

Discovering Relationships in Text

NLP Informs SNA

James M. Keiser

Applied Research Mathematician
Mathematics Research, NSA

24 April 2019

Outline

- **Questions**
- **Data**
- **Static and Dynamic Processing**
 - **Static graphs**
 - **Time dependent graphs**
 - **Linkages, Centrality, Sentiment**
- **Results**
- **Conclusions and Future Directions**
 - **Discoverable Relationships**

Questions

- **Three Broad Questions**
 - **What Happened? (forensic)**
 - **What is Happening? (explanatory)**
 - **What will Happen? (anticipatory)**
- **Data (for this talk)**
 - **Text**
 - **Novels**
 - **Scripts**

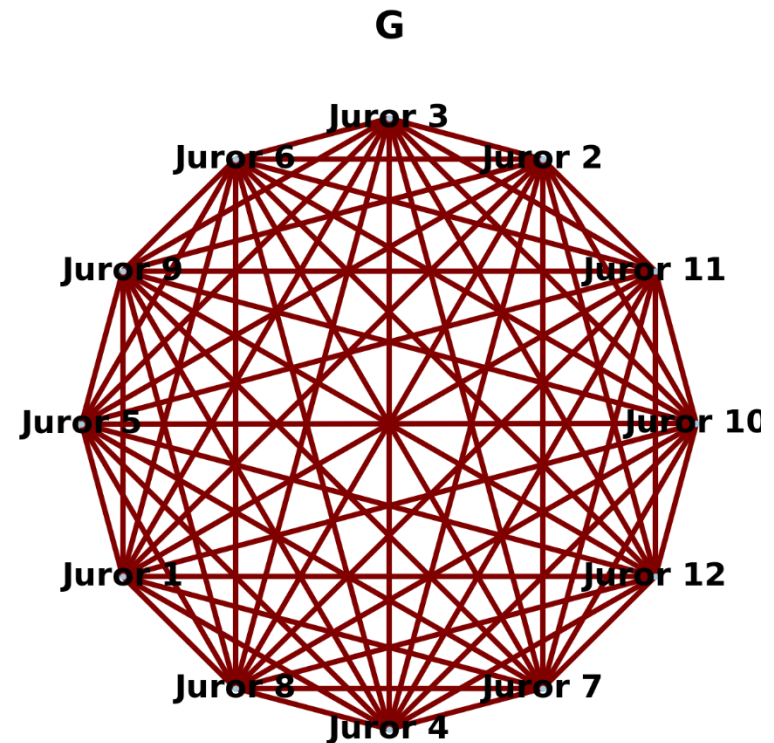
Exemplar Data

- **Historical**
 - Knuth—Stanford GraphBase: entity co-occurrence
- **Project Gutenberg**
 - Lots of texts: les Miserables, Homer, Huck Finn
- **Scripts**
 - Movies
 - Episodic series

Aggregate Processing

Linkages: *12 Angry Men*

- **Vertex**
 - juror
- **Edge**
 - at the table
- **Vertex size**
 - degree
- **Edge thickness**
 - importance



An Aside

The Question: How do characters interact throughout a narrative?

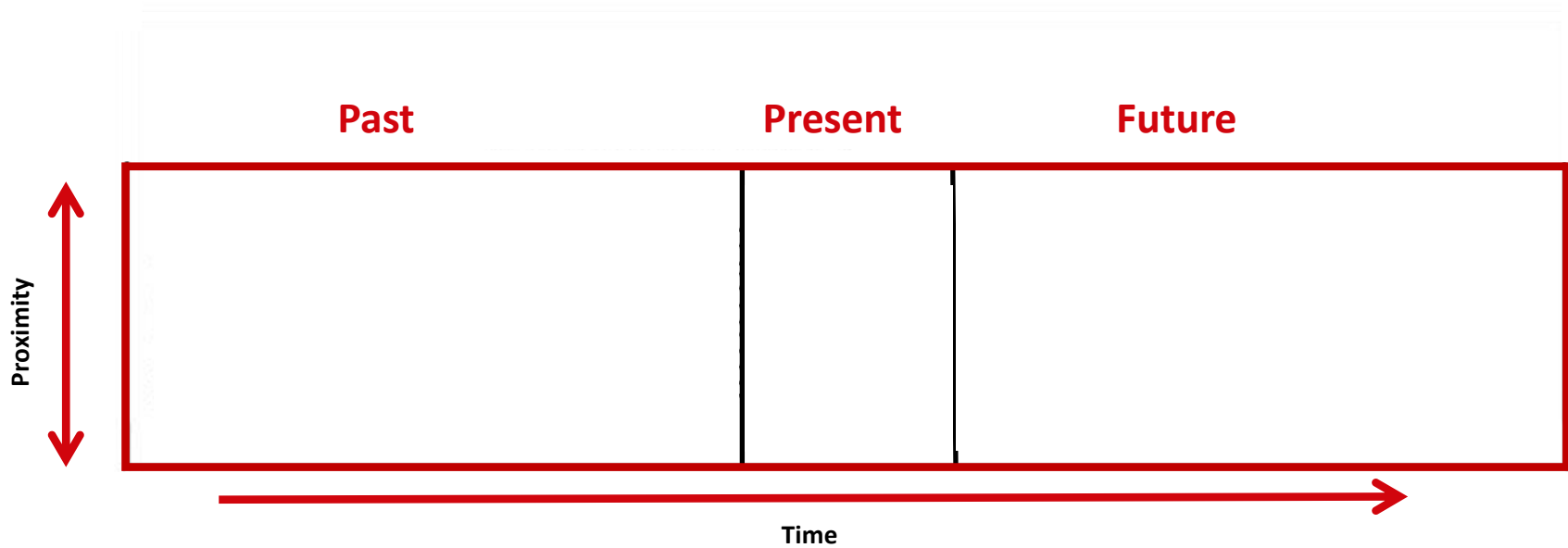
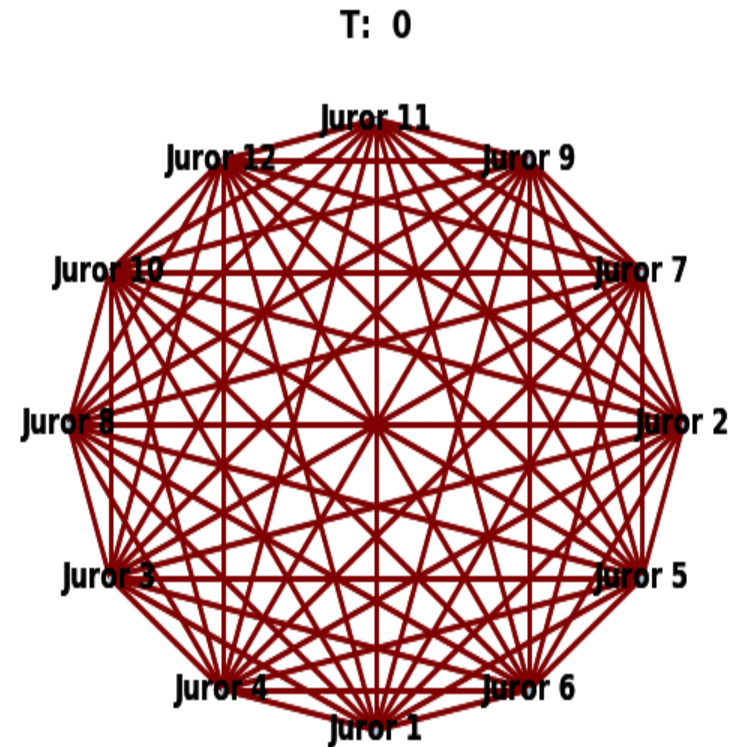


Image From: http://thumbnails.visually.netdna-cdn.com/character-interaction_50290b06e56dc.png

Dynamic Processing

Linkages: *12 Angry Men*

- **Vertex**
 - juror
- **Edge**
 - at the table
- **Vertex size**
 - degree
- **Edge thickness**
 - importance



Dynamic Processing

Sentiment

VADER: Valence Aware Dictionary for sEntiment Reasoning

Lexicographic, 7500 emojis, acronyms, words (English)

☹️ -1.9	bff 2.9	Amazon 0.7	Amortise 0.5	Love 3.2
😊 2.0	lol 2.9	Abduction -2.8	Amortised -0.2	Hate -2.7
☹️ -1.5		Abductions -2.0	Amortises 0.1	
😊 1.3				

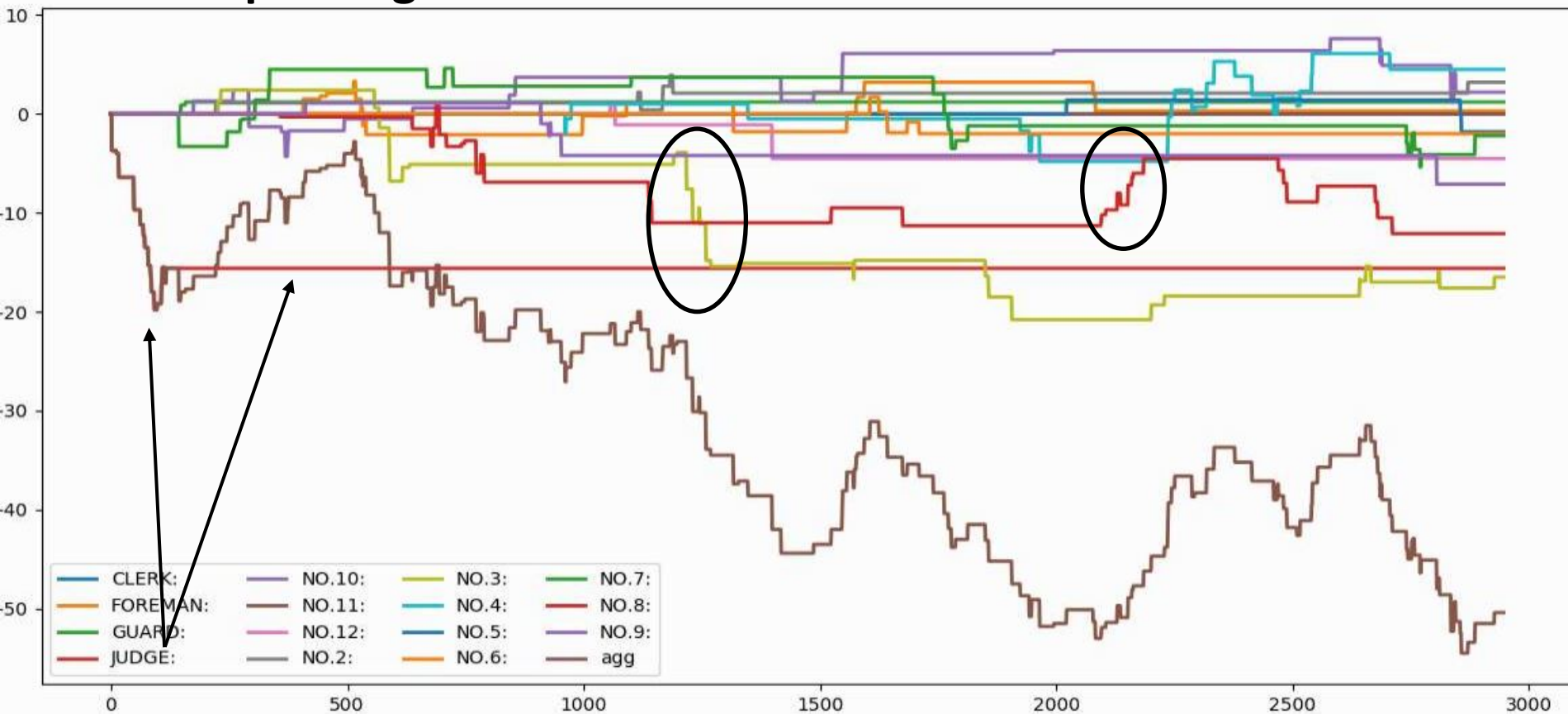
Negative, neutral, positive, combined sentiments

Uses lexicon and sentence structure

Part of NLTK

Dynamic Processing

Accumulated sentiment per actor per spoken line
No speaking => "Coast" sentiment



Dynamic Processing

Sentiment

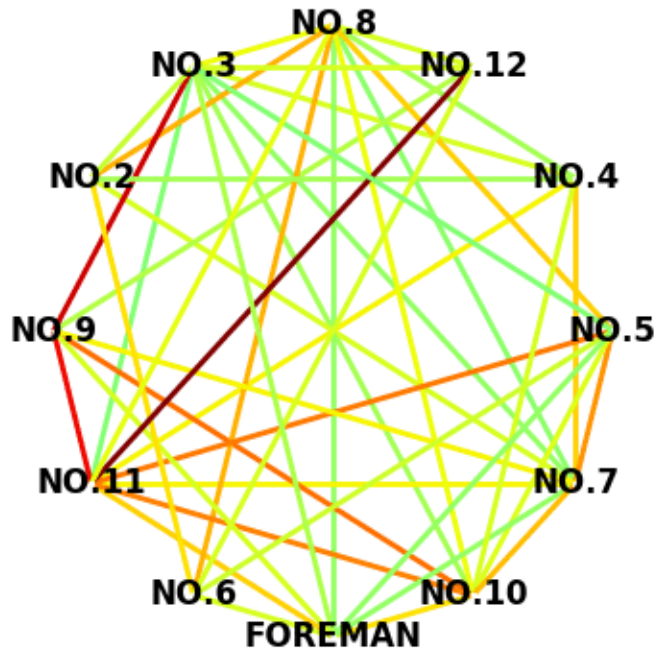
Let juror p make a statement, followed by juror q.

Further, suppose that the sentiment of juror q's statement is in response to the statement of juror p.

Let the edge between juror q and juror p be assigned a weight equal to the sentiment of juror p.

Lots of options: sentiment of juror q, sum of p&q, some other function, etc.

Dynamic Processing



Topological Metrics

Centrality

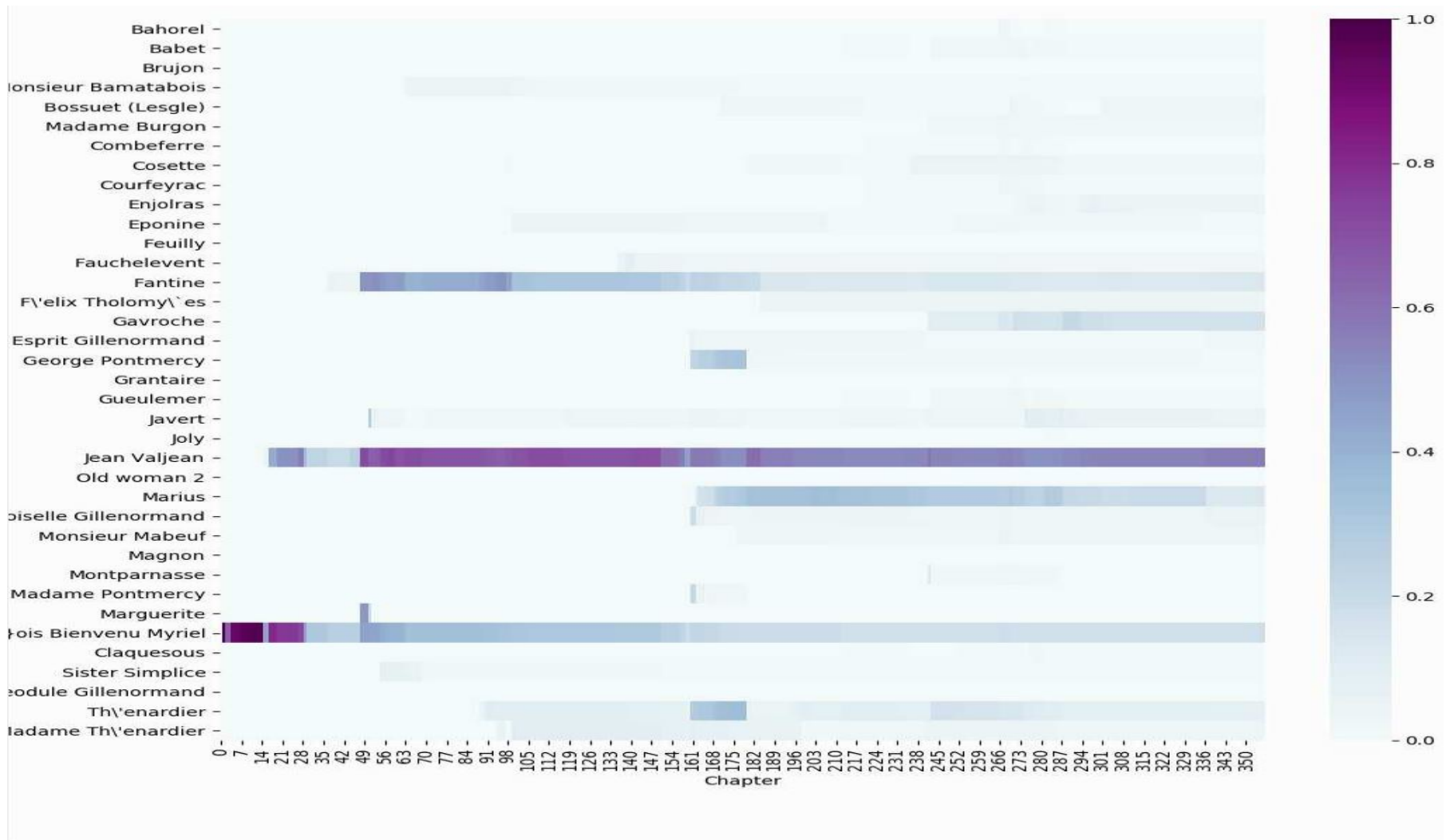
Measure the fraction of all pairs of shortest paths through a given vertex.

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

Consider the graph of entities (vertices) as a function of the narrative.

Compute the centrality for each entity in the aggregate graph up to sentence n in the narrative.

Topological Metrics



Future Explorations

Discoverable Relationships

Named Entity Recognition

Analyze Connecting Words

Enrich: dictionaries, thesauri, keyword lists,

Cluster, topic model

discover underlying relationships

SNA on resulting multi-graphs

Entities have different and multiple relationships

Dynamic Analysis

Relationships are born and evolve (and die)

Predictive Analytics

Discover emerging relationships

comments

questions