Universität Stuttgart

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Qd

Nils Reiter,
Benjamin Krautter,
Janis Pagel,
Marcus Willand

G. E. Lessing
Emilia Galotti
Studienausgabe

Reclam

**Detecting Protagonists
in German Plays
around 1800
as a Classification Task**

Disclaimer

- We have progressed in our work
- We will present the current state of our research
- **Not** only what is in the submitted paper
- Updated results
- Talk includes analysis on dramas from 1700 to 1900
- See also Krautter et. al. 2018: Titelhelden und Protagonisten – Interpretierbare Figurenklassifikation in deutschsprachigen Dramen. *LitLab Pamphlets*, vol. 7, November 2018.

Quantitative Drama Analytics (QuaDramA) Qd

- Cooperation between German literary studies and computational linguistics

Quantitative Drama Analytics (QuaDramA) Qd

- Cooperation between German literary studies and computational linguistics
- Analyse (German) dramatic texts computationally

Quantitative Drama Analytics (QuaDramA) Qd

- Cooperation between German literary studies and computational linguistics
- Analyse (German) dramatic texts computationally
- Historical perspective

Quantitative Drama Analytics (QuaDramA) Qd

- Cooperation between German literary studies and computational linguistics
- Analyse (German) dramatic texts computationally
- Historical perspective
- Investigate character types

Quantitative Drama Analytics (QuaDramA) Qd

- Cooperation between German literary studies and computational linguistics
- Analyse (German) dramatic texts computationally
- Historical perspective
- Investigate character types
- Coreference resolution for dramatic texts

Introduction

1

What is a Drama?

- Use of action and dialogue

What is a Drama?

- Use of action and dialogue
- Typically divided into acts and scenes

What is a Drama?

- Use of action and dialogue
- Typically divided into acts and scenes
- Typically designed to be performed (on stage)

What is a Drama?

- Use of action and dialogue
- Typically divided into acts and scenes
- Typically designed to be performed (on stage)
- Stage directions

What is a Drama?

- Use of action and dialogue
- Typically divided into acts and scenes
- Typically designed to be performed (on stage)
- Stage directions
- Cast

What is a Drama?

- Use of action and dialogue
- Typically divided into acts and scenes
- Typically designed to be performed (on stage)
- Stage directions
- Cast
- Dramatic conflict

Excerpt from Lessing's Emilia Galotti

Fünfter Auftritt

Der Prinz. Emilia. Marinelli.

DER PRINZ.
Wo ist sie? wo? -
Wir suchen Sie überall,
schönstes Fräulein.
- Sie sind doch wohl?
- Nun so ist alles wohl!
Der Graf, Ihre Mutter, -

EMILIA.
Ah, gnädigster Herr!
wo sind sie?
Wo ist meine Mutter?

Excerpt from Lessing's Emilia Galotti

Fünfter Auftritt

Der Prinz. Emilia. Marinelli.

DER PRINZ.
Wo ist sie? wo? -
Wir suchen Sie überall,
schönstes Fräulein.
- Sie sind doch wohl?
- Nun so ist alles wohl!
Der Graf, Ihre Mutter, -

EMILIA.
Ah, gnädigster Herr!
wo sind sie?
Wo ist meine Mutter?

```xml
<div type="scene">
  <div>
    <desc>
      <title>5. Auftritt</title>
    </desc>
  </div>
  <div type="text">
    <div type="h4">
      <head>Fünfter Auftritt</head>
      <stage>
        <hi>Der Prinz. Emilia. Marinelli.</hi>
      </stage>
      <sp who="#der_prinz">
        <speaker>DER PRINZ.</speaker>
          <p> Wo ist sie? wo? -
              Wir suchen Sie überall,
              schönstes Fräulein.-
                  Sie sind doch wohl?-
                Nun so ist alles wohl!
              Der Graf, Ihre Mutter, -</p>
      </sp>
      <sp who="#emilia">
        <speaker>EMILIA.</speaker>
          <l> Ah, gnädigster Herr!
              wo sind sie?
              Wo ist meine Mutter?</l>
      </sp>
```

Goals

- Classify all figures in play regarding the classes: *Protagonist - Not Protagonist*

Goals

- Classify all figures in play regarding the classes: *Protagonist - Not Protagonist*
- Analyse results w.r.t. literary interpretation

Definition of Being a Protagonist

- Difficult from theoretical point of view

Definition of Being a Protagonist

- Difficult from theoretical point of view
- We settled on:

Protagonist

- Causes or solves the central dramatic conflict
- This can happen actively or passively

Definition of Being a Protagonist

- Difficult from theoretical point of view
- We settled on:

Protagonist
- Causes or solves the central dramatic conflict
- This can happen actively or passively

- From this follows:
  - There can be more than one protagonist per drama
  - Not only "heroes" in a positive sense, but also "anti-heroes" allowed

Experiments

2

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas
Actives	Number of scenes where figure is present

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas
Actives	Number of scenes where figure is present
Passives	Number of Scenes where figure is mentioned

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas
Actives	Number of scenes where figure is present
Passives	Number of Scenes where figure is mentioned
LastAct	Is figure present in the last act?

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas
Actives	Number of scenes where figure is present
Passives	Number of Scenes where figure is mentioned
LastAct	Is figure present in the last act?
NumberFig	In respective drama

Features

Feature Name	Description
Tokens	Character speech, normalised on whole text
Centrality	Different measures: *(weighted) degree*, *closeness*, *betweenness*, *eigenvector*
Topic Modelling	10 Topics trained on the dramas
Actives	Number of scenes where figure is present
Passives	Number of Scenes where figure is mentioned
LastAct	Is figure present in the last act?
NumberFig	In respective drama
Genre	e.g. *Weimar Classicism*, *Bourgeois Tragedy*, *Naturalism*, etc.

Centrality
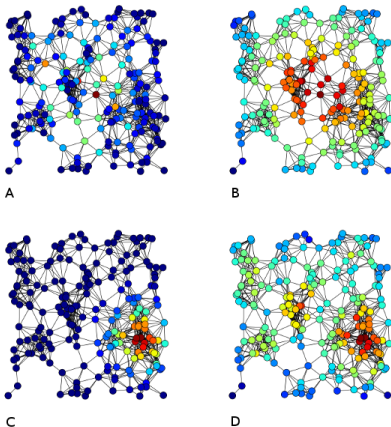
Figure 1: A) Betweenness centrality, B) Closeness centrality, C) Eigenvector centrality, D) Degree centrality. Source:
`https://en.wikipedia.org/wiki/Centrality`

Experimental Setup

- 114 dramas in our corpus

Experimental Setup

- 114 dramas in our corpus
- Three annotators

Experimental Setup

- 114 dramas in our corpus
- Three annotators
- Each data point represents a character in a play

Experimental Setup

- 114 dramas in our corpus
- Three annotators
- Each data point represents a character in a play
  - Arrays of feature values
  - Class: P (Title character) / C (Not title character)

Experimental Setup

- 114 dramas in our corpus
- Three annotators
- Each data point represents a character in a play
  - Arrays of feature values
  - Class: P (Title character) / C (Not title character)
  - Example Emilia: {`tokens=0.09`, `actives=0.16`, `...`, `class=P`}

Experimental Setup

- 114 dramas in our corpus
- Three annotators
- Each data point represents a character in a play
  - Arrays of feature values
  - Class: P (Title character) / C (Not title character)
  - Example Emilia: {`tokens=0.09`, `actives=0.16`, ..., `class=P`}
- Random forest

Results

3

Results

	Precision	Recall	F1	Accuracy
Majority BL	-	0.00	-	0.86
Tokens BL	0.62	0.99	0.76	0.93
Random Forest	**0.72**	**1.00**	**0.83**	**0.95**

Table 1: Results for classification of protagonists plus baselines. Shown are the average values of the experiments for each annotator.
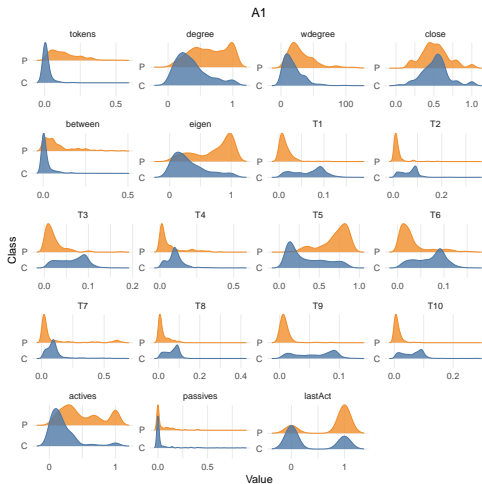
Feature Distribution

Going a step back...

Figure 2: Feature distribution for one annotator's data set.

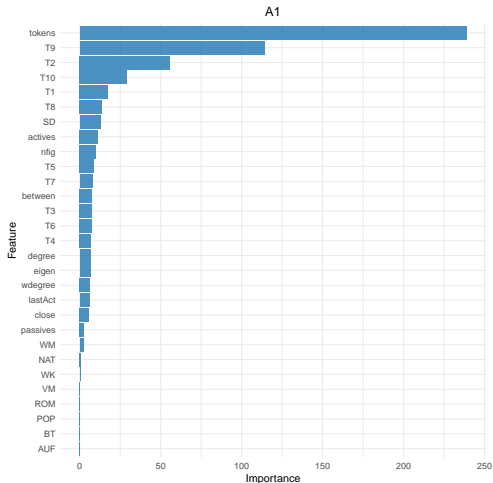# Feature Importance

Going back to the model...

Figure 3: Relative Feature Importance for one model.

Character Analysis

4

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani
- The prince wants Emilia for himself

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani
- The prince wants Emilia for himself
- His chamberlain Marinelli assassinates Appiani

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani
- The prince wants Emilia for himself
- His chamberlain Marinelli assassinates Appiani
- Orsina was the prince's mistress and plots to kill him

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani
- The prince wants Emilia for himself
- His chamberlain Marinelli assassinates Appiani
- Orsina was the prince's mistress and plots to kill him
- Odoardo kills Emilia at her wish

Analysis of Single Characters

Example: Emilia Galotti by Gotthold Ephraim Lessing
Bourgeois tragedy

- **Emilia**: Young bourgeois woman
- **Der Prinz**: "The prince", ruler of the region
- **Marinelli**: Chamberlain of the prince
- **Odoardo & Claudia**: Emilia's parents
- **Appiani & Orsina**: Count/Countess

Summary of plot

- Emilia is engaged to Count Appiani
- The prince wants Emilia for himself
- His chamberlain Marinelli assassinates Appiani
- Orsina was the prince's mistress and plots to kill him
- Odoardo kills Emilia at her wish
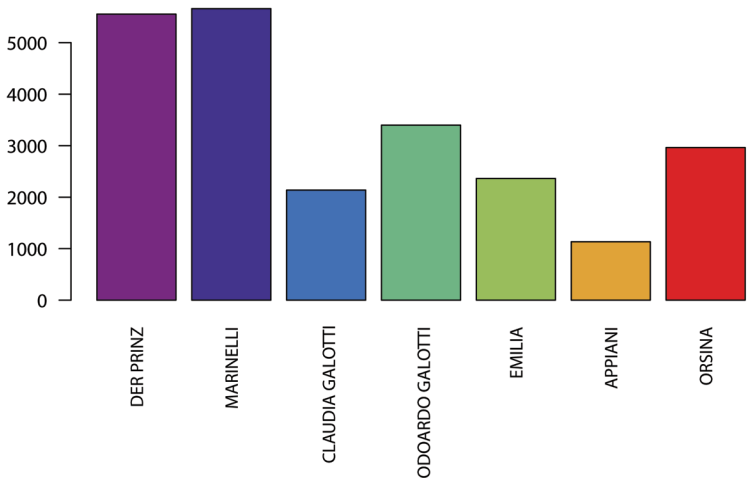- The prince blames Marinelli for her death

Tokens

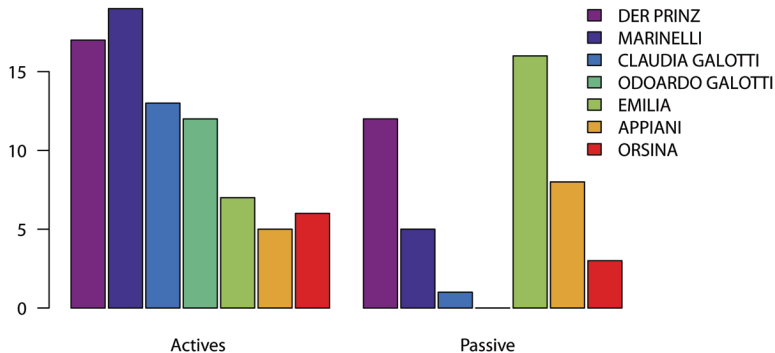Figure 4: Top seven figures with the highest token number in *Emilia Galotti*.

Presence

Figure 5: Active and passive presence in *Emilia Galotti*. A figure is only passively present in a scene if they are not actively present.

Shapley Graphs
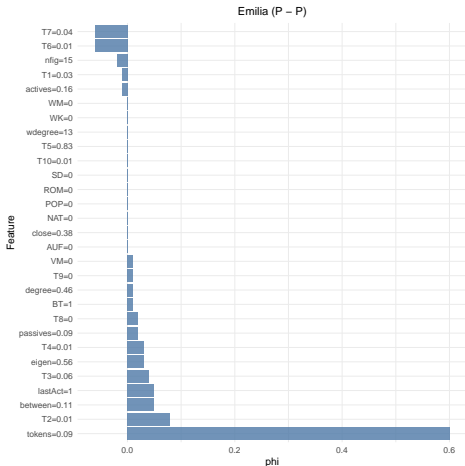
What has the model learnt about these characters?

Figure 6: Shapley graph for single figures in *Emilia Galotti*. Brackets mean:
(Actual class – Predicted class).

Shapley Graphs

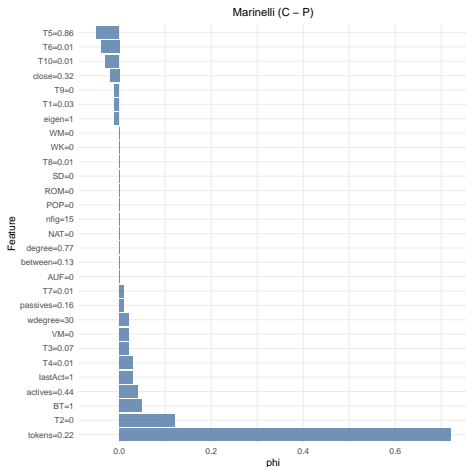What has the model learnt about these characters?

Figure 7: Shapley graph for single figures in *Emilia Galotti*. Brackets mean: (Actual class – Predicted class).

Shapley Graphs

What has the model learnt about these characters?

Figure 8: Shapley graph for single figures in *Emilia Galotti*. Brackets mean:
(Actual class – Predicted class).

Shapley Graphs

What has the model learnt about these characters?
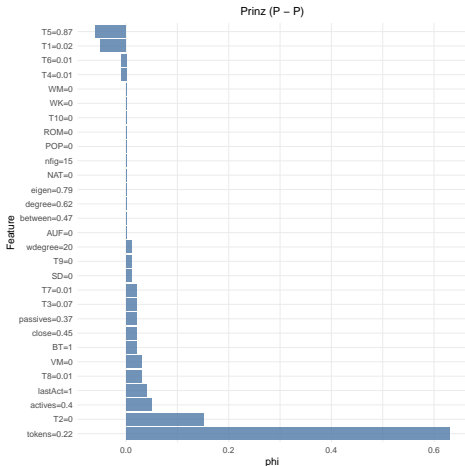

Figure 9: Shapley graph for single figures in *Emilia Galotti*. Brackets mean: (Actual class – Predicted class).

Shapley Graphs

What has the model learnt about these characters?

Figure 10: Shapley graph for single figures in *Emilia Galotti*. Brackets mean: (Actual class – Predicted class).

Shapley Graphs

What has the model learnt about these characters?
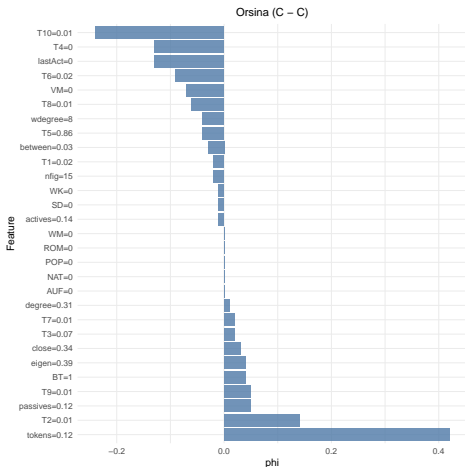
Figure 11: Shapley graph for single figures in *Emilia Galotti*. Brackets mean: (Actual class – Predicted class).

Shapley Graphs

What has the model learnt about these characters?
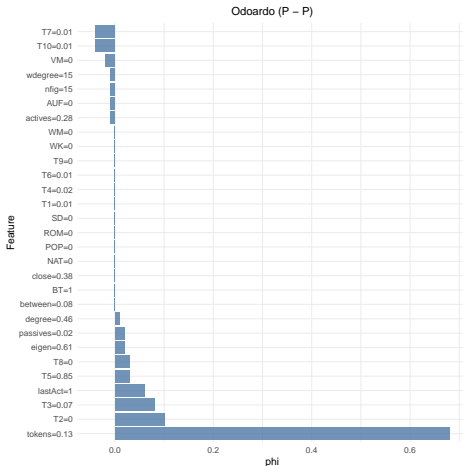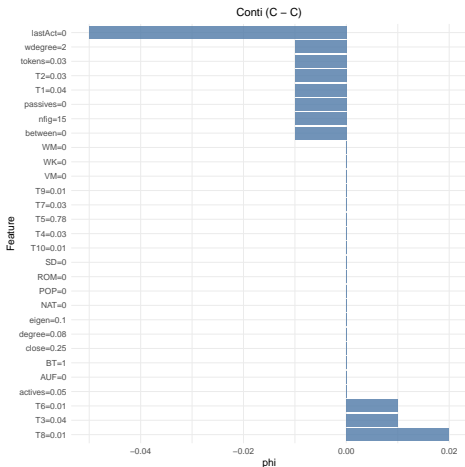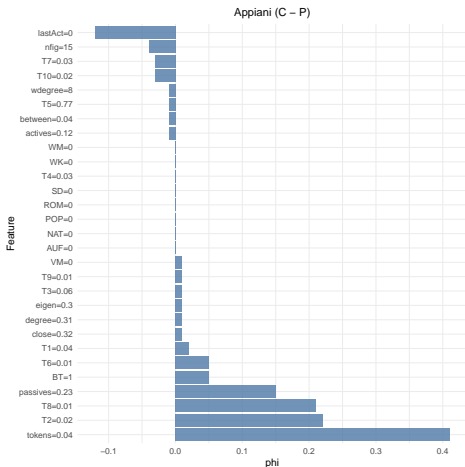

Figure 12: Shapley graph for single figures in *Emilia Galotti*. Brackets mean:
(Actual class – Predicted class).

Take-away

- High performance for protagonist classification

Take-away

- High performance for protagonist classification
- Tendency to produce False Positives

Take-away

- High performance for protagonist classification
- Tendency to produce False Positives
- Tokens feature is strong but not sufficient

Take-away

- High performance for protagonist classification
- Tendency to produce False Positives
- Tokens feature is strong but not sufficient
- Analysis of single characters yields interesting insides

Appendix

5

Results

Experiment 2

	Precision	Recall	F1	Accuracy
Majority BL	-	0.00	-	**0.97**
Tokens BL	0.38	**1.00**	0.55	0.95
Random Forest	**0.46**	**1.00**	**0.63**	0.96

Table 2: Results for classification of title characters plus baselines.

Results

Experiment 3

	Precision	Recall	F1	Accuracy
A1woTokens	0.82	0.98	0.89	0.96
A2woTokens	0.78	1.00	0.88	0.96
A3woTokens	0.51	1.00	0.67	0.93
TFwoTokens	0.37	1.00	0.54	0.95

Table 3: Results without using *tokens* feature.
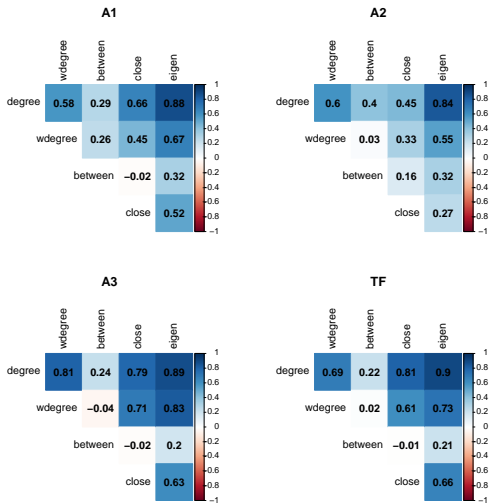
Centrality Correlation

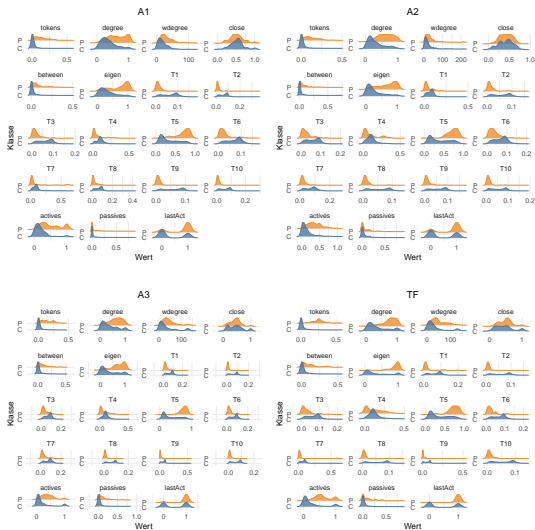Figure 13: Correlation for centrality features.

Feature Distribution

Figure 14: Feature distribution.

Feature Importance
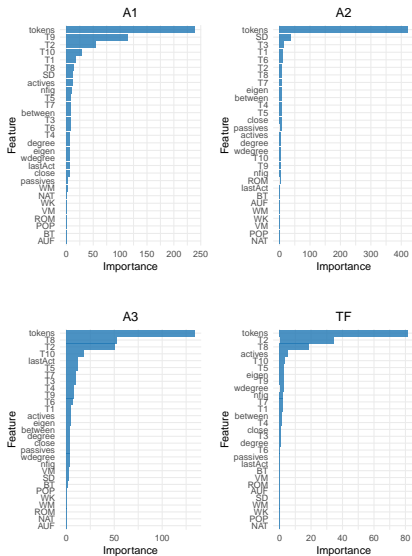
Figure 15: Relative Feature Importance.

Confusion Matrix

		Ref	
		C	**P**
Pred	**C**	878	0
	P	32	171

Table 4: Confusion matrix for A1.

		Ref	
		C	**P**
Pred	**C**	1196	0
	P	100	106

Table 6: Confusion matrix for A3.

		Ref	
		C	**P**
Pred	**C**	883	0
	P	45	176

Table 5: Confusion matrix for A2.

		Ref	
		C	**P**
Pred	**C**	1456	0
	P	57	49

Table 7: Confusion matrix for TF.

Annotation

- Three annotators with overlapping and unique dramas

Annotator	# Dramas	# Protag-onists (%)	# Non-Protagonists (%)	# Figures Total
A1	34	171 (16)	910 (84)	1081
A2	37	176 (16)	928 (84)	1104
A3	36	106 (8)	1296 (92)	1402

Table 8: Distribution of annotations.

Annotation

- Three annotators with overlapping and unique dramas

Annotator	# Dramas	# Protag- onists (%)	# Non- Protagonists (%)	# Figures Total
A1	34	171 (16)	910 (84)	1081
A2	37	176 (16)	928 (84)	1104
A3	36	106 (8)	1296 (92)	1402

Table 8: Distribution of annotations.

Combination	# Dramas	Cohen's κ
A1+A2	6	0.83
A1+A3	6	0.46
A2+A3	7	0.43

Table 9: Cohen's κ for different combinations of annotations.

Results

Experiment 1

	Data	Precision	Recall	F1	Accuracy
Majority Baseline	A1	-	0.00	-	0.84
	A2	-	0.00	-	0.84
	A3	-	0.00	-	0.92
Tokens Baseline	A1	0.72	1.00	0.84	0.94
	A2	0.70	0.99	0.82	0.93
	A3	0.44	1.00	0.61	0.91
Random Forest	A1	0.84	1.00	0.91	0.97
	A2	0.80	1.00	0.89	0.96
	A3	0.51	1.00	0.68	0.93

Table 10: Precision, Recall and F1 for classifying protagonists and accuracy.