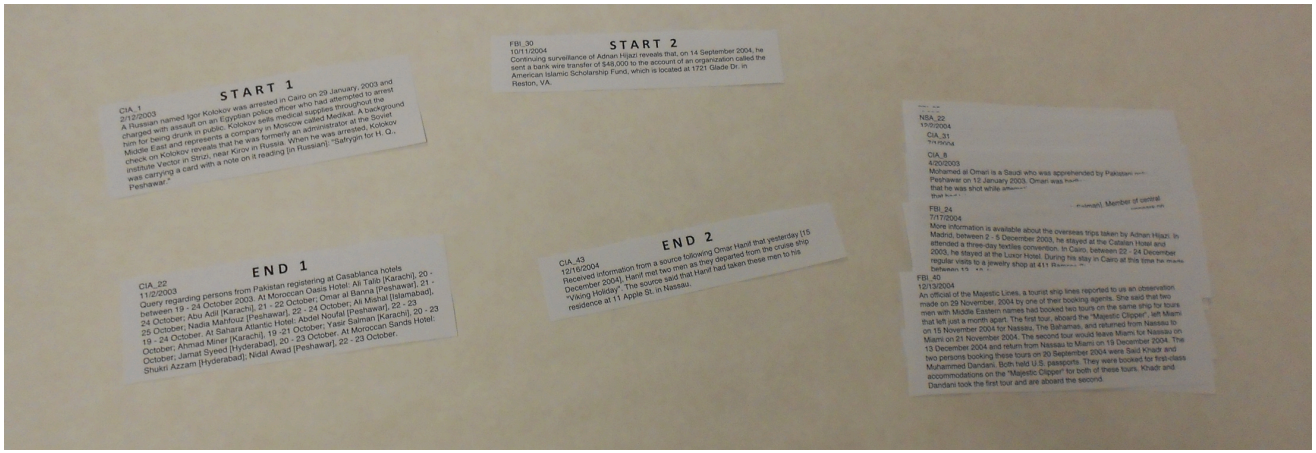
HOW ANALYSTS COGNITIVELY "CONNECT THE DOTS"

Lauren Bradel
Jessica Zeitz Self
Alex Endert
M. Shahriar Hossain
Chris North
Naren Ramakrishnan

Virginia Tech

The User Study

- 10 participants
- 47 documents with two pairs containing a start and an end
- *Task*: “connect the dots” between the starts and ends
- Individual think-aloud sessions with observation



Connecting Document Pairs

Low-Level Connections

- Entity-Entity Links

High-Level Connections

- Conceptual Connections
  - Temporal
  - Speculative
  - Domain Knowledge

	1	2	3	4	5	6	7	8	9	10
entity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
conceptual		✓	✓	✓	✓	✓	✓	✓	✓	✓
temporal		✓		✓					✓	✓
speculative			✓	✓	✓	✓	✓	✓	✓	✓
domain knowledge			✓	✓	✓	✓	✓	✓	✓	✓

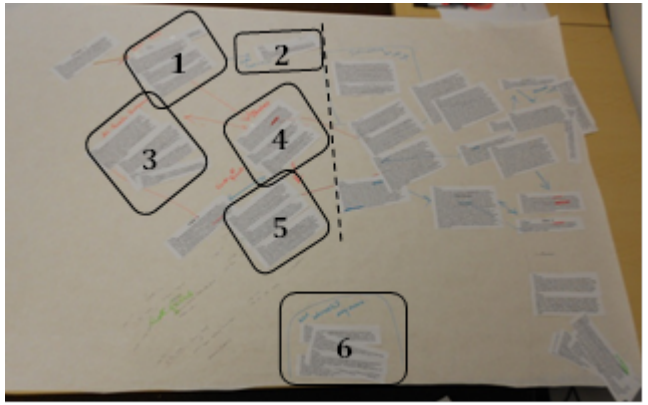
Constructing Conceptual Stories

- Storytelling strategies
  - Data-driven, hypothesis-driven
  - Work start to end, end to start, or inward from the start and end
  - Work on entire plot, work on separate stories
- Suspicion Sensor
  - Repeated occurrences of an entity
  - Two events are unbelievable as a coincidence
  - Domain Knowledge of what is suspicious

Spatial Representations

- Intermediate Representations

	1	2	3	4	5	6	7	8	9	10
clusters	✓	✓	✓		✓		✓	✓		✓
concept map		✓			✓		✓	✓	✓	✓
timeline	✓	✓	✓	✓		✓				



Numbers indicate distinct clusters, dotted line separates story 1 and story 2

- Final Story Shapes

	1	2	3	4	5	6	7	8	9	10
linear with branching	✓	✓		✓		✓				✓
web			✓				✓	✓		
disorganized					✓				✓	

Disorganized layout, documents overlaying note-based concept map

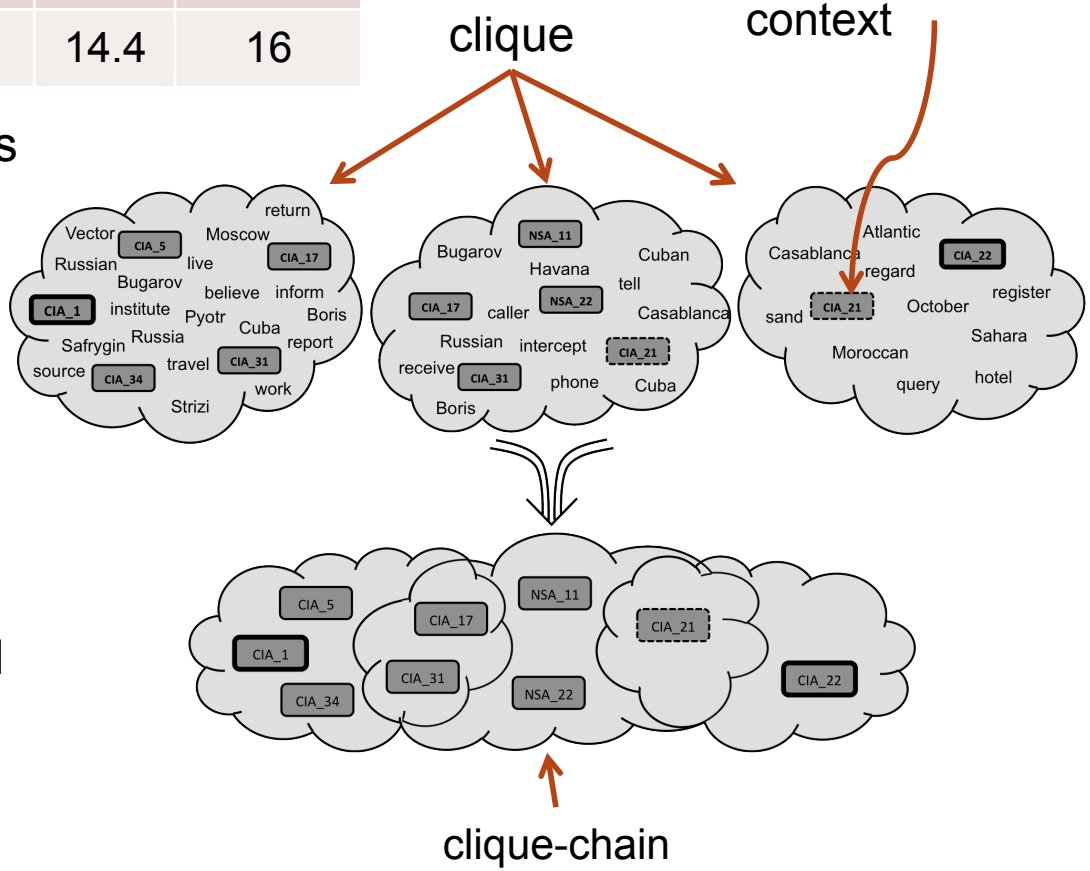
Comparison to Algorithm-Generated Stories

Number of documents included in stories

	algorithm	user min	user max	mean	median
story 1	7	14	26	18.9	19
story 2	4	9	18	14.4	16

- Participants who included all docs from algorithm
  - 50%+ for story 1
  - 75%+ for story 2
- Participants who included docs not included in algorithm
  - 50%+ for story 1
  - 63%+ for story 2
- Participants included background information, subplots, and side plots

- Left out by 2 participants
- Lots of entities & info, but little context



Implications for the Storytelling Algorithm

- Include relevant documents despite whether the date fits within a timeframe
- Provide multiple stories with distinct documents to encourage user to investigate multiple hypotheses
- Include relevant entity-rich documents lacking context since these are more likely to be left out by user
- Present documents in a web-like structure with off-shooting documents
- Allow user to quickly access documents containing background information
- Have algorithm add entities provided by user through the linking of documents