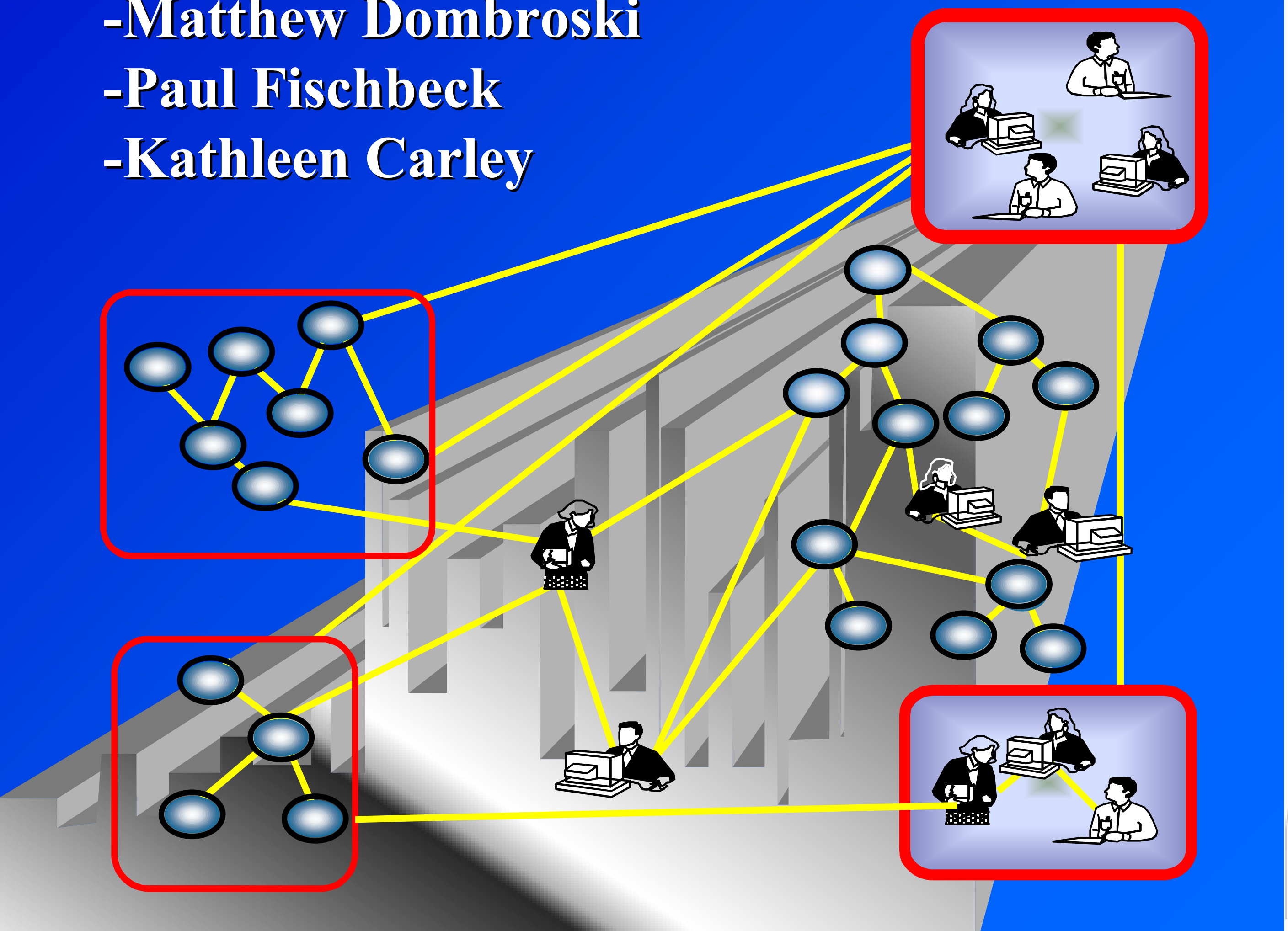
Estimating the Shape of Covert Networks

-Matthew Dombroski
-Paul Fischbeck
-Kathleen Carley



Carnegie Mellon

CASOS

Center for Computational Analysis of Social and Organizational Systems

Matthew J. Dombroski

Porter Hall

Pittsburgh, PA 15213

Tel: - 1-412-268-1876

Fax: 1-412-268-6938

Email: mdombros @andrew.cmu.edu

This work was supported by the NSF IGERT 9972762 in CASOS

2003

Agenda

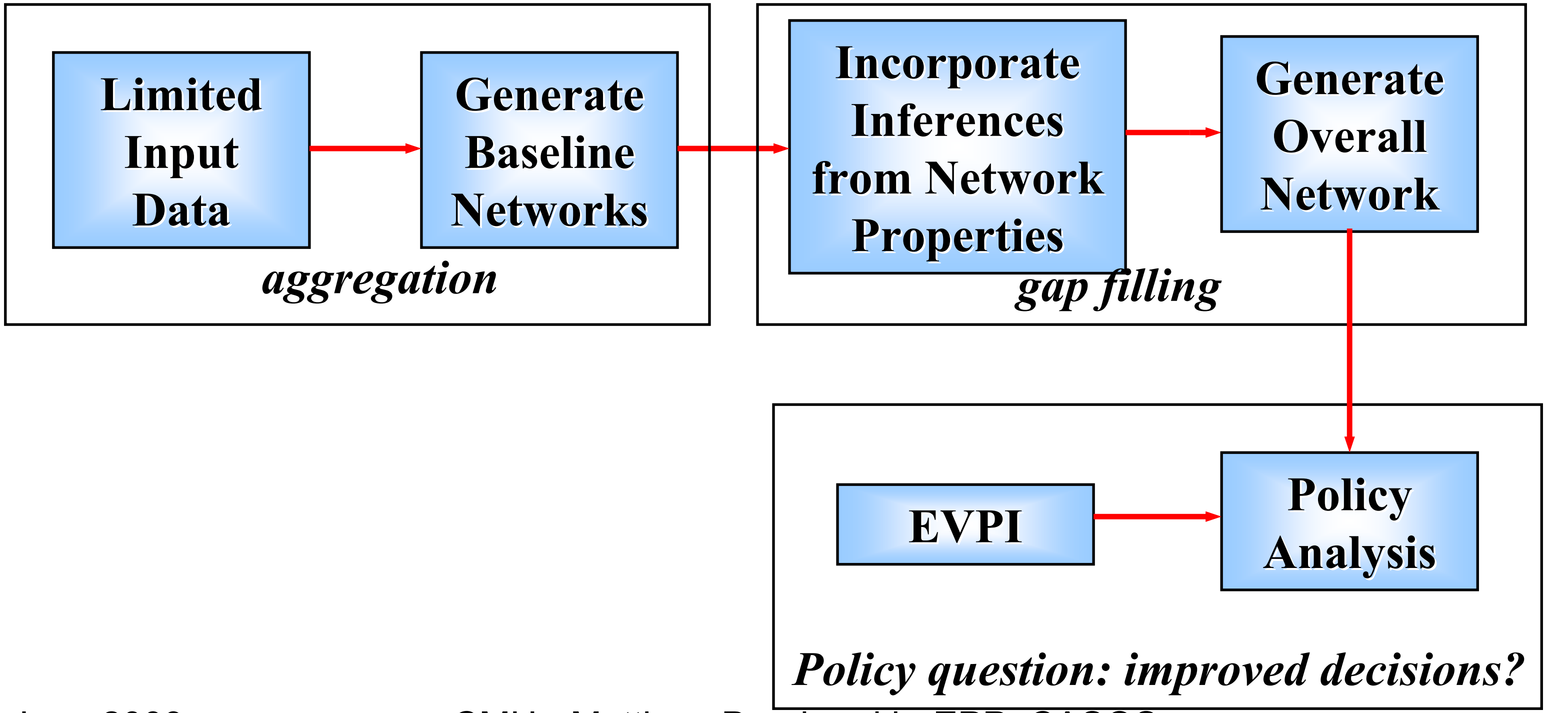
- Problem Statement
- Model Building
  - Relationship of Interest (ROI) and Aggregation
  - Network Properties
  - Empirical Data Transformation
  - Dyadic Dependency Models
- Comparison of updating models
- Command and Control Policy Analysis
  - Covert network destabilization
  - Comparison of Decision Performance Metrics versus Aggregate Network Metrics
- Conclusion and Future Research

Problem Statement and Motivation

- Command and Control may be able to make better decisions by determining
  - Who is in organization
  - Who is important in organization
  - How information flows through the organization
- Important applications in improving
  - Destabilization of covert and illicit organizations
  - Communication in relief and military operations with uncertain communication structure
- Most organizations have an unknown communication structure
  - Data available are often disparate and incomplete
  - High degree of uncertainty in the data
  - May not be able to fully characterize its structure

Proposed Approach

- Goal: Using a predefined relationship (dyad), construct a model that infers relationships and builds a prediction of the network quickly using limited data



Context Determines The Relationship of Interest (ROI)

- Clearly define the relationship that is of interest
  - Context dependent
    - Covert organizations (conspiring to commit an illicit act)
    - Military operations (request for resources or assistance)
  - Limits set of empirical data available for model
- Input data arrives according to Poisson process
- Quantify and distinguish input data informing ROI
  - Reliability
  - Strength of data towards relationship

Aggregation by Bayes Rule

- If L_{ij} is the event that persons *i* and *j* share the ROI
- And I_{ij} is the event that certain data is observed about the relationship between persons *i* and *j*
- Then Define: $P(I_{ij} \mid L_{ij})$ and $P(I_{ij} \mid \overline{L_{ij}})$

$$P(L_{ij} \mid I_{ij}) = \frac{P(I_{ij} \mid L_{ij})P(L_{ij})}{P(I_{ij} \mid L_{ij})P(L_{ij}) + P(I_{ij} \mid \overline{L_{ij}})P(\overline{L_{ij}})}$$

- Network priors can be assigned a priori

Incorporating Inference Using Empirical Data (cont.)

- Before constructing inference model
  - Transform empirical frequency networks into probabilities of ROI
  - Assume frequency of interaction indicates strength of relationship
- Transforming Empirical Network Data
  - Define ROI Threshold – number of interactions required to attain relationship of interest (x_{max})
  - Define marginal increase of each additional interaction to probability of ROI Transforming Empirical Network Data (λ)
- Transform interaction frequency into probabilities—

$$P(ROI) = \frac{(1 - e^{-\lambda x})}{(1 - e^{-\lambda x_{\max}})}$$

Network Transformation

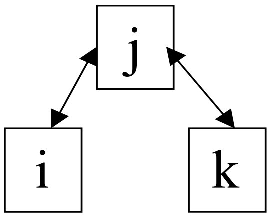
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0	2	1	0	0	2	0	0	0	1	1	2	0	0	0	1	0	1	0
2	0	0	10	0	0	2	1	0	2	0	0	0	6	2	0	1	0	0	0	1
3	2	10	0	6	11	14	15	4	12	0	5	4	3	8	10	8	11	0	2	19
4	1	0	6	0	2	3	9	1	8	0	0	5	0	0	2	4	3	2	2	6
5	0	0	11	2	0	2	8	1	1	1	0	0	2	0	1	1	0	0	0	3
6	0	2	14	3	2	0	30	2	8	0	4	4	1	6	2	14	9	0	1	51
7	2	1	15	9	8	30	0	10	4	2	7	3	0	12	9	10	9	2	3	40
8	0	0	4	1	1	2	10	0	3	0	2	0	1	3	3	3	5	0	0	6
9	0	2	12	8	1	8	4	3	0	0	5	5	2	2	4	5	6	1	0	5
10	0	0	0	0	1	0	2	0	0	0	0	0	0	0	1	2	0	0	0	0
11	1	0	5	0	0	4	7	2	5	0	0	0	0	1	3	3	5	3	0	7
12	1	0	4	5	0	4	3	0	5	0	0	0	0	0	0	0	0	0	0	3
13	2	6	3	0	2	1	0	1	2	0	0	0	0	2	1	3	3	0	1	0
14	0	2	8	0	0	6	12	3	2	0	1	0	2	0	3	8	11	1	4	8
15	0	0	10	2	1	2	9	3	4	1	3	0	1	3	0	9	14	0	6	9
16	0	1	8	4	1	14	10	3	5	2	3	0	3	8	9	0	26	3	1	12
17	1	0	11	3	0	9	9	5	6	0	5	0	3	11	14	26	0	3	0	9
18	0	0	0	2	0	0	2	0	1	0	3	0	0	1	0	3	3	0	0	0
19	1	0	2	2	0	1	3	0	0	0	0	0	1	4	6	1	0	0	0	5
20	0	1	19	6	3	51	40	6	5	0	7	3	0	8	9	12	9	0	5	0

Network Transformation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.00	0.00	0.26	0.14	0.00	0.00	0.26	0.00	0.00	0.00	0.14	0.14	0.26	0.00	0.00	0.00	0.14	0.00	0.14	0.00
2	0.00	0.00	0.80	0.00	0.00	0.26	0.14	0.00	0.26	0.00	0.00	0.00	0.60	0.26	0.00	0.14	0.00	0.00	0.00	0.14
3	0.26	0.80	0.00	0.60	0.83	0.91	0.93	0.45	0.86	0.00	0.53	0.45	0.36	0.71	0.80	0.71	0.83	0.00	0.26	0.98
4	0.14	0.00	0.60	0.00	0.26	0.36	0.76	0.14	0.71	0.00	0.00	0.53	0.00	0.00	0.26	0.45	0.36	0.26	0.26	0.60
5	0.00	0.00	0.83	0.26	0.00	0.26	0.71	0.14	0.14	0.14	0.00	0.00	0.26	0.00	0.14	0.14	0.00	0.00	0.00	0.36
6	0.00	0.26	0.91	0.36	0.26	0.00	1.00	0.26	0.71	0.00	0.45	0.45	0.14	0.60	0.26	0.91	0.76	0.00	0.14	1.00
7	0.26	0.14	0.93	0.76	0.71	1.00	0.00	0.80	0.45	0.26	0.66	0.36	0.00	0.86	0.76	0.80	0.76	0.26	0.36	1.00
8	0.00	0.00	0.45	0.14	0.14	0.26	0.80	0.00	0.36	0.00	0.26	0.00	0.14	0.36	0.36	0.36	0.53	0.00	0.00	0.60
9	0.00	0.26	0.86	0.71	0.14	0.71	0.45	0.36	0.00	0.00	0.53	0.53	0.26	0.26	0.45	0.53	0.60	0.14	0.00	0.53
10	0.00	0.00	0.00	0.00	0.14	0.00	0.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.26	0.00	0.00	0.00	0.00
11	0.14	0.00	0.53	0.00	0.00	0.45	0.66	0.26	0.53	0.00	0.00	0.00	0.00	0.14	0.36	0.36	0.53	0.36	0.00	0.66
12	0.14	0.00	0.45	0.53	0.00	0.45	0.36	0.00	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36
13	0.26	0.60	0.36	0.00	0.26	0.14	0.00	0.14	0.26	0.00	0.00	0.00	0.00	0.26	0.14	0.36	0.36	0.00	0.14	0.00
14	0.00	0.26	0.71	0.00	0.00	0.60	0.86	0.36	0.26	0.00	0.14	0.00	0.26	0.00	0.36	0.71	0.83	0.14	0.45	0.71
15	0.00	0.00	0.80	0.26	0.14	0.26	0.76	0.36	0.45	0.14	0.36	0.00	0.14	0.36	0.00	0.76	0.91	0.00	0.60	0.76
16	0.00	0.14	0.71	0.45	0.14	0.91	0.80	0.36	0.53	0.26	0.36	0.00	0.36	0.71	0.76	0.00	1.00	0.36	0.14	0.86
17	0.14	0.00	0.83	0.36	0.00	0.76	0.76	0.53	0.60	0.00	0.53	0.00	0.36	0.83	0.91	1.00	0.00	0.36	0.00	0.76
18	0.00	0.00	0.00	0.26	0.00	0.00	0.26	0.00	0.14	0.00	0.36	0.00	0.00	0.14	0.00	0.36	0.36	0.00	0.00	0.00
19	0.14	0.00	0.26	0.26	0.00	0.14	0.36	0.00	0.00	0.00	0.00	0.00	0.14	0.45	0.60	0.14	0.00	0.00	0.00	0.53
20	0.00	0.14	0.98	0.60	0.36	1.00	1.00	0.60	0.53	0.00	0.66	0.36	0.00	0.71	0.76	0.86	0.76	0.00	0.53	0.00

Network Properties

- Triad Closure Property (Skvoretz, 1990)

If 

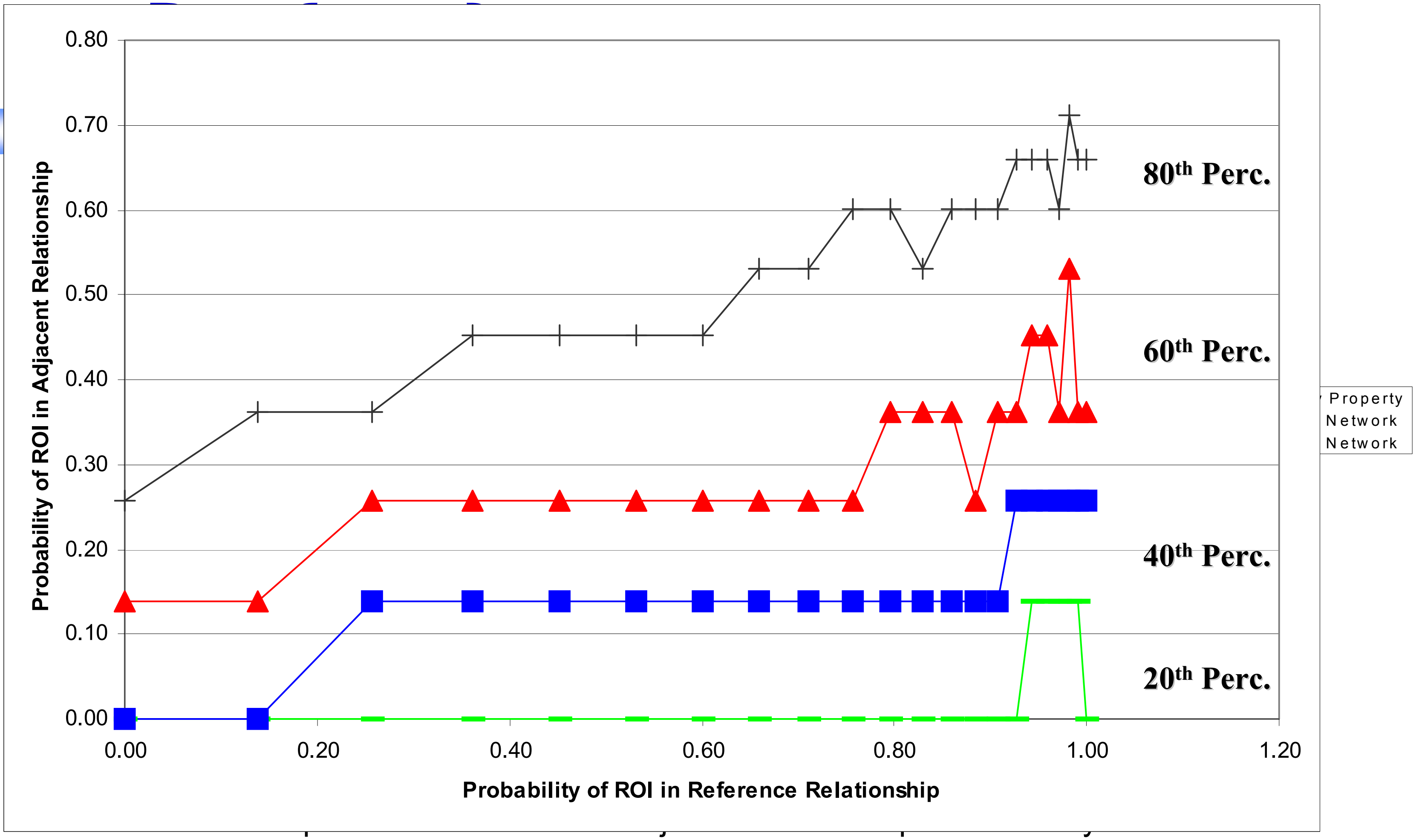
then $i \leftrightarrow k$ occurs with a probability greater than chance

- Adjacency Property
  - Corollary of triad closure that is modeled

If $i \leftrightarrow j$ occurs with probability greater than chance then $i \leftrightarrow k$

and $j \leftrightarrow k$ occur with a probability greater than chance

Incorporating Inference Using Empirical



- Model fit using neural network of 9 nodes and 3 layers (percent current error on 4000 training data points is 0.9%)

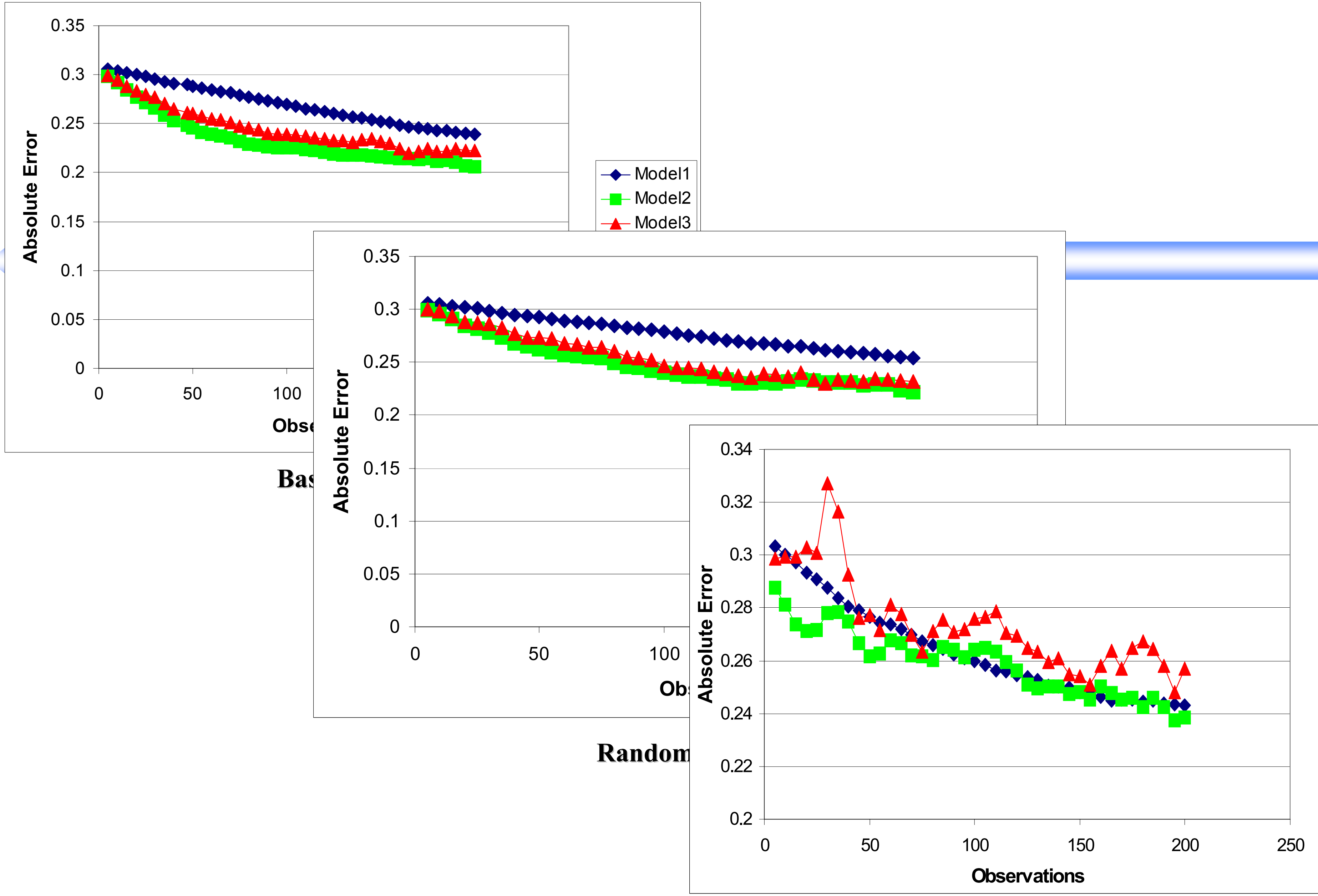
Several Possible Implementations of Model

- Three inference models constructed:
  - Model 1-Direct update dyads using incoming information and Bayes Rule only (control model)
  - Model 2-Model 1 plus indirect inferencing only on immediately adjacent relationships using inference model
  - Model 3-Model 2 plus indirect inferencing on all other relationships using inference model

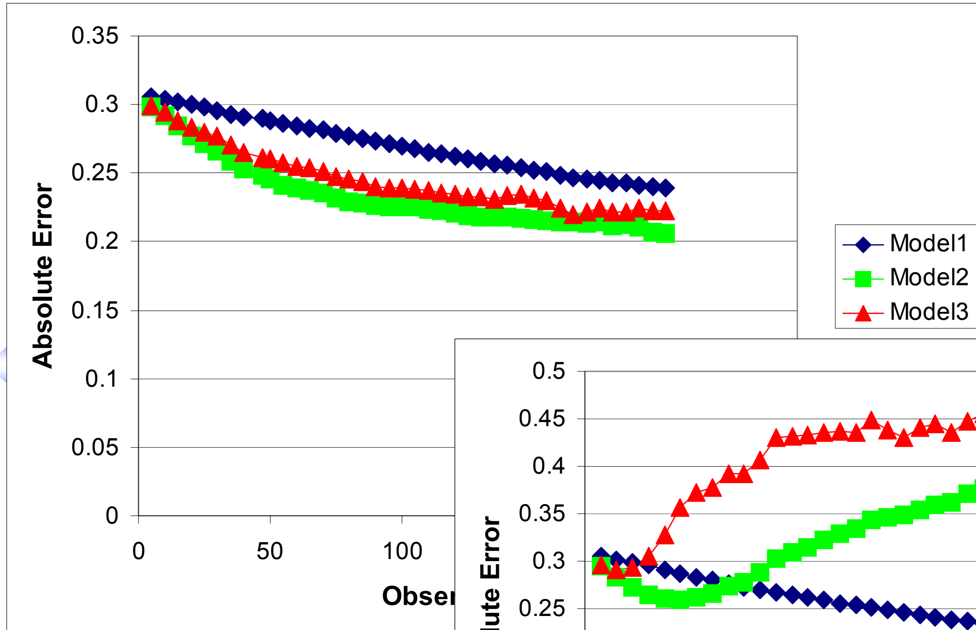
Performance of Models Compared Using Simulation

- Sample data (20 of 58 nodes), perform an update, calculate performance, another update
- Constant uniform prior (0.2)
- Reconstruct by varying:
  - Input data accuracy
  - Proportion of information supporting existence of ROI (0.5)
- Performance Metric
  - Absolute Error (sum of prediction probability minus actual probability)

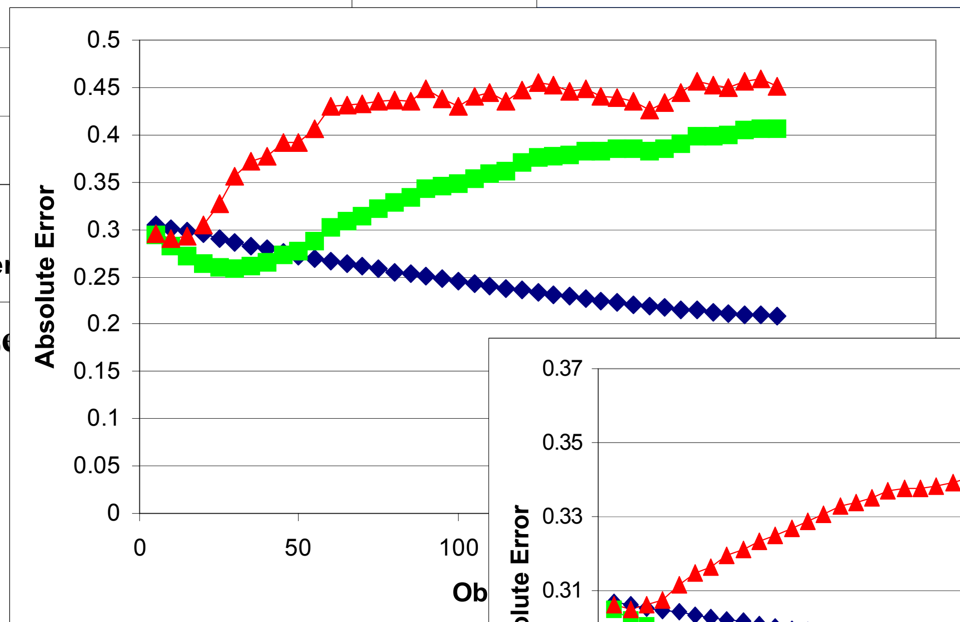
$$\text{Absolute Error} = \left(\sum_{i>j}\sum_{j=1}^{N} abs(p_{ij} - A_{ij})\right)$$



Sensor Reliable Conditionals (0.8, 0.1)

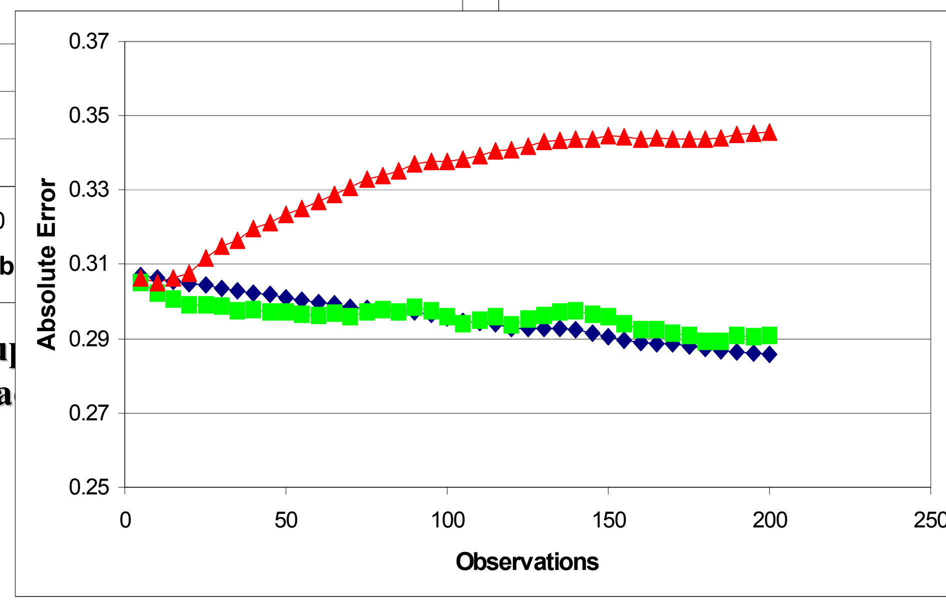


Base



0.8 P(observing dyad su

0.2 P(observing non dya



0.2 P(observing dyad supporting information)

0.8 P(observing non dyad supporting information)

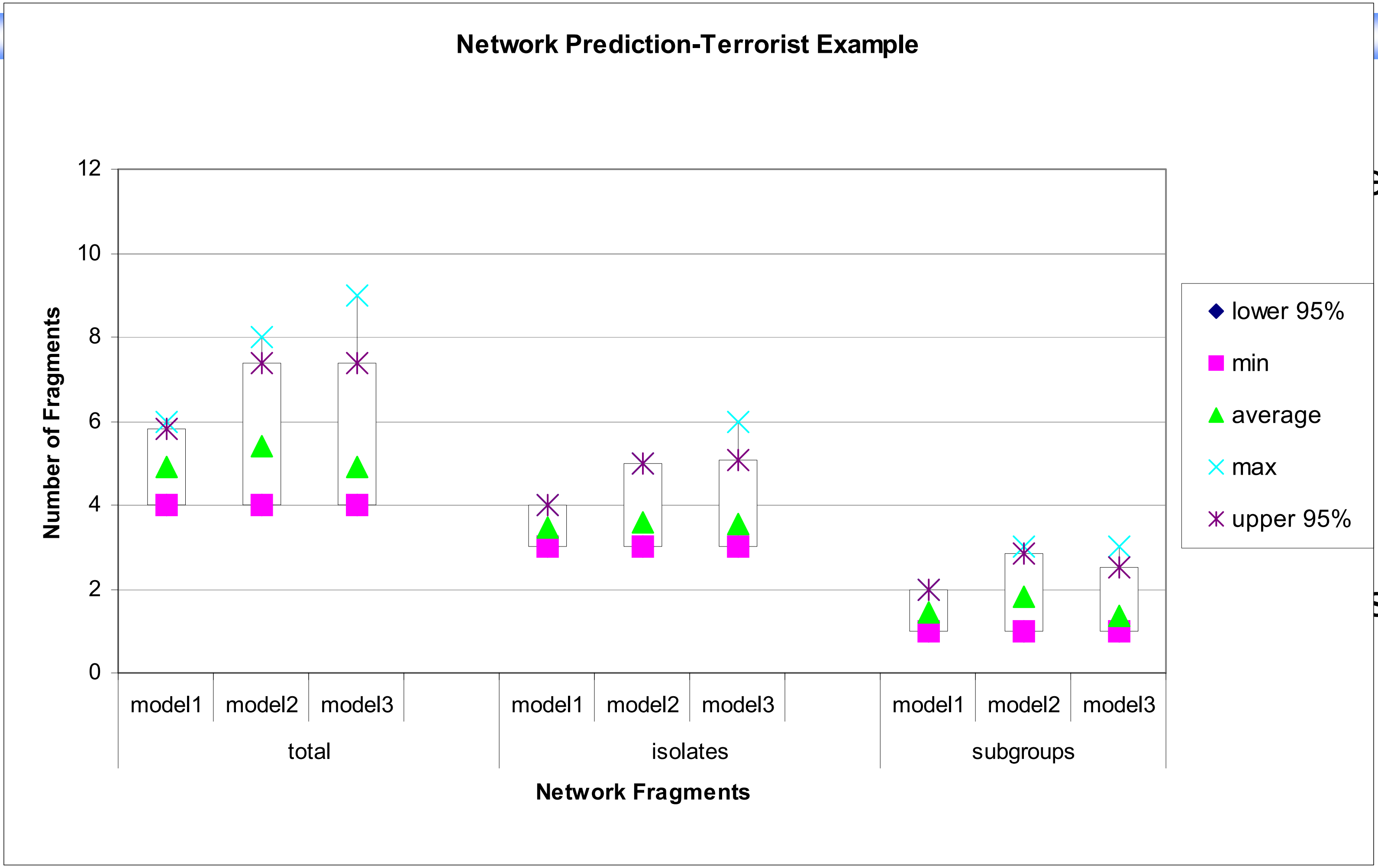
Conclusions from Simulations

- Inference model results are mixed
- Under conditions of balance between positive and negative tie input data and random or moderately accurate conditionals
  - Performance on Absolute Error is relatively good for inference models
  - After about 400 updates, base model outperforms inference models
- Under conditions of imbalance
  - Base model outperforms inference models for nearly all updates
  - Inference over/under predicts the network
- Under conditions of accurate conditionals
  - Models are indistinguishable

Policy Analysis

- What does this mean for command and control decisions?
- C2 Terrorist Example
  - C2 must make a decision of who to remove from network of conspirators using models
  - Get Network prediction after *x* updates
  - Estimate the *n* most critical individuals to remove from network
  - Remove them from network
  - Assess how many fragments the network is broken into
- Compare the number of fragments across different model implementations

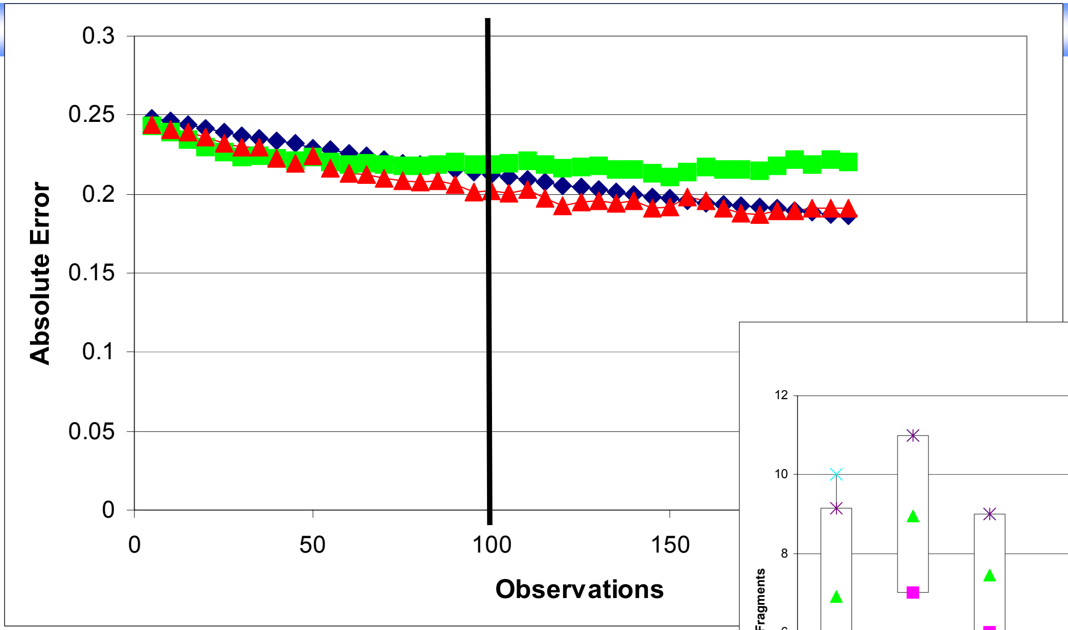
Policy Analysis
Scenario-Covert Networks (cont.)



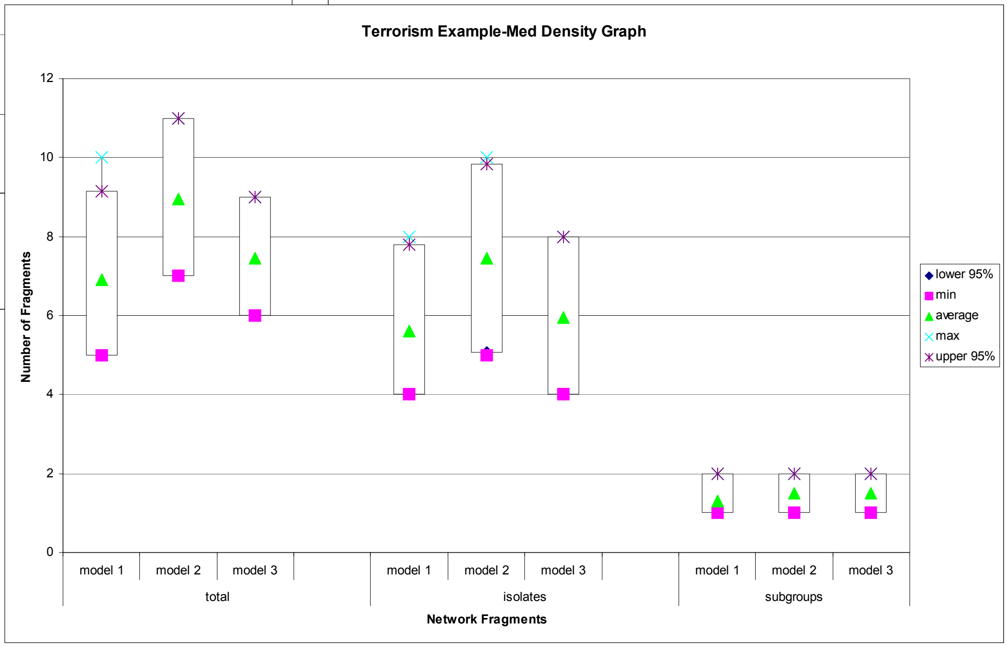
en

st

A Comparison of Simulation Performance Metrics to Decision Metrics



Mid Density Network



Network Match Performance Versus Decision Performance

- Conduct a paired t-test for each network prediction for absolute error and fragments
- Null Hypotheses:
  - E(absolute error$_1$ - absolute error$_2$) = 0, E(fragments$_1$ – fragments$_2$) = 0
  - E(absolute error$_1$ - absolute error$_3$) = 0, E(fragments$_1$ - fragments$_3$) = 0
  - E(absolute error$_2$ - absolute error$_3$) = 0, E(fragments$_2$ – fragments$_3$) = 0

Hypothesis	Mean Value	Standard Deviation	T-value	Dof	Significance Level
$-(Abs_1 - Abs_2)$	0.95	16.8	0.25	19	> 0.1
$Frag_1 - Frag_2$	-2.05	1.19	-7.70	19	< 0.001
$-(Abs_1 - Abs_3)$	-1.67	15.3	-0.48	19	> 0.1
$Frag_1 - Frag_3$	-0.55	1.19	-2.07	19	0.025 < S < 0.05
$-(Abs_2 - Abs_3)$	-2.60	3.01	-3.87	19	< 0.001
$Frag_2 - Frag_3$	1.50	1.15	5.85	19	< 0.001

Conclusions

- Good/bad performance on simulation metric does not imply good/bad performance in C2 decision space
  - Absolute Error recommends 3 over 2, but 2 = 1 and 3 = 1
  - Fragments recommends 2, 3, then 1
- Decision metrics may predict performance better than aggregate network metrics
  - Fragments is a better measure of the actual decision we will make
  - Absolute error is one step abstracted from the decision
- C2 decisions can benefit from a model
  - Care must be taken in evaluating performance of such models
  - Slight improvement in the number of fragments generated
  - Errors when too many updates occur

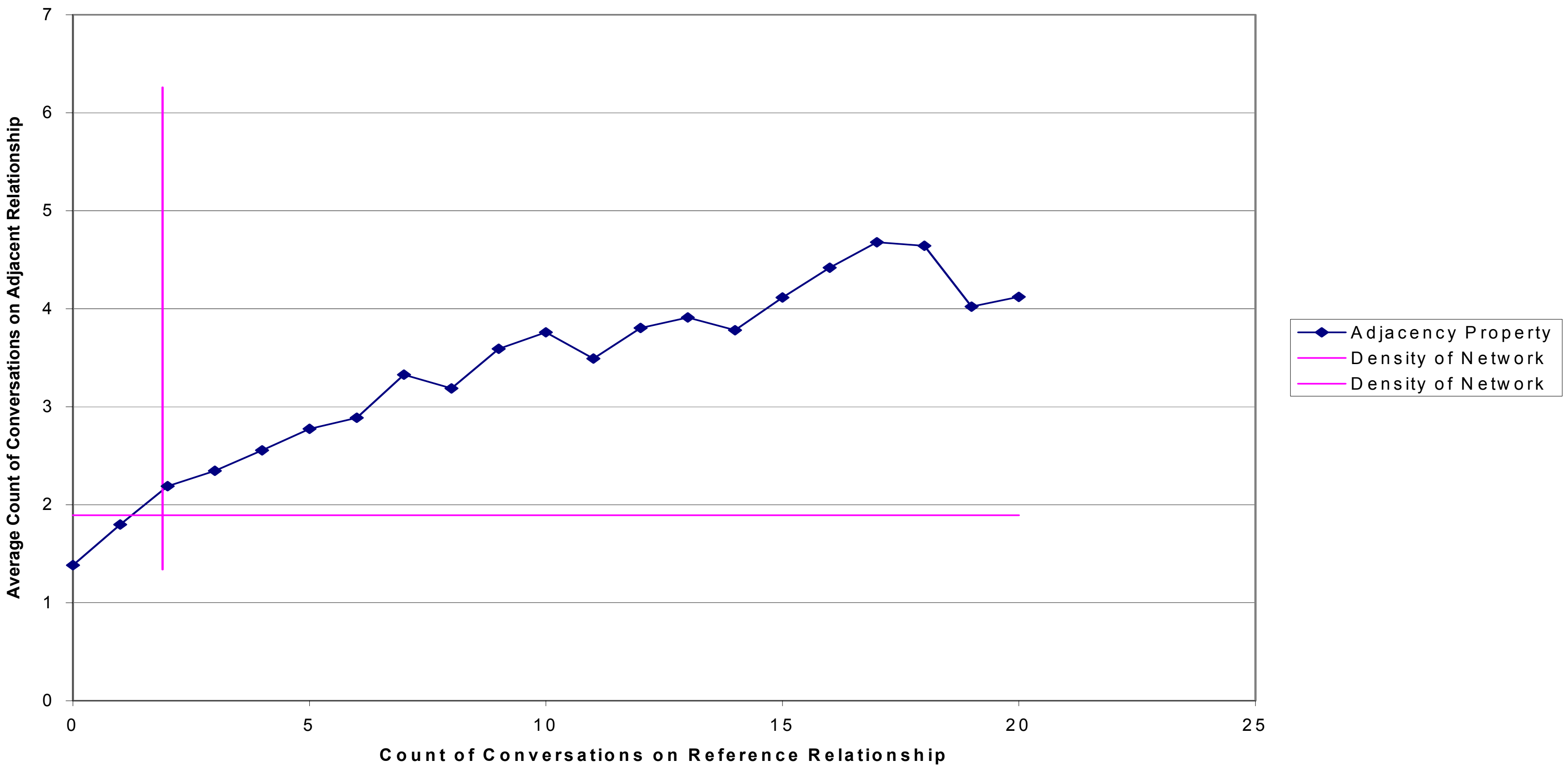
Future Work

- Explore the relationship between other C2 decision metrics and aggregate network metrics
  - Improving Communication for Military Operations
    - Network prediction indicates where communication is weak
    - Add dyads to force individuals to communicate better
    - Expected number of individuals to learn a new fact
  - Vaccination Strategies Against a Contagious Disease
- Develop more rigorous model using triad closure
- Characterize several different types of models for different types of networks
  - Work relationships
  - Friend relationships

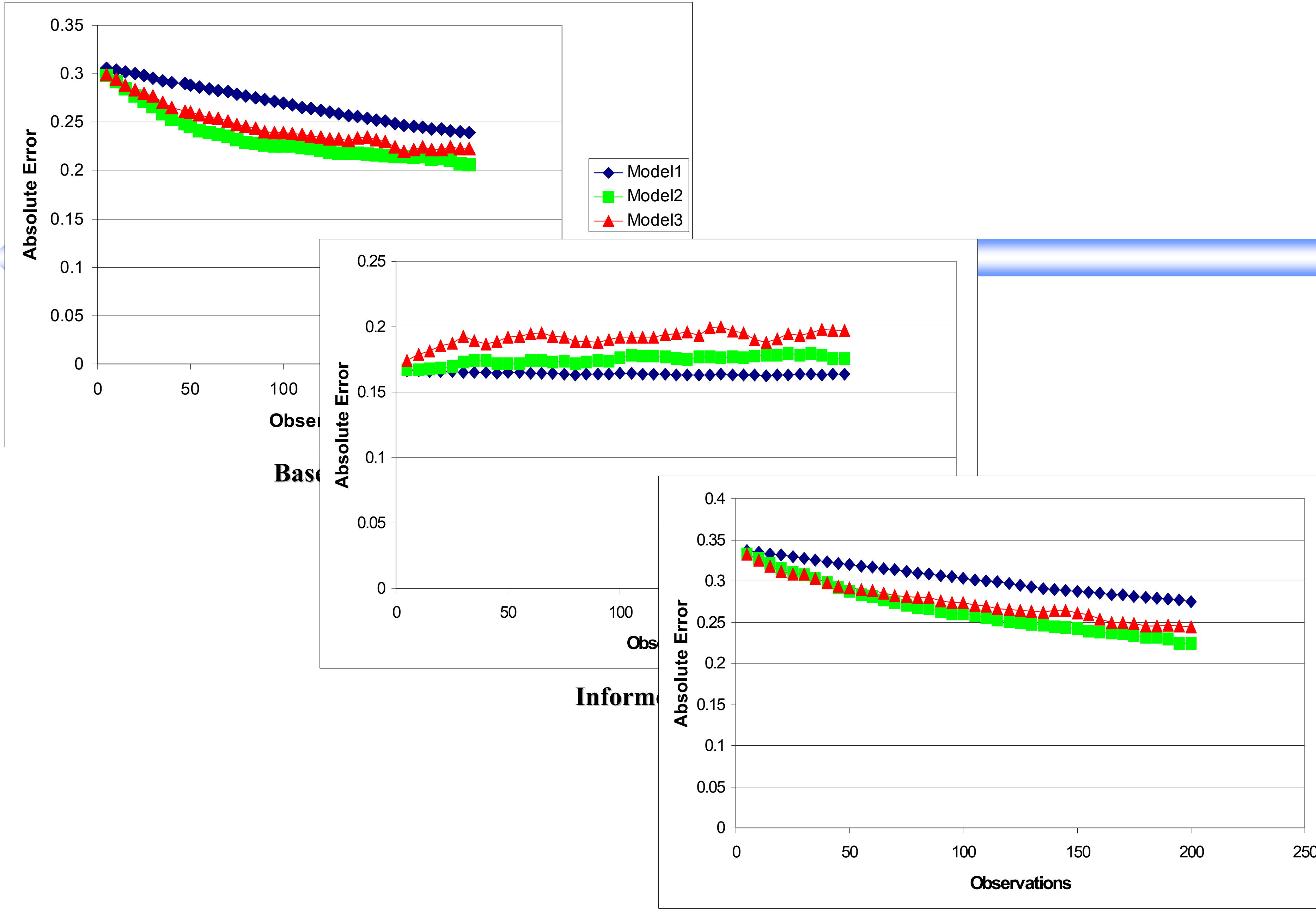
Questions and Comments

Acknowledgements: This research has been supported, in part, by National Science Foundation under the IGERT program for training and research in CASOS and the ONR. Additional support was also provided by the Center for Computational Analysis of Social and Organizational Systems and the Department of Engineering and Public Policy at Carnegie Mellon. The views and conclusions contained in this document are those of the author and should not be interpreted as representing official policies, either expressed or implied, of the Office of Naval Research grant number 1681-1-1001944, the National Science Foundation, or the U.S. government.

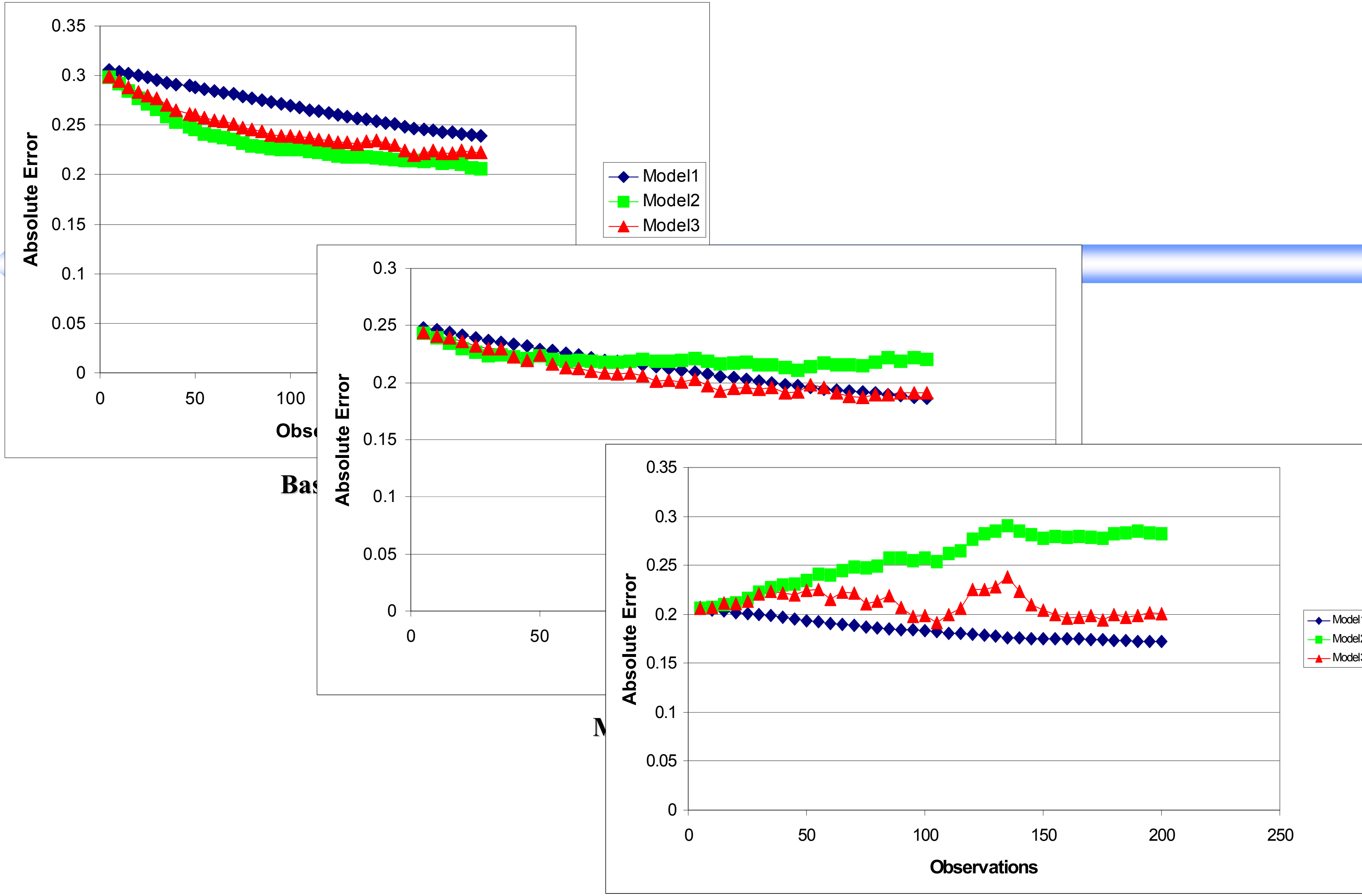
Incorporating Inference Using Empirical Data



- Bernard and Killworth (1976) fraternity brother network
  - Observed number of interactions between 58 individuals over a week
  - Average interactions between individuals-1.89, min-0, max-51

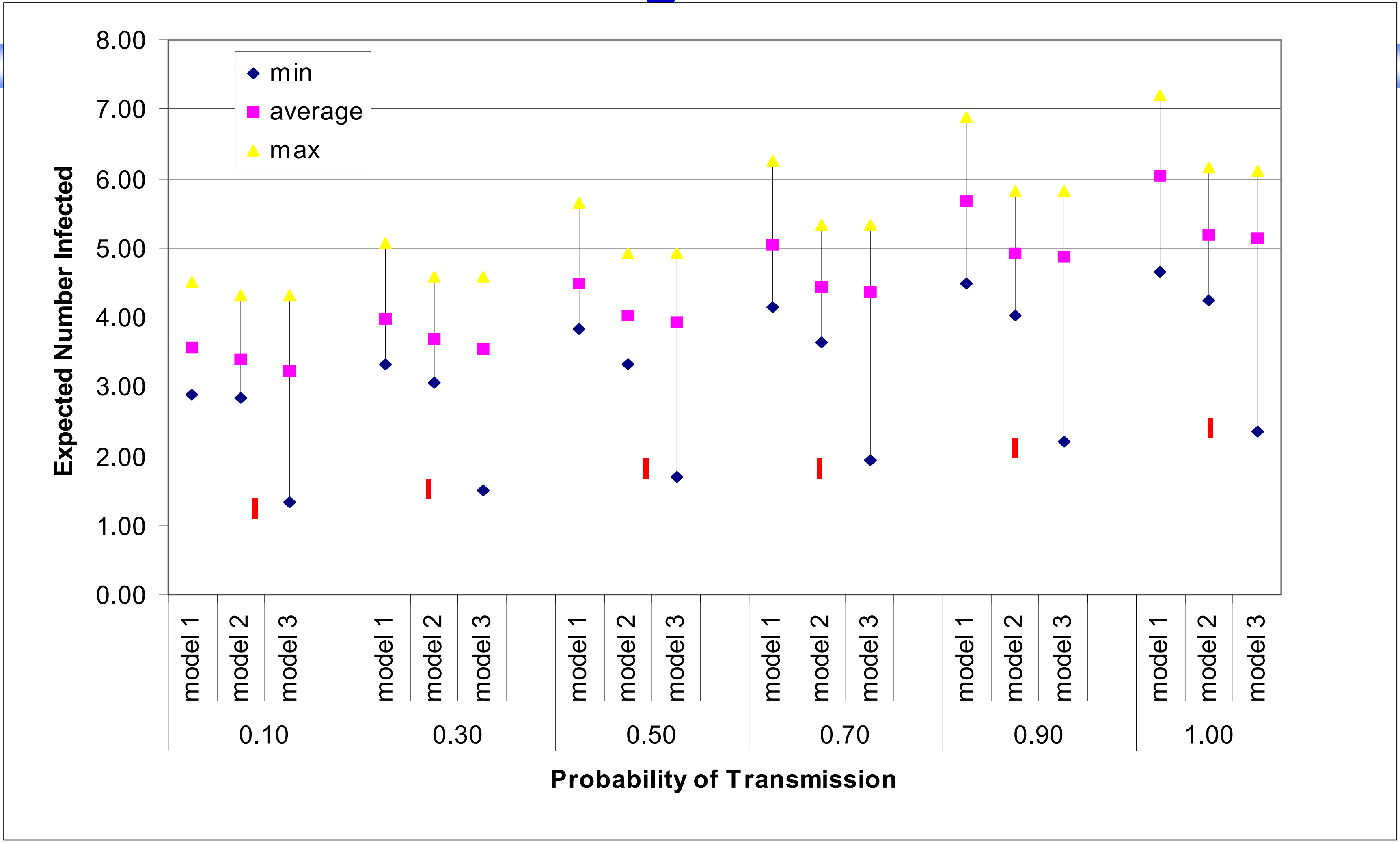


Misinformed Prior



Low Density Network

Policy Analysis
Scenario-Contagious Diseases



Policy Analysis
Scenario 3-Relief Organizations