



University of Stuttgart  
Germany

# QuaDrama

CRETA   
CENTER FOR REFLECTED TEXT ANALYTICS

## Klassifikation von Titelfiguren in deutschsprachigen Dramen

Evaluation am Beispiel von  
Lessings *Emilia Galotti*

Janis Pagel,  
Benjamin Krautter



Volkswagen **Stiftung**

„Man spricht vom ‚Helden‘ eines literarischen Werkes, wenn man die jeweilige Hauptfigur, den Protagonisten bezeichnen will und passt sich damit der eingebürgerten Sprachkonventionen an, die einen neutralen, wertfreien Gebrauch des Wortes erlaubt“

Plett, Bettina: *Problematische Naturen? Held und Heroismus im realistischen Erzählen*. Paderborn 2002, S. 21.

# Helden, Protagonisten und Hauptfiguren?

## Figuren im Drama

- **Manfred Pfister** (1977): „Quantitative Dominanzrelationen“
- Zwei Kriterien, die zwischen Haupt- und Nebenfiguren im Drama unterscheiden sollen:
  - (1) Dauer der Bühnenpräsenz einer Figur
  - (2) Anteil der Figurenrede am Haupttext

# Titelfiguren als Zielkategorie

## Figuren im Drama

- Beispiele für titelgebende Figuren:

Schillers *Maria Stuart*, Kleists *Penthesilea*, Goethes *Faust*, ...

- Vorteile:

(1) Definitionsproblem (insb. HeldIn und ProtagonistIn) erübrigt sich

(2) kein aufwändiges Annotationsverfahren für Goldstandard nötig

(3) *a priori* Annahme: Titelfiguren verkörpern und repräsentieren

Eigenschaften im Text, die sie in vielen Fällen auch zu Hauptfiguren der Handlung machen

Titelfiguren im Drama

# Experimente

# Versuchsaufbau

## 2 Experimente

- Ursprüngliches Korpus (Alt):
  - 38 Dramen
- Aktuelles Korpus (Neu):
  - 75 Dramen

Korpus	#Dramen	#Titelfiguren	#Figuren Total
Alt	38	42 (3%)	1208
Neu	75	81 (4%)	2211

# Versuchsaufbau

## Klassifikation

- Datenpunkte: Figuren, repräsentiert als Feature-Vektoren
- Features:
  - Tokens
  - Aktive und Passive Präsenz
  - Topic Models
  - Netzwerkmaße (Degree, WDegree, Betweenness, Closeness, Eigenvector)
  - Metadaten (Anzahl der Figuren, Gattung/Epoche etc.)
- 2 Klassen: Titelfigur (TF) oder Nicht-Titelfigur (C)
- Random Forest + 2 Baselines
- Sampling: SMOTE; 10-Fold-Cross Validation

# Features

## Passive Präsenz

- Aktive Präsenz: (normalisierte) Anzahl an Szenen, in der eine Figur auftritt
- Passive Präsenz: (normalisierte) Anzahl an Szenen, in der eine Figur von einer anderen Figur erwähnt wird, ohne selber auf der Bühne präsent zu sein
- Da nicht für alle Stücke eine Koreferenzauflösung vorliegt wird eine Szene gezählt, falls die Figur namentlich erwähnt wurde



# Ergebnisse

## 2 Experimente

Alt	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL	-	0,00	-	0,97	1,00	0,98	0,00
Tokens BL	0,34	1,00	0,51	1,00	0,93	0,96	0,56
Alle Feat.	0,45	1,00	0,62	1,00	0,96	0,98	0,66

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL	-	0,00	-	0,96	1,00	0,98	0,00
Tokens BL	0,35	1,00	0,52	1,00	0,93	0,96	0,57
Alle Feat.	0,48	1,00	0,65	1,00	0,96	0,98	0,68

# Ergebnisse

## 2 Experimente

Alt	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							0,00
Tokens BL							0,56
Alle Feat.							0,66

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							0,00
Tokens BL							0,57
Alle Feat.							0,68

# Ergebnisse

## 2 Experimente

Alt	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL				0,97	1,00	0,98	
Tokens BL				1,00	0,93	0,96	
Alle Feat.				1,00	0,96	0,98	

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL				0,96	1,00	0,98	
Tokens BL				1,00	0,93	0,96	
Alle Feat.				1,00	0,96	0,98	

# Ergebnisse

## 2 Experimente

Alt	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							
Tokens BL		1,00					
Alle Feat.		1,00					

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							
Tokens BL		1,00					
Alle Feat.		1,00					

# Ergebnisse

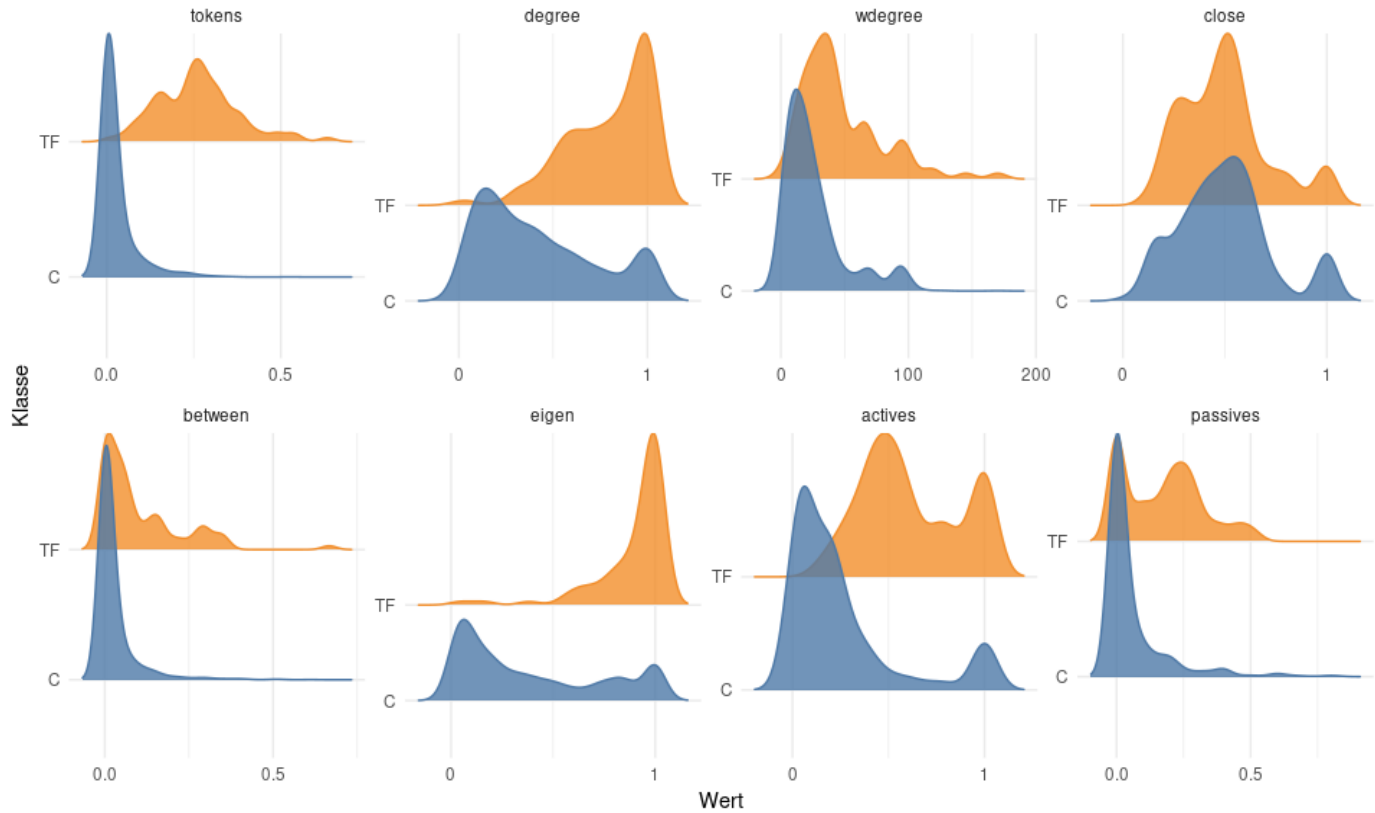
## 2 Experimente

Alt	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							
Tokens BL	0,34						
Alle Feat.	0,45						

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL							
Tokens BL	0,35						
Alle Feat.	0,48						

# Ergebnisse

## Feature-Distribution

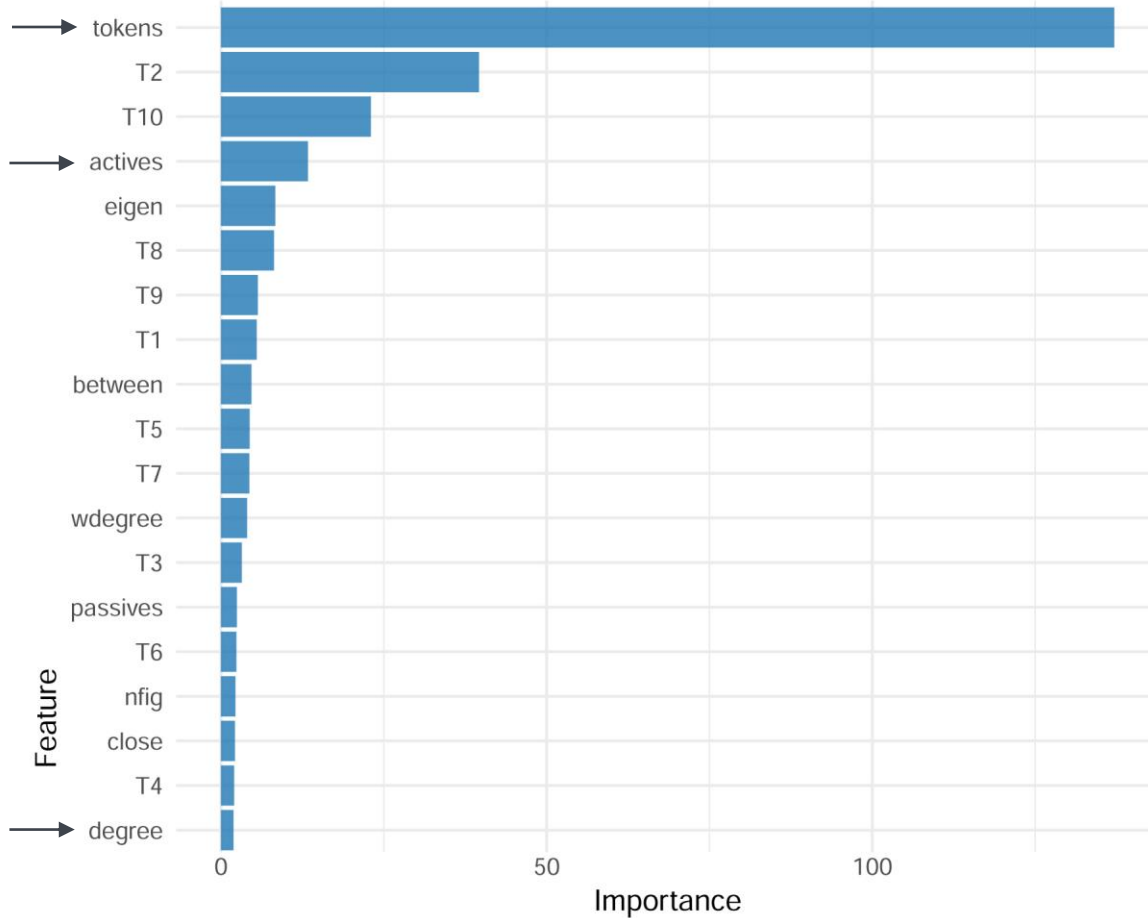


Titelfiguren im Drama

# Evaluation der Ergebnisse

# Evaluation der Ergebnisse

## Feature Importance

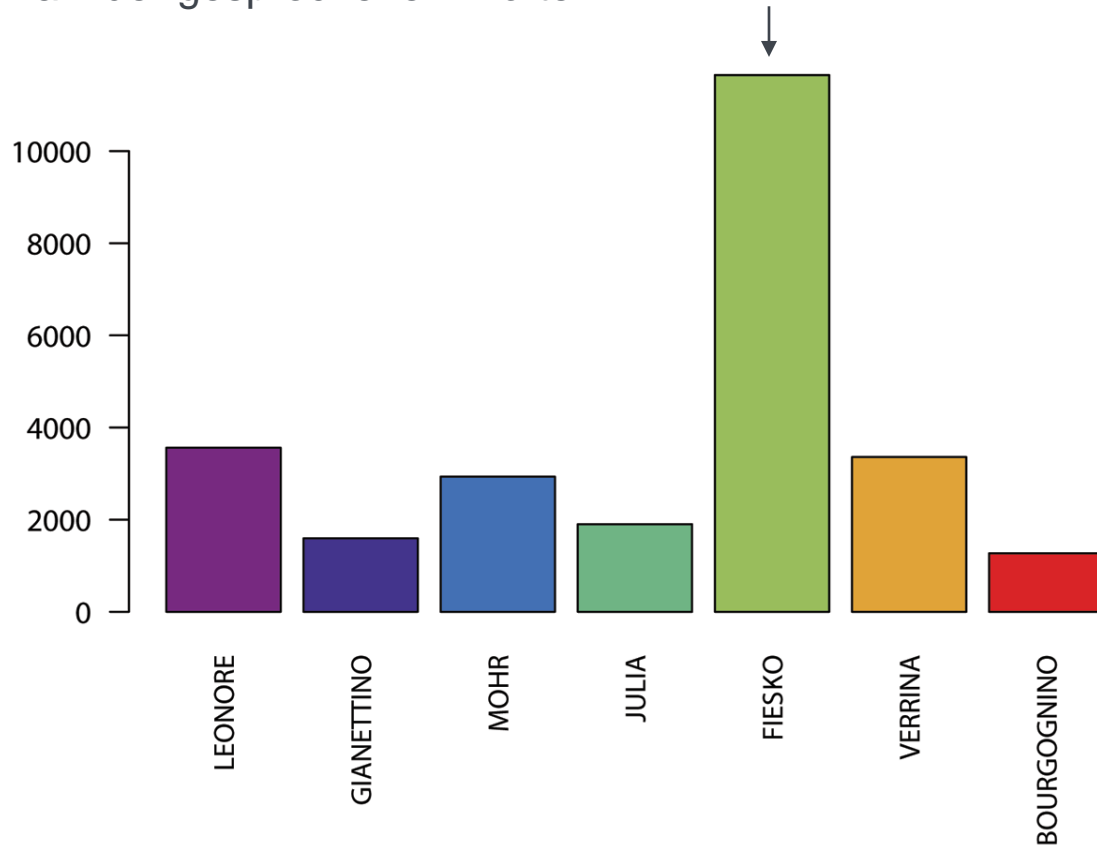




# Evaluation der Ergebnisse

Schillers *Die Verschwörung des Fiesko zu Genua* (1783)

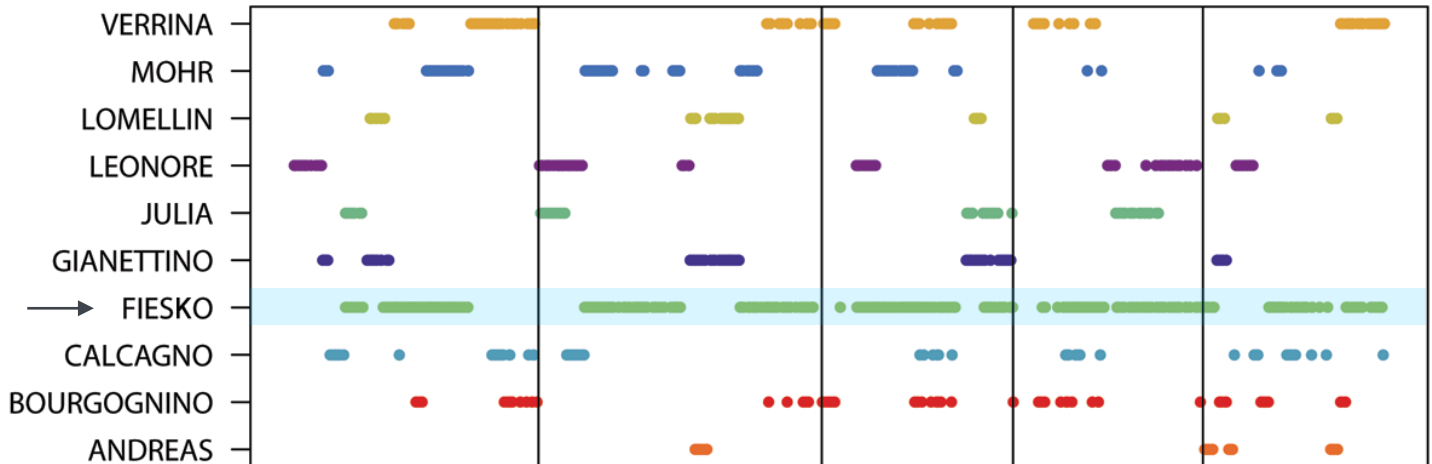
- Zahl der gesprochenen Wörter



# Evaluation der Ergebnisse

Schillers *Die Verschwörung des Fiesko zu Genua* (1783)

- Kopräsenz



# Evaluation der Ergebnisse

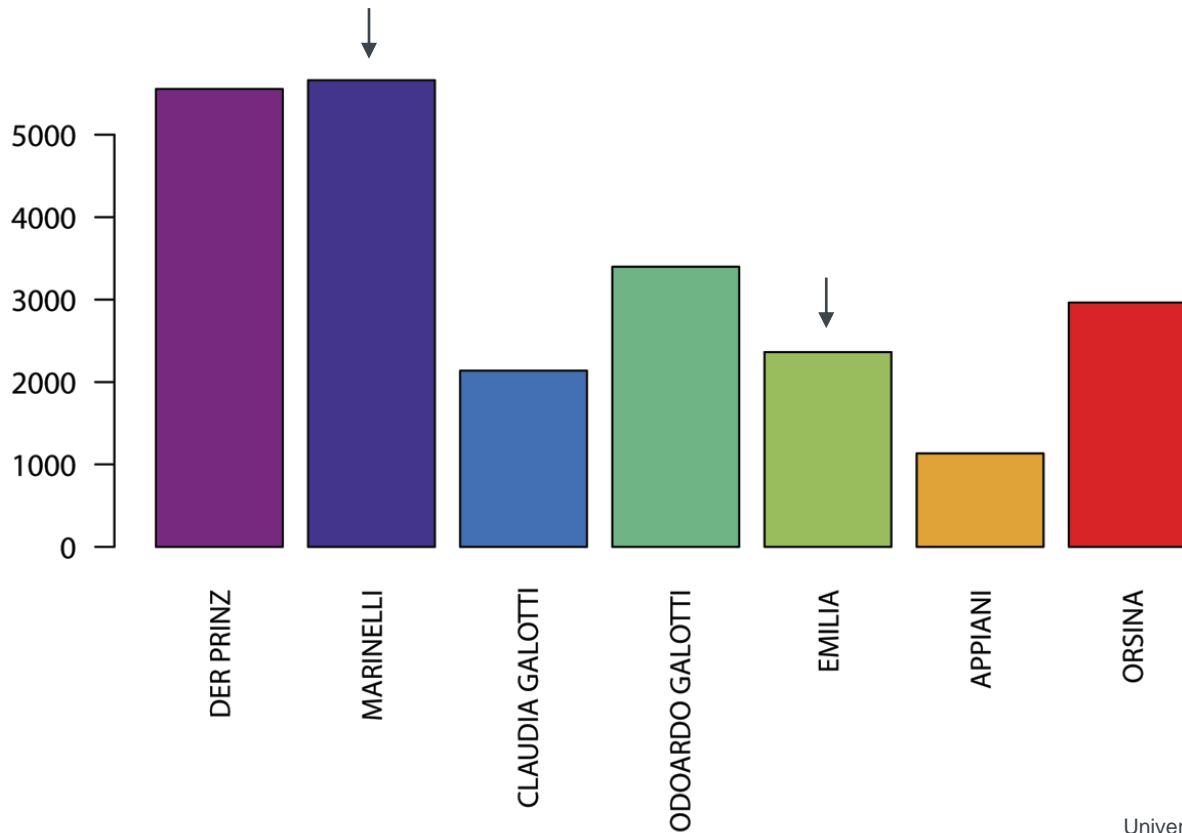
Lessings *Emilia Galotti* (1772)

- Durch Korpuserweiterung verändert sich das Klassifikationsergebnis:
- (1) ursprüngl. Korpus (38 Dramen): 4 Figuren als titelgebend ausgezeichnet
  - Emilia Galotti
  - Odoardo Galotti (Emilias Vater)
  - Prinz Hettore Gonzaga
  - Marinelli (Kammerdiener des Prinzen)
- (2) aktuelles Korpus (75 Dramen): 2 Figuren als titelgebend ausgezeichnet
  - Emilia Galotti
  - Prinz Hettore Gonzaga

# Evaluation der Ergebnisse

Lessings *Emilia Galotti* (1772)

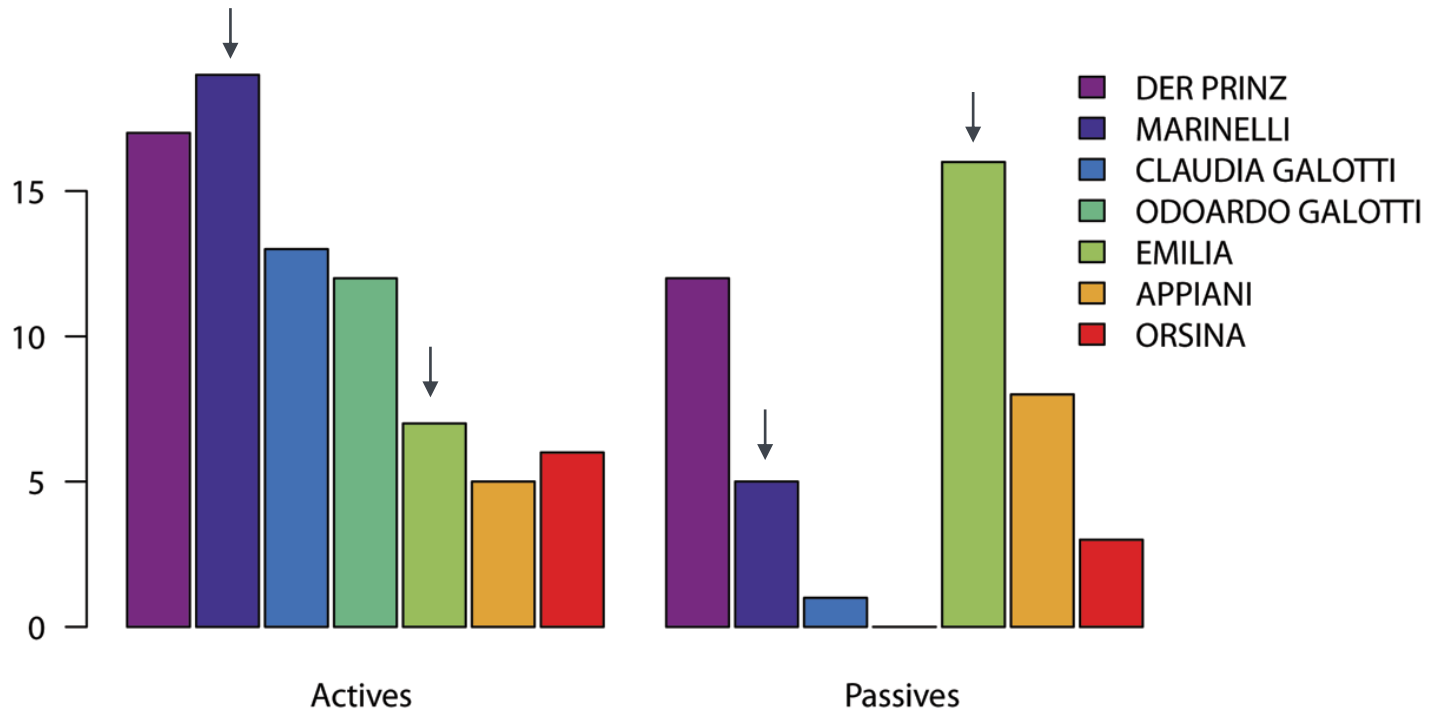
- Zahl der gesprochenen Wörter



# Evaluation der Ergebnisse

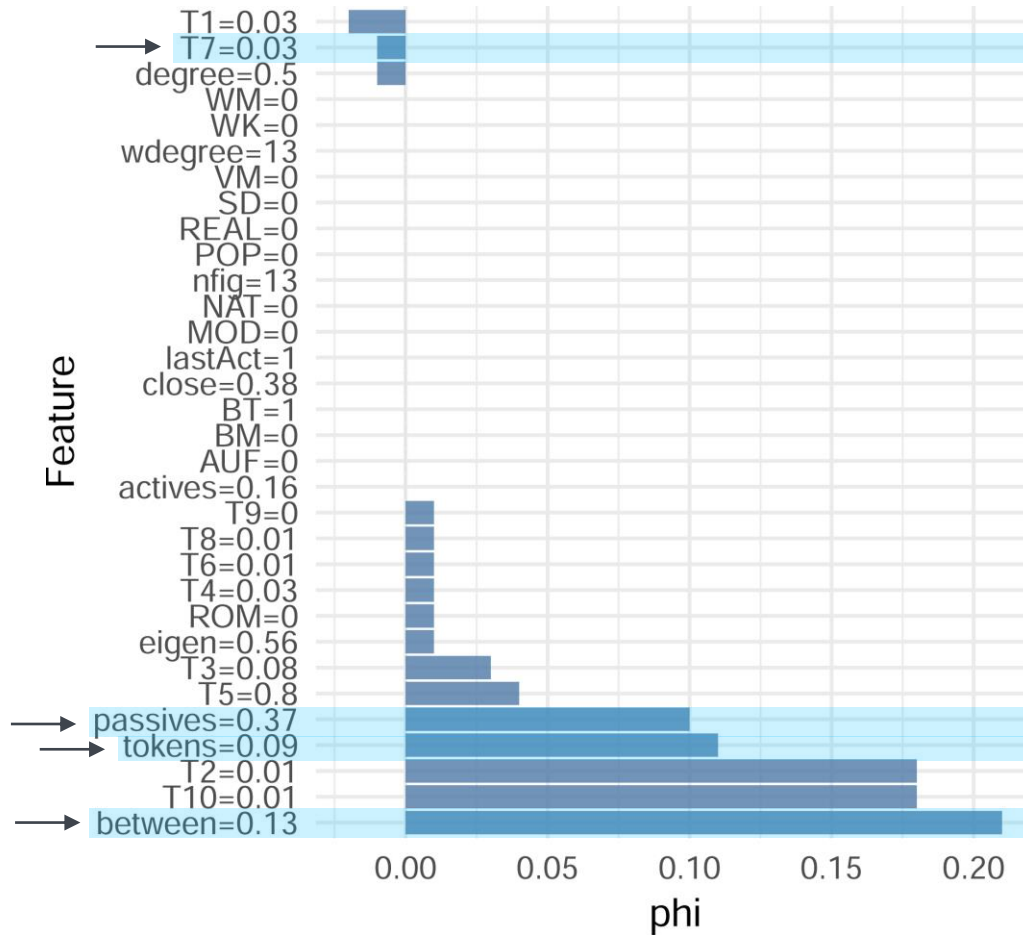
Lessings *Emilia Galotti* (1772)

- Aktive und passive Präsenz



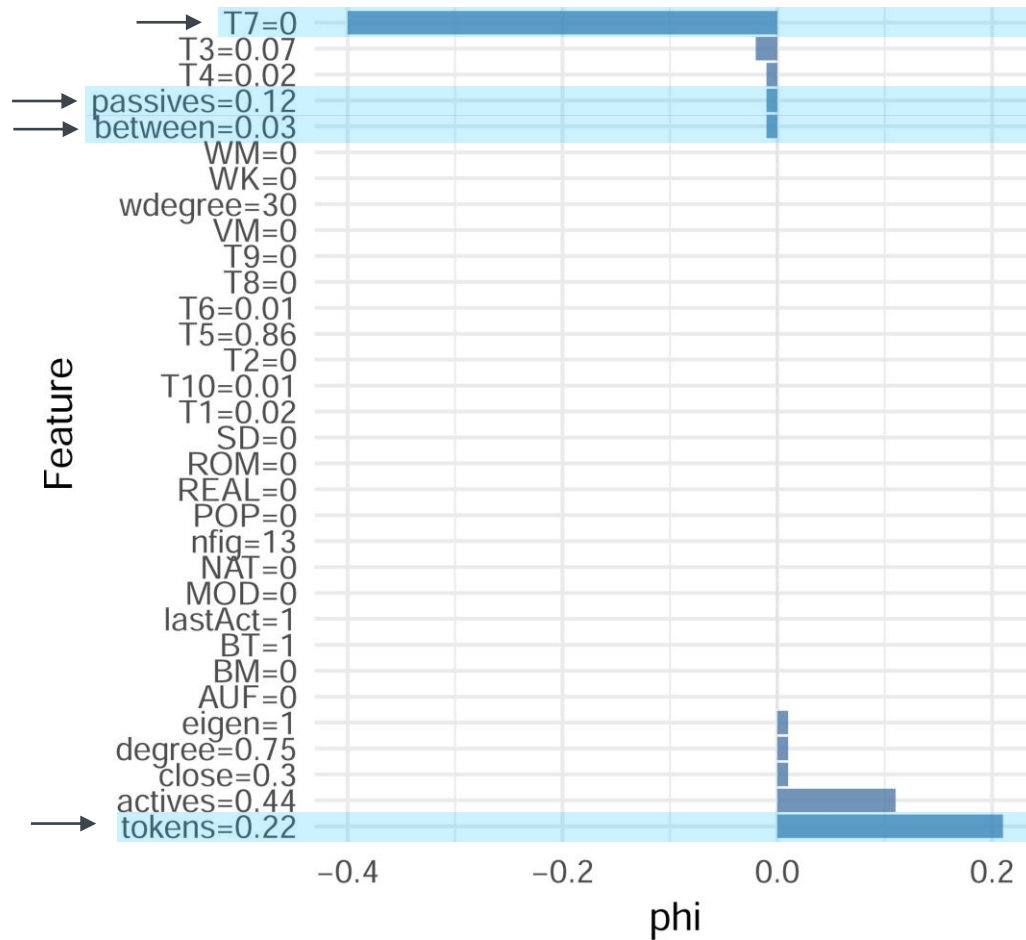
# Featureanalyse

Emilia



# Featureanalyse

Marinelli



Figurenklassifikation im Drama

# **Kurzes Fazit**



# Figurenklassifikation im Drama

## Abschlussbetrachtung

- Machine Learning erzielt gute Ergebnisse für die Klassifikation von Titelfiguren
- Titelfiguren als Zielkategorie:
  - vermeidet Probleme
  - aber: welche Erkenntnis? Evaluation von weiteren Dramen nötig!
- Einfluss des Topic Modeling auf Klassifikation muss nachvollziehbarer werden

**Vielen Dank für die  
Aufmerksamkeit**

<https://quadrama.github.io/>

**Anhang**

# Ergebnisse

Neu	P (TF)	R (TF)	F1 (TF)	P (C)	R (C)	F1 (C)	MCC
Maj. BL	-	0,00	-	0,96	1,00	0,98	0,00
Tokens BL	0,35	1,00	0,52	1,00	0,93	0,96	0,57
Ohne Tokens	0,34	1,00	0,50	1,00	0,93	0,96	0,56
Ohne Epochen	0,46	1,00	0,63	1,00	0,96	0,98	0,67
Alle Feat.	0,48	1,00	0,65	1,00	0,96	0,98	0,68

# Ergebnisse

## Feature-Distribution

