'Shall I compare thee to a network?'

Visualizing the Topological Structure of Shakespeare's Plays

Bastian Rieck   Heike Leitte
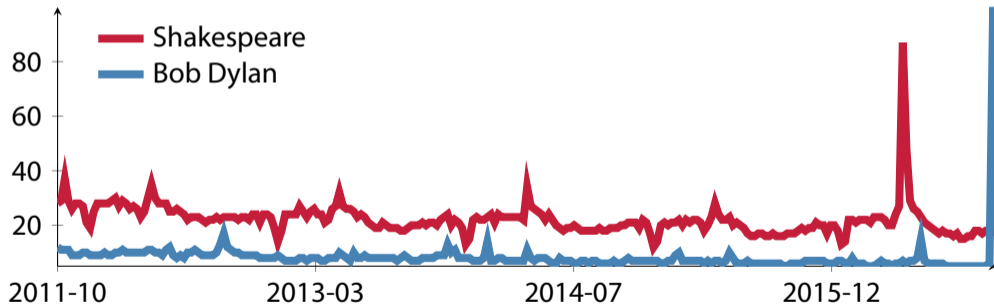
Visual Information Analysis Group
TU Kaiserslautern
Germany

TECHNISCHE UNIVERSITÄT
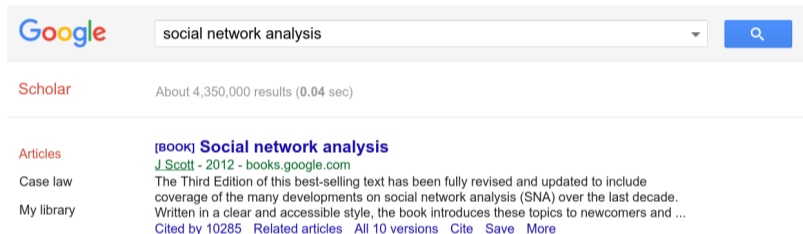KAISERSLAUTERN

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

- 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

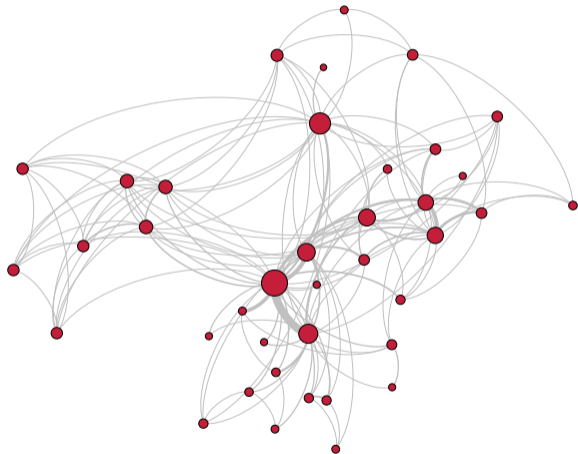Text Graph Descriptor Comparison Embedding
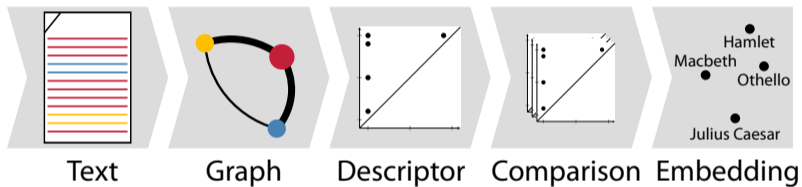
From plays to graphs

- Use freely-available tagged corpus[1]
- 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
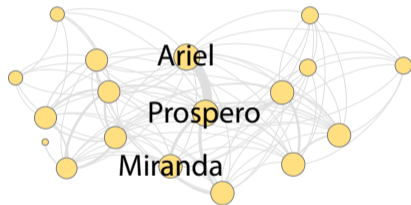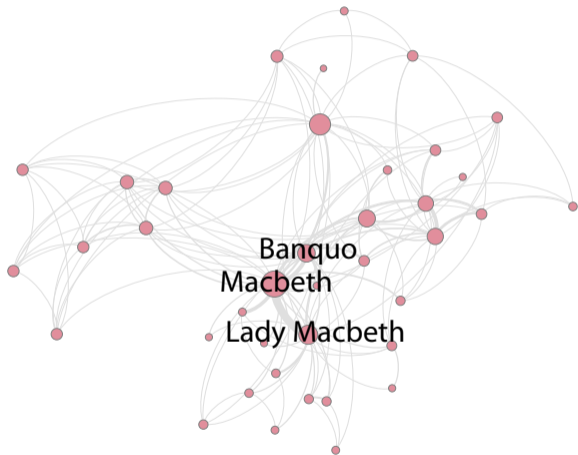
Results

'Macbeth' vs. 'The Tempest'

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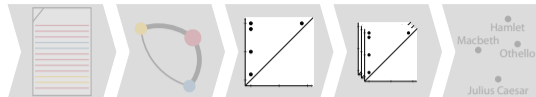The graphs *appear* to be different.
How to quantify this?

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

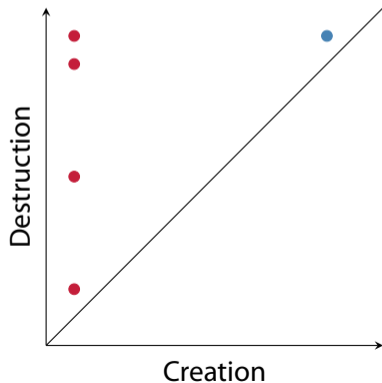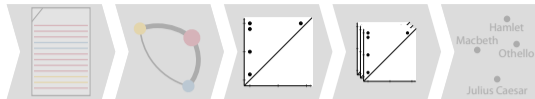- Features in graphs: number of connected components β_0, number of cycles β_1
- *Key insight*: both numbers depend on the 'scale' at which we view the graph
- Obtain scale information by edge weights; analyse how β_0 and β_1 change

Example

$$\epsilon_0 = 0.0 \quad \epsilon_1 = 0.1 \quad \epsilon_2 = 0.5 \quad \epsilon_3 = 0.9 \quad \epsilon_4 = 1.0$$
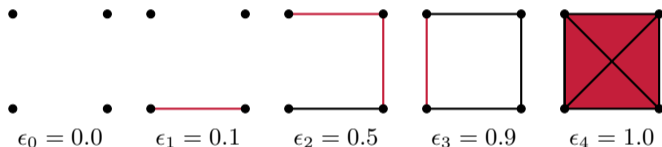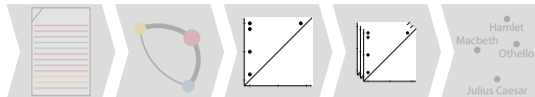
- Observation: Some thresholds *create* a feature, other thresholds *destroy* a feature
- ϵ_0 creates four connected components
- ϵ_1 destroys one connected component by merging it with another one
- Add (ϵ_0, ϵ_1) to the *persistence diagram*

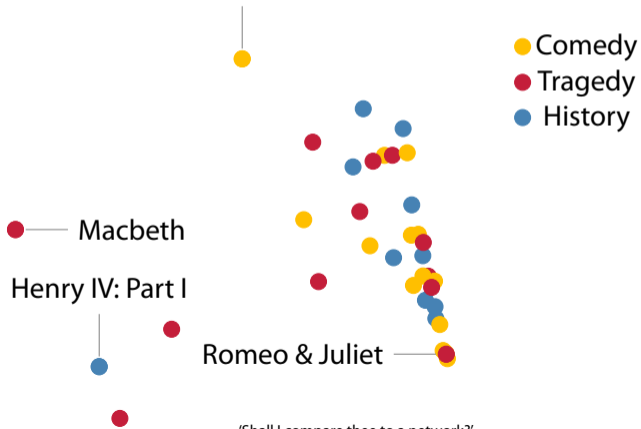# Persistence diagram

- Connected components
- Cycles

Destruction

Creation

$$\mathrm{W}_2(\mathbf{X}, \mathbf{Y}) := \left(\inf_{\eta \colon \mathbf{X} \to \mathbf{Y}} \sum_{x \in \mathbf{X}} \|x - \eta(x)\|_\infty^2 \right)^{\frac{1}{2}}$$

Results

All's Well that Ends Well

- ● Comedy
- ● Tragedy
- ● History

Macbeth

Henry IV: Part I

Romeo & Juliet

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, …

   :octocat: https://github.com/Submanifold/Shakespeare