

# Tutorial: Quantitative Dramenanalyse

Ruhr-Universität Bochum

Digitale Analyse großer Textkorpora

17. Dezember 2021

<https://quadrama.github.io>

# QuaDramA: Quantitative Drama Analytics

## QuaDramA



- Test hypotheses about dramatic characters quantitatively
- Three dimensions
  - Character types
  - Relations between characters
  - Development of character types

## QuaDramA: Tracking Character Knowledge (Q:TRACK)

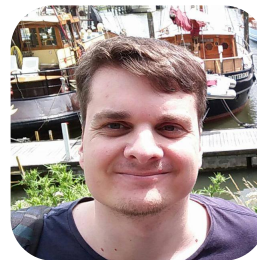
- How is character knowledge transported during the drama?
- When and how do characters learn about important facts?



Computational  
LITERARY STUDIES

# QuaDramA

- 2 Doktoranden: Janis Pagel (Computerlinguistik),  
Benjamin Krautter (Literaturwissenschaft),  
Universität zu Köln



- 1 Postdoc: Melanie Andresen (Korpuslinguistik),  
Universität Stuttgart



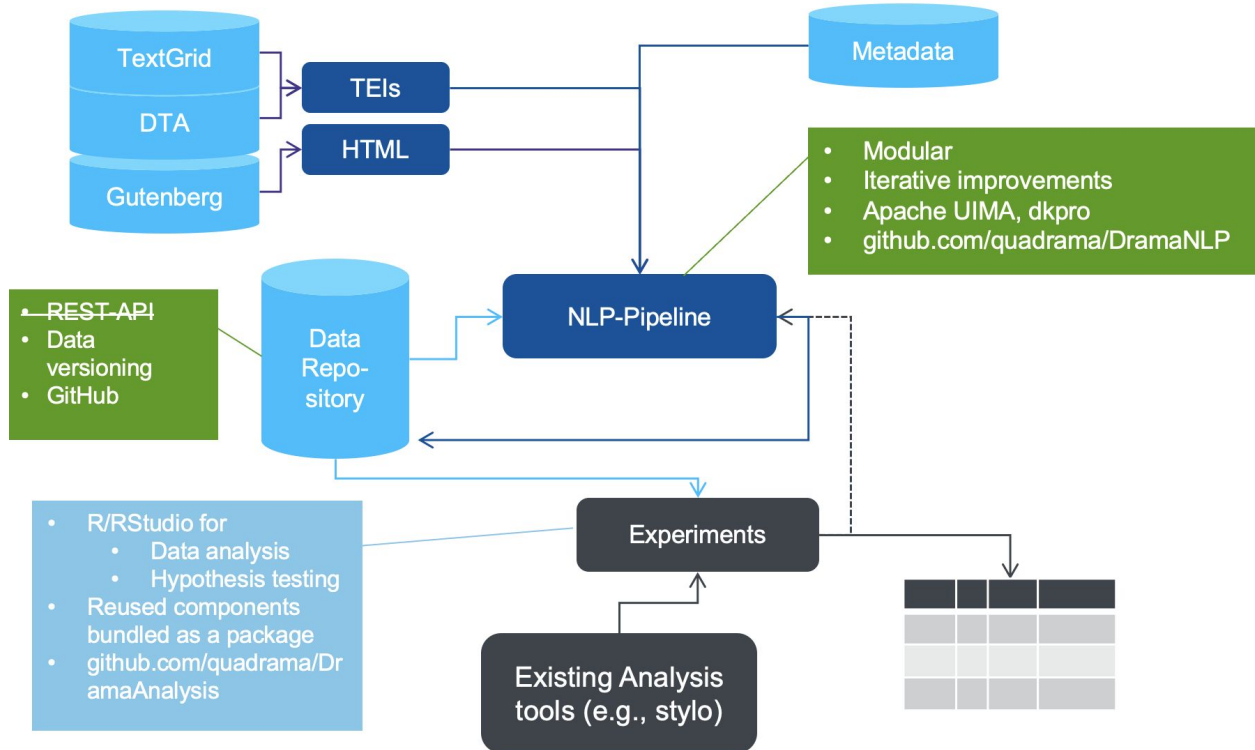
- 1 Projektleiter: Nils Reiter (Digital Humanities/Computerlinguistik),  
Universität zu Köln

- Ex-Projektleiter: Marcus Willand

Offene Promotionsstelle in Köln  
(65% TV-L 13, ab 01.03.2021)



# Technische Projektarchitektur



# Challenges

- Finding common ground between literary studies and computational linguistics
  - Different research methods, goals and ways to speak about things
  - Conscious compromises at the acceptable for everyone
- Processing structured texts
  - NLP assumption: Text is an unstructured stream of tokens
  - Not realistic for many use cases



# Outline

## Teil 1: Allgemeine Einleitung

- a) Projektvorstellung ✓
- b) Operationalisierung als zentrale Herausforderung

## Teil 2: Erster Use-Case: Bühnenpräsenz und Netzwerkanalyse

- a) Inhaltliche Einführung und mögliche Fragestellungen
- b) R-Einführung für Pythonistas
- c) Hands-On-Session in breakout-Rooms

## Pause

## Teil 3: Zweiter Use-Case: Passive Präsenz

- a) Inhaltliche Einführung
- b) R-Einführung in Collections
- c) Hands-On-Session in breakout-Rooms

## Teil 4: Abschlussrunde

- a) Offene Fragen und Diskussion
- b) Take-Home-Messages und Ausblick

# Operationalisierung als zentrale Herausforderung in den Digital Humanities



# Operationalisierung

## Frage: Wie definiert man einen theoretischen Begriff?

P. W. Bridgman: *Logic of Modern Physics* (1927)

“In general, we mean by any concept nothing more than a set of operations; the concept is synonymous with the corresponding set of operations [...] the proper definition of a concept is not in terms of its properties but in terms of actual operations.”  
(S. 5)

Beispiel: Länge

“The concept of length is therefore fixed when the operations by which length is fixed are fixed: that is, the concept of length involves as much and nothing more than the set of operations by which length is determined.” (S. 6)



# Operationalisierung

## Übertragung auf geisteswissenschaftliche Begriffe?

Franco Moretti: *“Operationalizing”: or, the function of measurement in modern literary theory* (2013)

“[...] it describes the process whereby concepts are transformed into a series of operations—which, in their turn, allow to measure all sorts of objects. **Operationalizing means building a bridge from concepts to measurement, and then to the world. In our case: from the concepts of literary theory, through some form of quantification, to literary texts.**” (S. 1)

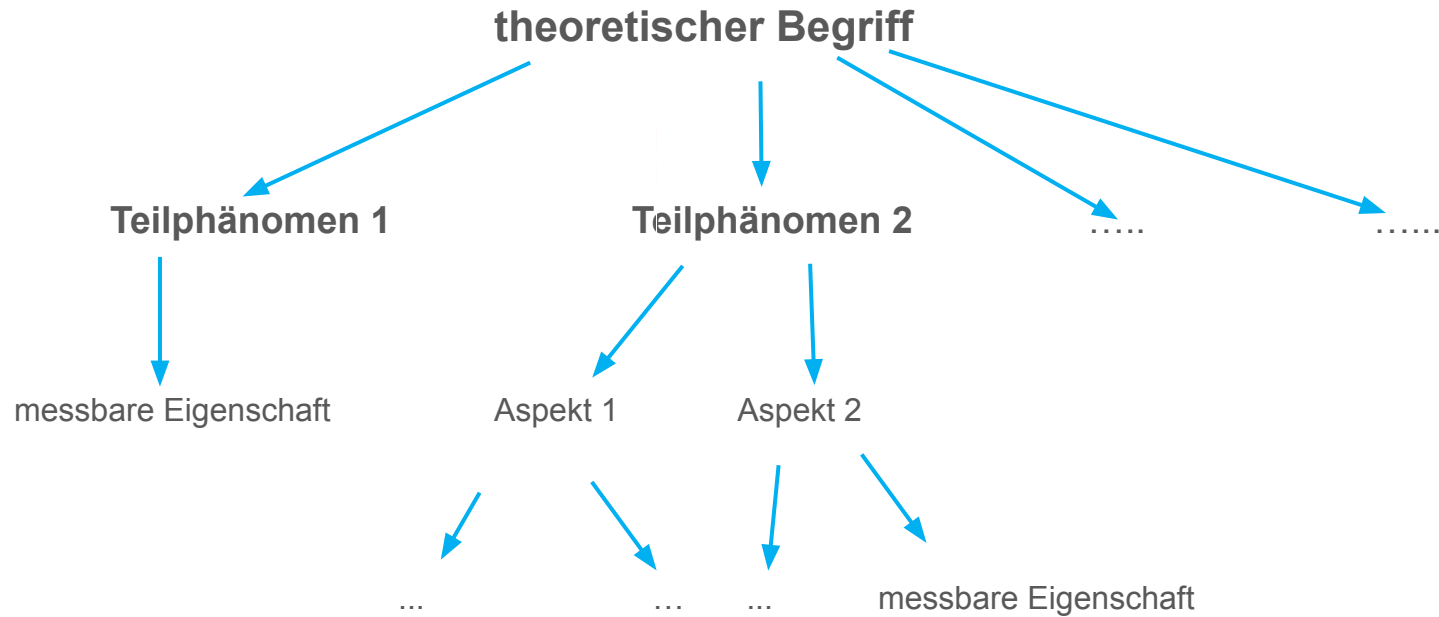
# Operationalisierung

Axel Pichler, Nils Reiter: *Zur Operationalisierung theoretischer Begriffe in der algorithmischen Textanalyse* (2021)

Operationalisierung als “Entwicklung von Verfahren, die einen Begriff über potentiell mehrere Teilschritte oder -begriffe explizit und regelgeleitet auf Textoberflächenphänomene zurückführen.” (S. 4)

Mensch	manuelle Annotation: Annotationsguideline als Regelset, das Begriffe und Textoberflächenphänomene zusammenführt
Computer	computergestützte Operationalisierung: etwa maschinelle Lernverfahren, die auf Basis annotierter Daten trainiert werden

# Operationalisierung geisteswissenschaftlicher Begriffe



# Operationalisierung

## **Beispiel: Protagonisten im Drama**

Was könnten (messbare) Teilphänomene sein, die zusammengekommen Protagonist:innen eines Dramas definieren?

- Zahl der gesprochenen Wörter
- Länge der Redebeiträge?
- Bühnenpräsenz: Wie häufig steht eine Figur auf der Bühne?
- Wie häufig wird über eine Figur gesprochen, die nicht auf der Bühne steht?
- Figurenrelationen: Mit wie vielen anderen Figuren steht die Figur gemeinsam auf der Bühne?
- Über welche Themen spricht die Figur?
- ...

# Operationalisierung

## Beispiel: Protagonisten im Drama

Was könnten (messbare) Teilphänomene sein, die zusammengenommen Protagonisten definieren?

- Zahl der gesprochenen Wörter
- Länge der Redebeiträge?
- **Bühnenpräsenz: Wie häufig steht eine Figur auf der Bühne?**
- **Figurenrelationen: Mit wie vielen anderen Figuren steht die Figur gemeinsam auf der Bühne?**
- **Wie häufig wird über eine Figur gesprochen, die nicht auf der Bühne steht?**
- Über welche Themen spricht die Figur?
- ...

# Aktive Bühnenpräsenz und Netzwerkanalyse

Trilcke u.a. (2015)



# Warum quantitative Dramenanalyse?

## Struktur von Dramen:

Beispiel: Lessings *Emilia Galotti* (1772)

Segmentierung durch Akt- und  
Szenenangaben (Aufzüge und Auftritte)

**Vierter Auftritt**

*Der Prinz. Conti, mit den Gemälden, wovon er das eine verwandt gegen  
einen Stuhl lehnet.*

Markierung von Haupt- und  
Nebentext (Regieanweisungen)

CONTI *indem er das andere zurecht stellet.* Ich bitte, Prinz, daß Sie die  
Schranken unserer Kunst erwägen wollen. Vieles von dem Anzüglich-  
sten der Schönheit liegt ganz außer den Grenzen derselben. – Treten  
Sie so! –

Die jeweils sprechenden Figuren sind  
markiert

DER PRINZ *nach einer kurzen Betrachtung.* Vortrefflich, Conti; – ganz  
vortrefflich! – Das gilt Ihrer Kunst, Ihrem Pinsel. – Aber geschmeichelt,  
Conti; ganz unendlich geschmeichelt!

CONTI. Das Original schien dieser Meinung nicht zu sein. Auch ist es  
in der Tat nicht mehr geschmeichelt, als die Kunst schmeicheln muß.  
Die Kunst muß malen, wie sich die plastische Natur, – wenn es eine  
gibt – das Bild dachte: ohne den Abfall, welchen der widerstrebende  
Stoff unvermeidlich macht; ohne das Verderb, mit welchem die Zeit  
dagegen an kämpfet.

# Warum quantitative Dramenanalyse?

```
<div type="scene">
```

```
<head>Vierter Auftritt</head>
```

```
<stage>Der Prinz. Conti, mit den Gemälden, wovon er  
das eine verwandt gegen einen Stuhl lehnet.</stage>
```

```
<sp who="#conti">
```

```
<speaker>CONTI</speaker>
```

```
<stage>indem er das andere zurecht stellt.</stage>
```

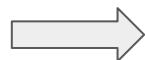
```
<p>
```

Ich bitte, Prinz, daß Sie die Schranken unserer Kunst erwägen wollen. Vieles von dem Anzüglichsten der Schönheit liegt ganz außer den Grenzen derselben. – Treten Sie so! –

```
</p>
```

```
</sp>
```

[...]



Fischer u.a.: *German Drama Corpus*  
<https://dracor.org>

## Vierter Auftritt

*Der Prinz. Conti, mit den Gemälden, wovon er das eine verwandt gegen einen Stuhl lehnet.*

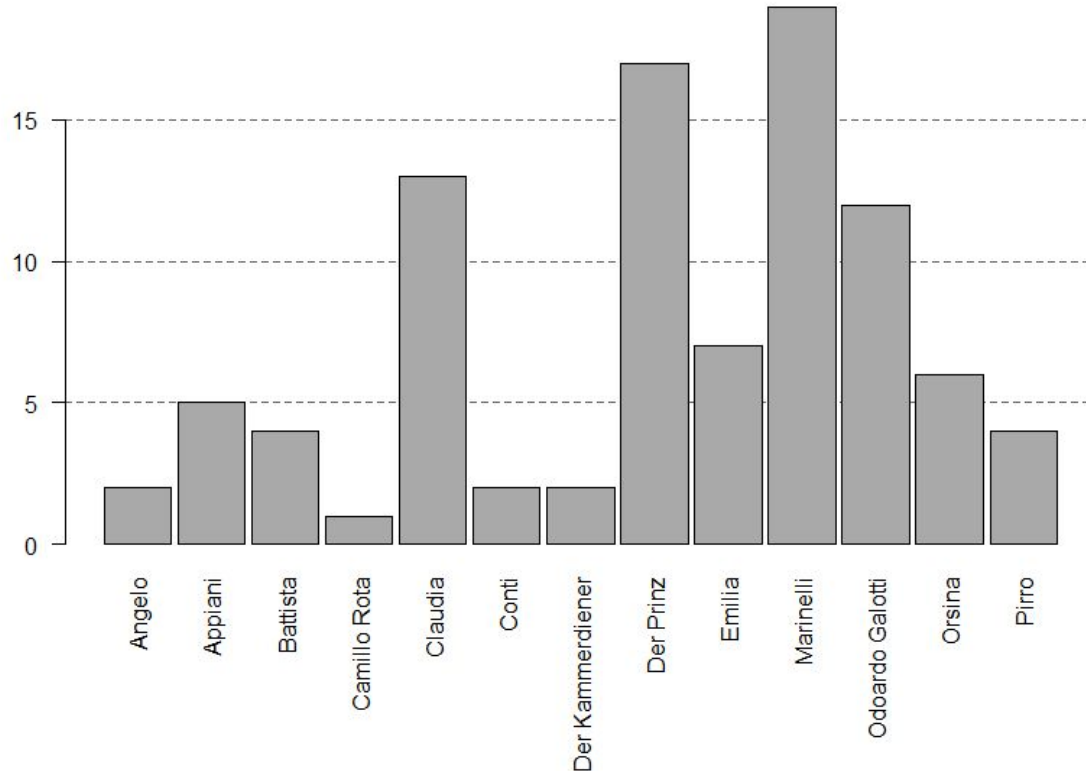
CONTI *indem er das andere zurecht stellt.* Ich bitte, Prinz, daß Sie die Schranken unserer Kunst erwägen wollen. Vieles von dem Anzüglichsten der Schönheit liegt ganz außer den Grenzen derselben. – Treten Sie so! –

DER PRINZ *nach einer kurzen Betrachtung.* Vortrefflich, Conti; – ganz vortrefflich! – Das gilt Ihrer Kunst, Ihrem Pinsel. – Aber geschmeichelt, Conti; ganz unendlich geschmeichelt!

CONTI. Das Original schien dieser Meinung nicht zu sein. Auch ist es in der Tat nicht mehr geschmeichelt, als die Kunst schmeicheln muß. Die Kunst muß malen, wie sich die plastische Natur, – wenn es eine gibt – das Bild dachte: ohne den Abfall, welchen der widerstrebende Stoff unvermeidlich macht; ohne das Verderb, mit welchem die Zeit dagegen an kämpfet.



# Aktive Bühnenpräsenz



Manfred Pfister: *Das Drama* (2001),  
S. 226f.

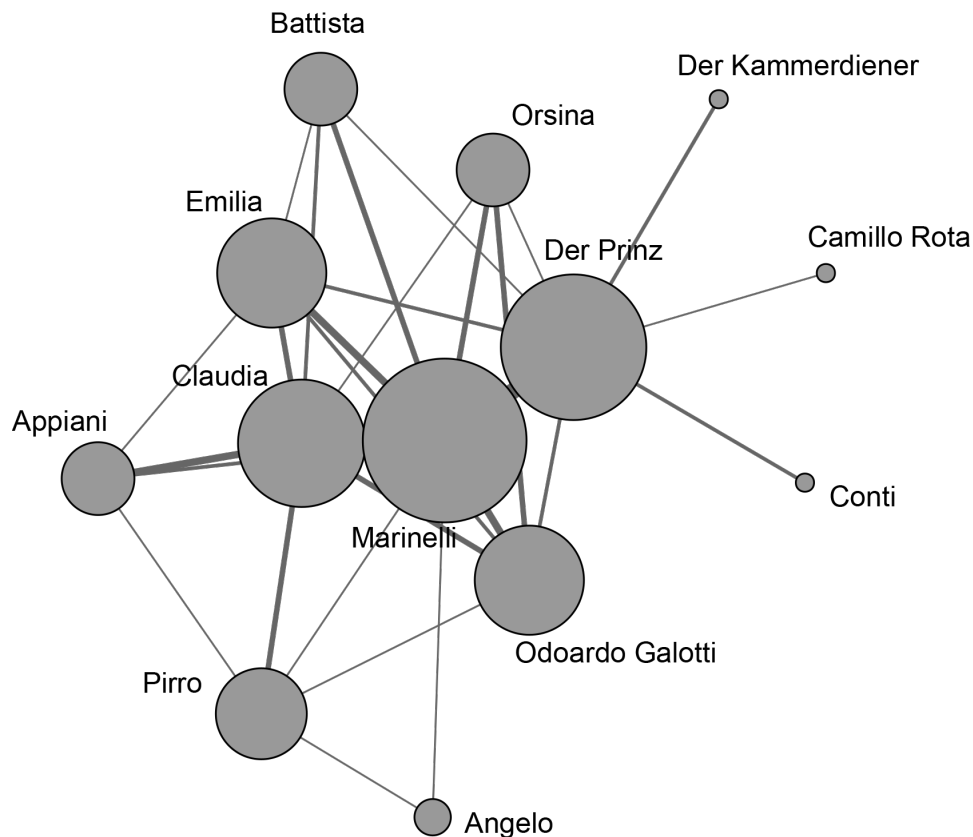
“Rein quantitativ erfaßbar sind auch die Dominanzrelationen innerhalb des Personals nach den Kriterien der **Dauer der Bühnenpräsenz** einer Figur und ihres **Anteils am Haupttext**. Nach beiden Kriterien ergeben sich Abstufungen zwischen den Figuren des Personals, die sich jedoch nicht unbedingt zu decken brauchen.”

# Figurenkonfiguration

	1	2	3	4	5	6	7	8	9	10
Dirne	1	0	0	0	0	0	0	0	0	1
Soldat	1	1	0	0	0	0	0	0	0	0
Stubenmädchen	0	1	1	0	0	0	0	0	0	0
Junger Herr	0	0	1	1	0	0	0	0	0	0
Junge Frau	0	0	0	1	1	0	0	0	0	0
Ehemann	0	0	0	0	1	1	0	0	0	0
Süßes Mädel	0	0	0	0	0	1	1	0	0	0
Dichter	0	0	0	0	0	0	1	1	0	0
Schauspielerin	0	0	0	0	0	0	0	1	1	0
Graf	0	0	0	0	0	0	0	0	1	1

Arthur Schnitzler: Der Reigen (1900), aus Manfred Pfister: Das Drama (2001), S. 239.

# Figurennetzwerk: Emilia Galotti



## Konsituierende Elemente eines Netzwerks:

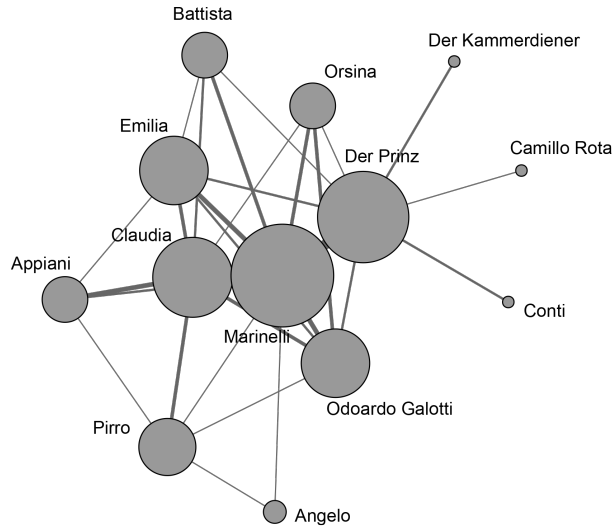
- 1) Knoten: in unserem Fall die Figuren eines Dramas
- 2) Kanten (Linien, die die Knoten verbinden): in unserem Fall die gemeinsame Bühnenpräsenz zweier Figuren

## Gewichtung von Knoten und Kanten:

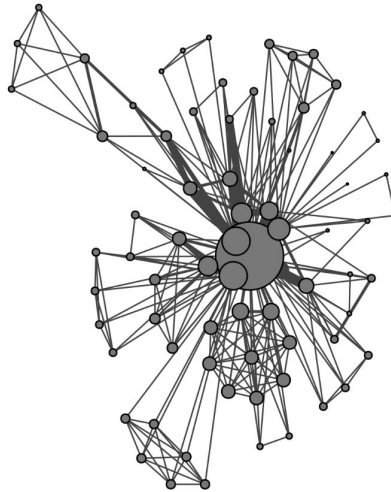
- 1) Knoten: *Degree*, Zahl der Kanten, die ein Knoten auf sich vereint; je höher das *Degree*, desto größer der Knoten
- 2) Kanten: Wie häufig sind zwei Figuren gemeinsam auf der Bühne? Je häufiger, desto dicker die Kante

# Vergleich verschiedener Figurennetzwerke

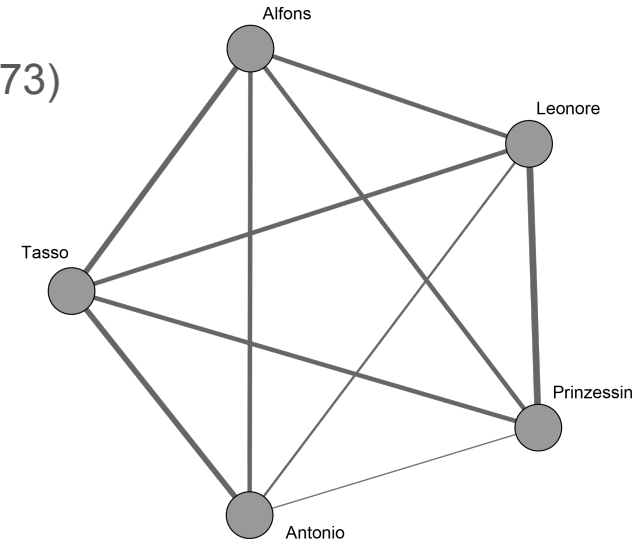
*Emilia Galotti* (1772)



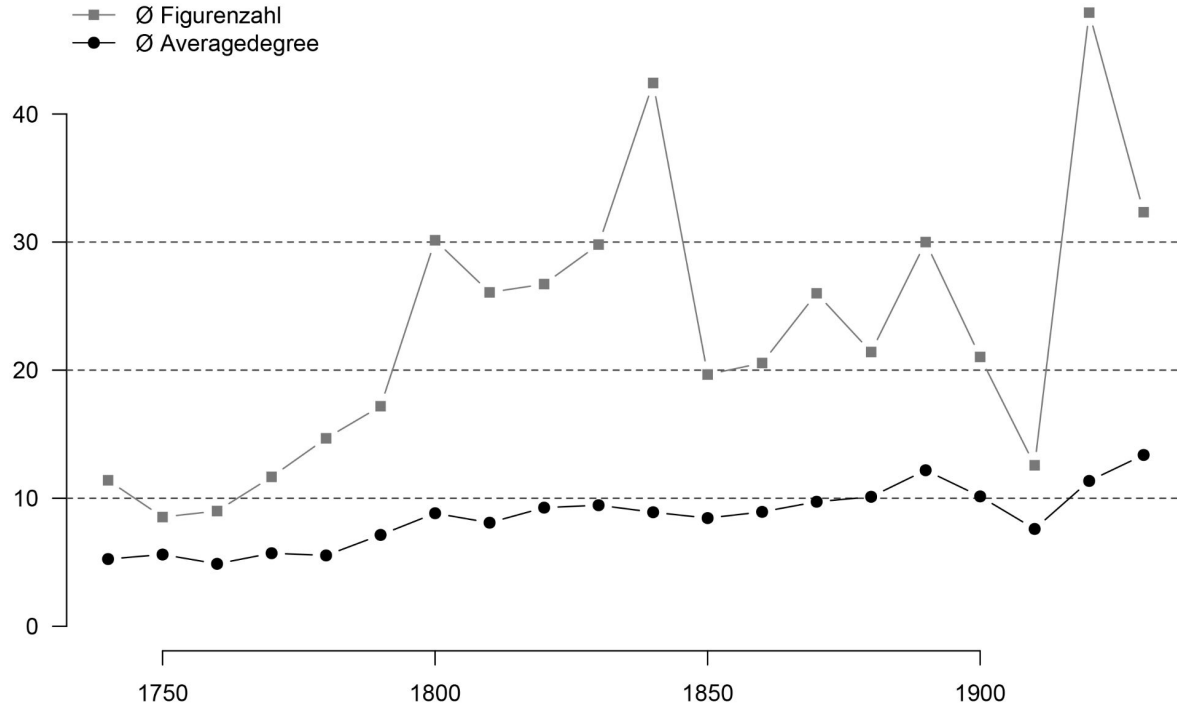
*Goetz von Berlichingen* (1773)



*Torquato Tasso* (1790)

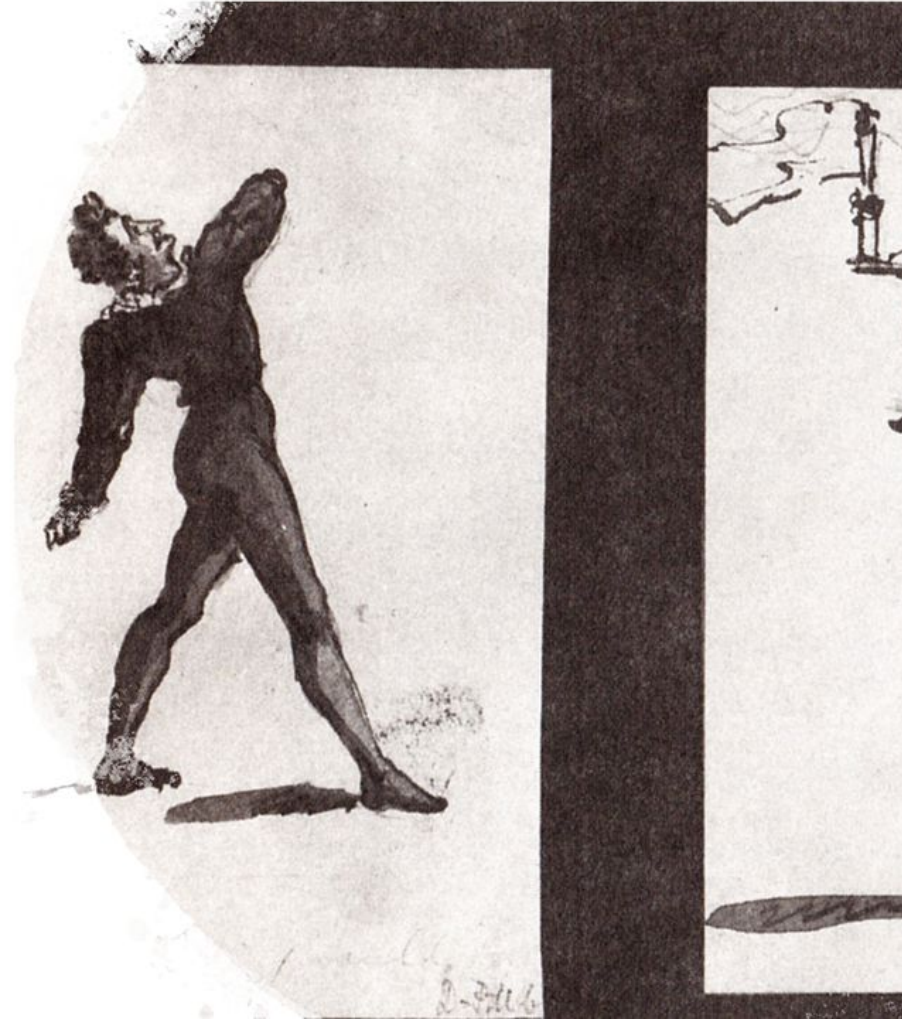


# Rechnen mit Netzwerkmetriken



- 525 Dramen zwischen 1730 und 1930
- Intervalle von zehn Jahren
- Recht hohe positive Korrelation (0,6) zwischen Figurenzahl und Degree

# Passive Präsenz von Figuren



# Passive Präsenz: Was meinen wir damit?

Lessing: *Emilia Galotti*

[... ] Erster Auftritt

**DER PRINZ** *an einem Arbeitstische, voller Briefschaften und Papiere, deren einige er durchläuft.* Klagen, nichts als Klagen! Bittschriften, nichts als Bittschriften! – Die traurigen Geschäfte; und man beneidet uns noch! – Das glaub' ich; wenn wir allen helfen könnten: dann wären wir zu beneiden. – Emilia? *Indem er noch eine von den Bittschriften aufschlägt, und nach dem unterschriebnen Namen sieht.* Eine Emilia? – Aber eine Emilia Bruneschi – nicht Galotti. Nicht Emilia Galotti! – Was will sie, diese Emilia Bruneschi? *Er lieset.* Viel gefodert; sehr viel. – Doch sie heißt Emilia. Gewährt! *Er unterschreibt und klingelt; worauf ein Kammerdiener hereintritt.* Es ist wohl noch keiner von den Räten in dem Vorzimmer? [...]

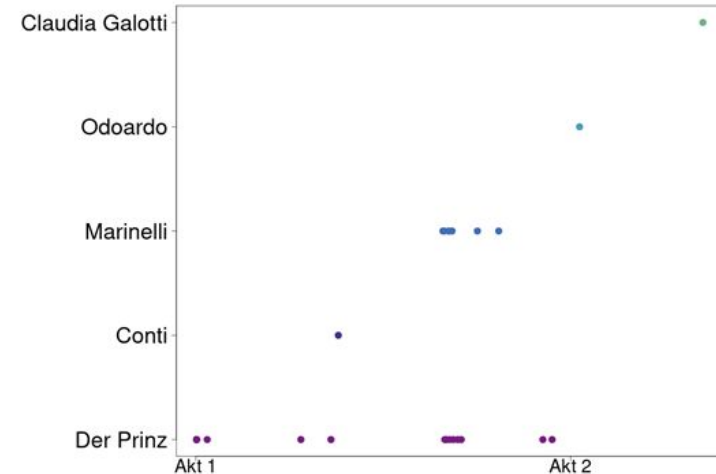
# Passive Präsenz: Was meinen wir damit?

## [... ] Erster Auftritt

**DER PRINZ** *an einem Arbeitstische, voller Briefschaften und Papiere, deren einige er durchläuft. Klagen, nichts als Klagen! Bittschriften, nichts als Bittschriften! – Die traurigen Geschäfte; und man beneidet uns noch! – Das glaub' ich; wenn wir allen helfen könnten: dann wären wir zu beneiden. – Emilia? Indem er noch eine von den Bittschriften aufschlägt, und nach dem unterschriebnen Namen sieht. Eine Emilia? – Aber eine Emilia Bruneschi – nicht Galotti. Nicht Emilia Galotti! – **Was will sie, diese Emilia Bruneschi?** Er lieset. Viel gefodert; sehr viel. – **Doch sie heißt Emilia.** Gewährt! Er unterschreibt und klingelt; worauf ein Kammerdiener hereintritt. Es ist wohl noch keiner von den Räten in dem Vorzimmer?*

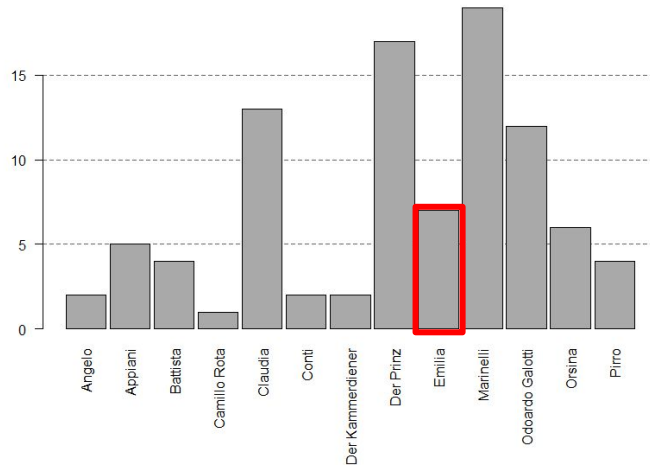
[...]

→ Koreferenzproblem

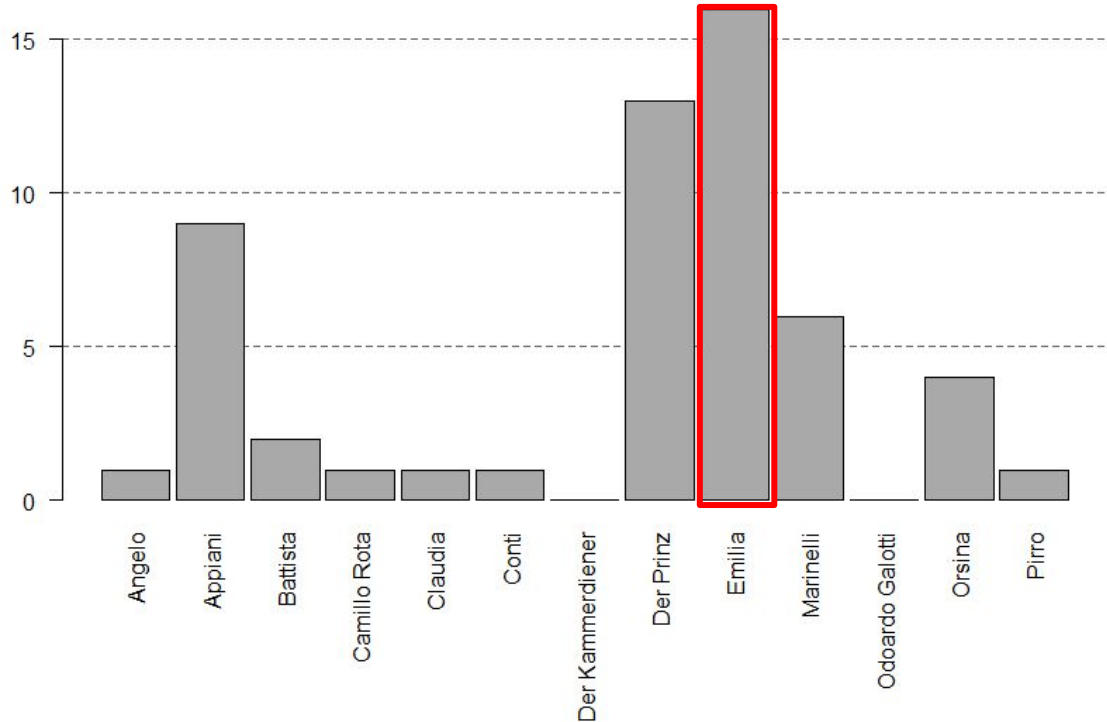




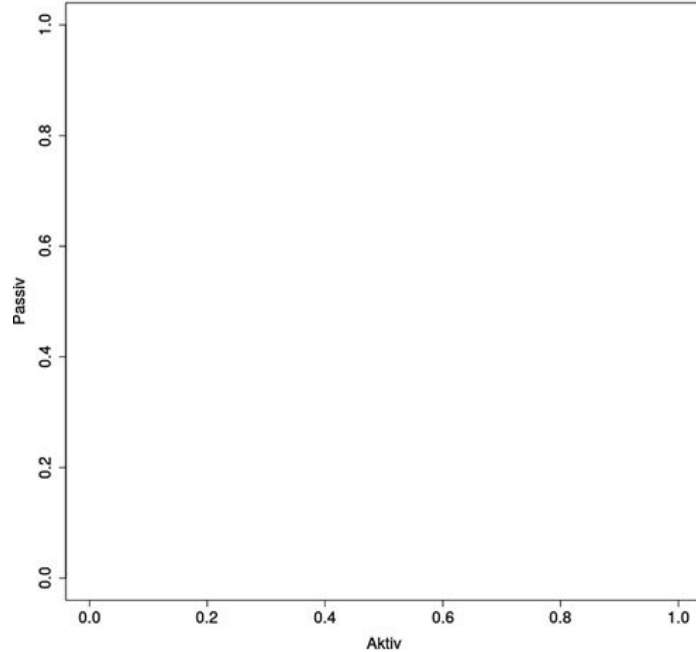
# Vergleich: aktive und passive Präsenz



**Heuristik:** Eine Figur ist passiv präsent, wenn in einer Szene namentlich über sie gesprochen wird, ohne dass sie selbst auf der Bühne steht

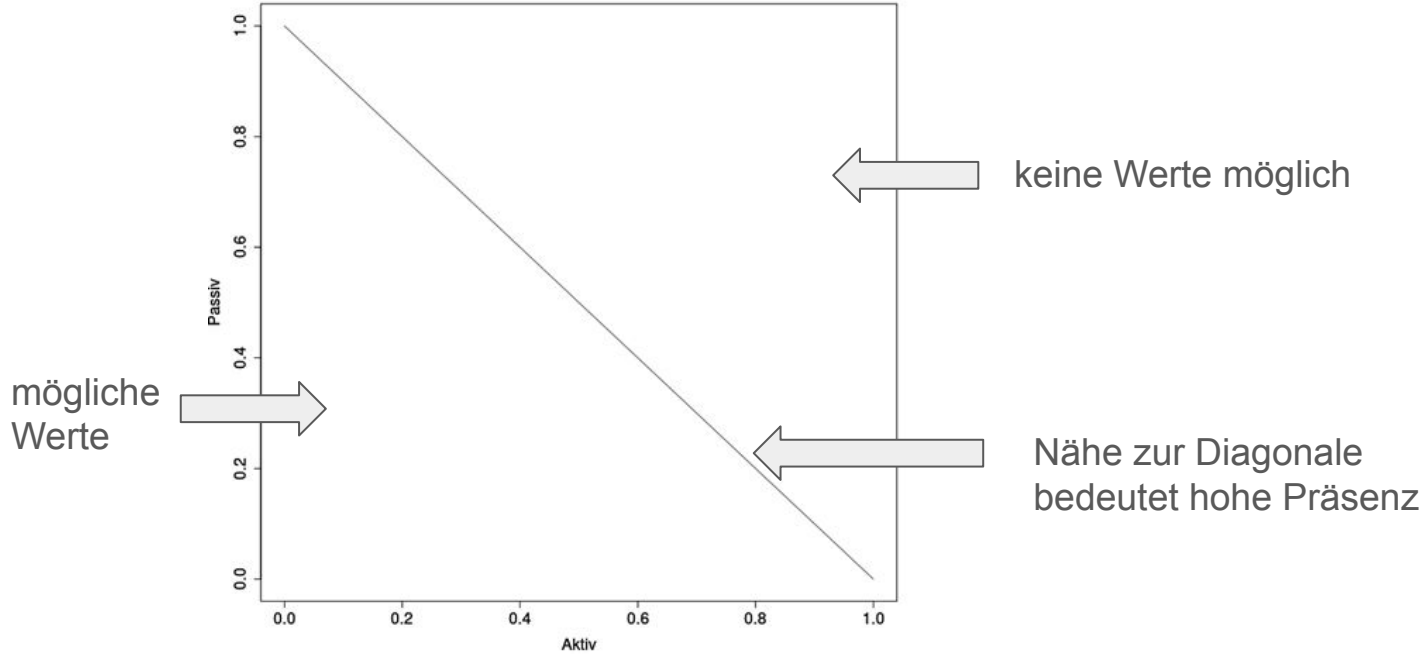


# Korpusuntersuchung: Passive Präsenz

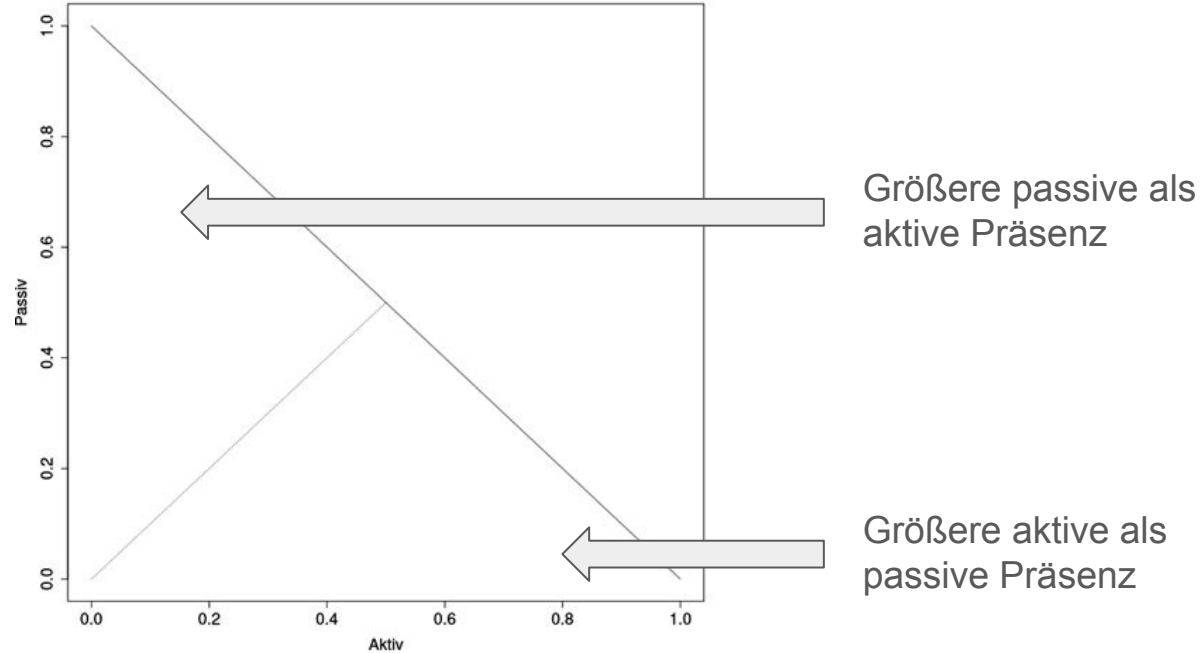


Werte sind normalisiert, werden also durch die Gesamtzahl der Szenen des Dramas geteilt: Wertebereich  $[0;1]$

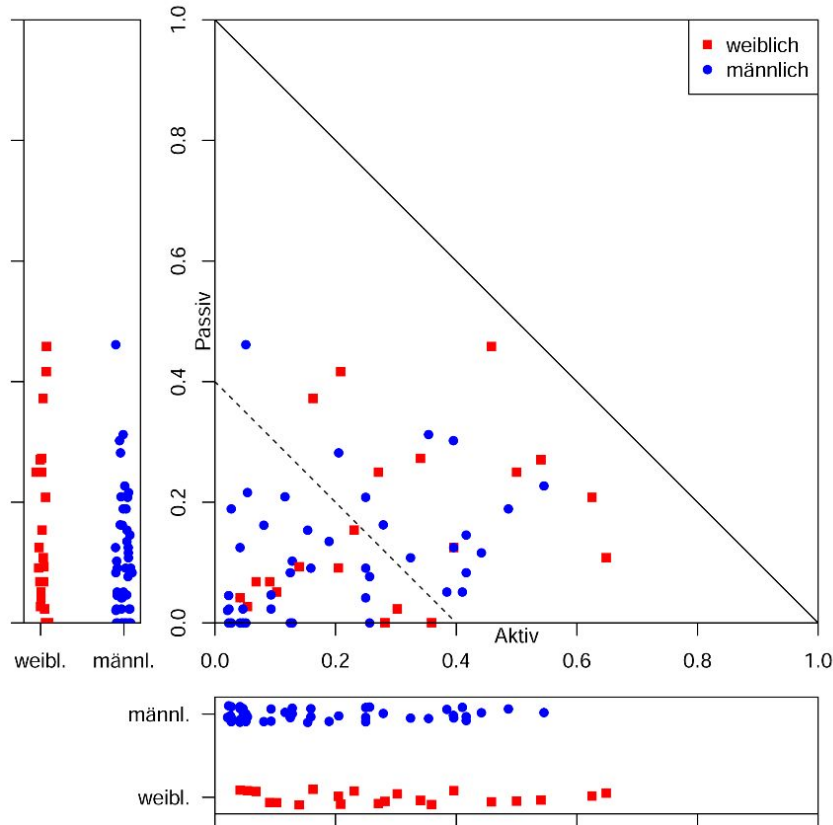
# Korpusuntersuchung: Passive Präsenz



# Korpusuntersuchung: Passive Präsenz

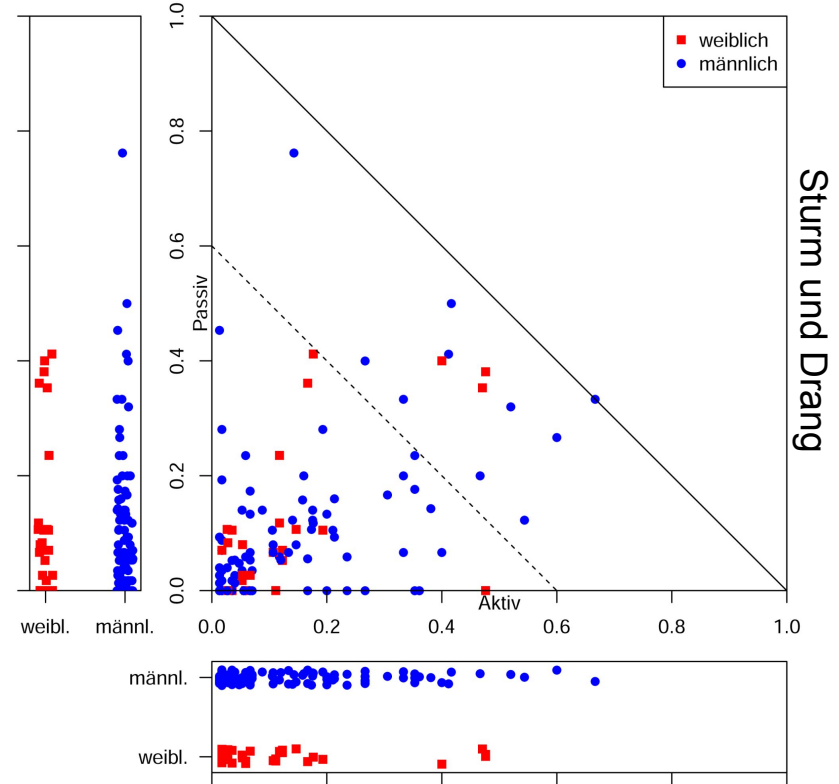
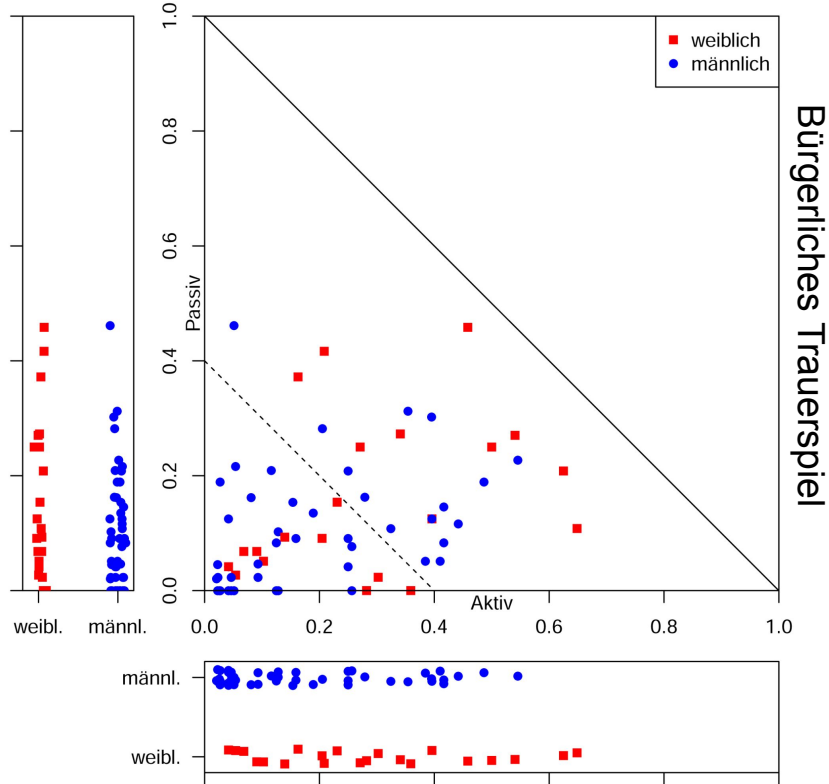


# Korpusuntersuchung: Passive Präsenz



- **Heuristik:** Eine Figur ist passiv präsent, wenn in einer Szene namentlich über sie gesprochen wird, ohne dass sie selbst auf der Bühne steht
- Sechs bürgerliche Trauerspiele
- 73 Figuren
- 236 Szenen

# Korpusvergleich: Passive Präsenz



Abschlussrunde

# Take-Home Messages

- Computational Literary Studies: Spannendes Forschungsfeld mit vielen offenen Fragen
- Kernherausforderung Operationalisierung: Messbarmachung literatur- und geisteswissenschaftlicher Konzepte
  - Woran erkennt man eigentlich Protagonisten? Erzählebenen? Fokalisierung? Intertextuelle Referenz? ...
- (Ein bisschen) Programmieren können zählt sich aus, weil es unabhängig macht
- Unbestelltes Feld: Kombination aus strukturellen und textlichen Informationen

Interesse an Anwendungen von Sprachtechnologie auf  
Fragen aus den Geisteswissenschaften?

Promotionsstelle in Köln!  
(Ausschreibung kommt)



# Ausblick: Netzwerkanalyse 2.0

## Aktuelle Arbeit im Projekt Q:TRACK

- Netzwerk zeigt, wer wie viele Informationen ausgibt/empfängt
- Informationen über Familienrelationen
  - X ist Kind von Y

