Learning Models of Macrobehaviors in Complex Adaptive Systems

Andrew Fast

afast@cs.umass.edu
http://www.cs.umass.edu/~afast

Knowledge Discovery Laboratory
Department of Computer Science
University of Massachusetts Amherst

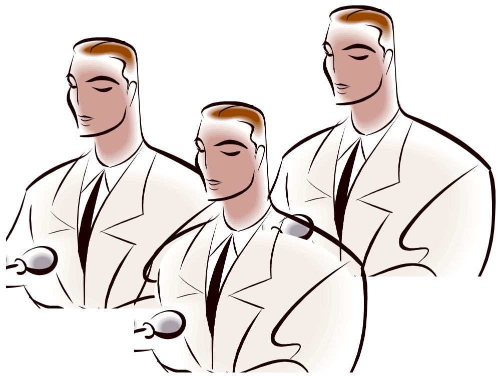
Advisor: David Jensen

Complex Adaptive Systems

Complex : many diverse interconnected entities

Adaptive : capacity to learn from experience and change over time

Example: The NFL



Head Coach

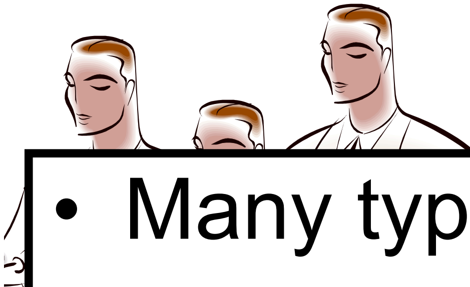
Owners/
Front Office

Team

Players

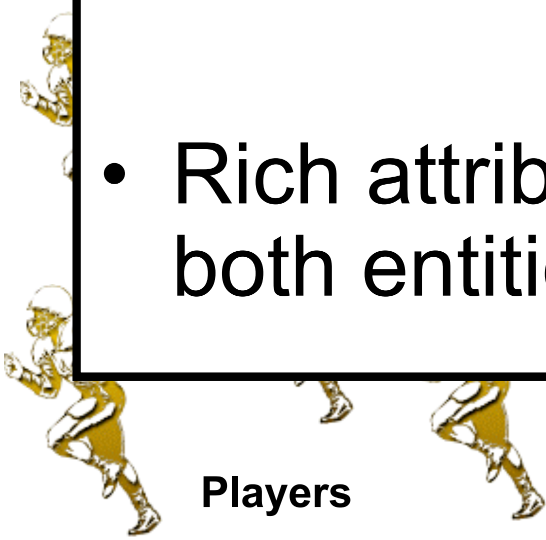
Assistant
Coaches

Example: The NFL



oach

- Many types of entities
- Non-uniform complex relationships among entities
- Rich attribute information about both entities and relationships

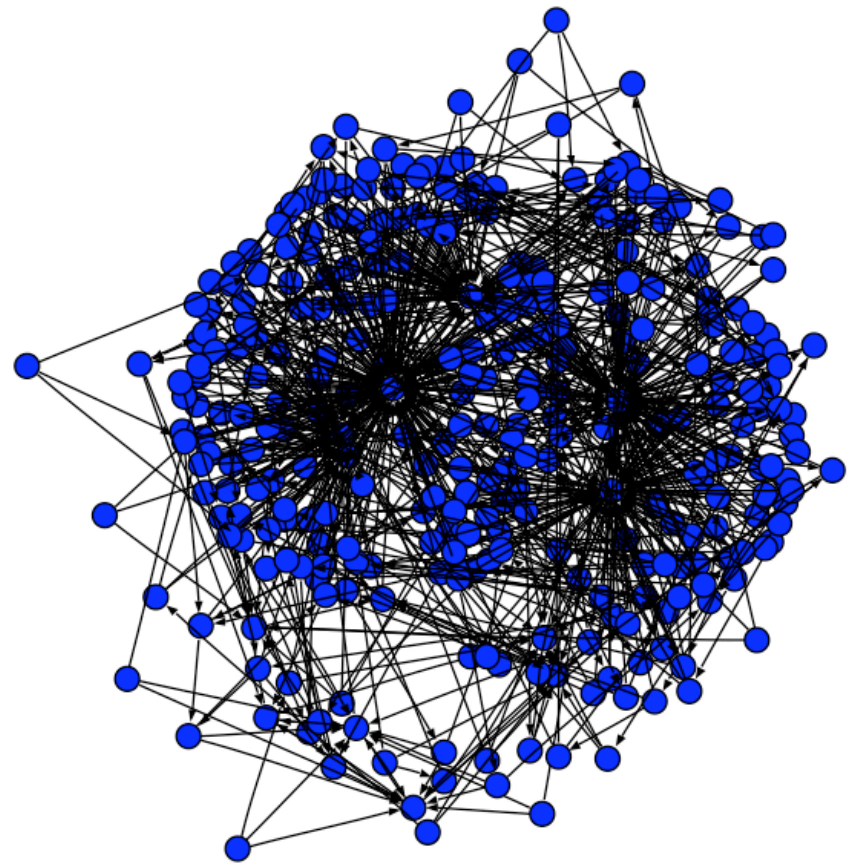


Players

Assistant Coaches

More Examples

- Social Network Among Fraudulent Stock Brokers
- A multi-agent UAV system for monitoring ground targets
- Citation patterns in academic fields
- **Usage patterns of networked computers and users**



(Fast, Jensen, and Levine 2005)

Outline

1. Overview of Complex Adaptive Systems
2. The Challenge of Macrobehaviors
3. Objective: Empirical Methods for Modeling Macrobehaviors
    1. Choosing the Proper Representation
    2. Flexible Models of Temporal Phenomena
    3. Flexible Models of Group Structure
4. Potential Pitfalls
5. Conclusions

Macrobehaviors

- Characteristics or behaviors of aggregated entities that arise over time, often unexpectedly, from interactions among the individuals and their attributes (Schelling 1978).
- Macrobehaviors can be either beneficial or pathological

Beneficial

- Wisdom of Crowds (Suroweicki 2004)
- "Invisible Hand" (Smith 1776)

Pathological

- Tragedy of the Commons (Turner 1993)
- Arms Race (Etcheson 1989)
- Monopolies (Arthur 1990)

The Challenge Of Macrobehaviors

Due to the complexity of the systems, macrobehaviors are difficult to **identify** and **predict** in real systems…

…but, the success or failure of these systems often depend on **timely understanding** of these behaviors

Macrobehaviors occur:

- At differing time scales and intervals
- Within different size groups of entities

… but,

- the correct time scale
- the correct group size

are not known *a priori.*

The Challenge Of Macrobehaviors

Due to the complexity of the

…but, the success or failure of

Hypothesis

Models of Time	+	Models of Groups	=	Models of Macrobehaviors

Macrobehaviors occur:

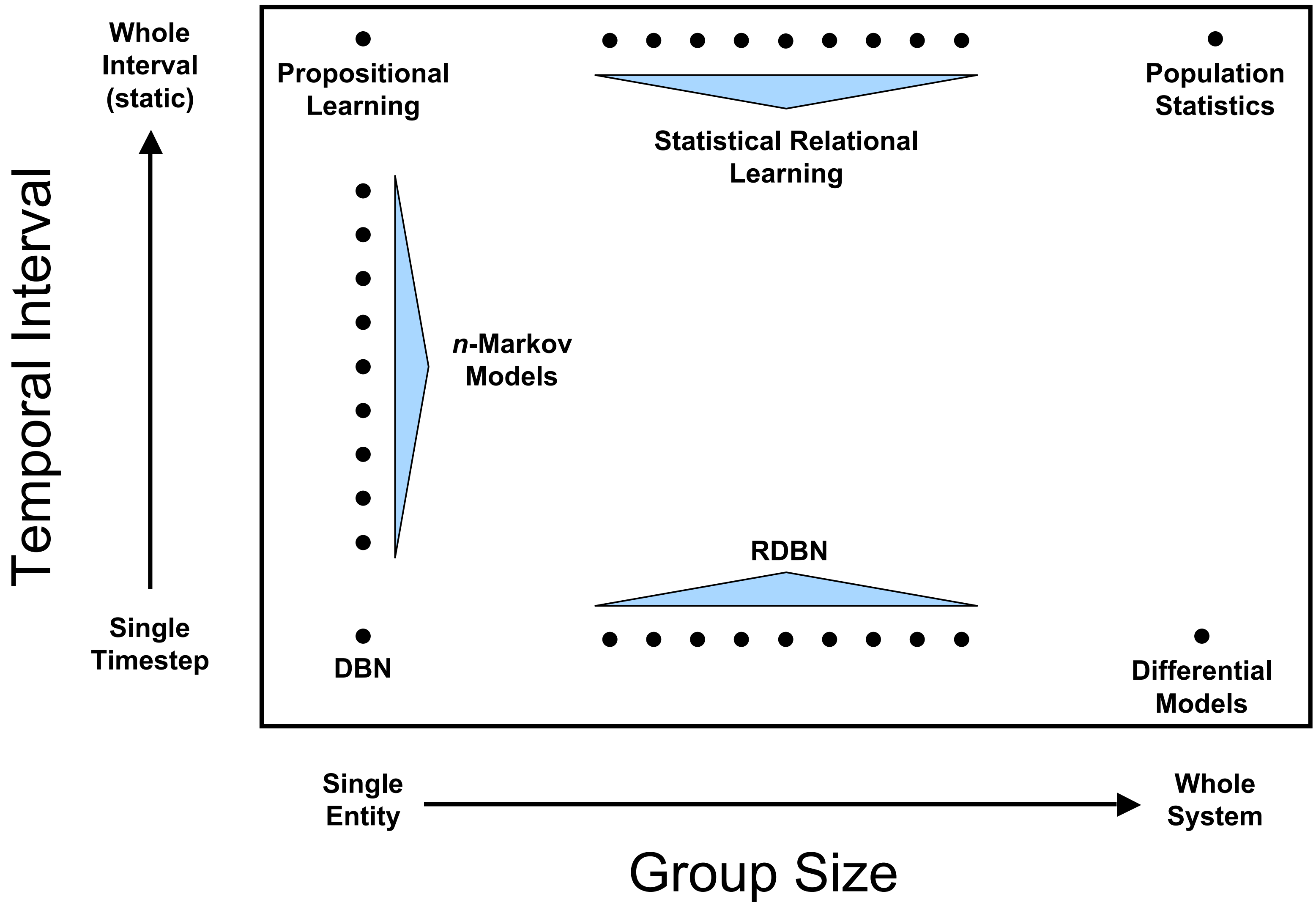
- At differing time scales and intervals
- Within different size groups of entities

… but,

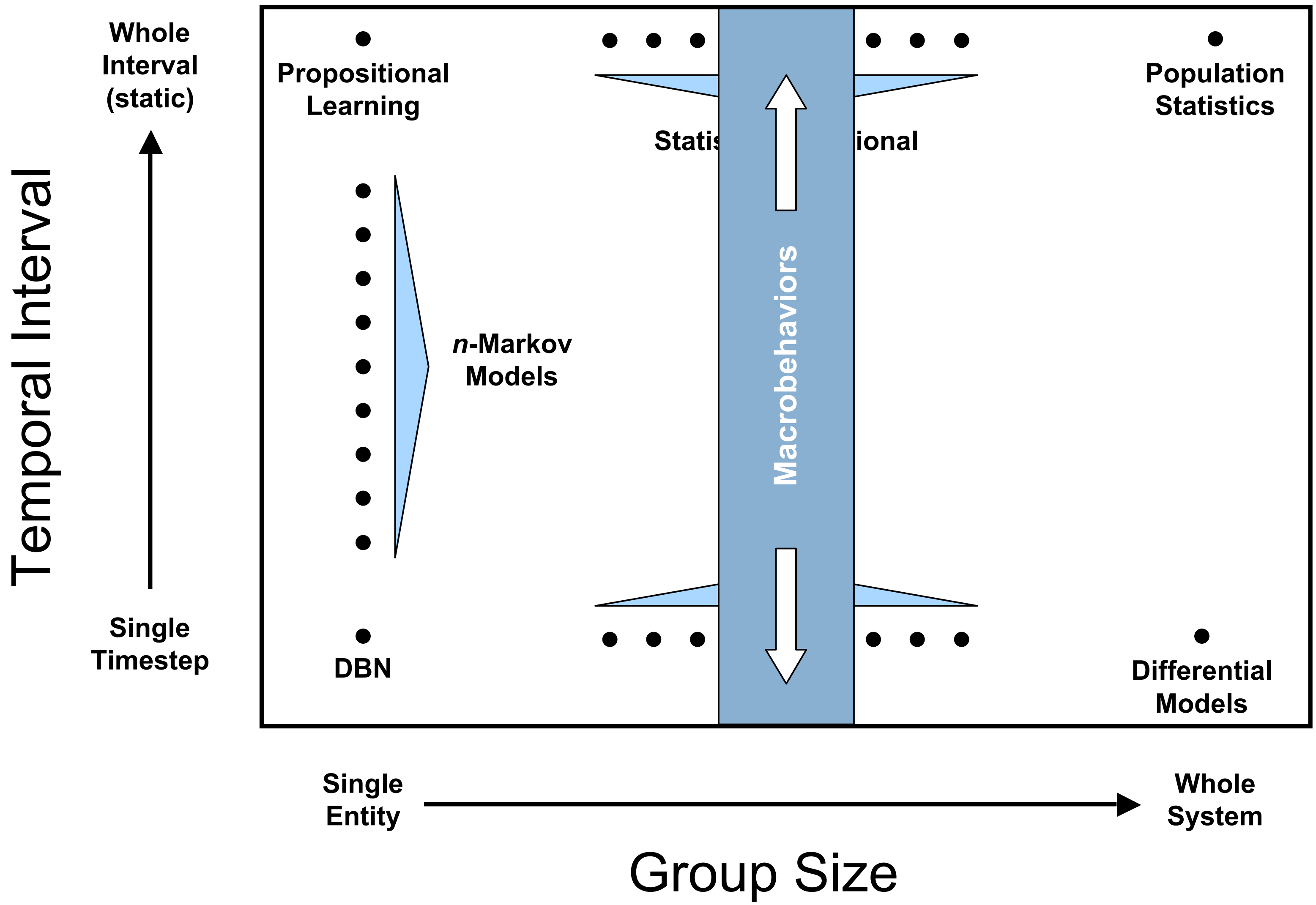
- the correct time scale
- the correct group size

are not known *a priori.*

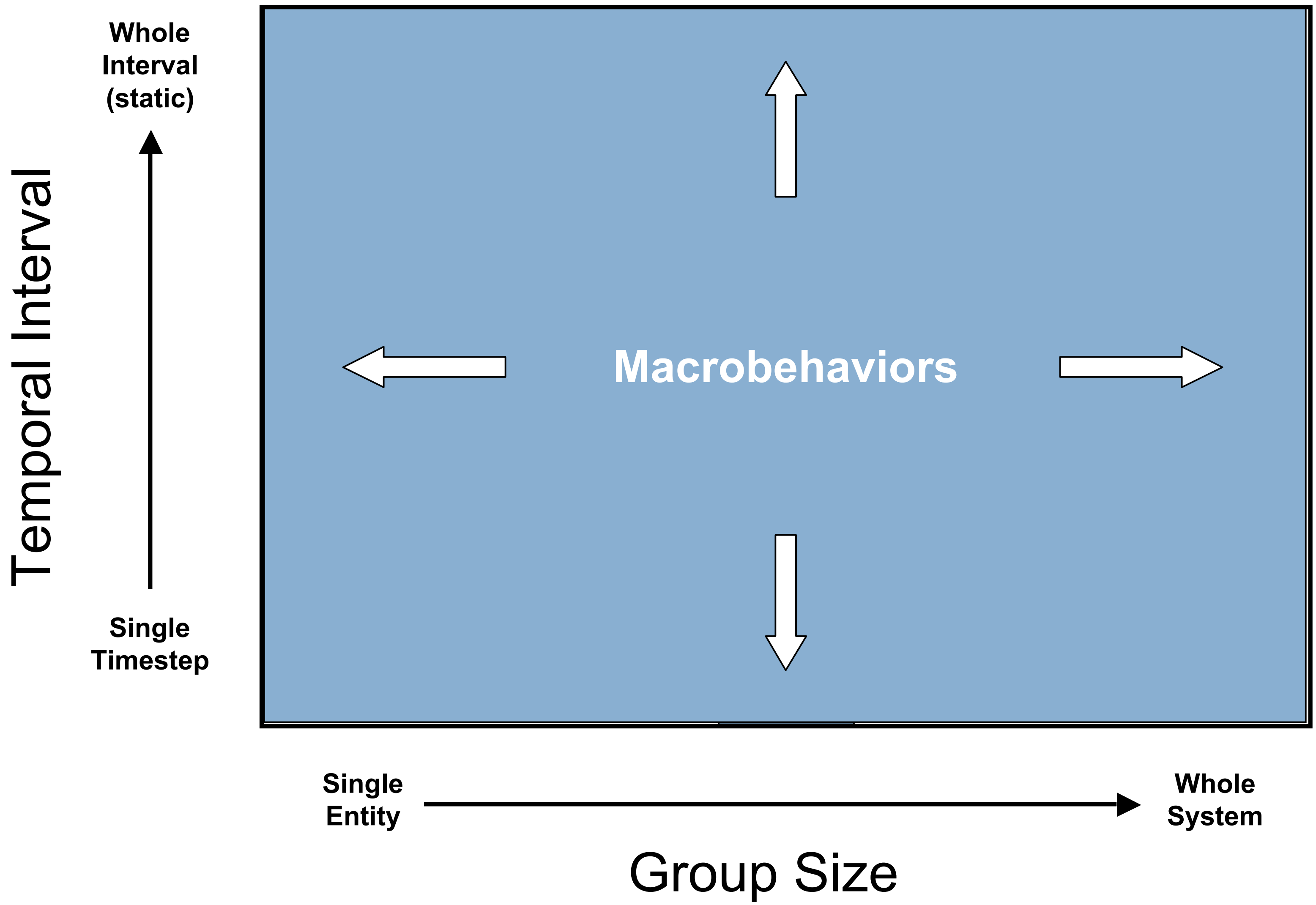
Modeling Macrobehaviors



Modeling Macrobehaviors



Modeling Macrobehaviors



Outline

1. Overview of Complex Adaptive Systems
2. The Challenge of Macrobehaviors
3. Objective: Empirical Methods for Modeling Macrobehaviors
    1. Choosing the Proper Representation
    2. Flexible Models of Temporal Phenomena
    3. Flexible Models of Group Structure
4. Potential Pitfalls
5. Conclusions

Research Objectives

Develop models of macrobehaviors that are:

Data Driven

–Automatically and efficiently learned from data

Understandable

–especially to human observers

Actionable

–Able to guide future decisions

Correct

–An accurate and unbiased representation of the true state of the world

1) Choosing the Proper Representation

Graph Representation

- Neural Network of *C. Elegans* (Watts and Strogatz 1998)
- Power Grid (Watts and Strogatz 1998)
- Broadway musical teams (Guimera, Uzzi, Spiro and Amaral 2005)
- Spread of Influence (Kempe, Kleinberg, Tardös 2003)
- Stock Fraud (Neville et al. 2005)
- Hollywood movie industry (Neville et al. 2003)
- Scientific co-authorship (Guimera, Uzzi, Spiro and Amaral 2005)
- …

All can be represented as a graph with attributes on the vertices and edges.

Modeling Graphs

- Statistical Relational Learning
  - Common representation is a graph with attributes
- Methods designed for data that :
  - Heterogeneous
  - Non-Independent
- Relational Dependency Networks (Neville and Jensen 2004)
  - Joint model of relational data
  - Able to learn cyclic dependencies
  - Allows for Collective Inference
  - Uses a relational decision tree model as a CPD (Neville et al. 2003)

2) Flexible Models of Temporal Phenomena

Temporal Relational Probability Tree

Will the Patriots win the Super Bowl in 2006?

Temporal Relational Probability Tree

Goal: Use past information to predict future results.

Temporal Relational Probability Tree

But how much past information to use?

Temporal Relational Probability Tree

But how much past information to use?

P(Playoffs) = 3/45 = 0.67

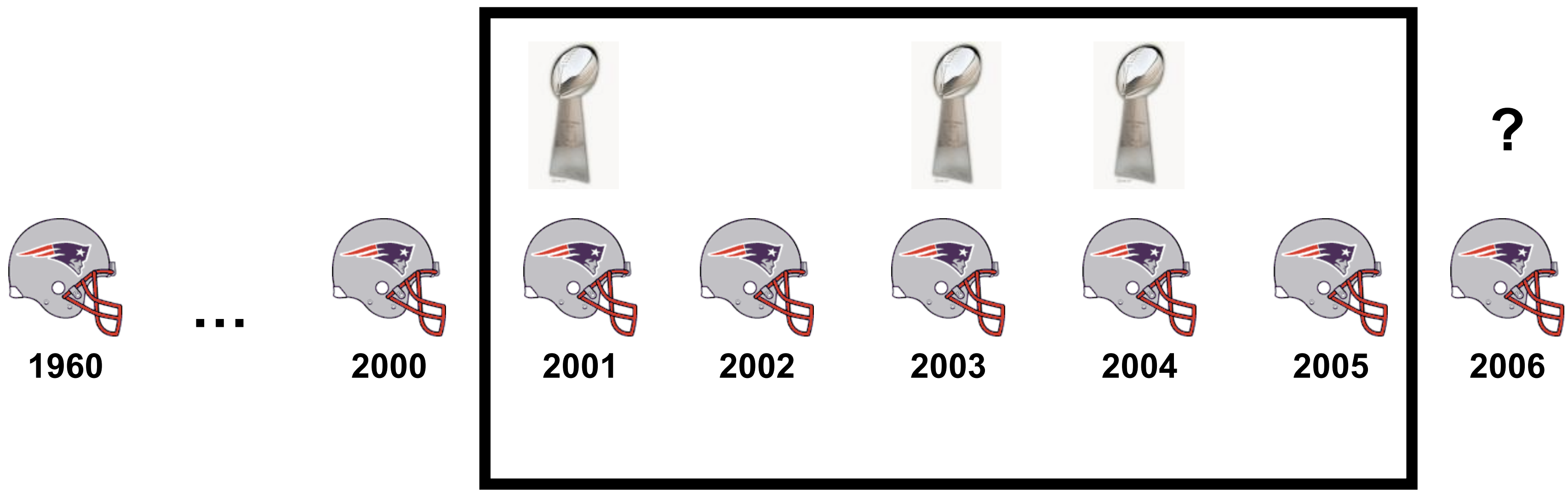
Temporal Relational Probability Tree

But how much past information to use?

P(Playoffs) = 0/1 = 0.0

Temporal Relational Probability Tree

But how much past information to use?

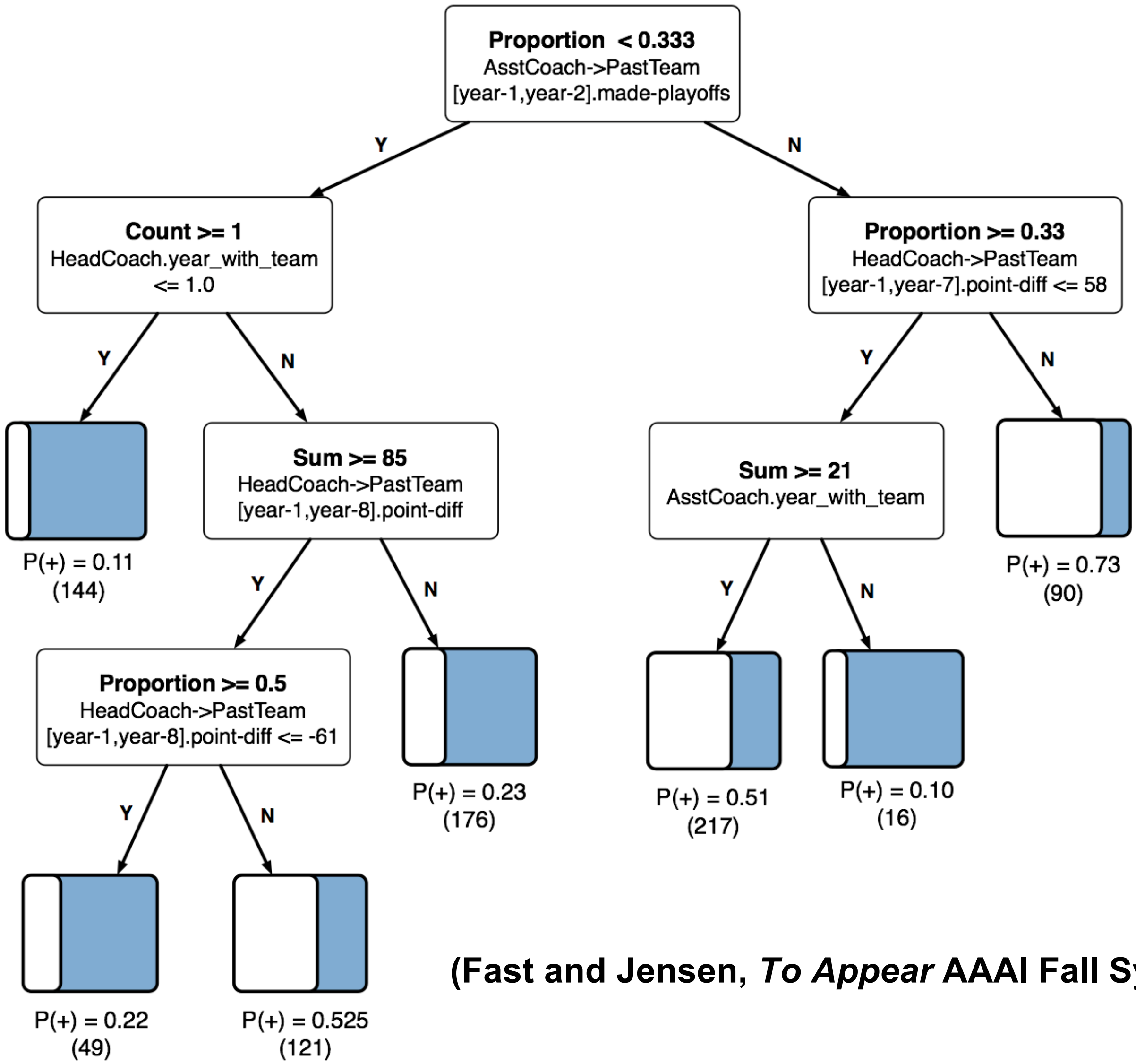


P(Playoffs) = 3/5 = 0.60

Temporal Relational Probability Tree

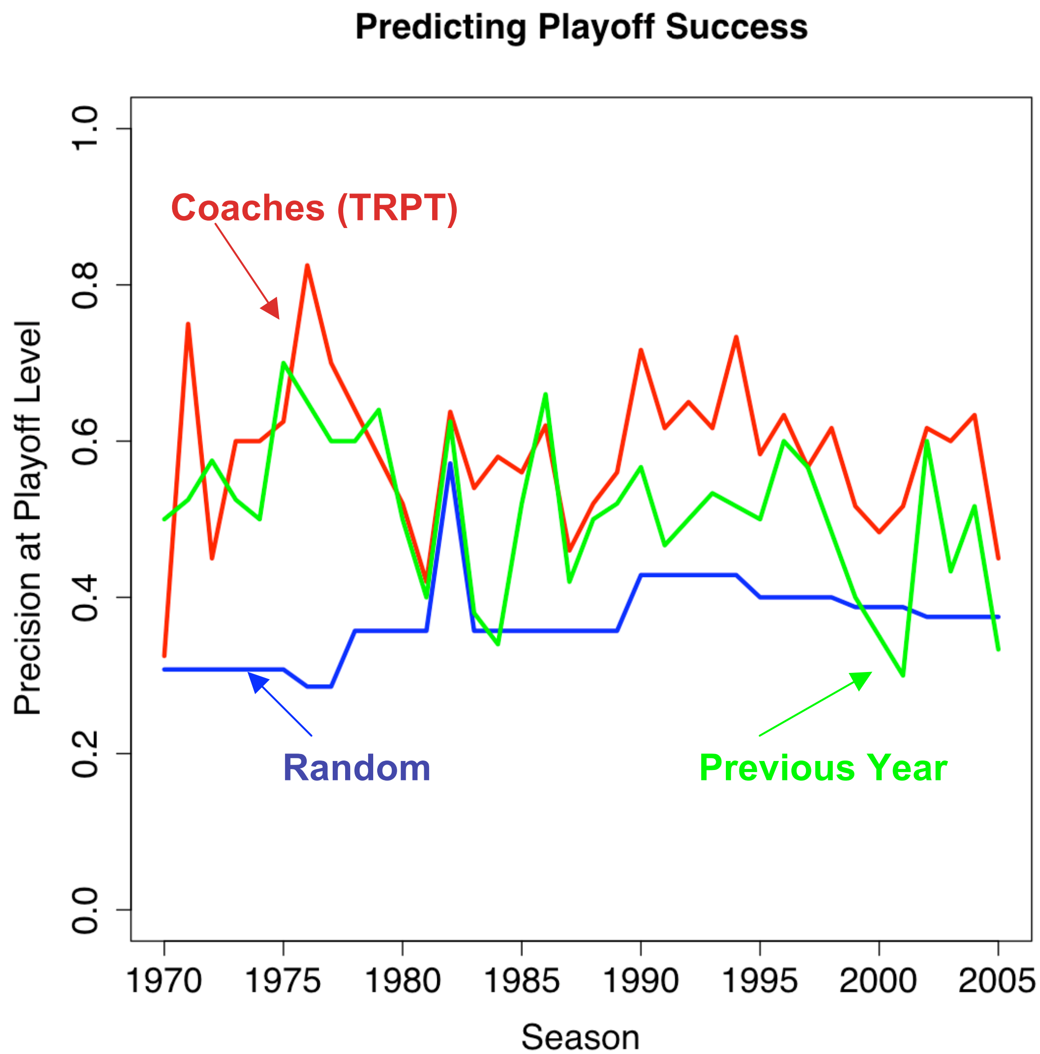
- Use a recursive partitioning algorithm for probability estimation
- Search over both aggregations of entities and temporal intervals.
- Incurs a large feature space expansion
  - $\frac{n^2+n}{2}$ possible contiguous intervals

NFL Playoff Models



(Fast and Jensen, *To Appear* AAAI Fall Symposium 2006)

Evaluating NFL Playoff Models



Coaches significantly greater than Previous year ($p < 0.01$)

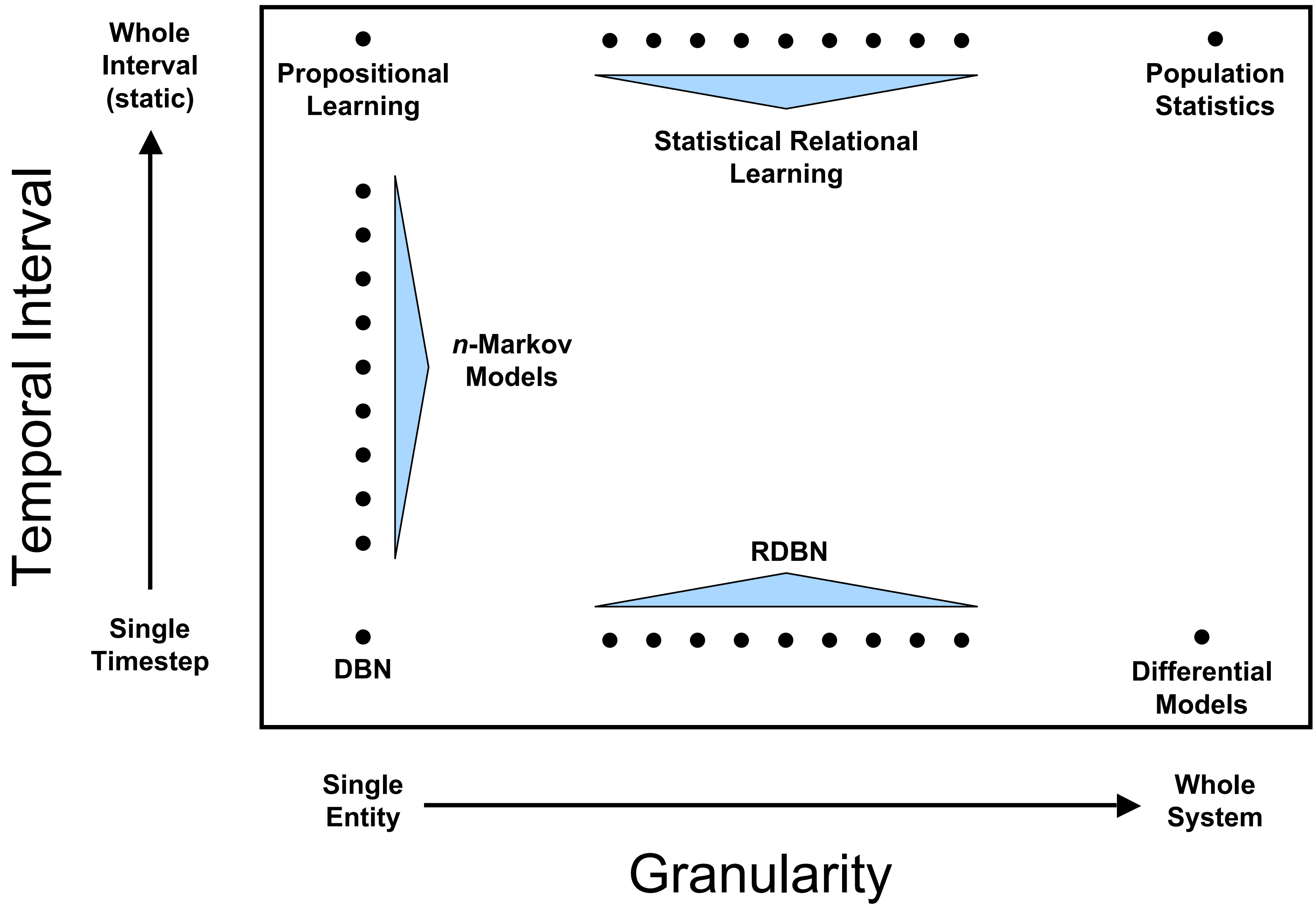
1982 - Player Strike

1990 - Playoffs Expanded to 12 teams

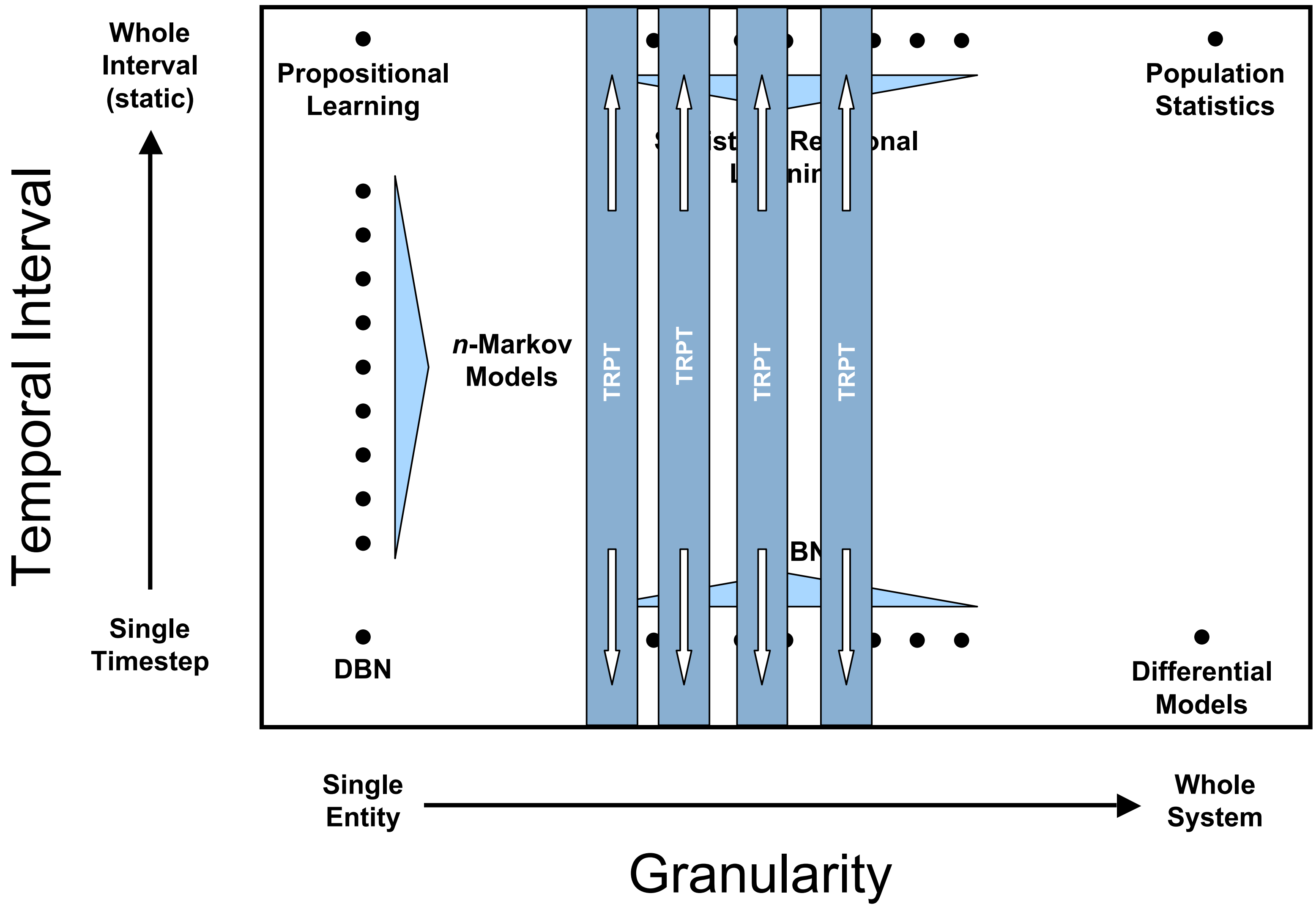
1992 - Salary Cap

1994 - Free Agency

Modeling Macrobehaviors



Modeling Macrobehaviors



3) Flexible Models of Group Structure

Not Group Finding

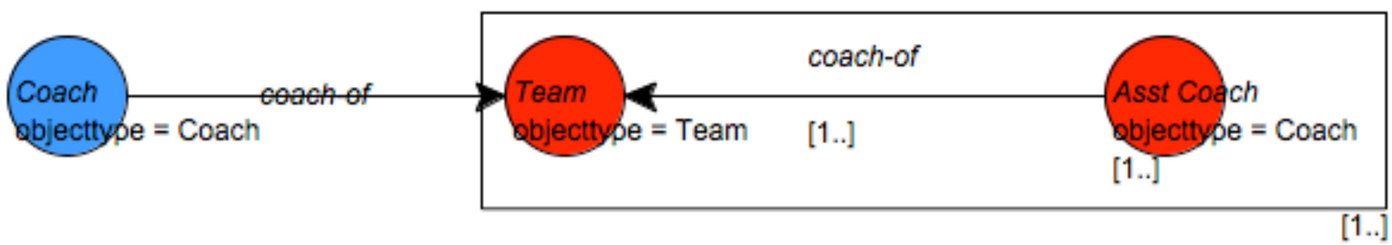
- Many algorithms exist for finding groups in data.
  - **Latent Dirichlet Allocation** (Blei, Ng, and Jordan 2003)
  - **Hierarchical Dirichlet Process**(Teh, Jordan, Beal, and Blei 2004)

    (See Fast, Jensen, and Levine 2005)
  - **Group Detection Algorithm** (Kubica, Moore, Schneider and Yang 2002)
  - **Community Finding Algorithm** (Girvan and Newman 2002)
  - …
- Instead, learn which group structures are correlated with macrobehaviors.

QGraph

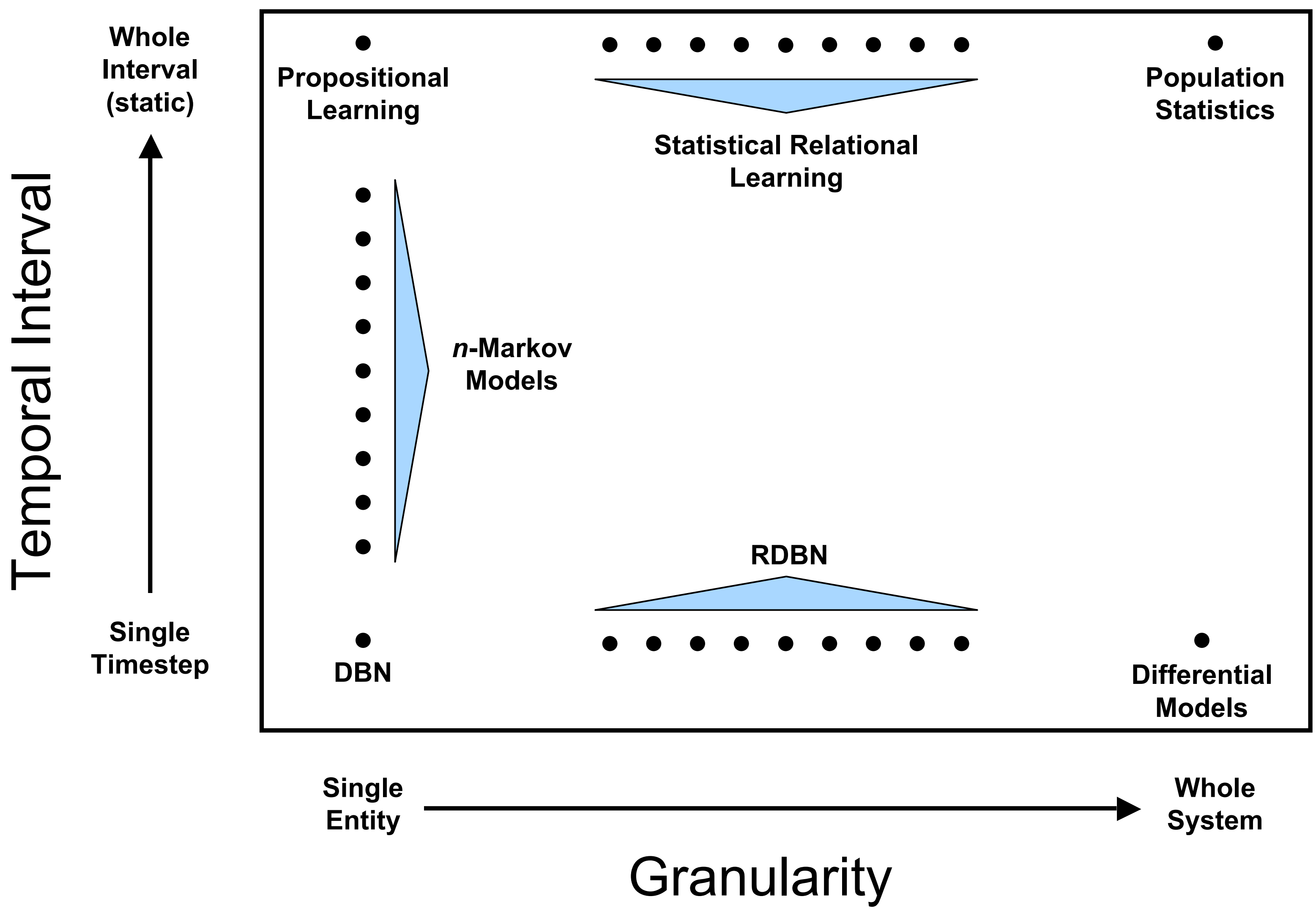
- Query language designed for relational data

  (Blau, Immerman, and Jensen 2002)

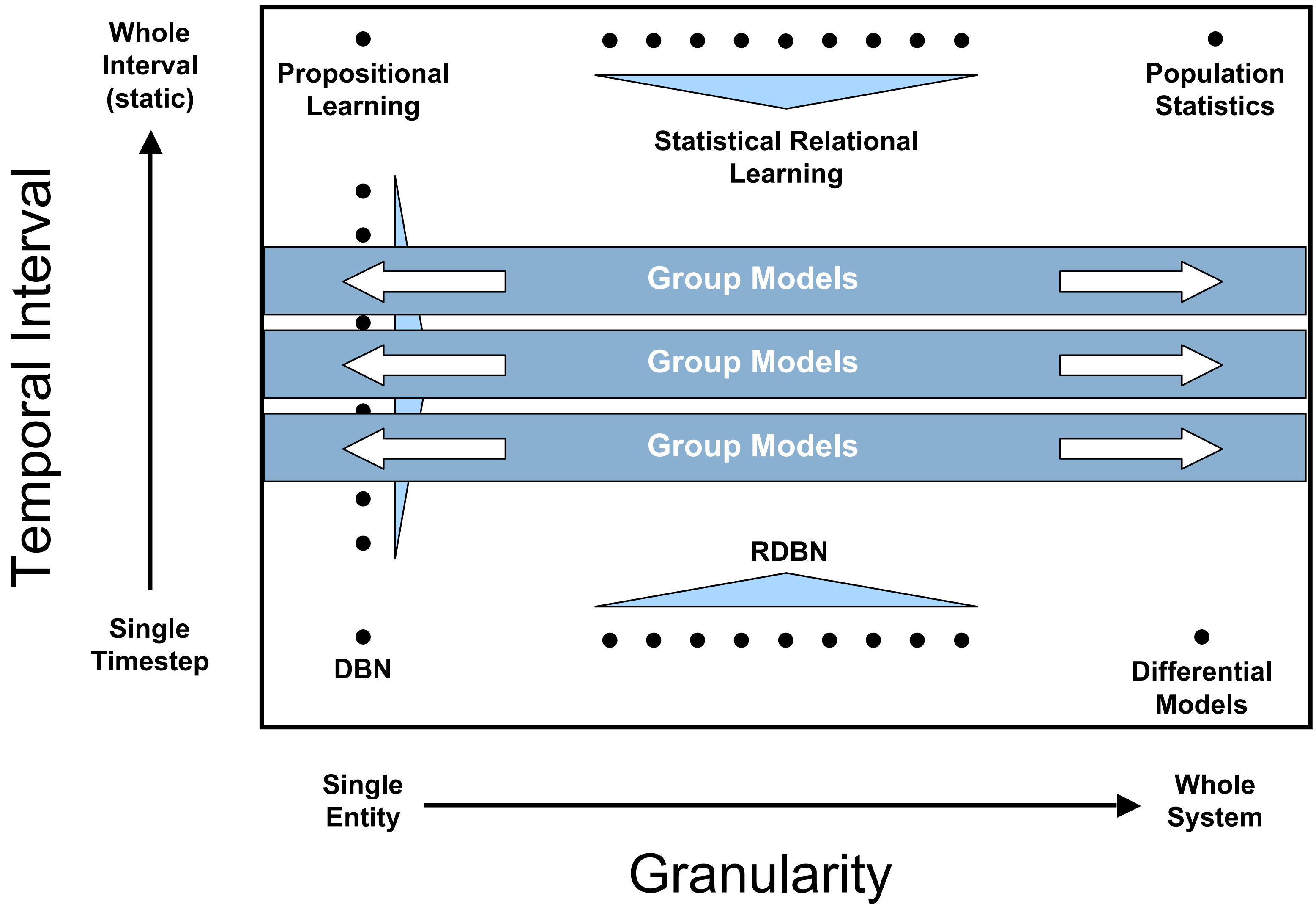


- Use QGraph to enumerate possible groups and calculate features based on those groups

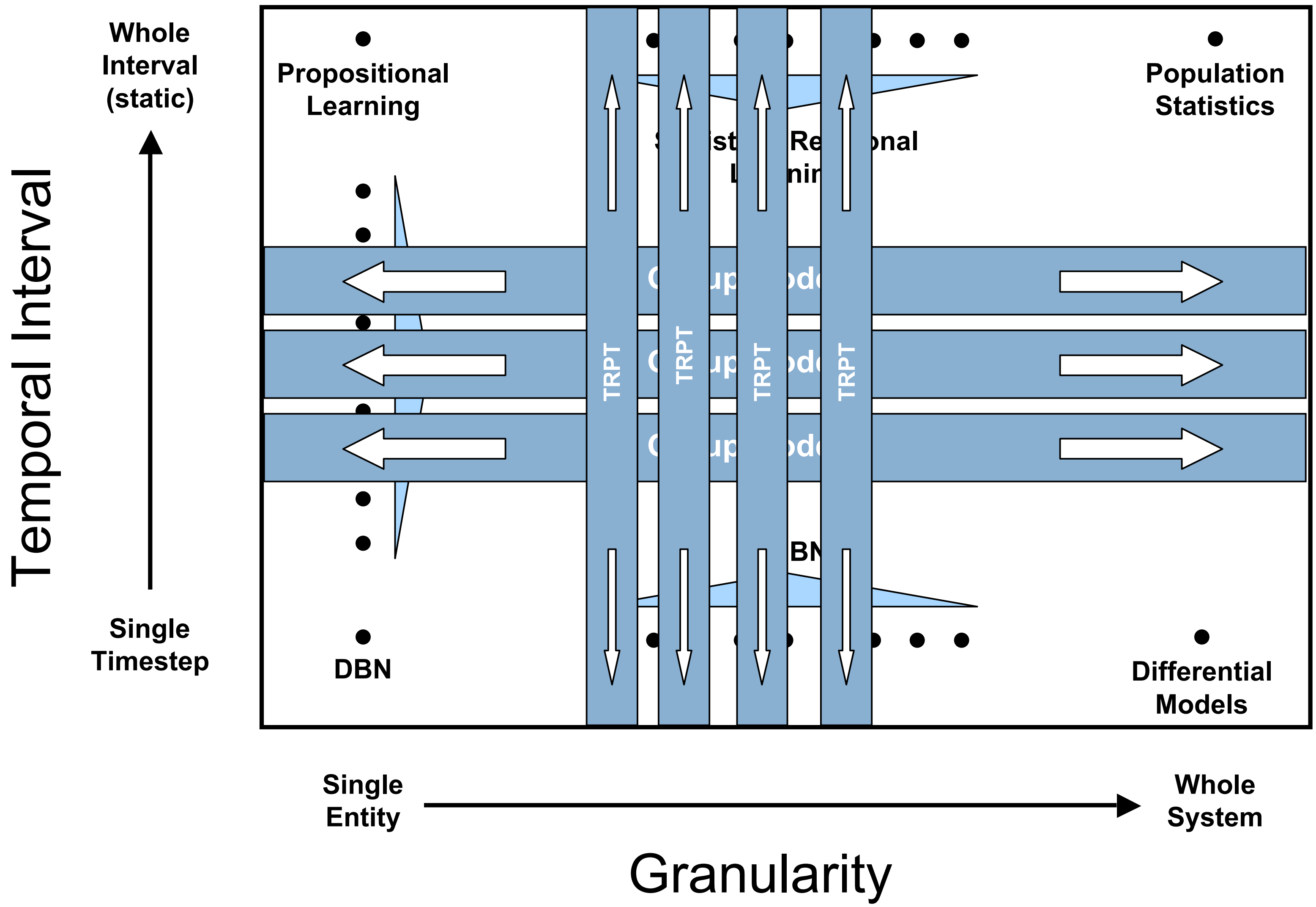
Modeling Macrobehaviors



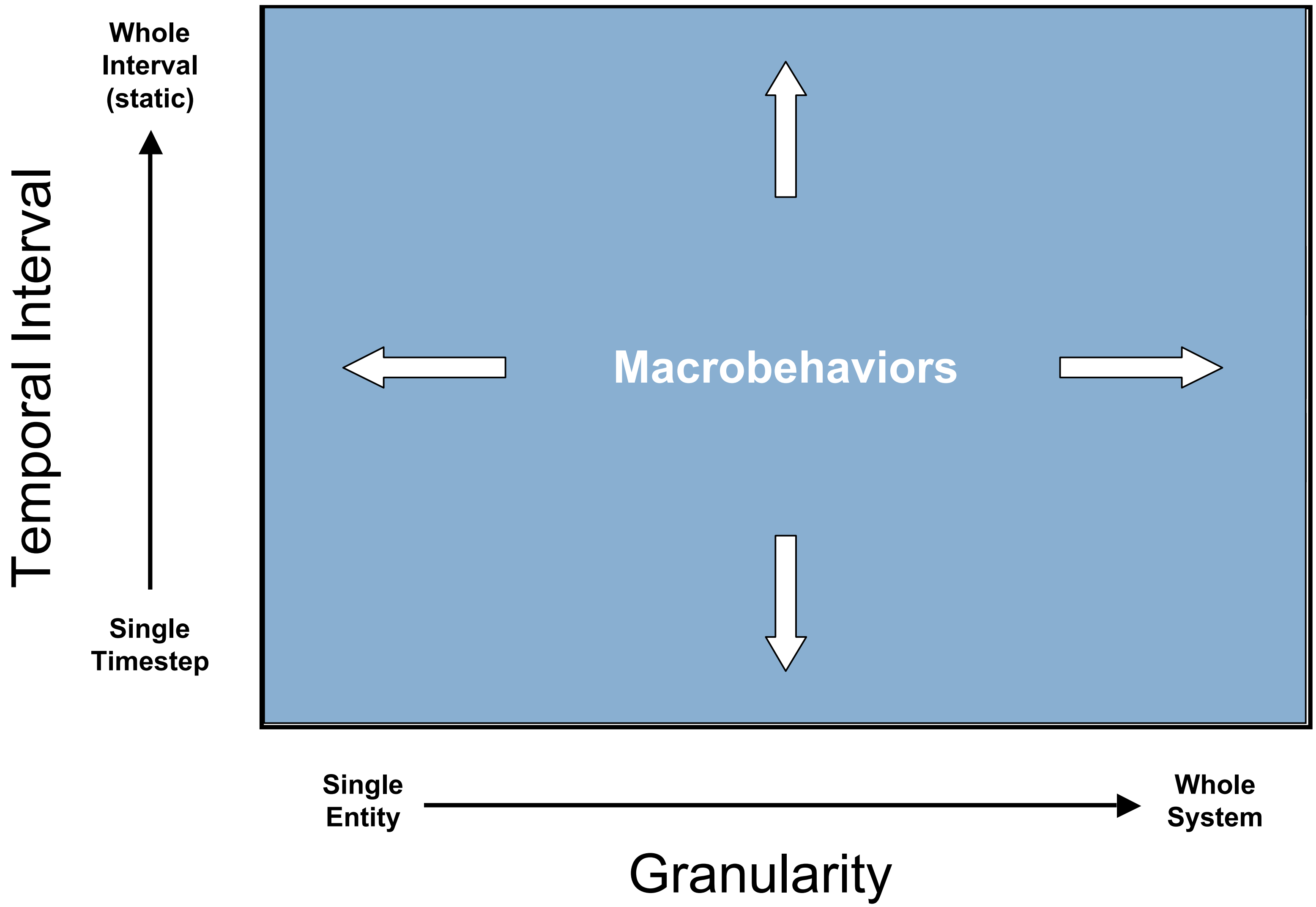
Modeling Macrobehaviors



Modeling Macrobehaviors



Modeling Macrobehaviors



Outline

1. Overview of Complex Adaptive Systems
2. The Challenge of Macrobehaviors
3. Objective: Empirical Methods for Modeling Macrobehaviors
    1. Choosing the Proper Representation
    2. Flexible Models of Temporal Phenomena
    3. Flexible Models of Group Structure
4. Potential Pitfalls
5. Conclusions

Drowning in Variance

- Feature space expansion could lead to problems with high variance models
- Potential Solutions:
  - **Collective Inference** - Use inferences on related entities and relations to constrain inferences on current entity
  - **Large Data Sets** - Complex Adaptive Systems tend to generate VERY large data sets.
  - **New Features Constrain Model Space -** Temporal and Group Structure constrain the space of models, effectively limiting variance

Learning Incorrect Models

- Expanded representation opens door for biases in structure learning.
  - Autocorrelation and Linkage
  - Degree Disparity
  - Some new bias?
- Multiple Comparisons Problem

Solution is the same!

New randomization tests for relational data that are able to deal with both time and group structure

Timeline

- Extend relational learning algorithms to search over all possible temporal intervals.
- Extend relational learning algorithms to search over possible group structure.
- Explore new learning algorithms for potential biases
- Explore options for improving initial models based in results of previous step

Thesis Contributions

- New relational learning algorithms able to learn both the temporal and complex structural dependencies needed to detect macrobehaviors in complex adaptive systems.
- Development of novel computationally intensive methods for evaluating the bias and variance of those new models to determine if the models accurately represent the true state of the world

Remaining Concerns

- Are these methods sufficient for modeling macrobehaviors?
  - Are there other dimensions I should consider?
- Other approaches for modeling groups?
- Other potential pitfalls?
- Contributions from other fields?

Questions?

Andrew Fast
afast@cs.umass.edu
http://www.cs.umass.edu/~afast

Acknowledgements

- Advisor
  - David Jensen
- Other Faculty
  - Victor Lesser
  - Amy McGovern
  - Brian Levine
- The Knowledge Discovery Laboratory at the University of Massachusetts Amherst