Online Designs for Metric Multidimensional Scaling

Prepaper talk

Overview

- Contributions
- Introduction
- Previous Work
- Technique
- Results

Contributions

- Technique for computing incomplete designs for metric-MDS in an online fashion
- Distance-Feeder Architecture for sampling-based MDS schemes

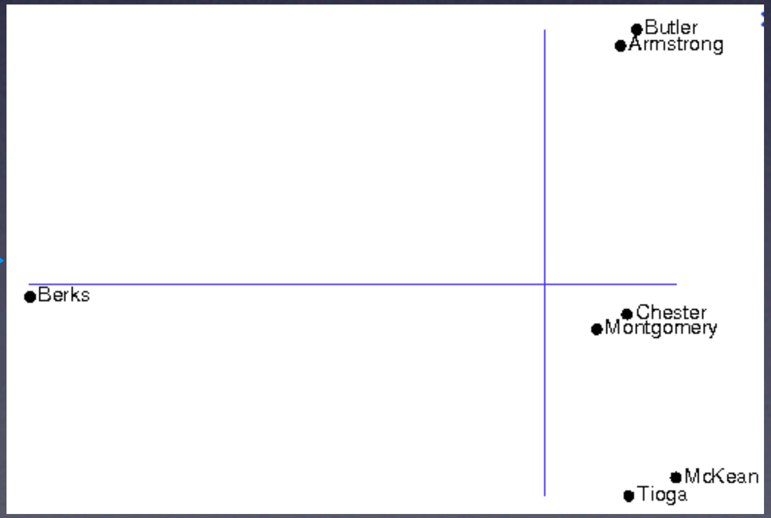
Introduction

Definitions: Multidimensional Scaling

- Family of techniques to compute coordinates for points based on their mutual distances
- Metric MDS is a popular and flexible variant

	Montgomery	Chester	Berks	Tioga	Butler	Armstrong	McKean
Montgomery	0,000	0,025	0,068	0,035	0,042	0,041	0,037
Chester	0,025	0,000	0,073	0,039	0,043	0,044	0,042
Berks	0,068	0,073	0,000	0,074	0,076	0,074	0,079
Tioga	0,035	0,039	0,074	0,000	0,056	0,055	0,030
Butler	0,042	0,043	0,076	0,056	0,000	0,021	0,055
Armstrong	0,041	0,044	0,074	0,055	0,021	0,000	0,053
McKean	0,037	0,042	0,079	0,030	0,055	0,053	0,000

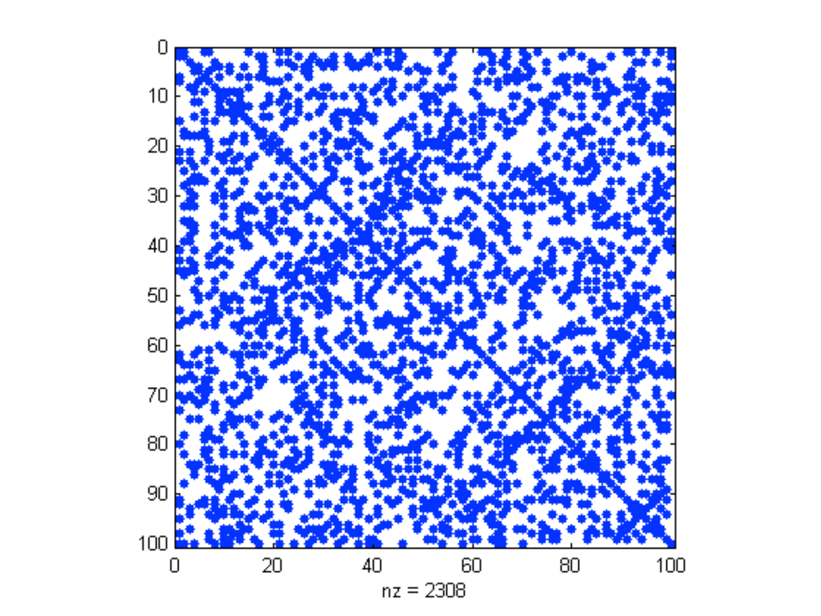
Distance Matrix D



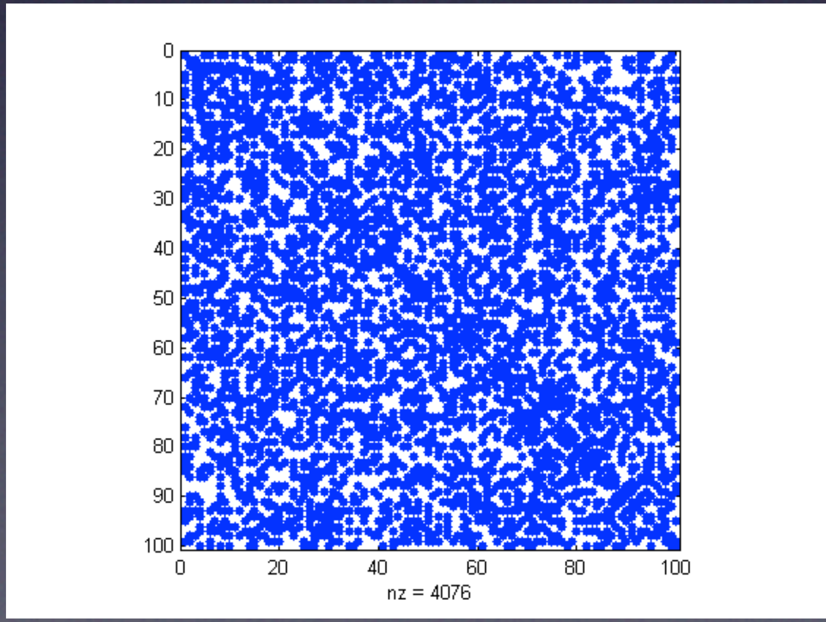
Coordinates

Definitions: MDS Designs

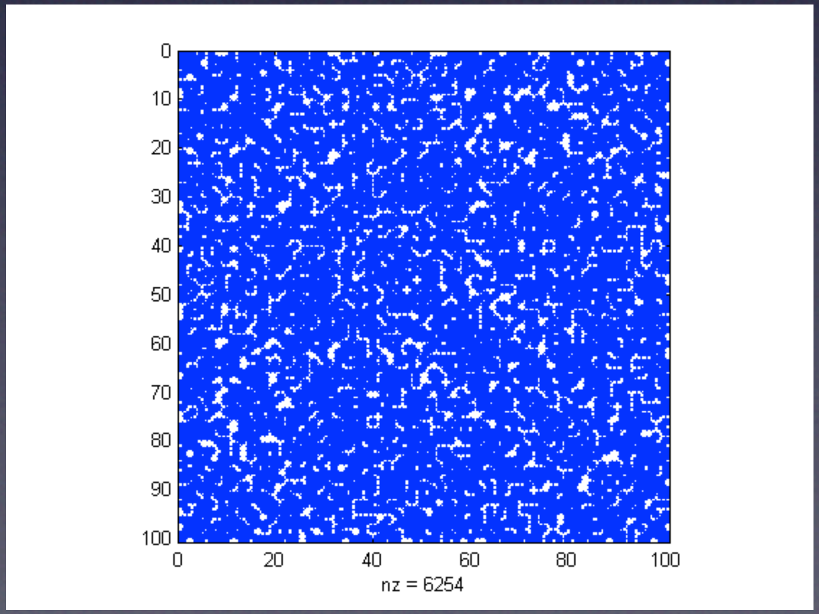
- Input distance matrix often <u>*overdetermines*</u> layout coordinates.
- **Full design**: use entire dist. matrix
- **Incomplete design**: sparsify dist. matrix



25%



50%



75%

Why incomplete designs?

- Full distance matrix may be very expensive to compute:
  - N is large, dist matrix is O(N^2)
  - and/or D(i,j) is costly
    - expensive function
    - gathered from real humans

Definition: Online Design

- Incomplete Design that is **not known** in advance and determined at run time
- Some previous work used *static* Incomplete designs:
  - D calculations done in advance of layout
- Online designs start with an incomplete design and add to it until terminating

Space of Incomplete Design Solutions

	D Cheap	D Expensive
N SMALL	Complete design with SMACOF	??
N LARGE	Online Design with Glimmer, LAMP, etc.	??

Space of Incomplete Design Solutions

	D Cheap	D Expensive
N SMALL	Complete design with SMACOF	??
N LARGE	Glimmer, LAMP, etc.	??

Most of the research focuses here

Cost(Iteration) = Cost(D)

Space of Incomplete Design Solutions

	D Cheap	D Expensive
N SMALL	Complete design with SMACOF	??
N LARGE	Incomplete design w/ Glimmer, LAMP,	??

Cost(Iteration)<<Cost(D)

Not optimally handled

Algorithm Choices

Cost Relationship	Optimal Objective	Algorithm Design
Cost(Iteration) ~ Cost(D)	Minimize Iterations	Iteration + D Coupled
Cost(Iteration) << Cost(D)	Minimize Distance Calculations	Iteration + D Independent

Algorithm Choices

Cost Relationship	Optimal Objective	Algorithm Design
Cost(Iteration) ~ Cost(D)	Minimize Iterations	Iteration + D Coupled
Cost(Iteration) << Cost(D)	Minimize Distance Calculations	Iteration + D Decoupled

Approach introduced in this paper

“Cheap” D Examples

- D is Euclidean O(m) where m << N
- D is Jaccard/Cosine/etc.

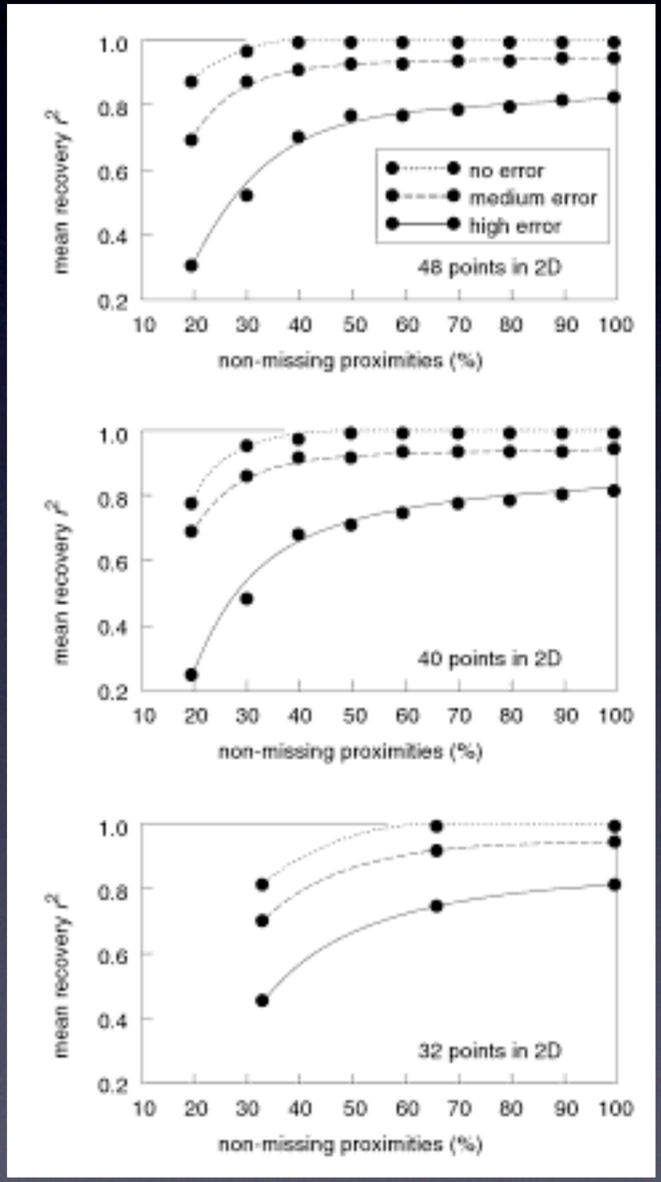
Costly D Examples

- D is human sourced
  - marketing, sociology, psychophysics
- D is computationally costly
  - database query
  - String edit distance
  - Earth mover's distance

Previous Work

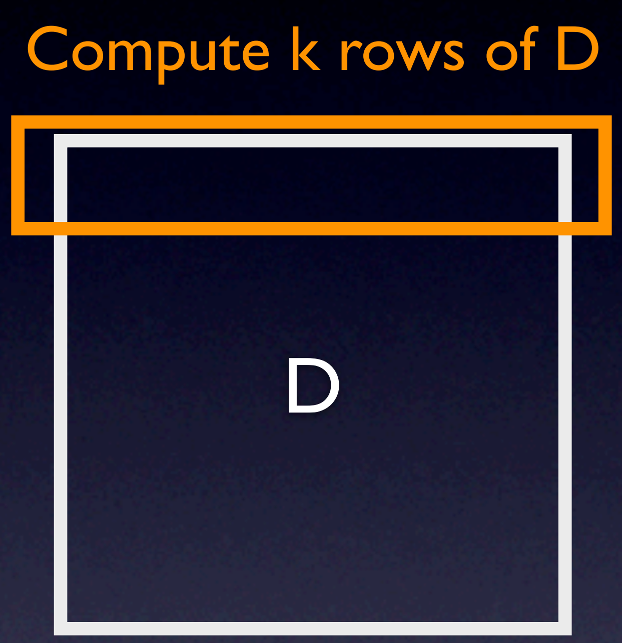
Previous Work: Static Incomplete Designs

- Spence and Domoney '74
  - randomly eliminated fraction of distance matrix
  - Measured correlation of distances in low-d with distances in high-d
  - Recovery depends on error in data



Previous Work: Static Incomplete Designs

- LMDS, PMDS
  - select K "control points"
  - Classical MDS (no weights or missing values)
- PLMP, LAMP, etc.
  - Also control-point based
  - Require points to be coordinate-based (Euclidean)



Previous Work: Online Designs

- Chalmers 96 and Glimmer09
- Force-based simulations with flexible energy function, dealing naturally with missing entries
- At each iteration sample from D

Technique

Glimmer Modification: Distance Feeding

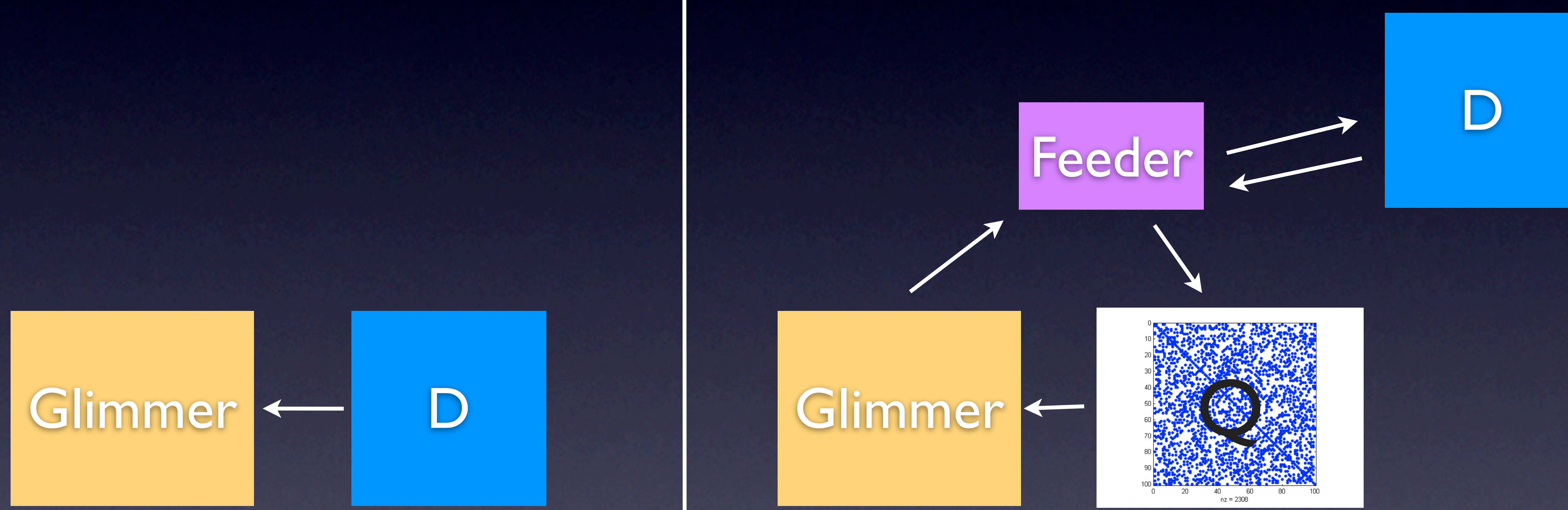
Glimmer (in a nutshell)

- For each point:
    - Sample K random distances from D
    - Compute residuals R
    - Simulate Forces proportional to R
- Check termination

Glimmer (with feeder)

- Request sparse random distance matrix Q
- For each point:
    - Sample K random distances from ~~D~~ Q
    - Compute residuals R
    - Simulate Forces proportional to R
- Check termination

Distance Feeder Diagram



Glimmer

Glimmer (with feeder)

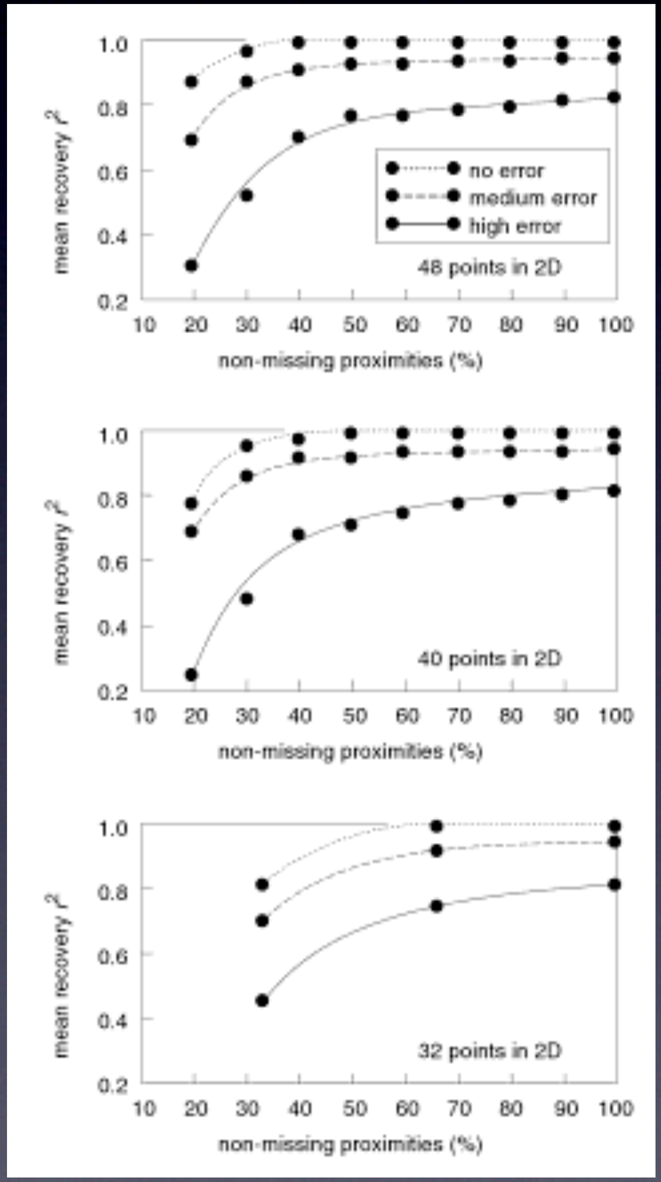
Distance Feeder Diagram

Glimmer

Glimmer (with feeder)

Online Design Outer Loop

- Idea: embed fixed incomplete design glimmer within an outer loop
- Inner Loop: Glimmer with fixed design
- Outer Loop: Slowly increase the fixed design size until "convergence"

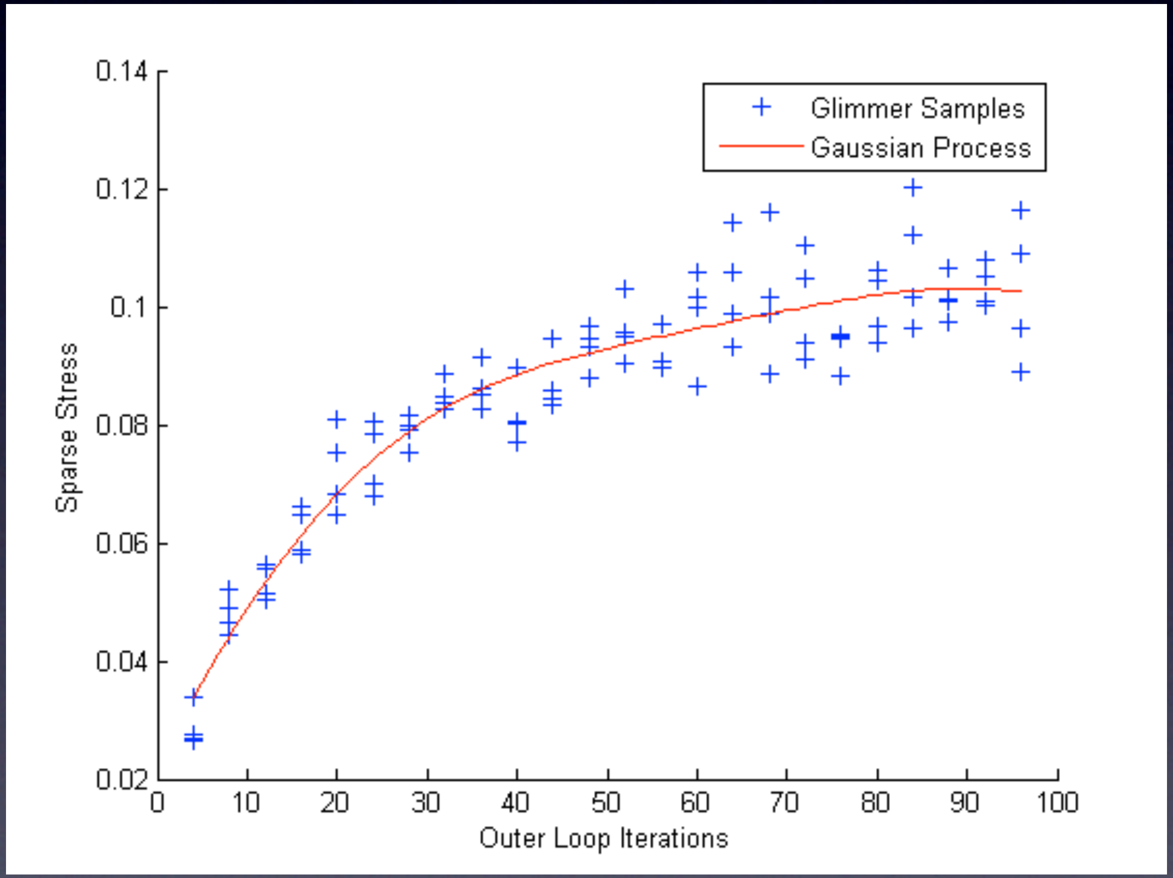


Convergence

- What we really want: Detect when termination stress converges
- What we have: Sparse stress
- Use Stress as a proxy convergence criterion

Smoothing Noise w/Gaussian Process Regression

- What about noise?
- Glimmer iterations are cheap relative to D calculation: run several times
- Use Gaussian process as a smoother
  - works because series conforms to def of GP
  - Use mean of process as obj. function

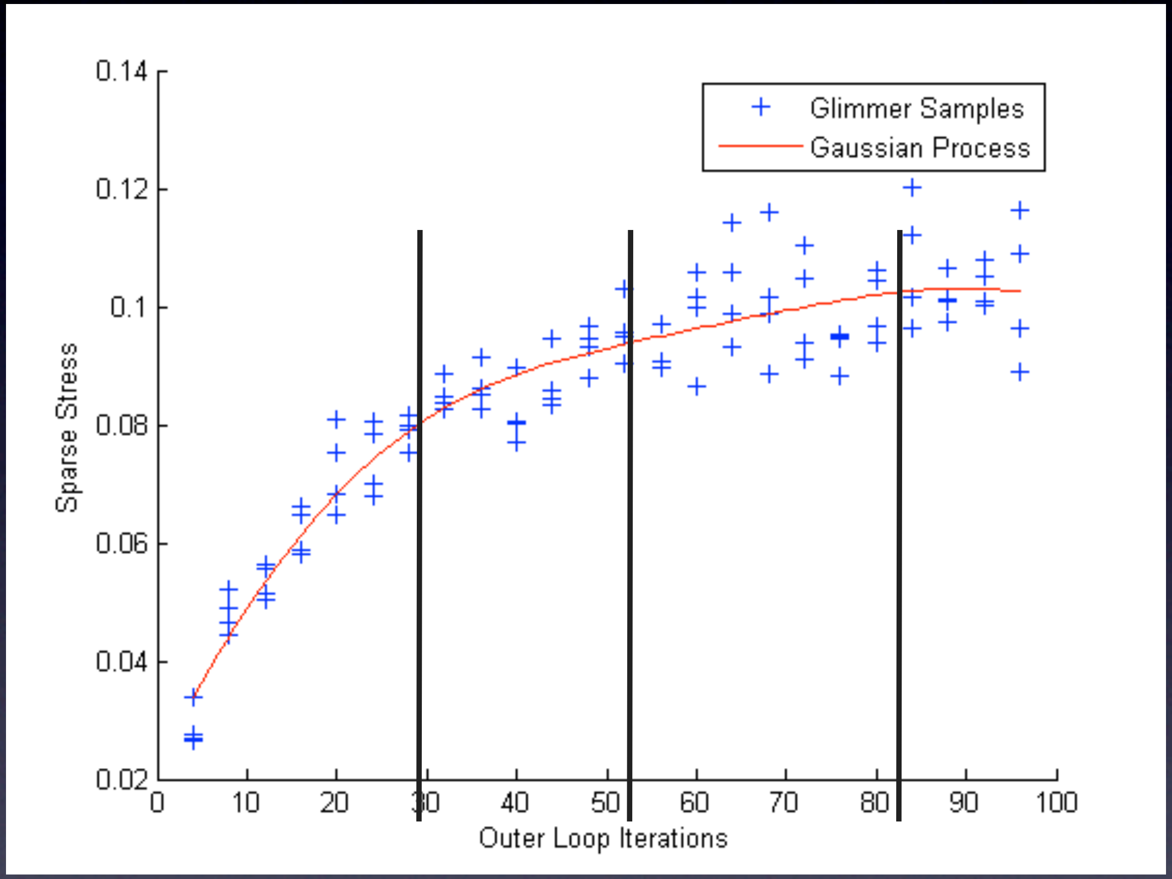


Results

Results

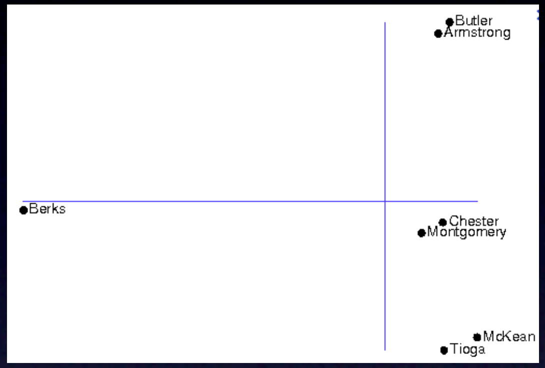
Name	N	D
videogame	96	Human
concepts	9600	DB Query
molecules	661	Euclidean O(n)
coastline	??	??
chicken	446	string edit
chromosome	4200	string edit
seaanimals	1100	string edit

Results Proposal

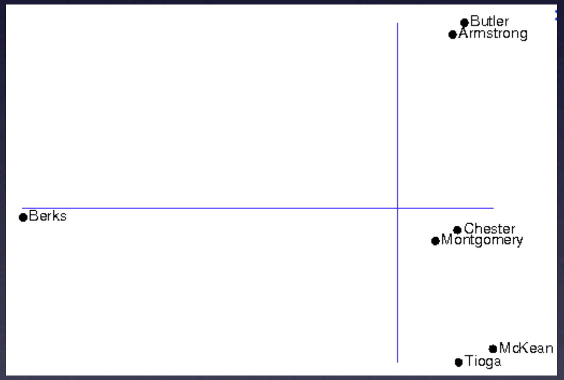


A B C

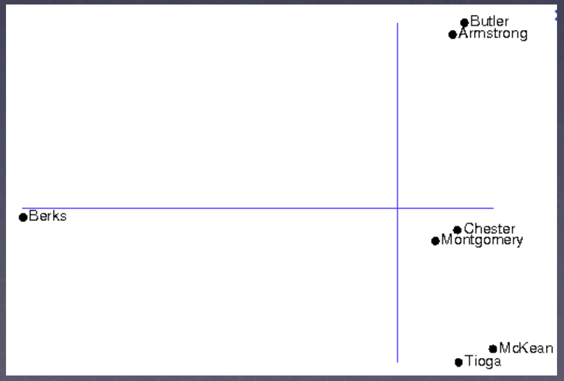
Different convergence thresholds



A



B



C

Their corresponding layouts
Superimposed on full layout

Conclusion

- Notified an MDS use case poorly served by existing tools; problems with costly distance functions
- Modified the Glimmer algorithm to work with a constrained input of distance data
- Proposed a technique for slowly growing an online design until layout quality converges