

'Shall I compare thee to a network?'

Visualizing the Topological Structure of Shakespeare's Plays

Bastian Rieck Heike Lette

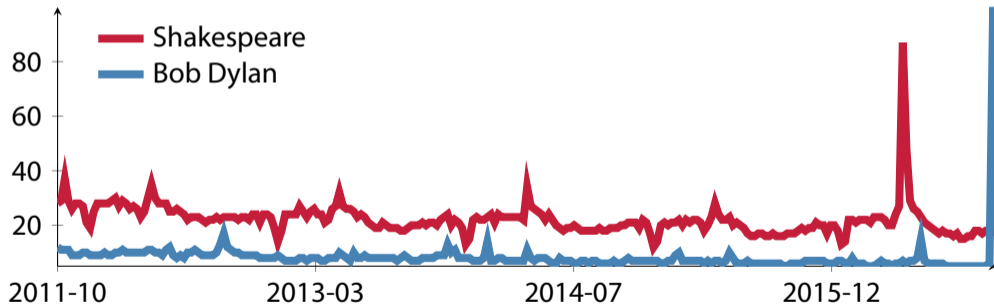
Visual Information Analysis Group
TU Kaiserslautern
Germany



Why Shakespeare?

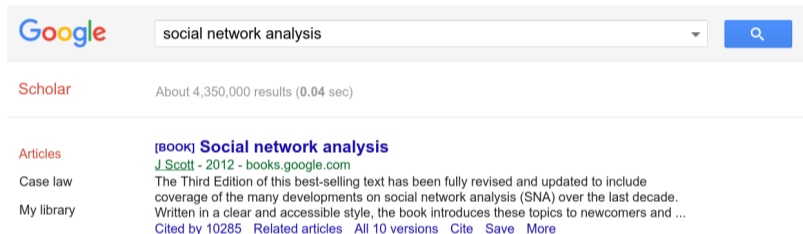
1564–1616

Tons of reasons: 38 plays, 154 sonnets, more than 1,000 words and 50 idioms...



Agenda

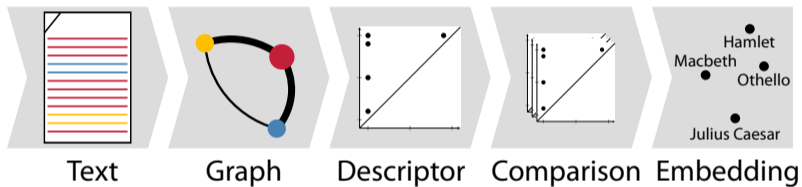
- Social network analysis: well-established for 'real-world' social networks



The screenshot shows a Google Scholar search interface. At the top left is the Google logo. To its right is a search bar containing the text "social network analysis" and a search button with a magnifying glass icon. Below the search bar, the word "Scholar" is displayed in red, followed by the text "About 4,350,000 results (0.04 sec)". A horizontal line separates this from the search results. The first result is a book entry: "[BOOK] Social network analysis" in blue, followed by the author "J.Scott - 2012 - books.google.com" in green. Below the author is a short description: "The Third Edition of this best-selling text has been fully revised and updated to include coverage of the many developments on social network analysis (SNA) over the last decade. Written in a clear and accessible style, the book introduces these topics to newcomers and ...". At the bottom of the entry are several links: "Cited by 10285", "Related articles", "All 10 versions", "Cite", "Save", and "More".

- Create networks for Shakespeare's 38 plays
- *Distant reading* paradigm: Compare relations or structures
- Use novel methods to derive a structural descriptor of each play

Workflow



From plays to graphs

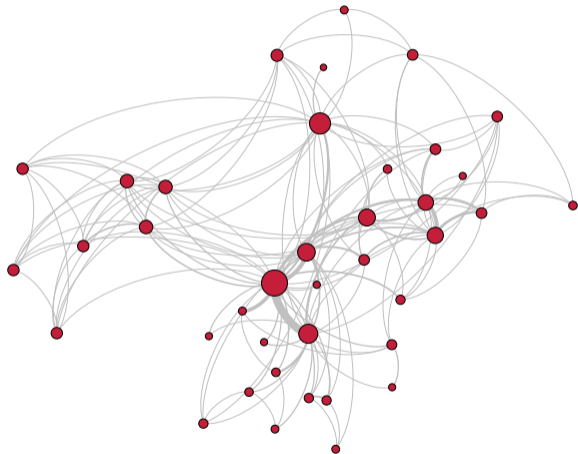


- Use freely-available tagged corpus¹
- Extract speakers by scene and count their words
- Connect speakers u and v that co-occur in a scene by an edge
- Different edge weight schemes

¹<http://lexically.net/wordsmith/support/shakespeare.html>

Results

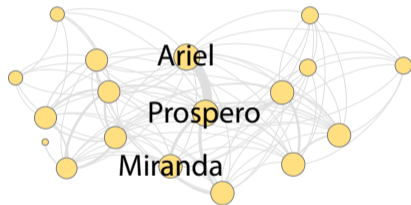
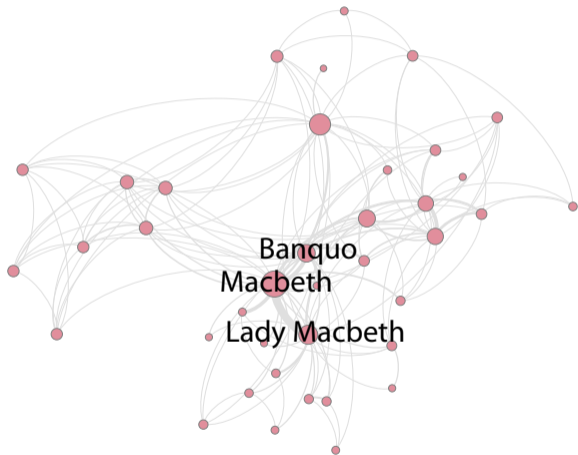
'Macbeth' vs. 'The Tempest'



Layout created with Gephi and ForceAtlas2.
The graphs *appear* to be different.
How to quantify this?

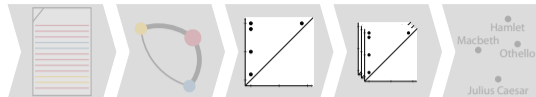
Results

'Macbeth' vs. 'The Tempest'



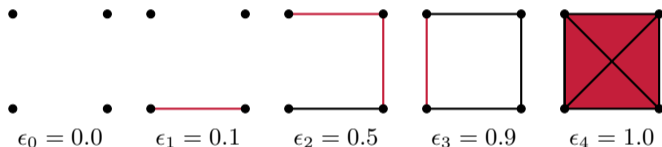
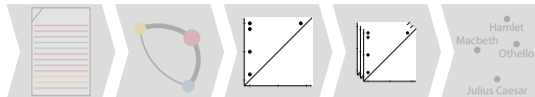
Layout created with Gephi and ForceAtlas2.
The graphs *appear* to be different.
How to quantify this?

A topological feature descriptor



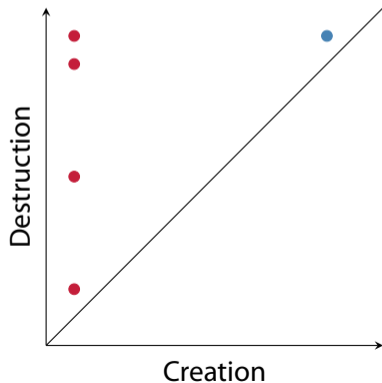
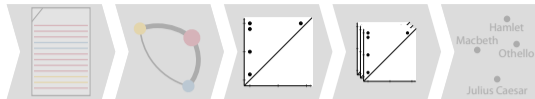
- Features in graphs: number of connected components β_0 , number of cycles β_1
- *Key insight*: both numbers depend on the 'scale' at which we view the graph
- Obtain scale information by edge weights; analyse how β_0 and β_1 change

Example



- Observation: Some thresholds *create* a feature, other thresholds *destroy* a feature
- ϵ_0 creates four connected components
- ϵ_1 destroys one connected component by merging it with another one
- Add (ϵ_0, ϵ_1) to the *persistence diagram*

Persistence diagram



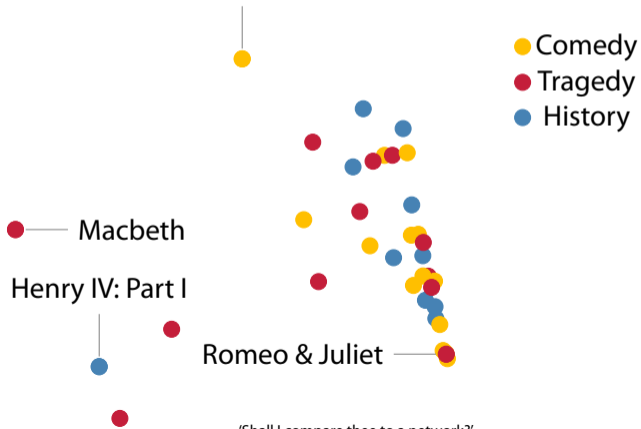
- Connected components
- Cycles

$$W_2(\mathbf{X}, \mathbf{Y}) := \left(\inf_{\eta: \mathbf{X} \rightarrow \mathbf{Y}} \sum_{x \in \mathbf{X}} \|x - \eta(x)\|_{\infty}^2 \right)^{\frac{1}{2}}$$

Results



All's Well that Ends Well



'Shall I compare thee to a network?'

Conclusion



- Proof-of-concept
- Extraction of *co-occurrence networks* from Shakespeare's plays
- Structural analysis based on topological features
- Future work:
 - Emotional content
 - Play recommendation
 - Other types of texts, other authors, ...



<https://github.com/Submanifold/Shakespeare>