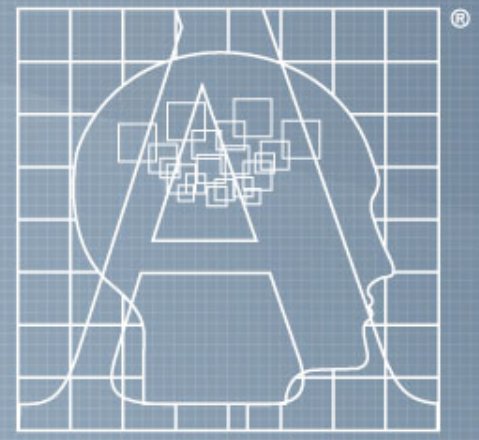


APTIMA®

HUMAN-CENTERED ENGINEERING

AARDVARK

Use Case Thoughts

Concepts

- Integrated Analysis
  - An RFI goes out to a team of analysts
  - Multiple analysts, multiple models
  - Integrate for better collaboration and decisionmaking

- Challenges
  - Combining qualitative and quantitative data
  - Collaborative visual analysis
  - Data management

Integrated Analysis

- If an adversary makes some decision (or is prevented from making a decision), what are the implications?
- Possible integrated/comprehensive analysis methods:
  - Social/economic/etc. pressures (economic, stats, AMB/SD model-based)
  - Game theoretic (state interests)
  - Social network analysis (and link analysis)
  - Network based
    - proliferation network
    - finance network
  - Leadership analysis (elite actors)
  - Influence analysis (Bayesian)

Use Case: VAST Challenge 2011

Initial Ideas: AARDVARK CONOPS

- *[this is out of scope since this is on the input side, but could be very useful so keeping it here for now] AARDVARK could help users select/create/operationalize variables for different analysis methods*
- Users go about doing their analyses in whatever tools they currently use (UCINet, R, ABM tool, ArcGIS, etc.)
- AARDVARK supports importing the data/results from those various tools
- AARDVARK supports relating those results/data/variables (some automation support, but probably more of a mixed-initiative human/machine task)
- AARDVARK uses built-in algorithms to look for trends/patterns across those results (which is pretty cool, since it's now operating across multiple data sets that have been related)
- A user analyzes the integrated data, as well as any recommendations that AARDVARK comes up with (e.g., pointing users to trends, etc.) using AARDVARK's built-in visual analytics

Initial Ideas: Use Case

- Should demonstrate use of multiple data models
  - Natural language processing/content analysis (Social Analytics Toolkit)
    - Opinions and sentiment of a population on key issues
  - Survey data
    - Demographics and opinions of a population
    - FACETS uses a ABM (SCIPR) to update “outdated” survey data to reflect more recent opinions
  - Social network analysis
    - Key leaders/influence
  - [optional fourth method] Cybersecurity modeling?

VAST 2011

- Mix of data similar to customer's multi-model RFIs
  - Mini-Challenge 1: microblog data to characterize an epidemic spread
  - Mini-Challenge 2: cyber security analysis for situational awareness of a corporate network infrastructure
  - Mini-Challenge 3: news analysis to investigate terrorist activity in the region
- For the Grand Challenge, you are charged with investigating the cause of the epidemic.
  - Are any terrorist activities related to the current epidemic?
  - Describe the series of events, planned or otherwise, that led to the current epidemic.

VAST 2011

- FY13: Build a tool that can analyze VAST Challenge data at a more useful level of detail
  - Actionable information, decision support
  - Not just “seems like this is suspicious”
- FY14: Evaluation exercise to compare AARDVARK to existing tools

VAST 2011: Building the Use Case

- More robust use case based on Grand Challenge
  - Mini-challenges as individual analyst questions
  - Combine with workflow information from JWAC
- Generating survey and demographic data
  - Necessary to model realistic JWAC RFI
  - Will artificially build based on real-world city of the appropriate size (e.g. Houston)
  - Map to opinions about terrorist groups, health infrastructure, city government
- Models based on data
  - Topic model of tweets and news stories – in progress
  - Social analysis of terrorist networks drawn from news stories
  - Agent-based models of generated survey & demographic data
  - Cybersecurity?

Development

Potential Tasks

- Requirements
  - Find connections between model data
  - Find discrepancies and disagreements
  - Produce a “big picture” view for analysis
- Possible components
  - Interactive information fusion
  - Visual analysis
  - Collaboration tools

Data Model and Integration

APTIMA HUMAN-CENTERED ENGINEERING

- We want to be able to import multiple data sets with different structures and granularities
- We want each imported data set to stay intact and exportable – the datasets shouldn't clobber each other
- We want to be able to integrate these datasets
  - To allow additional analysis/inference
  - To drive visualization

Dimensions of Integration

APTIMA HUMAN-CENTERED ENGINEERING

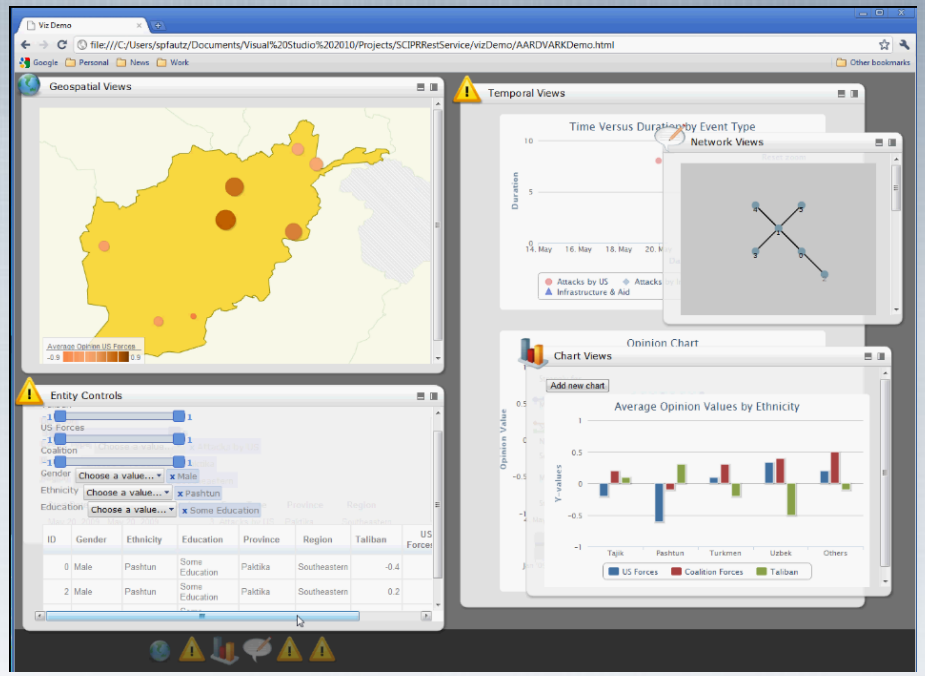
- Spatial integration
  - Infer geolocation relationships
- Temporal integration
  - Normalize to same timescale
  - Infer event relationships???
- Semantic integration
  - Infer topic relationship
  - Infer refers_to relationships

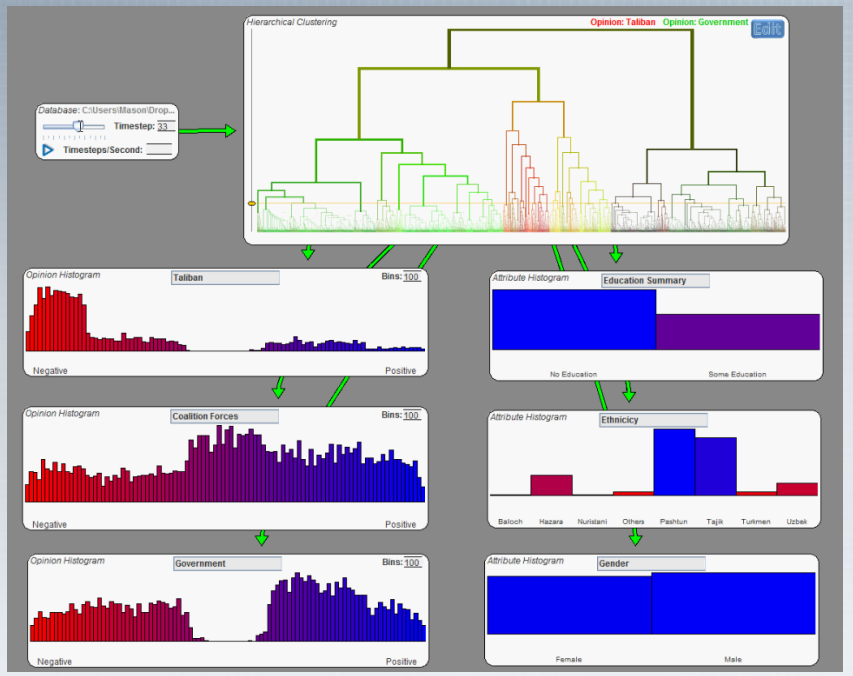
The Nuts and Bolts of Data Integration

APTIMA HUMAN-CENTERED ENGINEERING

- When integrating data there are several approaches we could take
  - Flat Graph Structure
    - All imported datasets populate a common set of entities and relationships
  - “Stratified” Graph
    - All imported data sets create their own entities and relationships.
    - Think of each graph as an overlay
    - Needs to be able to be flattened for consumption
- But how do we maintain provenance?
- How do we prevent clobbering?
- When importing a dataset that provides attributes, which entities do they get attached to?

Integrated Visual Analysis





- Phase 1 designs: combining views
  - Coordinated multiple views
  - Linked views

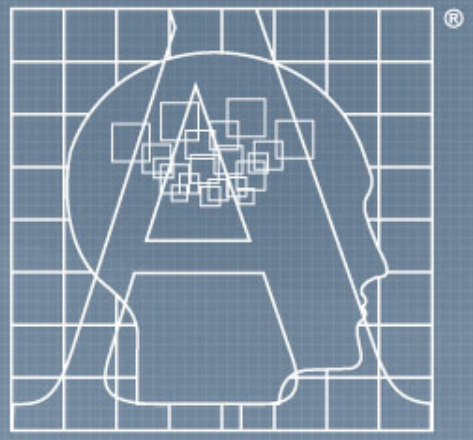
Brown University Partnership

- Can provide advanced research on general visualization questions
- User studies, research prototypes
- Ideas for focus
  - Integrated visualization of qualitative and quantitative data
  - Visualization to compare models: multi-window vs. layered views
  - Pattern recognition to guide attention; is it useful?
- With VAST 2011 use case, can work with the same data we do

Evaluation

Metrics

- Evaluation by Brown researchers
  - e.g., What is the best way to present multi-model data?
- Evaluation workshop at the end of phase 2 or in the option to test some of the metrics/anticipated benefits and collect some data
  - Red vs. Blue exercise using AARDVARK versus existing tools
  - JWAC is interested in setting this up
- Metrics / Research Questions
  - AARDVARK enables users to compare results across two or more data models more quickly than the current approach (which is manual)
    - Perhaps we'd time how long it takes two group to come up with the answer to a question that requires data from at least two different models to answer
  - Other metrics?



APTIMA®

HUMAN-CENTERED ENGINEERING