# '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



#### 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

Google	social network analysis
Scholar	About 4,350,000 results (0.04 sec)
Articles Case law My library	[BOOK] Social network analysis J Scott - 2012 - books.google.com 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 Cited by 10285 Related articles All 10 versions Cite Save 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 corpus<sup>1</sup>
- Extract speakers by scene and count their words
- Connect speakers  $\boldsymbol{u}$  and  $\boldsymbol{v}$  that co-occur in a scene by an edge
- Different edge weight schemes

Bastian Rieck, Heike Leitte

<sup>&</sup>lt;sup>1</sup>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?





- Features in graphs: number of connected components  $\beta_0$ , number of cycles  $\beta_1$
- Key insight: both numbers depend on the 'scale' at which we view the graph
- Obtain scale information by edge weights; analyse how  $\beta_0$  and  $\beta_1$  change

# Example



• Observation: Some thresholds create a feature, other thresholds destroy a feature

 $\epsilon_4 = 1.0$ 

- $\epsilon_0$  creates four connected components
- $\epsilon_1$  destroys one connected component by merging it with another one
- Add  $(\epsilon_0, \epsilon_1)$  to the *persistence diagram*

# Persistence diagram





Bastian Rieck, Heike Leitte

'Shall I compare thee to a network?'

#### Results





## 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