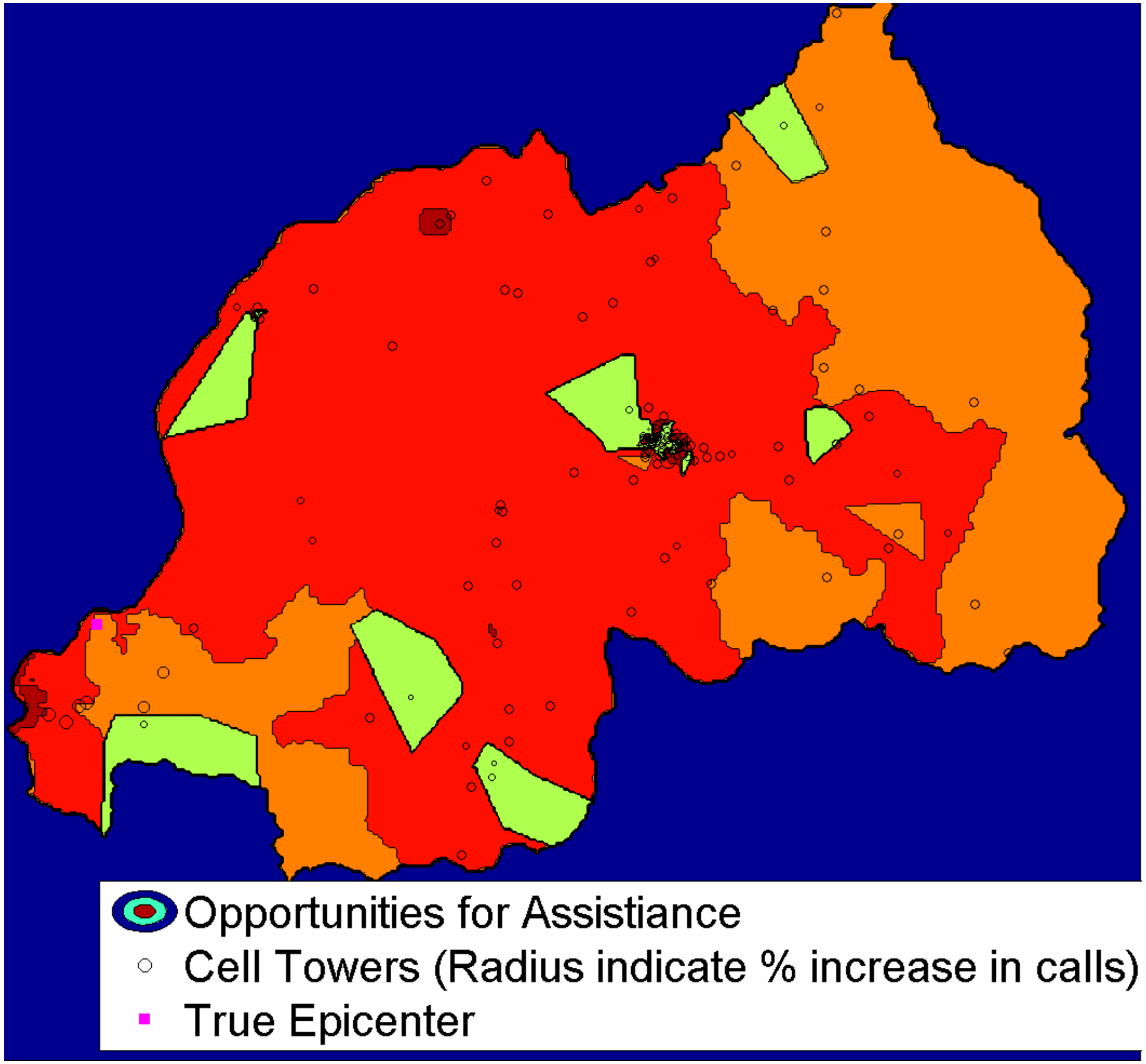


Inferring the Epicenter



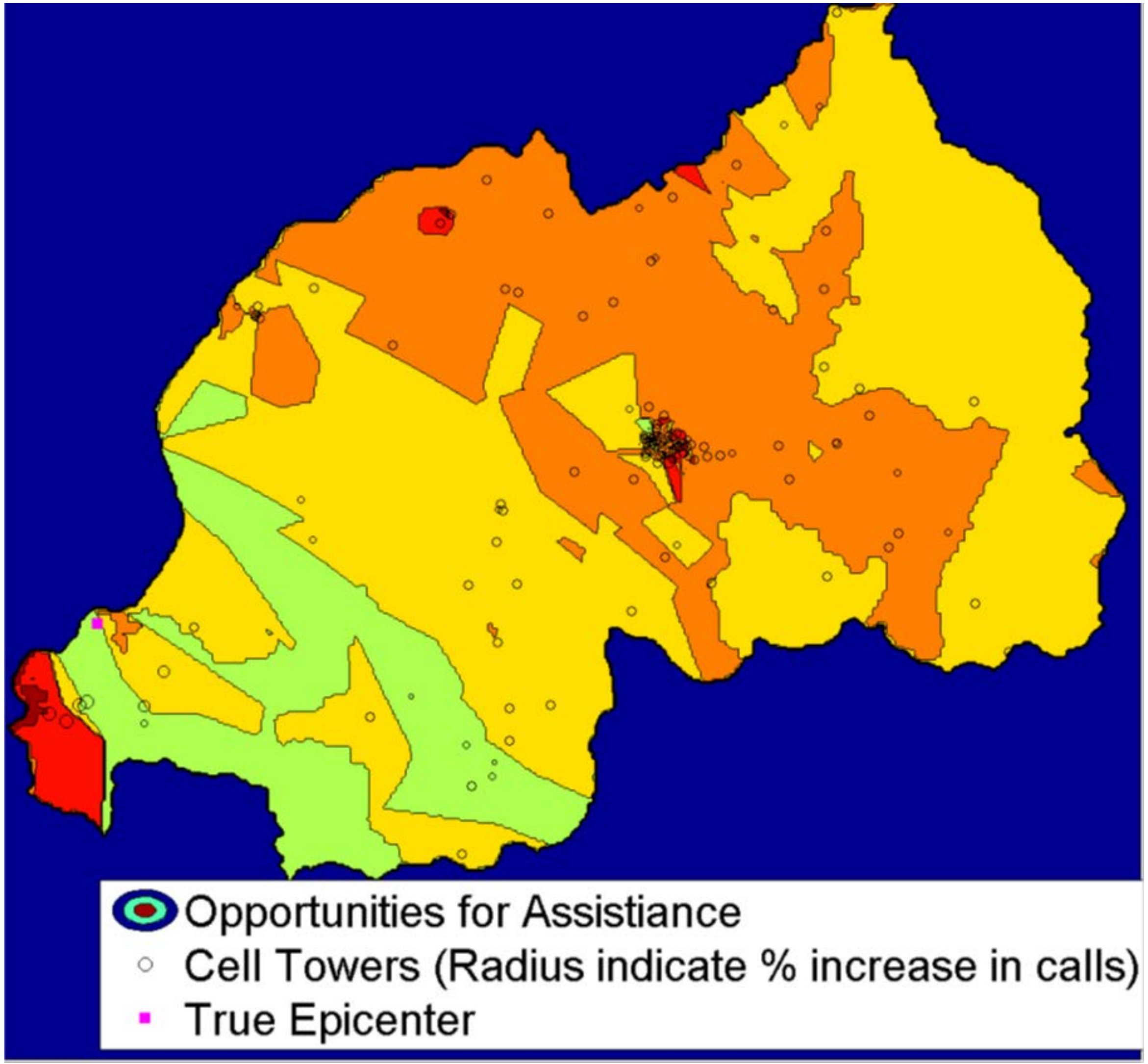
Inferring Opportunities to Assist

- Opportunities for Assistance Day 0



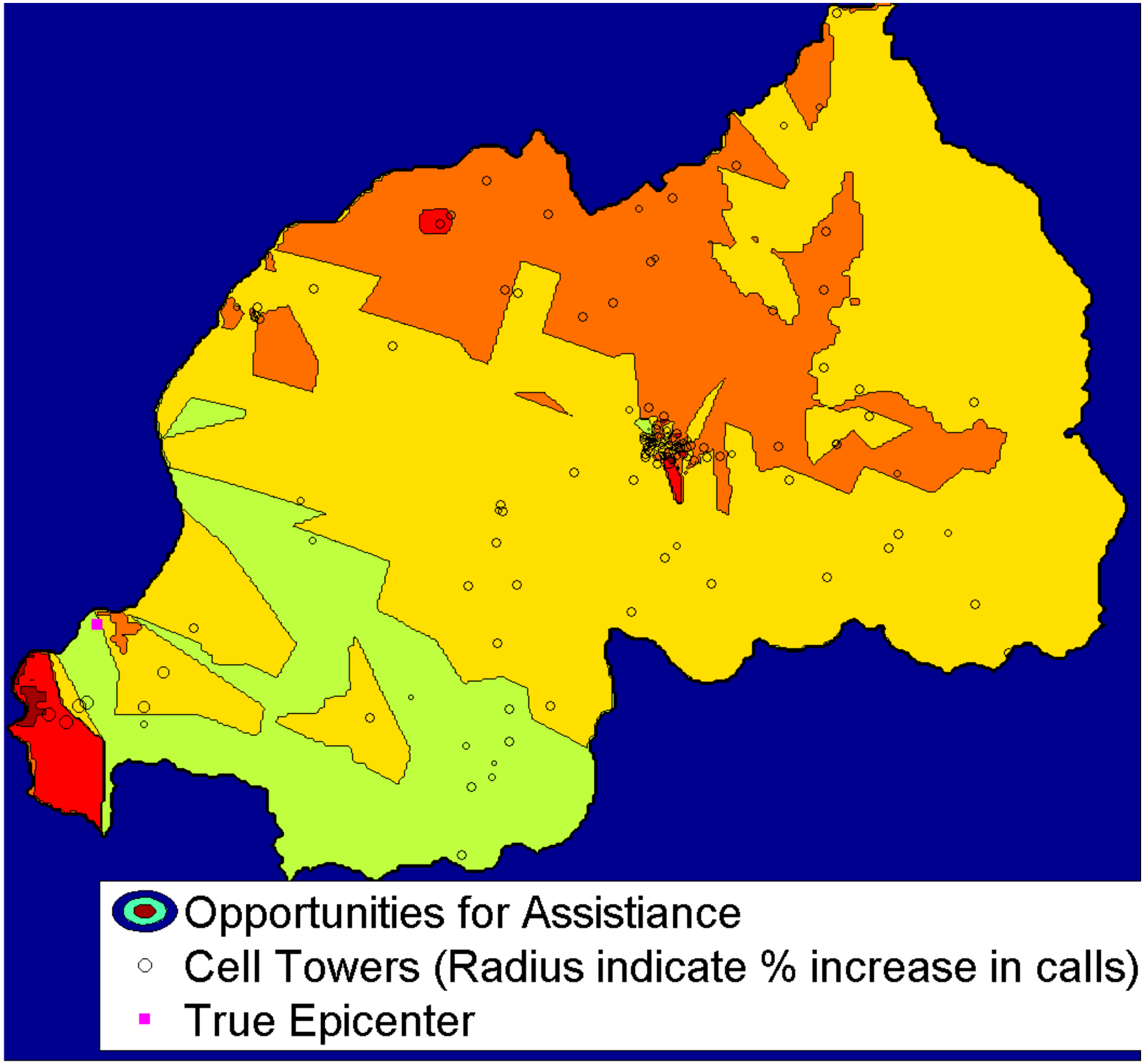
Inferring Opportunities to Assist

- Opportunities for Assistance Day 1



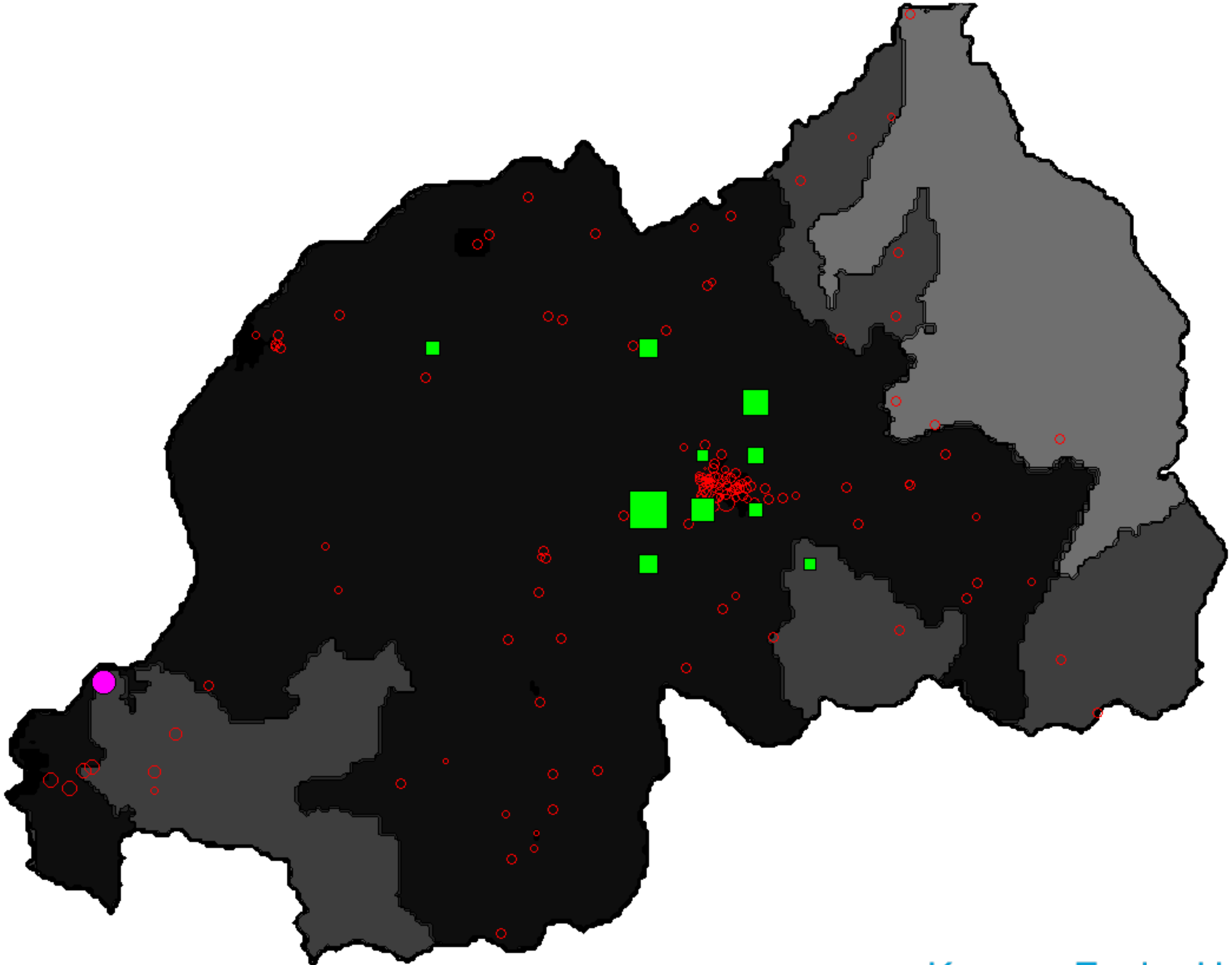
Inferring Opportunities to Assist

- Opportunities for Assistance Day 2



Value of Survey

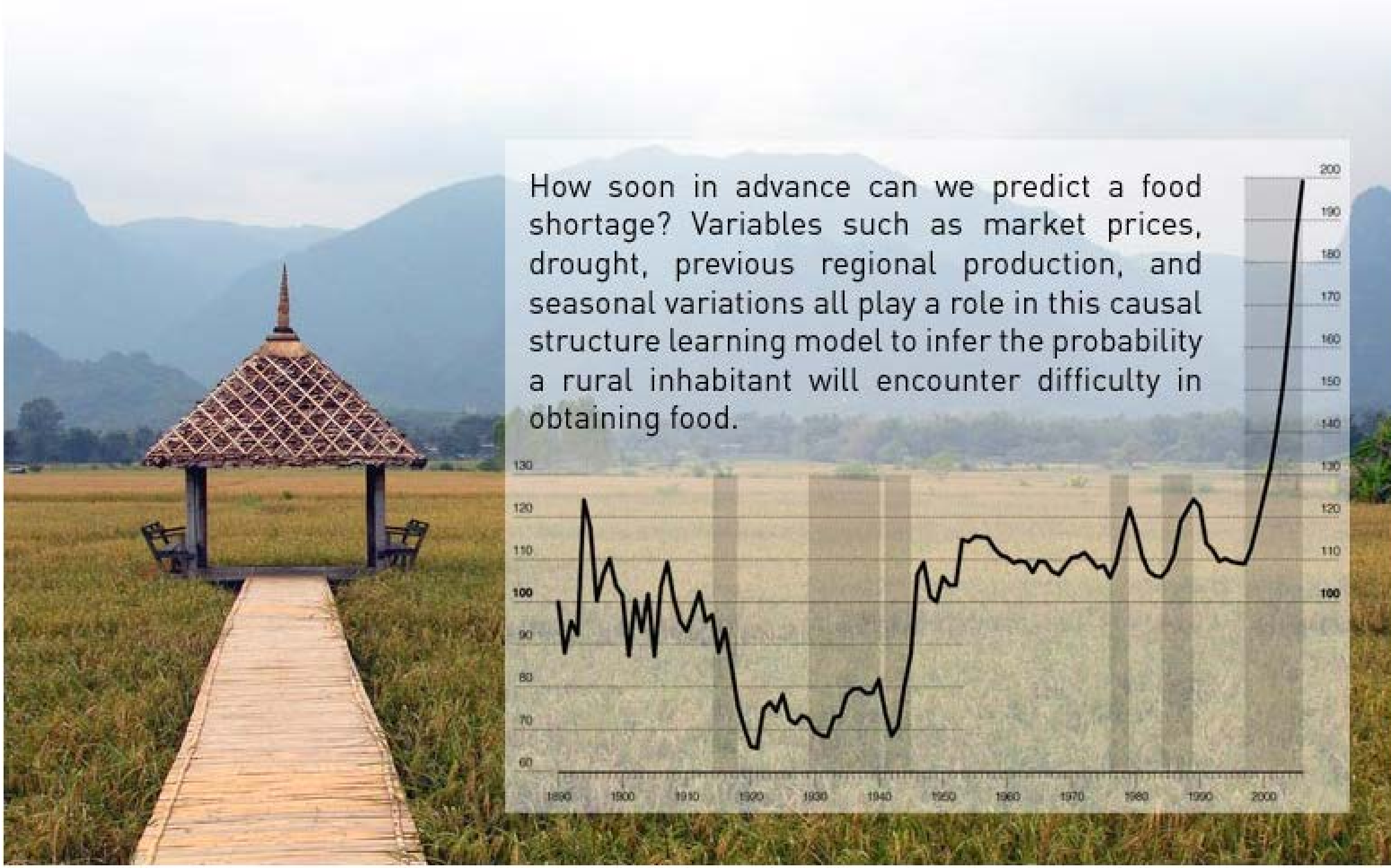
- Ideal Reconnaissance (Day 2)



Kapoor, Eagle, Horvitz, 2010

AI-D.ORG

AI-D ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT

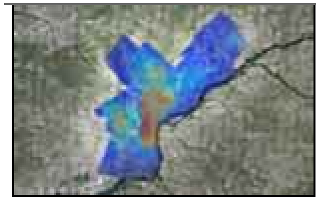


Data-Driven Development

The unprecedented volume of data currently being generated in the

AAAI AI-D Symposium

The AAAI AI-D Spring Symposium at Stanford is being



Can we quantify a crime wave? Is crime contagious? Given the time, place, and nature of a crime, we are attempting to infer casual relationships between crimes and locations across a city. - *J. TOOLE, J. PLOTKIN, N. EAGLE*

Quantifying the Stability of Society

Is there such a thing as a 'poverty trap'? Logistic classifiers applied on communication and census data point to a new mechanism for poverty that relates to the persistence of relationships. This analysis shows that economic exchanges flow primarily through these persistent edges and the inability to maintain these ties can prevent upward economic mobility. - *Y. DE MONTJOYE, A. CLAUSET, N. EAGLE*

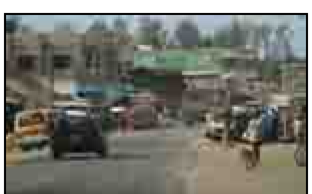
Economic Shocks in Rwanda

Do people react to economic shocks in a similar manner? Time-series analysis of anonymized mobile phone records coupled with random surveys, will hopefully lead to better insight about the dynamics of rural economies. - *J. BLUMENSTOCK, N. EAGLE*

Communication as a Lens into Poverty

How do communication patterns reflect poverty? We find the principal components of a wide range of diversity metrics, including Shannon entropy, explain over two-thirds the variance of regional socioeconomic status. - *N. EAGLE, M. MACY, R. CLAXTON*

Identifying Need and Risk



Can mobile phones identify high-risk behavior? A group of 10 male sex-workers in coastal Kenya where provided with mobile phones that logged communication, proximity and movement behavior. When coupled with self -report surveys, we are attempting to develop a system that can infer the onset of high-risk behavior and deliver salient information in real-time. - *E. SANDERS, N. EAGLE*

AI-D Sample Research Projects

Below are a list of active AI-D research projects. If you'd like to add your own project to this list, please feel free to get involved.

- Food Shortage
- Disease Surveillance
- Diffusion of Norms
- Mobility and Malaria
- Slum Dynamics
- Computational City Planning
- Urban Growth Models
- Expertise Inference
- Crime as Contagion
- Stability of Society
- Shock Modeling
- Entropy and Poverty
- Realtime Risk

Co-Location: Computational Epidemiology

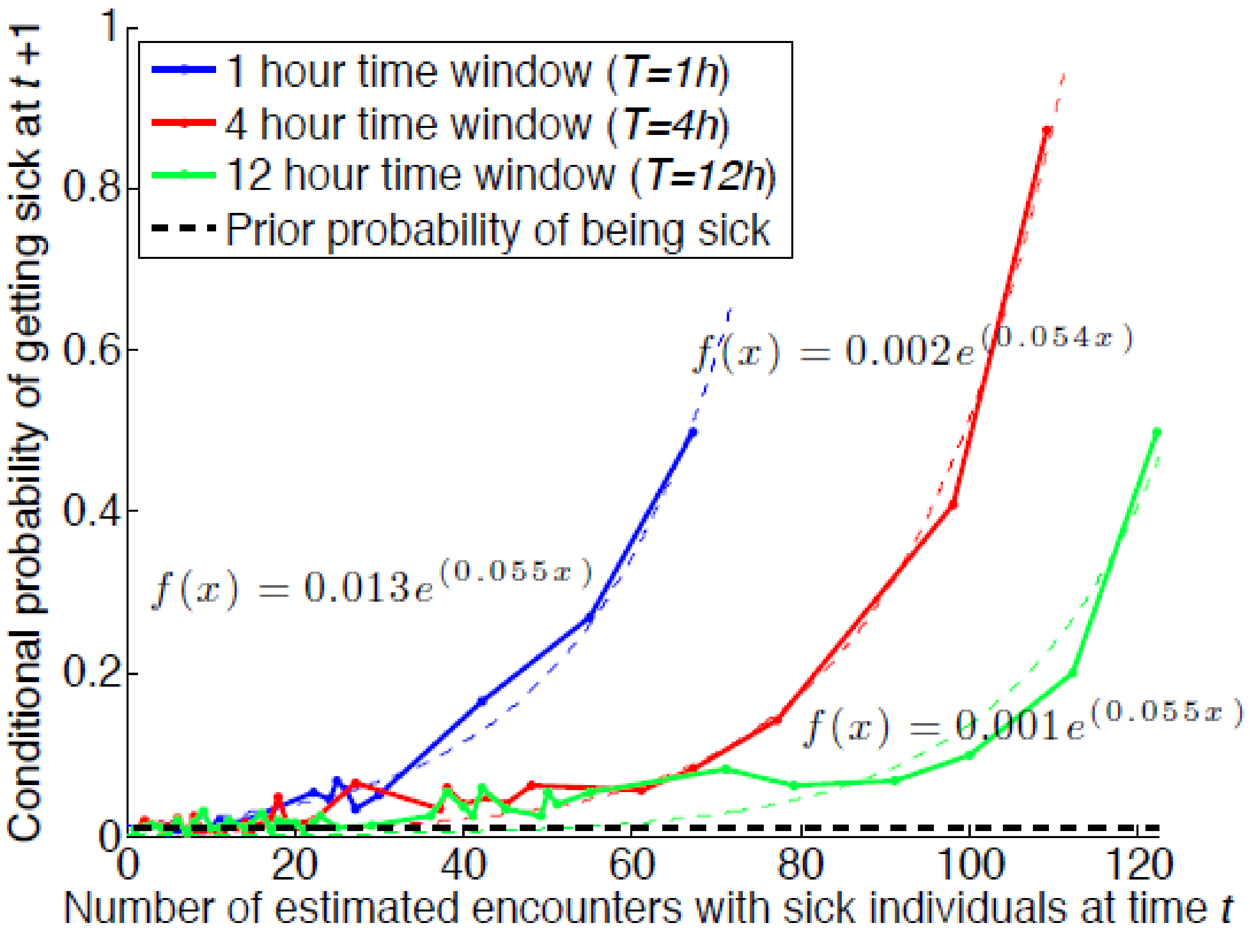
Understanding spread of illness

A. Sadilek, H. Kautz, V. Silenzio, Modeling Spread of Disease from Social Interactions, ICWSM 2012.

Identifying Illness from Tweet Terms

Positive Features		Negative Features	
Feature	**Weight**	**Feature**	**Weight**
sick	0.9579	sick of	−0.4005
headache	0.5249	you	−0.3662
flu	0.5051	of	−0.3559
fever	0.3879	your	−0.3131
feel	0.3451	lol	−0.3017
cough	0.3062	who	−0.1816
feeling	0.3055	u	−0.1778
coughing	0.2917	love	−0.1753
throat	0.2842	it	−0.1627
cold	0.2825	her	−0.1618
home	0.2107	they	−0.1617
still	0.2101	people	−0.1548
bed	0.2088	shit	−0.1486
better	0.1988	smoking	−0.0980
being	0.1943	i'm sick of	−0.0894
being sick	0.1919	so sick of	−0.0887
stomach	0.1703	pressure	−0.0837
and my	0.1687	massage	−0.0726
infection	0.1686	i love	−0.0719
morning	0.1647	pregnant	−0.0639

Collocation and Transmission

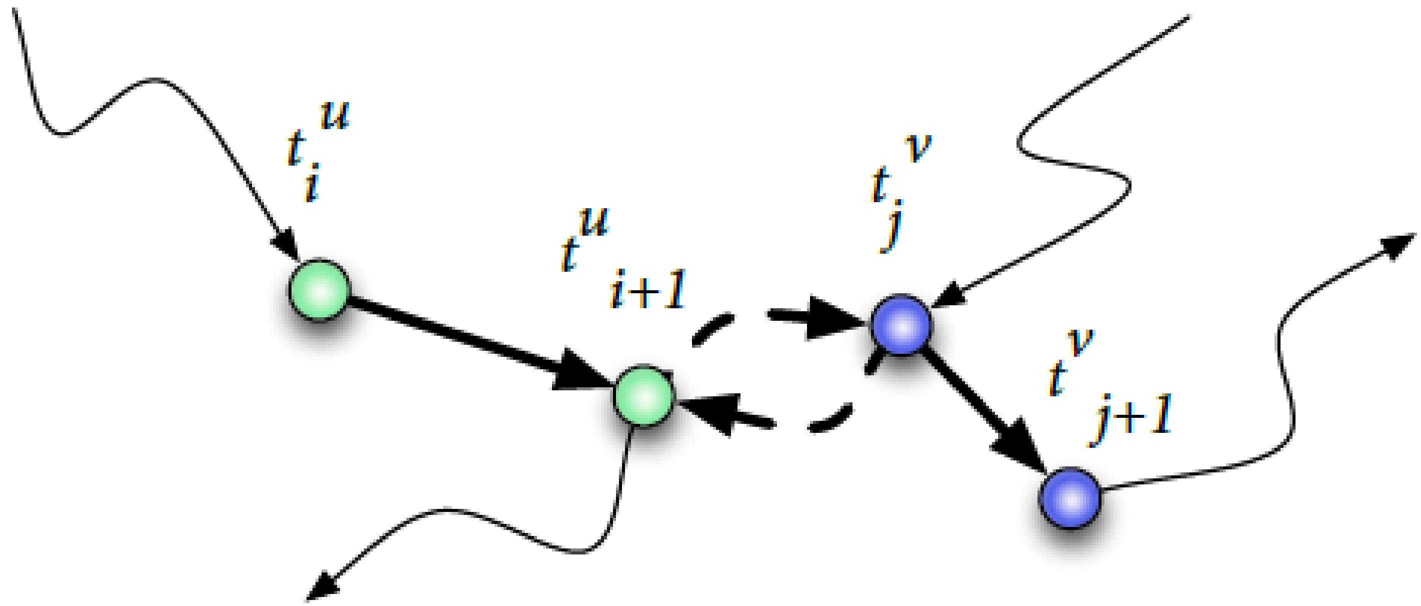


Directions for Disrupting Spread of Illness

"Crowd physics:" On physics of the crowd

e.g., Studies of flow through a population

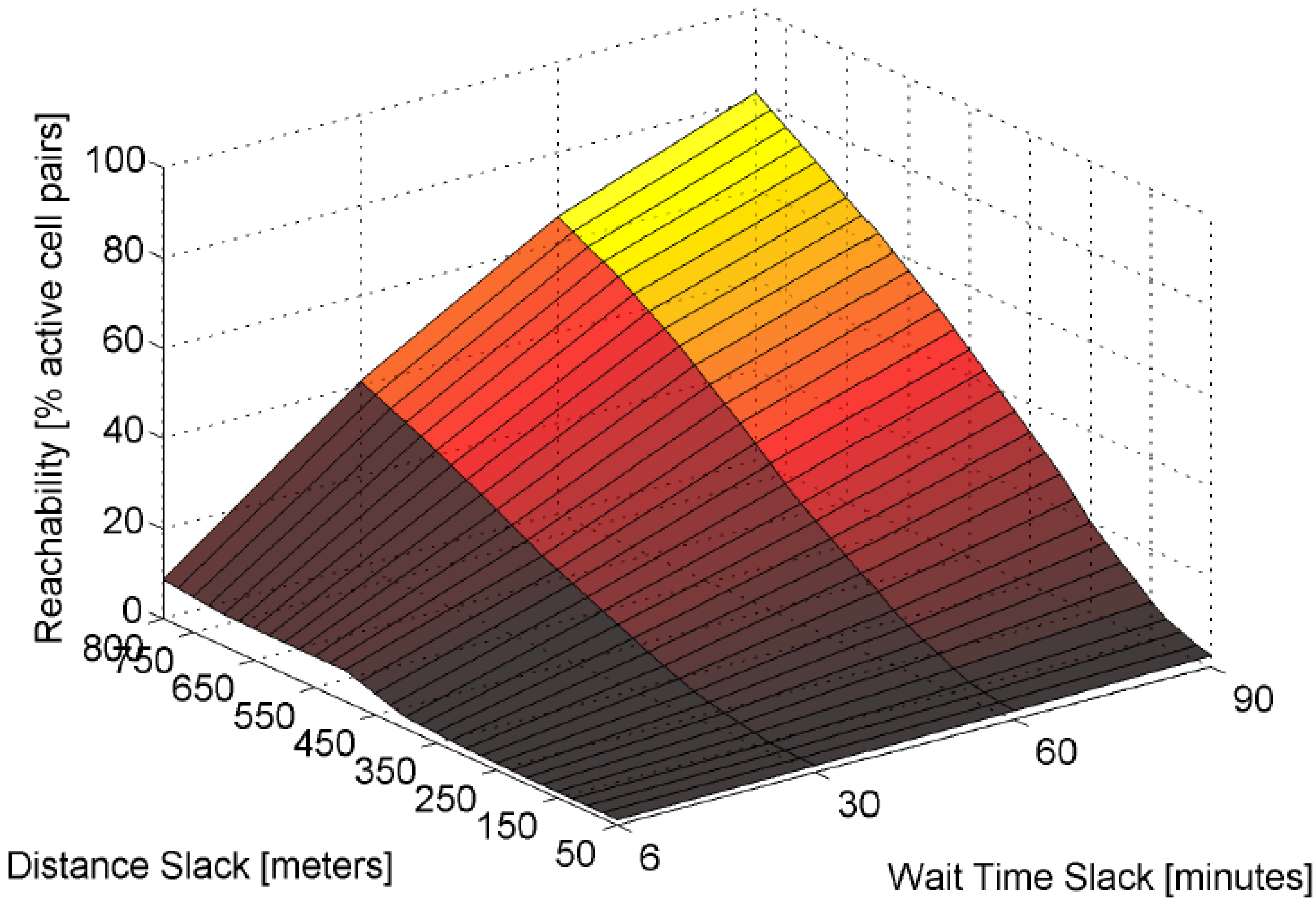
e.g., Routing graph per proximity & dwell



A. Sadilek, J. Krumm, and E. Horvitz. Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks, ICWSM 2013.

Reachability, Permeability, Phase Transitions

e.g., In Seattle




Click to toggle time slider animation.
12/31/2011 4:00 pm
/31/2011
Tipping Point
1/2/2012

Opportunities to Slow Spread of Disease

Study robustness & fragility of routing graph

Disruption of reachability and permeability