To Catch a Protagonist

Quantitative Dominance Relations in German-Language Drama (1730–1930)

Frank Fischer[1], Peer Trilcke[2], Christopher Kittel[3],
Carsten Milling, Daniil Skorinkin[1]

[1] Higher School of Economics, Moscow
[2] University of Potsdam
[3] University of Graz

DH2018, Ciudad de México, 28 June 2018

Quantitative dominance relations?

- »quantitative Dominanzrelationen«: »an important parameter to determine the central or peripheral position of a character« (Pfister 1997, p. 227)
- presence on stage and percentage of spoken text of a character in relation to full text of a play
- network analysis brought in new parameters, typically centrality values
- average distance (Moretti 2011, following an earlier study on the Marvel universe: Alberich et al. 2002)
- tension between character space and word space: different quantitative measures can contradict each other

- → multi-dimensional phenomenon calling for a multi-dimensional approach

Our Data

- **dracor.org** – curated in-house drama corpora in TEI (Russian and German, 18th–early 20th century)
- dracor.org includes an API for direct use
- this paper revolves around our German corpus featuring 465 plays from 1730–1930
- *dramavis*: Python script collection for all our research purposes
- extraction and calculation of quantitative and network-analytical data
- network extraction based on joint presence of characters on stage

Chapters

1. **Multi-dimensional approach**

2. **Distribution of major and minor characters**

3. **Towards a general characterisation of dominance relations in literary texts**

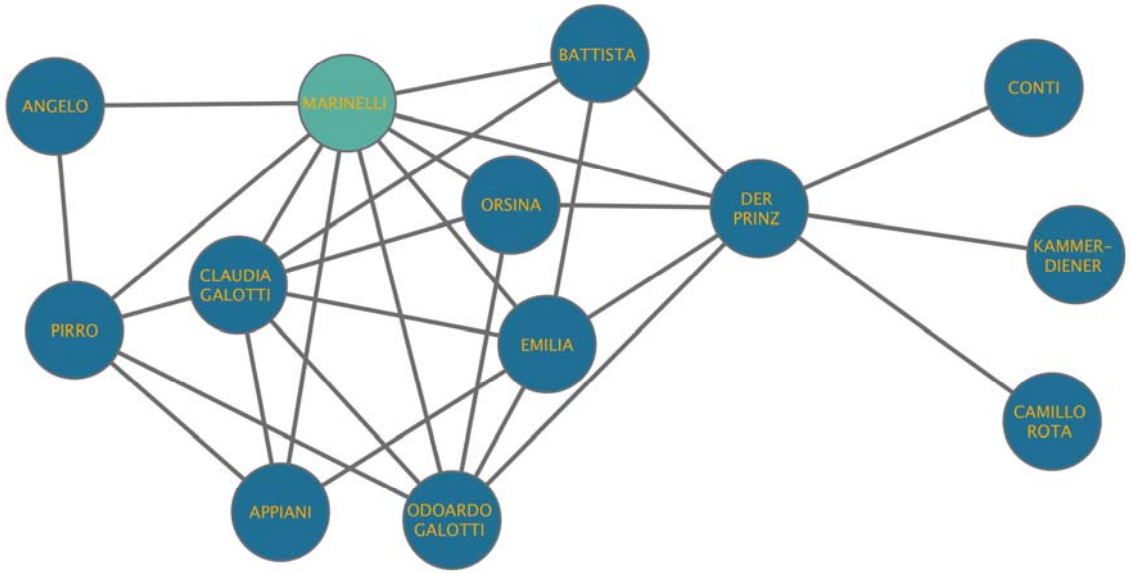
Chapter I

THE BIGGEST CHATTERBOX IN GERMAN LITERATURE

by **Dario Kampkaspar, Frank Fischer, Mathias Göbel and Peer Trilcke — 23 Jun 2015**

	Character	Title	Author	Chars	Words	Speech acts	Additional data
1	GEORG	Ignorabimus	Holz, Arno	133443	20859	952	http://dlina.github.io/390/
2	DUFROY	Ignorabimus	Holz, Arno	107534	16588	885	http://dlina.github.io/390/
3	FAUST	Faust. Der Tragödie erster Teil	Goethe, Johann Wolfgang	97546	9037	225	http://dlina.github.io/243/
4	MEPHISTOPHELES	Faust. Der Tragödie erster Teil	Goethe, Johann Wolfgang	92536	8408	257	http://dlina.github.io/243/
5	GOTHLAND	Herzog Theodor von Gothland	Grabbe, Christian Dietrich	86529	16325	508	http://dlina.github.io/158/
6	HOLLRIEDER	Sonnenfinsternis	Holz, Arno	85600	13544	663	http://dlina.github.io/174/
7	ONKEL LUDWIG	Ignorabimus	Holz, Arno	79066	12322	864	http://dlina.github.io/390/
8	LIBUSSA	Die Gründung Prags	Brentano, Clemens	70723	13139	308	http://dlina.github.io/384/
9	FRANZ	Franz von Sickingen	Lassalle, Ferdinand	67829	12445	219	http://dlina.github.io/287/
10	CARDENIO	Halle	Arnim, Ludwig Achim von	67167	12299	237	http://dlina.github.io/301/
11	MARIANNE	Ignorabimus	Holz, Arno	66707	10383	766	http://dlina.github.io/390/
12	ANATOL	Anatol	Schnitzler, Arthur	61885	11526	723	http://dlina.github.io/89/
13	FIESCO	Die Verschwörung des Fiesco zu Genua	Schiller, Friedrich	61633	10412	326	http://dlina.github.io/451/
14	MEPHISTOPHELES	Faust. Der Tragödie zweiter Teil	Goethe, Johann Wolfgang von	61231	10845	240	http://dlina.github.io/201/
15	CROMWELL	Ein Faust der That	Bleibtreu, Karl	61034	10581	257	http://dlina.github.io/322/
16	LA BELLA CENCI	Sonnenfinsternis	Holz, Arno	60956	10000	453	http://dlina.github.io/174/
17	DOKTOR FAUST	Doktor Faust	Soden, Julius von	60696	10640	543	http://dlina.github.io/450/
18	TASSO	Torquato Tasso	Goethe, Johann Wolfgang von	60095	11338	123	http://dlina.github.io/82/
19	FRANZ VON MOOR	Die Räuber	Schiller, Friedrich	57676	10303	172	http://dlina.github.io/8/
20	CARLOS	Don Carlos, Infant von Spanien	Schiller, Friedrich	55514	10444	333	http://dlina.github.io/217/

G. E. Lessing: »Emilia Galotti« (1772)



Character	Degree	Words
Marinelli	9	5.013
Der Prinz	8	4.799
Claudia Galotti	7	1.819
Emilia Galotti	6	2.009
Odoardo Galotti	6	2.894
Pirro	5	312
Orsina	4	2.522
Battista	4	176
Appiani	4	977
Angelo	2	575
Kammerdiener	1	37
Camillo Rotta	1	94
Conti	1	658

Character	Degree	Words	Degree Ranking	Words Ranking
Marinelli	9	5.013	1	1
Der Prinz	8	4.799	2	2
Claudia Galotti	7	1.819	3	6
Emilia Galotti	6	2.009	4	5
Odoardo Galotti	6	2.894	4	3
Pirro	5	312	6	10
Orsina	4	2.522	7	4
Battista	4	176	7	11
Appiani	4	977	7	7
Angelo	2	575	10	9
Kammerdiener	1	37	11	13
Camillo Rotta	1	94	11	12
Conti	1	658	11	8

Eight Quantitative Character-Dominance Measures

Count-based

- Words
- Speech Acts
- Frequency

Network-based

- Degree
- Closeness
- Betweenness
- Weighted Degree
- Eigenvector

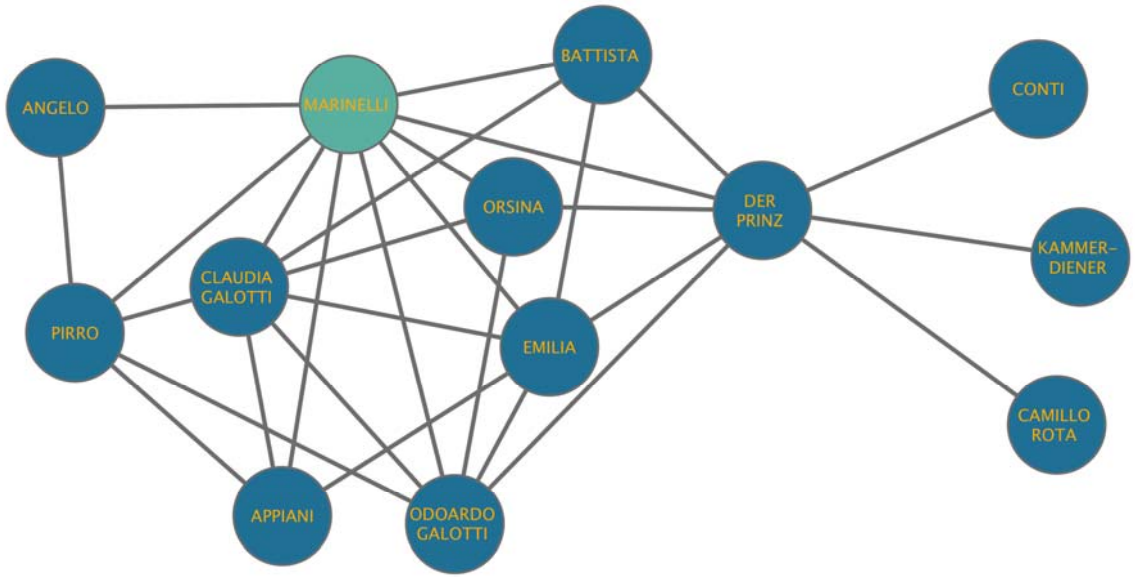
For each measure we created a ranking.

… plus meta-rankings

- Count-based ranking (out of all count-based measures)
- Network-based ranking (out of all network-based measures)
- Overall ranking (out of all eight measures)

… calculated for all 9.924 characters in our corpus of 465 plays

Rankings for character ›Marinelli‹
in G. E. Lessing‘s »Emilia Galotti« (1772)



Measure	Rank
Count	
Frequency	1
Speech Acts	1
Words	1
Network	
Degree	1
Closeness	1
Betweenness	2
Weighted Degree	1
Eigenvector	1
Meta	
Count-Raking	1
Network-Ranking	1
Overall-Ranking	1

Eight Quantitative Character-Dominance Measures

- There is no gold standard – no consensual list of clear main characters, …
- … but in many cases our algorithm yields the title character as ›protagonist‹

Eight Quantitative Character Dominance Measures

Author	Year	Title	Dominant Character
Gottsched	1731	Sterbender Cato	Cato
Wieland	1758	Lady Johanna Gray	Lady Johanna Gray
Lessing	1759	Philotas	Philotas
Goethe	1773	Götz von Berlichingen	Götz
Brandes	1775	Ariadne auf Naxos	Ariadne
Goethe	1788	Egmont	Egmont
Kleist	1810	Prinz Friedrich von Homburg	Prinz Friedrich
Uhland	1818	Ernst Herzog von Schwaben	Ernst
Grabbe	1835	Hannibal	Hannibal
Hofmannsthal	1906	Ödipus und die Sphinx	Ödipus
Schnitzler	1912	Professor Bernhardi	Professor Bernhardi
Wedekind	1912	Franziska	Franziska
			usw.

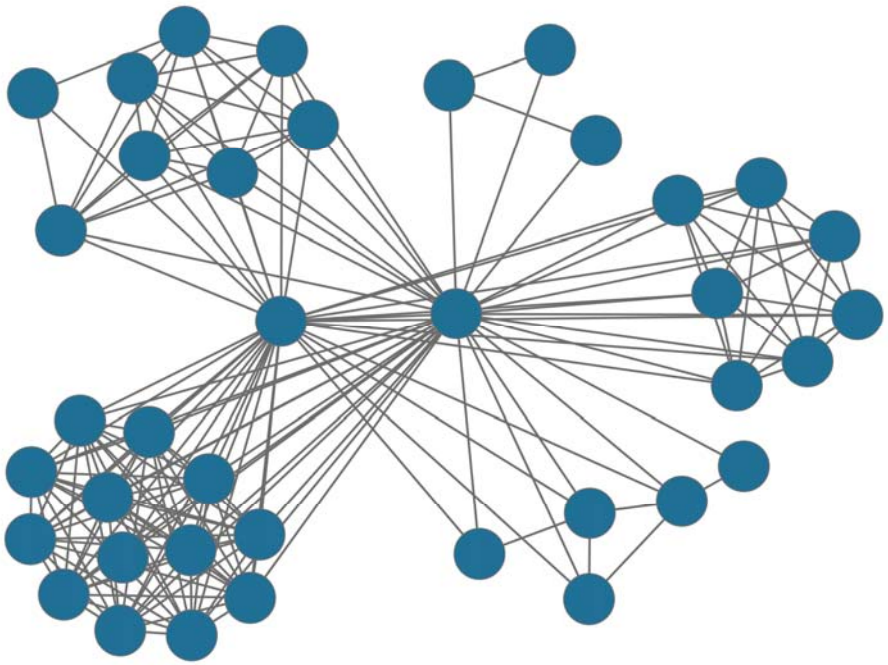
Eight Quantitative Character-Dominance Measures

How sure is our algorithm of its own findings?

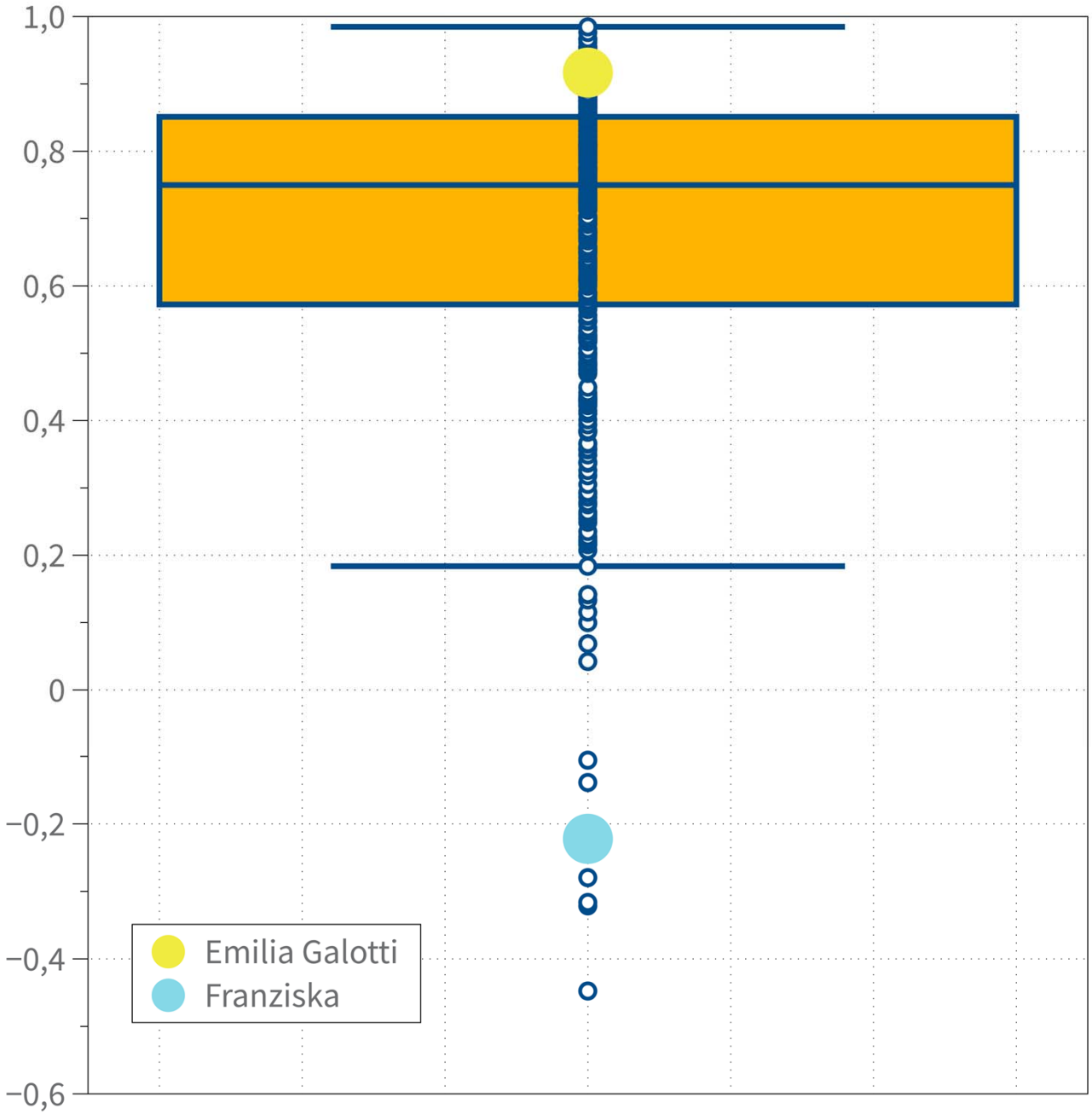
- in 33 % of the cases, the main character is ranking 1st in all eight rankings
- in 17 % of the cases, the main character is ranking 1st in seven rankings

- so, in half the cases our algorithm is very sure of itself at catching the protagonist

Frank Wedekind: »Franziska« (1912)



Character	Count Ranking	Network Ranking
Franziska	1	1
Veit Kunz	2	2
Herzog	15	3
Laurus Bein	3	16
Gespensterschreck	3	18
Breitenbach	16	5
Gislind	18	6
Sophie	20	4
Spreizfüsschen	3	23
Mausi	3	26
Herzogin	17	13
Rohrdommel	3	27
Fahrstuhl	24	9
…	…	…



Spearmans Rho (–1 to 1) for all plays

Median: 0,74

Emilia Galotti: »a great example for dramatic algebra« (F. Schlegel)

Franziska: ›outsider‹ characters

Chapter II

Character	Degree
Marinelli	9
Der Prinz	8
Claudia Galotti	7
Emilia Galotti	6
Odoardo Galotti	6
Pirro	5
Orsina	4
Battista	4
Appiani	4
Angelo	2
Kammerdiener	1
Camillo Rotta	1
Conti	1

Degree Range:
1 to 9 = 8

Cut into
4 equal groups
(in this case 8/4 = 2)

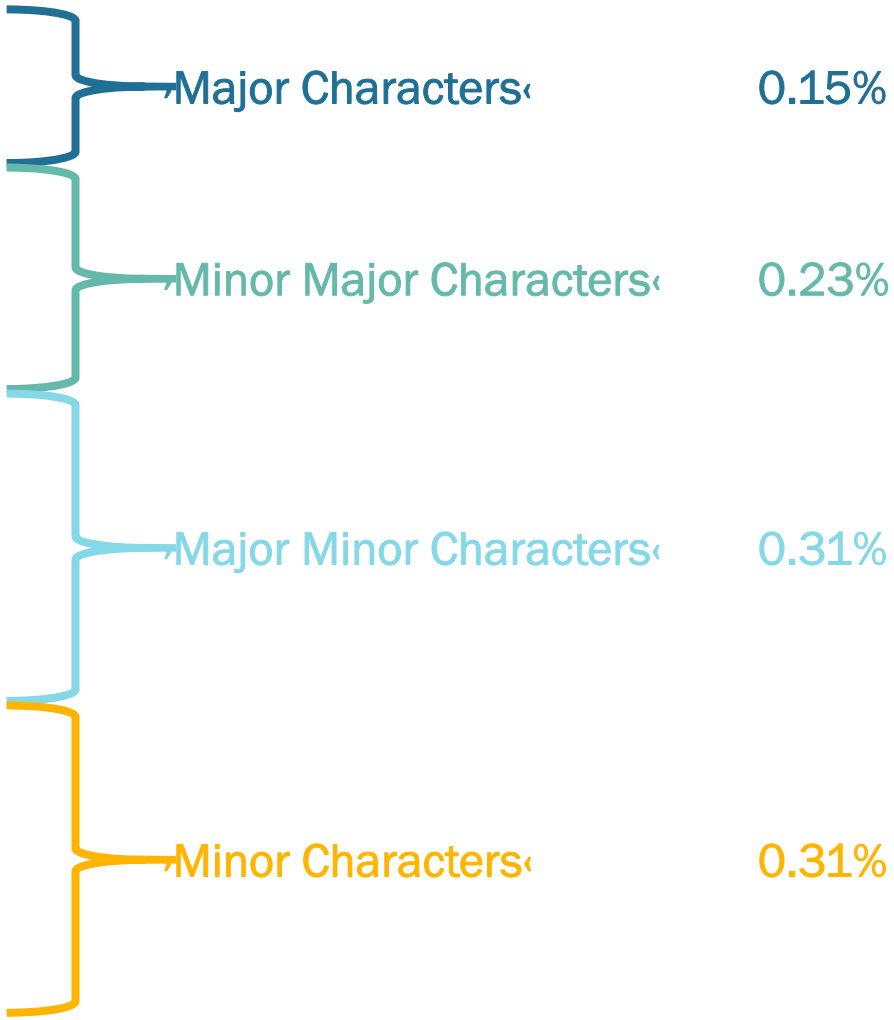
Group 1: 7.01 – 9.00

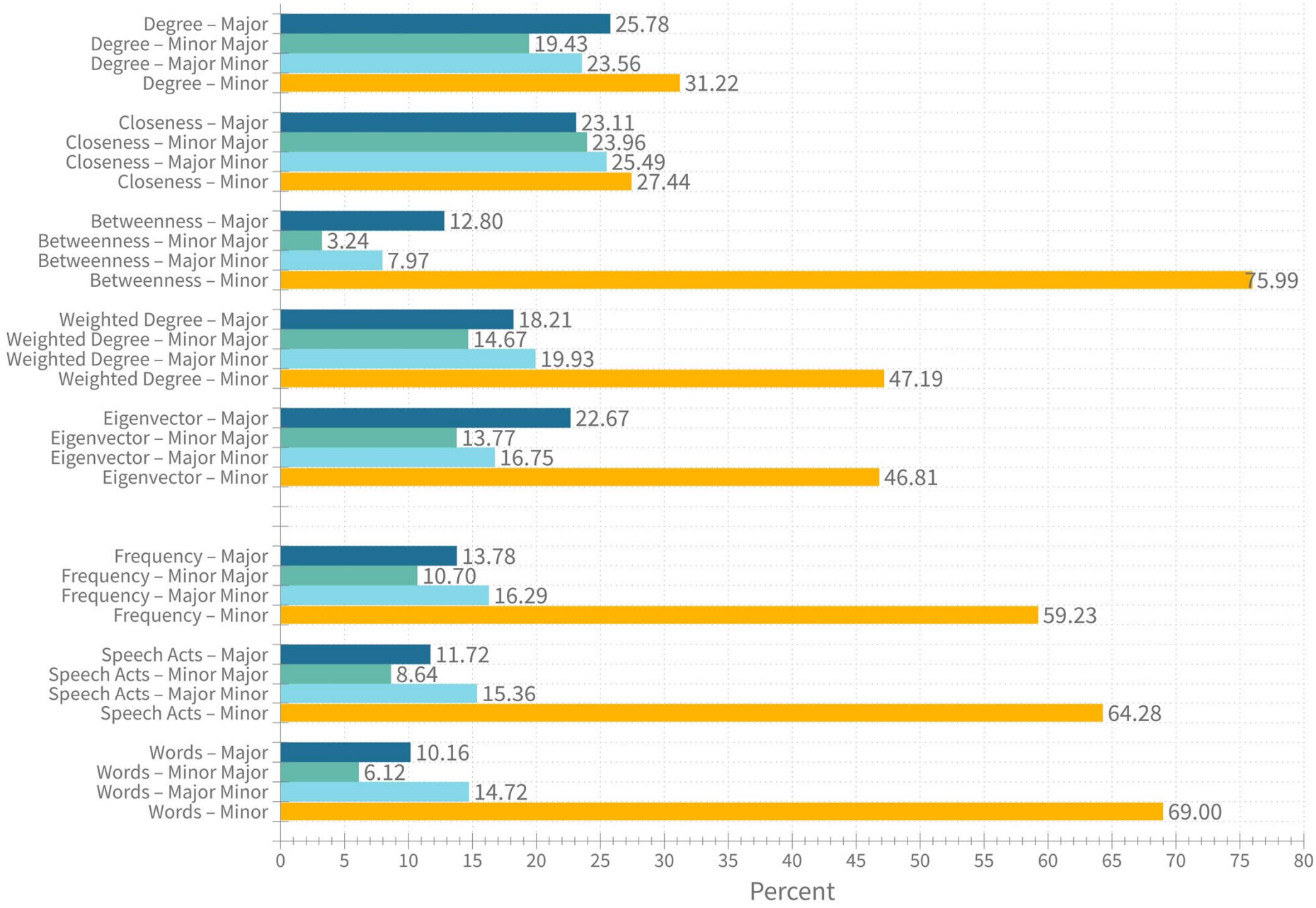
Group 2: 5.01 – 7.00

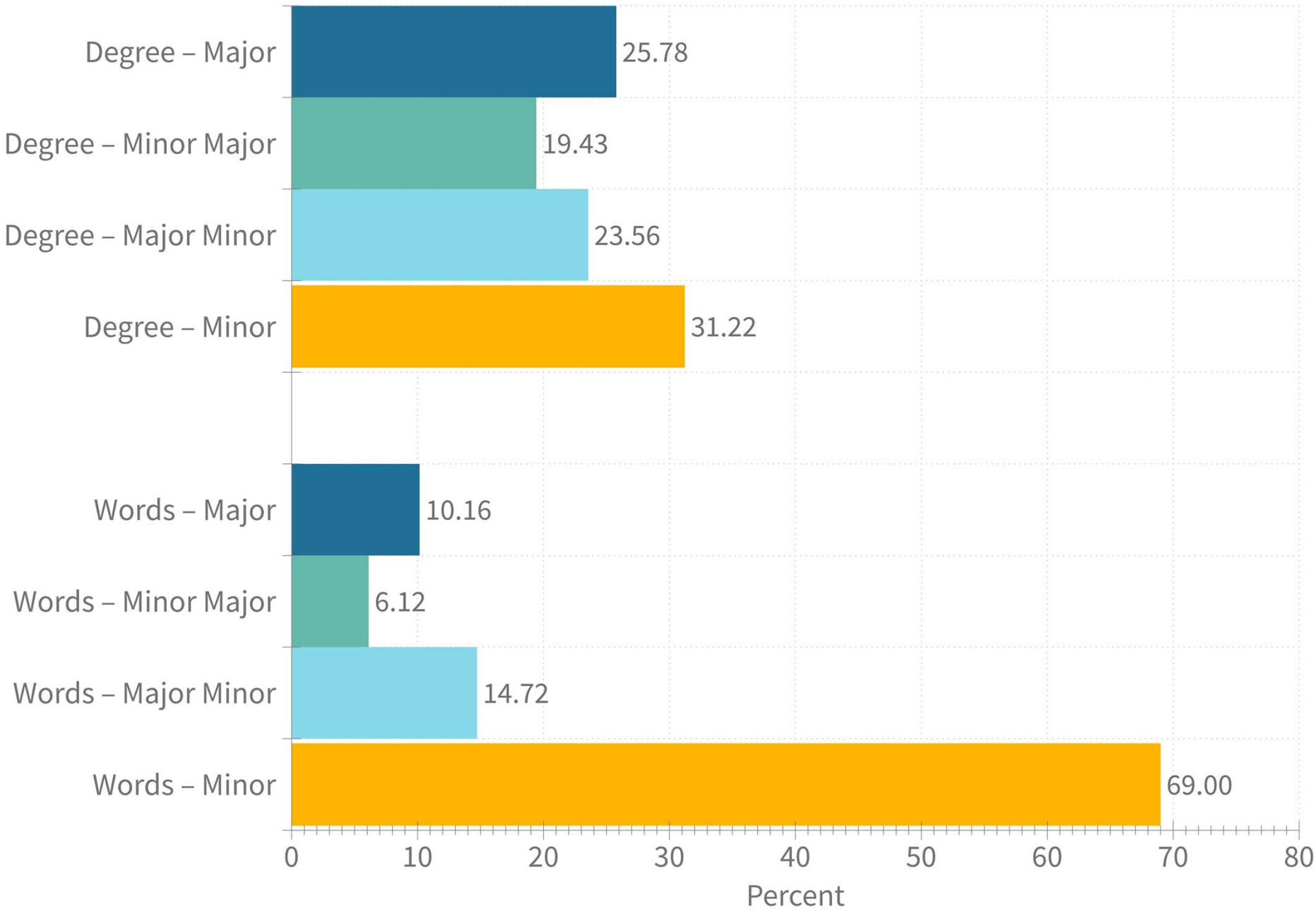
Group 3: 3.01 – 5.00

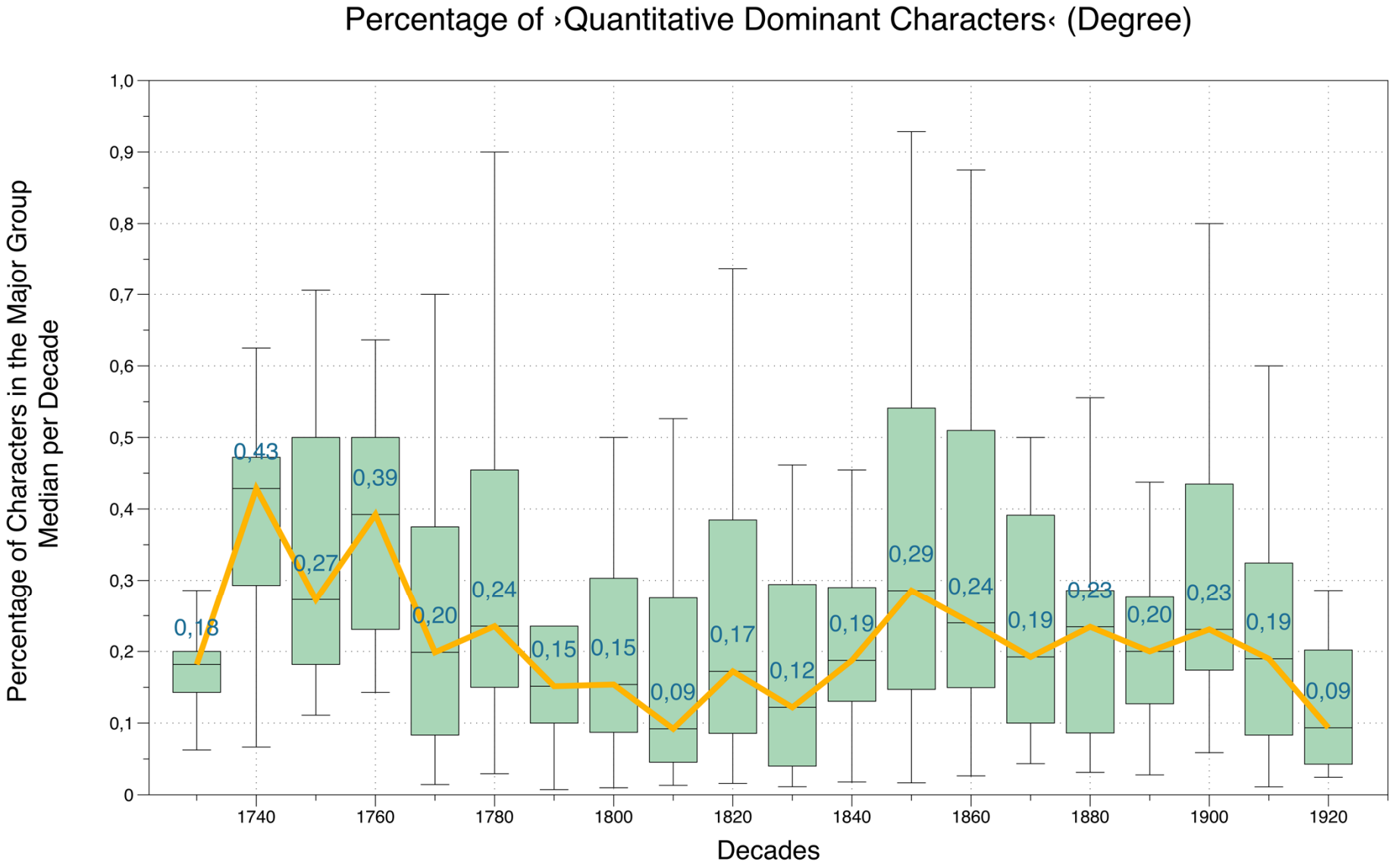
Group 4: 1.00 – 3.00

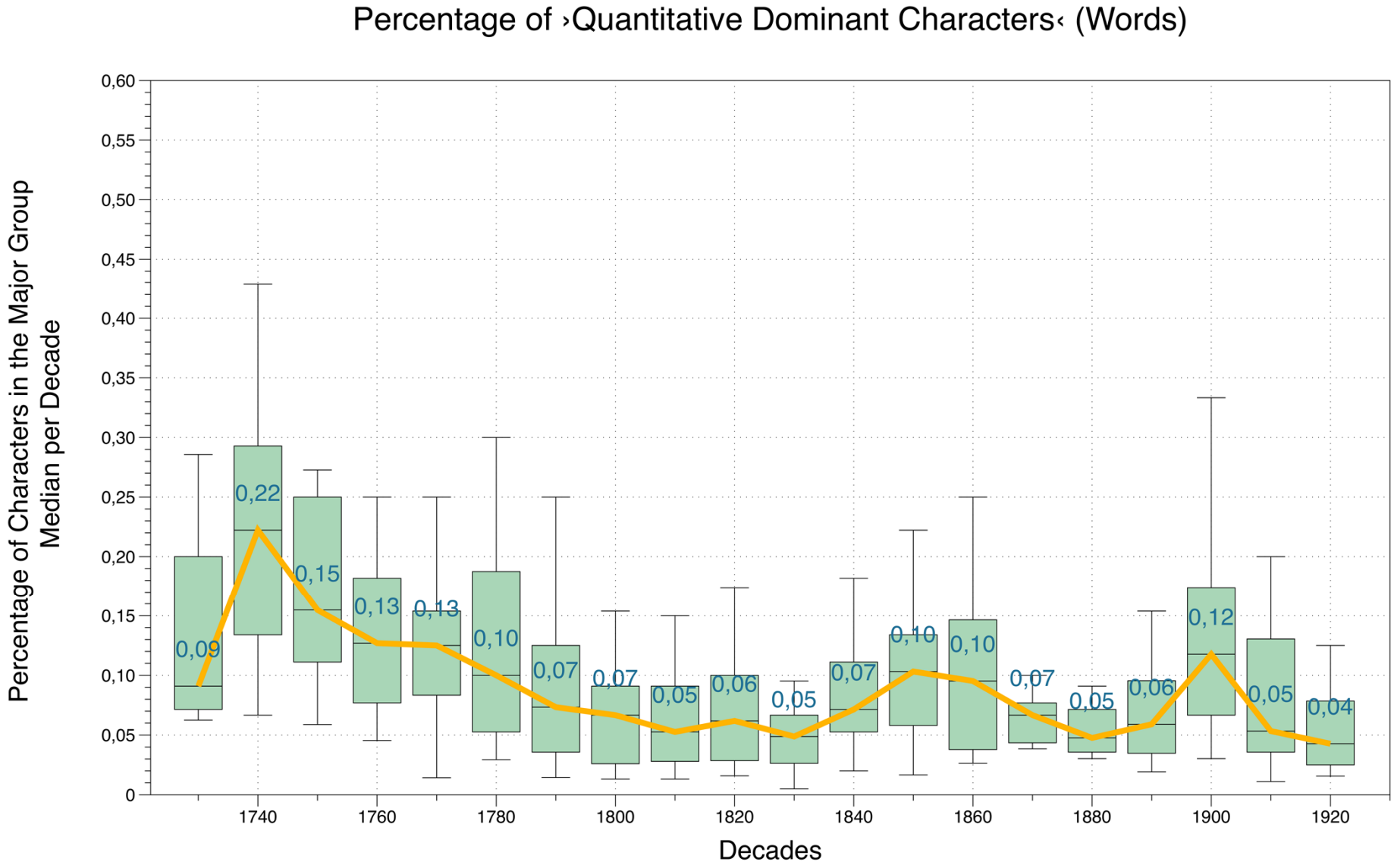
Character	Degree
Marinelli	9
Der Prinz	8
Claudia Galotti	7
Emilia Galotti	6
Odoardo Galotti	6
Pirro	5
Orsina	4
Battista	4
Appiani	4
Angelo	2
Kammerdiener	1
Camillo Rotta	1
Conti	1



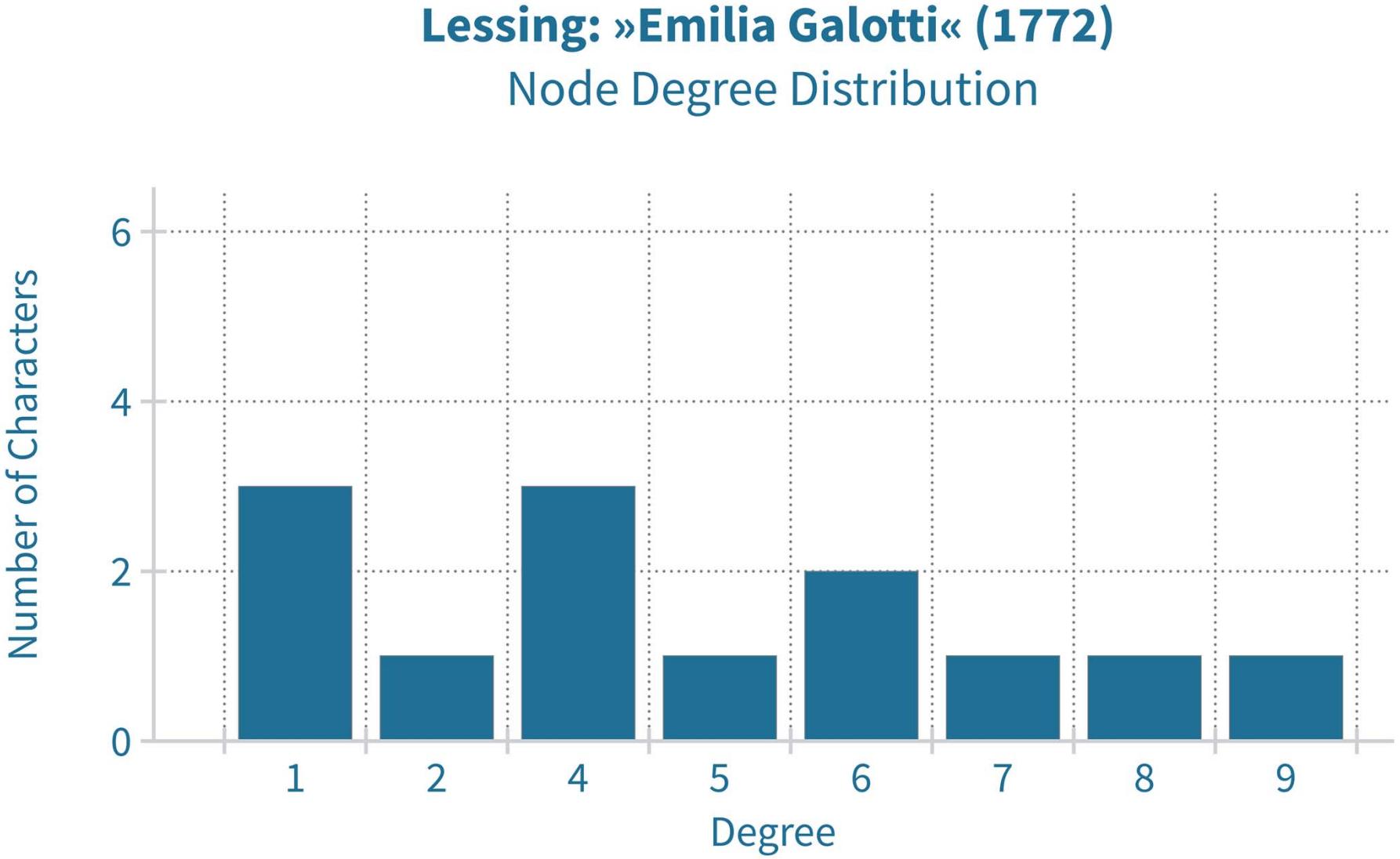

Degree – Major
Degree – Minor Major
Degree – Major Minor
Degree – Minor
25.78
19.43
23.56
31.22
Closeness – Major
Closeness – Minor Major
Closeness – Major Minor
Closeness – Minor
23.11
23.96
25.49
27.44
Betweenness – Major
Betweenness – Minor Major
Betweenness – Major Minor
Betweenness – Minor
12.80
3.24
7.97
75.99
Weighted Degree – Major
Weighted Degree – Minor Major
Weighted Degree – Major Minor
Weighted Degree – Minor
18.21
14.67
19.93
47.19
Eigenvector – Major
Eigenvector – Minor Major
Eigenvector – Major Minor
Eigenvector – Minor
22.67
13.77
16.75
46.81
Frequency – Major
Frequency – Minor Major
Frequency – Major Minor
Frequency – Minor
13.78
10.70
16.29
59.23
Speech Acts – Major
Speech Acts – Minor Major
Speech Acts – Major Minor
Speech Acts – Minor
11.72
8.64
15.36
64.28
Words – Major
Words – Minor Major
Words – Major Minor
Words – Minor
10.16
6.12
14.72
69.00
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
Percent

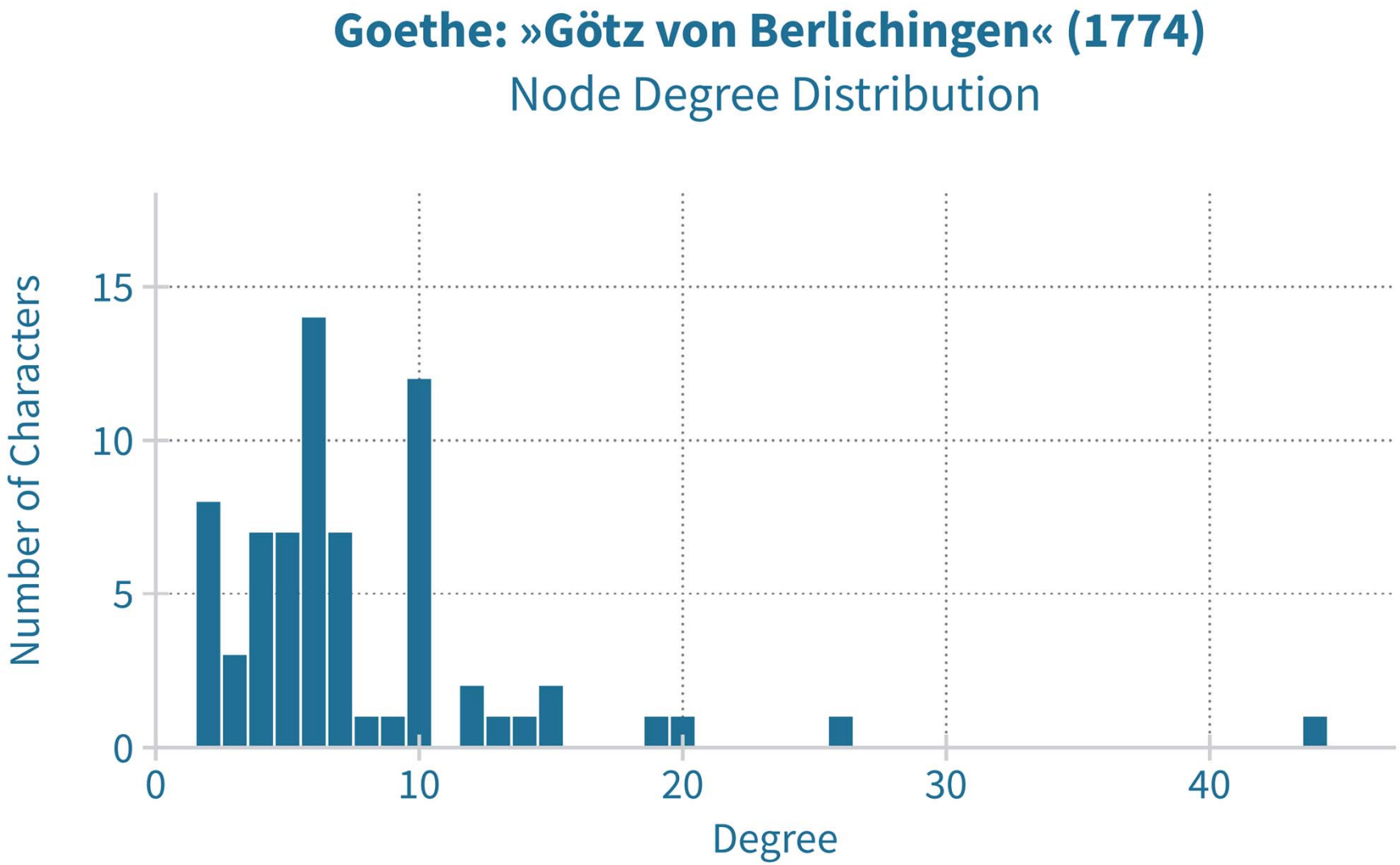

Degree – Major
25.78
Degree – Minor Major
19.43
Degree – Major Minor
23.56
Degree – Minor
31.22
Words – Major
10.16
Words – Minor Major
6.12
Words – Major Minor
14.72
Words – Minor
69.00
0
10
20
30
40
50
60
70
80
Percent

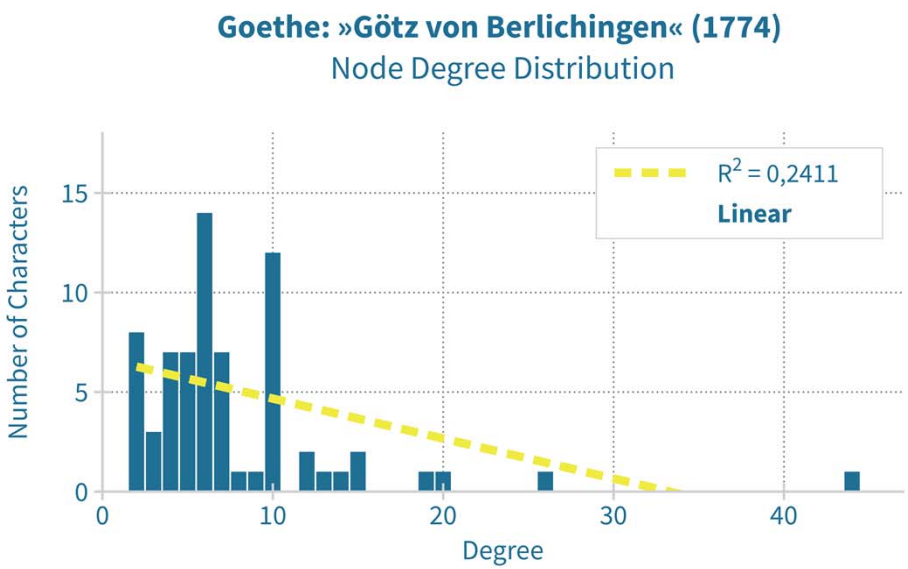

Percentage of ›Quantitative Dominant Characters‹ (Degree)
Percentage of Characters in the Major Group
Median per Decade
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0,18
0,43
0,27
0,39
0,20
0,24
0,15
0,15
0,09
0,17
0,12
0,19
0,29
0,24
0,19
0,23
0,20
0,23
0,19
0,09
1740
1760
1780
1800
1820
1840
1860
1880
1900
1920
Decades

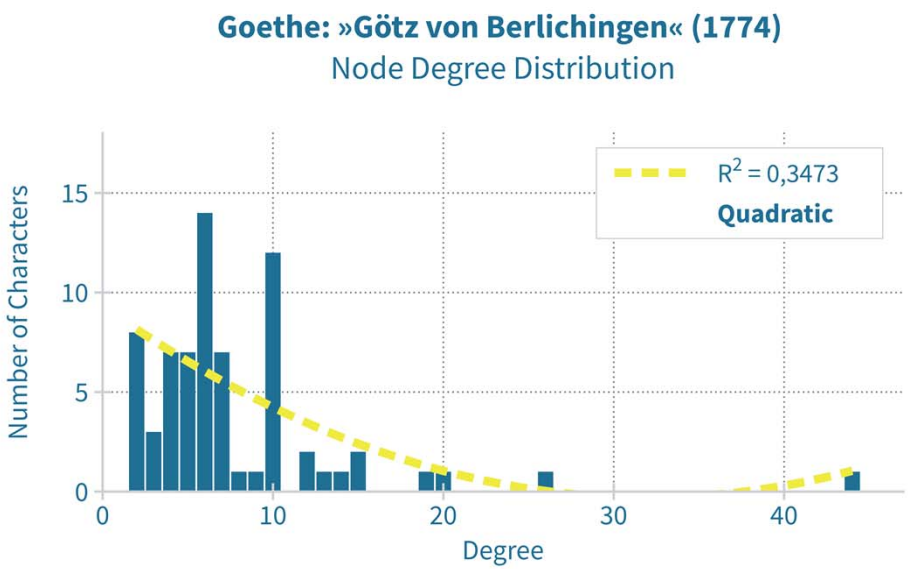

Percentage of ›Quantitative Dominant Characters‹ (Words)
Percentage of Characters in the Major Group
Median per Decade
0,60
0,55
0,50
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0
0,09
0,22
0,15
0,13
0,13
0,10
0,07
0,07
0,05
0,06
0,05
0,07
0,10
0,10
0,07
0,05
0,06
0,12
0,05
0,04
1740
1760
1780
1800
1820
1840
1860
1880
1900
1920
Decades

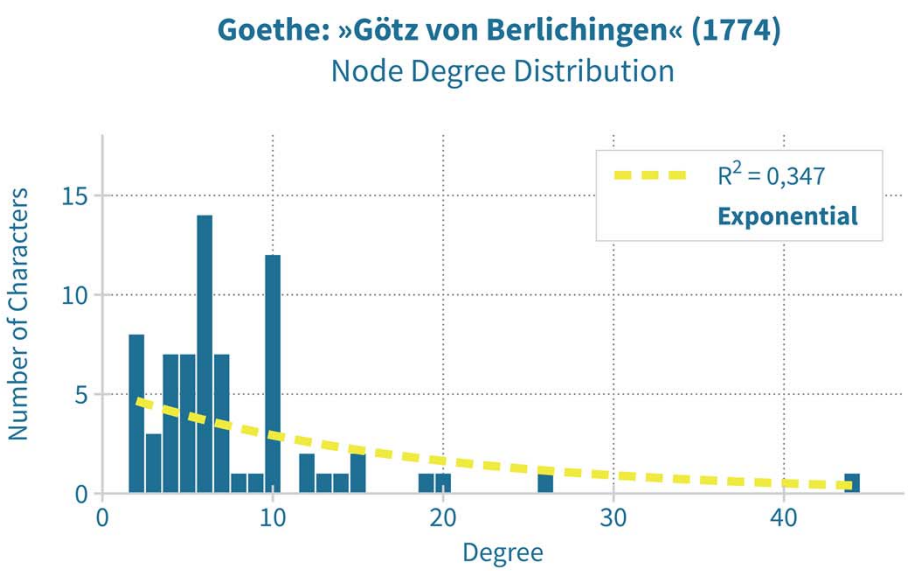
Chapter III

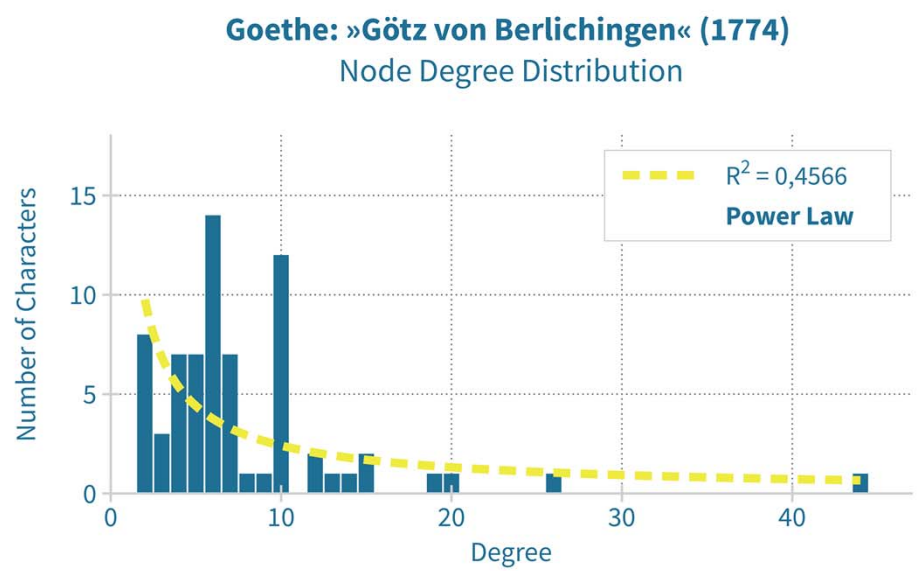

Lessing: »Emilia Galotti« (1772)
Node Degree Distribution
Number of Characters
0
2
4
6
1
2
4
5
6
7
8
9
Degree

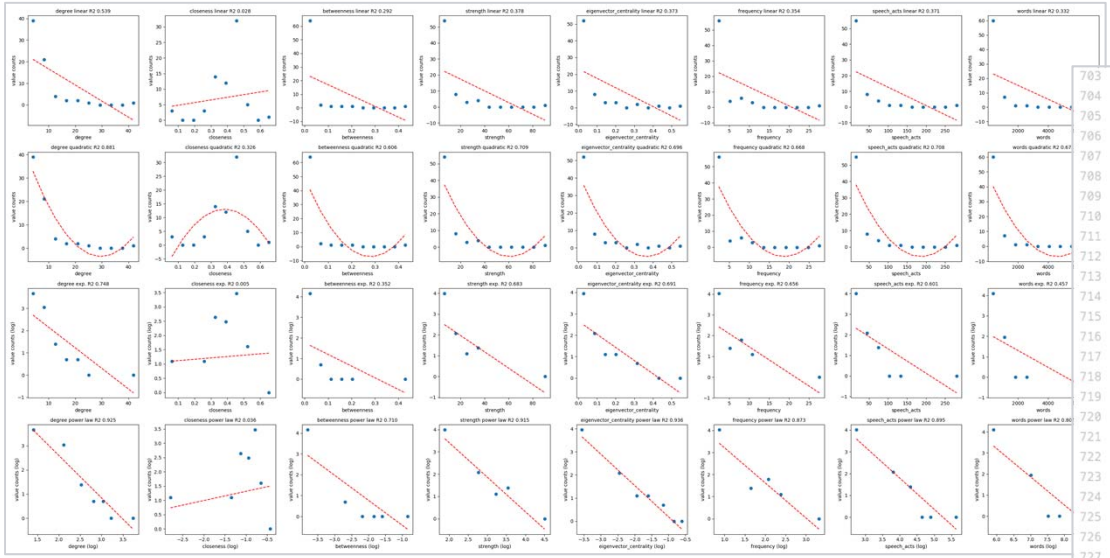

Goethe: »Götz von Berlichingen« (1774)
Node Degree Distribution
Number of Characters
15
10
5
0
0
10
20
30
40
Degree


Goethe: »Götz von Berlichingen« (1774)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,2411
Linear
0
5
10
15
0
10
20
30
40


Goethe: »Götz von Berlichingen« (1774)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,3473
Quadratic
0
5
10
15
0
10
20
30
40


Goethe: »Götz von Berlichingen« (1774)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,347
Exponential
0
5
10
15
0
10
20
30
40


Goethe: »Götz von Berlichingen« (1774)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,4566
Power Law
0
5
10
15
0
10
20
30
40



metrics	linear	exponential	powerlaw	quadratic	max_val	max_type
degree	0.314738141169	0.521023093503	0.585518147208	0.553864426346	0.585518147208232	powerlaw
closeness	0.407407407407	0.636363636364	0.683332388229	0.78151889263	0.7815188926300052	quadratic
betweenness	0.191653413319	0.273259021084	0.607990303965	0.609210619633	0.6092106196332882	quadratic
strength	0.219227688347	0.485068739305	0.668529250935	0.498556599168	0.6685292509350541	powerlaw
eigenvector_centrality	0.0601534102328	0.273536776822	0.41824507601	0.799492395734	0.799492395734045	quadratic
frequency	0.263623829182	0.273123948812	0.463500161467	0.627382904689	0.6273829046887349	quadratic
speech_acts	0.28841331276	0.430347679141	0.573673426792	0.587967140991	0.5879671409911709	quadratic
words	0.263623829182	0.273123948812	0.591713607114	0.627382904689	0.6273829046887347	quadratic

```python
def get_regression_metrics(self):
    metrics = ['degree', 'closeness', 'betweenness',
               'strength', 'eigenvector_centrality',
               'frequency', 'speech_acts', 'words']
    metrics_dfs = []
    for metric in metrics:
        temp_df = pd.DataFrame(columns=[metric])
        temp_df[metric+"_interval"] = [
                i.mid for i in pd.cut(self.centralities[metric], 10)
                                 .value_counts()
                                 .index.tolist()]
        temp_df[metric] = (pd.cut(self.centralities[metric], 10)
                           .value_counts()
                           .tolist())
        temp_df.sort_values(metric+"_interval", inplace=True)
        temp_df.reset_index(drop=True, inplace=True)
        metrics_dfs.append(temp_df)
    index = ["linear", "exponential", "powerlaw", "quadratic"]
    reg_metrics = pd.DataFrame(columns=metrics, index=index)
    # fit linear models
    fig = plt.figure(figsize=(len(metrics)*4, len(index)*4))
    gs = gridspec.GridSpec(len(index), len(metrics))
    i = 0  # subplot enumerator
    for metric, temp_df in zip(metrics, metrics_dfs):
        X = np.array(temp_df[metric+"_interval"]).reshape(-1, 1)
        y = np.array(temp_df[metric]).reshape(-1, 1)
        model = linear_model.LinearRegression()
        model.fit(X, y)
        score = model.score(X, y)
        reg_metrics.loc["linear", metric] = score
        ax = plt.subplot(gs[i])
        plt.scatter(X, y)
        plt.plot(X, model.predict(X), 'r--',
                 label='coeff: %.3f, intercept: %.3f' % (model.coef_[0][0],
                                                         model.intercept_[0]))
        # plt.legend(fontsize='x-small')
        ax.set_title(metric + " linear R2 %.3f" % score, size='medium')
        ax.set_xlabel(metric)
        ax.set_ylabel("value counts")
        i += 1
fit quadratic models
```

top left: dramavis regression-plot output: https://github.com/dlina/project/tree/master/results
bottom left: dramavis regression-value output: https://github.com/dlina/project/tree/master/data/textgrid-repository-dramas
right: dramavis code excerpt (regression analysis): https://github.com/lehkost/dramavis

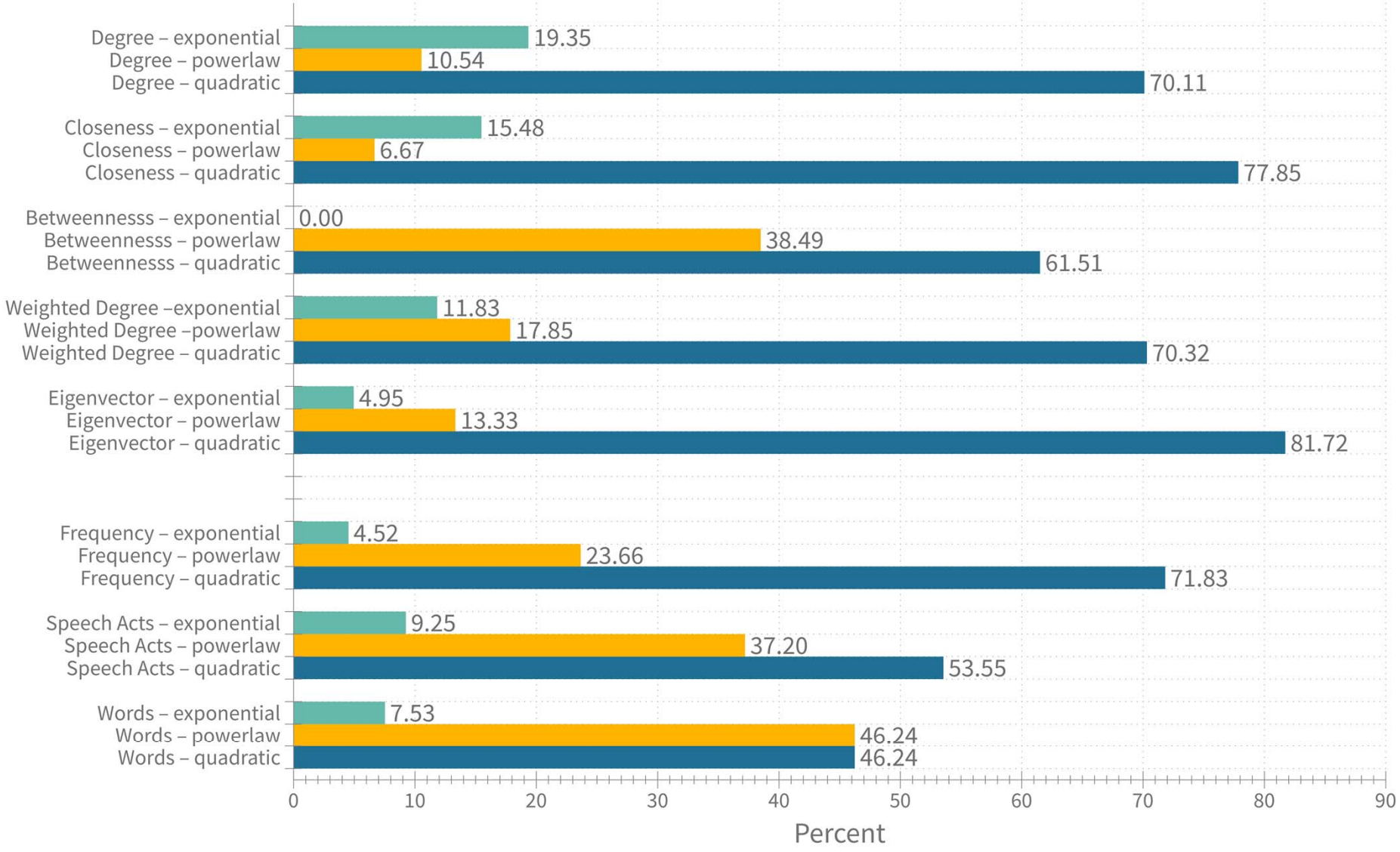
What to expect? Power laws!

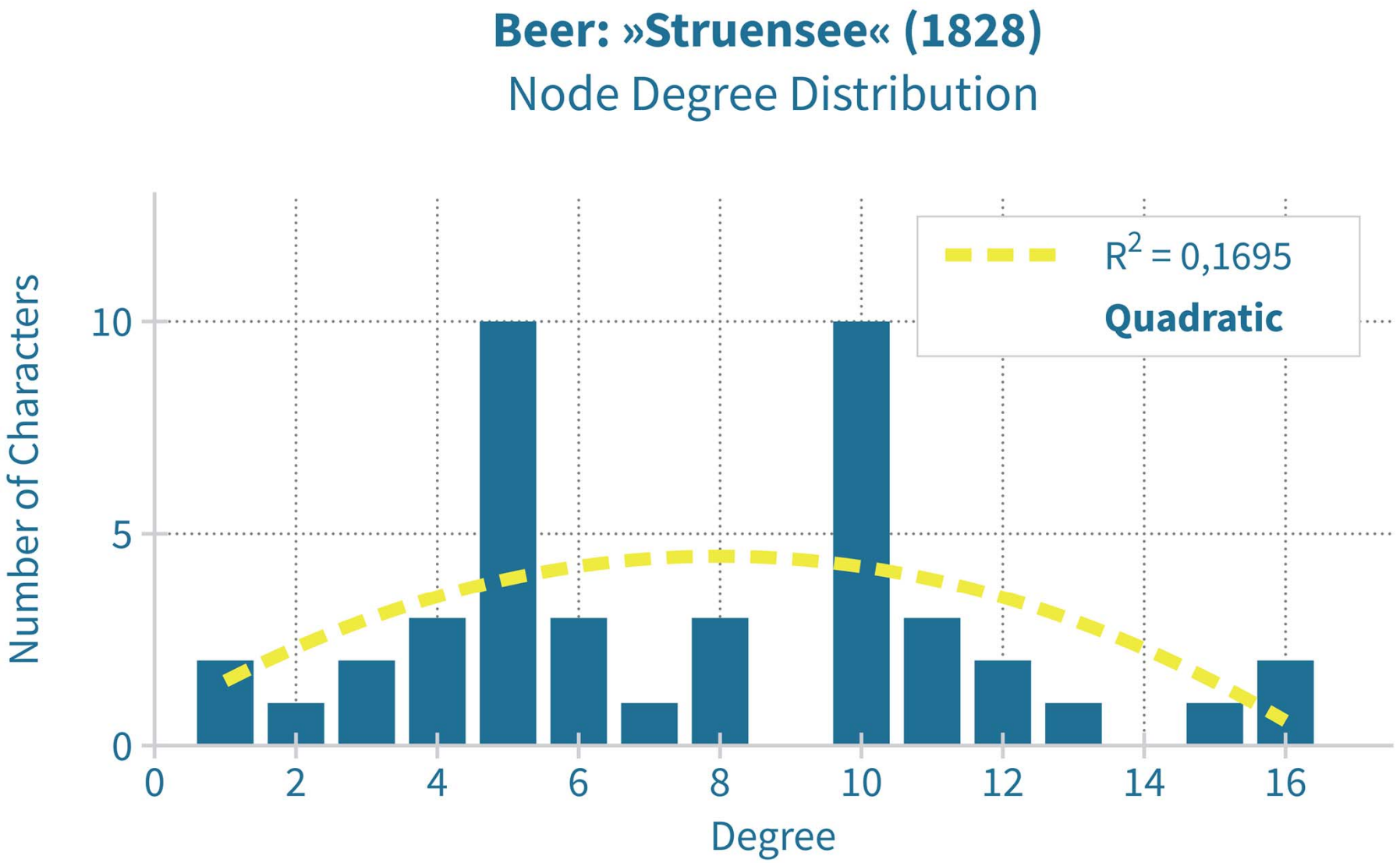
»[I]f we visualize these results in the form of a histogram […], **we find the power-law distribution that is characteristic of all networks** […].«

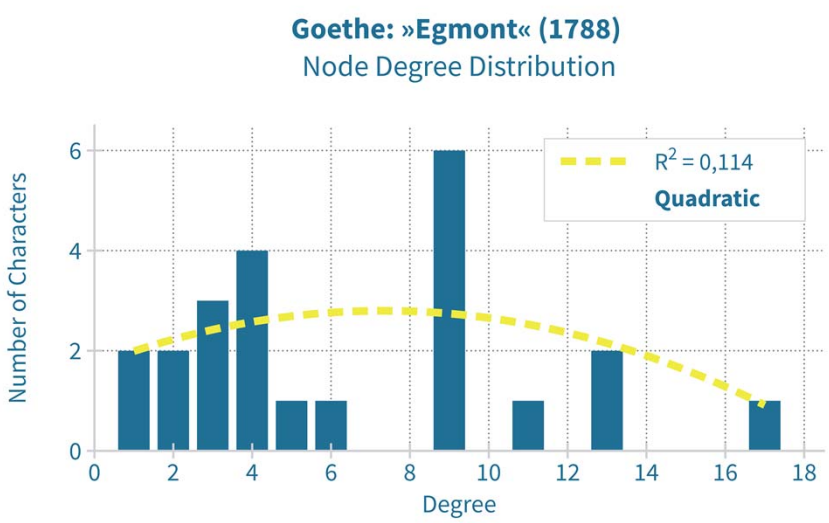
Moretti 2011, 4 on the Node Degree Distribution in »Hamlet«

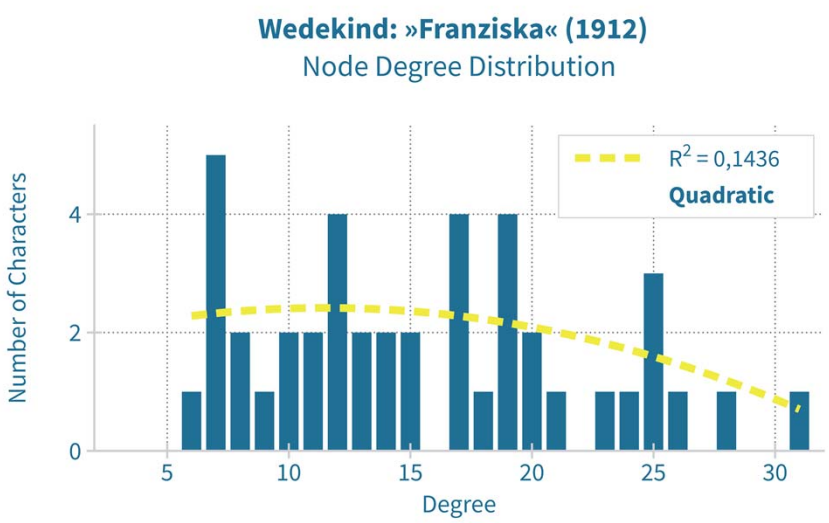
»The take-home message of this rant will be that **the universe counts in powers rather than linear progressions**, and thus in most cases a power law is not so much surprising as it is overwhelmingly expected. Reporting power laws in your data is a bit like reporting furry ears on your puppy; often true, but not terribly useful.«

Weingart 2012 in his »Power Law Rant«


Degree – exponential
Degree – powerlaw
Degree – quadratic
19.35
10.54
70.11
Closeness – exponential
Closeness – powerlaw
Closeness – quadratic
15.48
6.67
77.85
Betweennesss – exponential
Betweennesss – powerlaw
Betweennesss – quadratic
0.00
38.49
61.51
Weighted Degree –exponential
Weighted Degree –powerlaw
Weighted Degree – quadratic
11.83
17.85
70.32
Eigenvector – exponential
Eigenvector – powerlaw
Eigenvector – quadratic
4.95
13.33
81.72
Frequency – exponential
Frequency – powerlaw
Frequency – quadratic
4.52
23.66
71.83
Speech Acts – exponential
Speech Acts – powerlaw
Speech Acts – quadratic
9.25
37.20
53.55
Words – exponential
Words – powerlaw
Words – quadratic
7.53
46.24
46.24
0
10
20
30
40
50
60
70
80
90
Percent

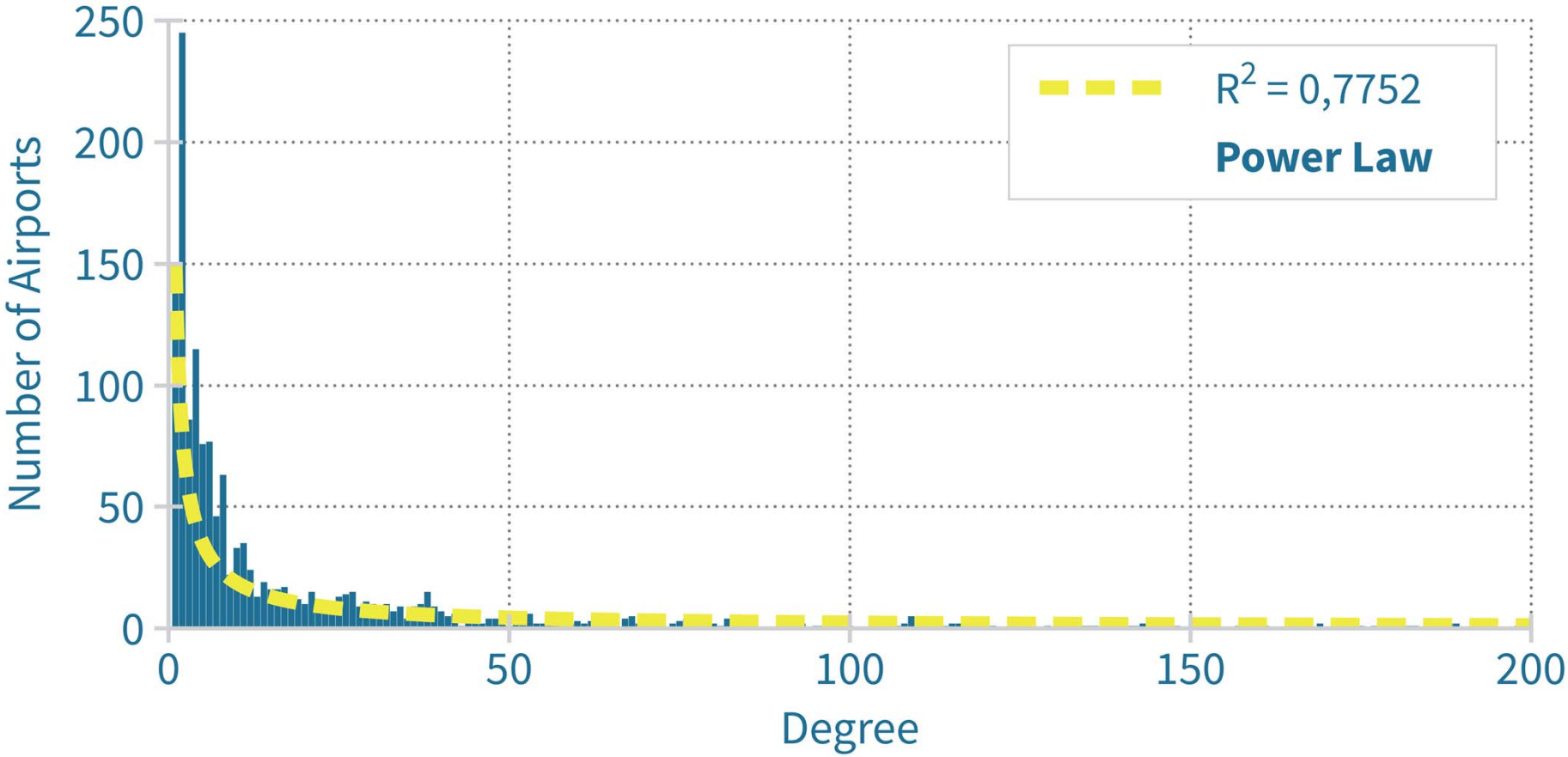

Beer: »Struensee« (1828)
Node Degree Distribution
Number of Characters
10
5
0
0
2
4
6
8
10
12
14
16
Degree
R2 = 0,1695
Quadratic

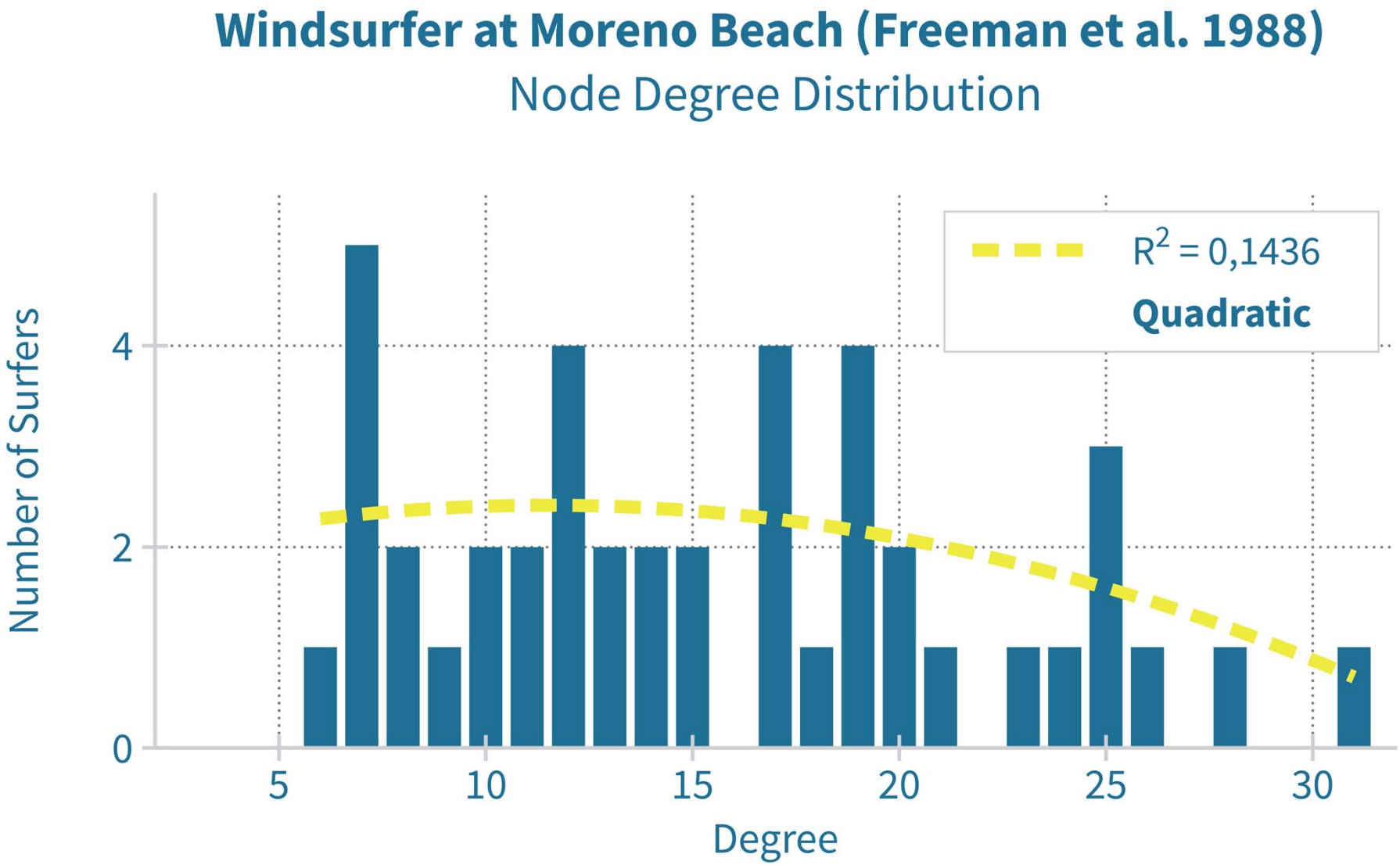

Goethe: »Egmont« (1788)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,114
Quadratic
0
2
4
6
8
10
12
14
16
18

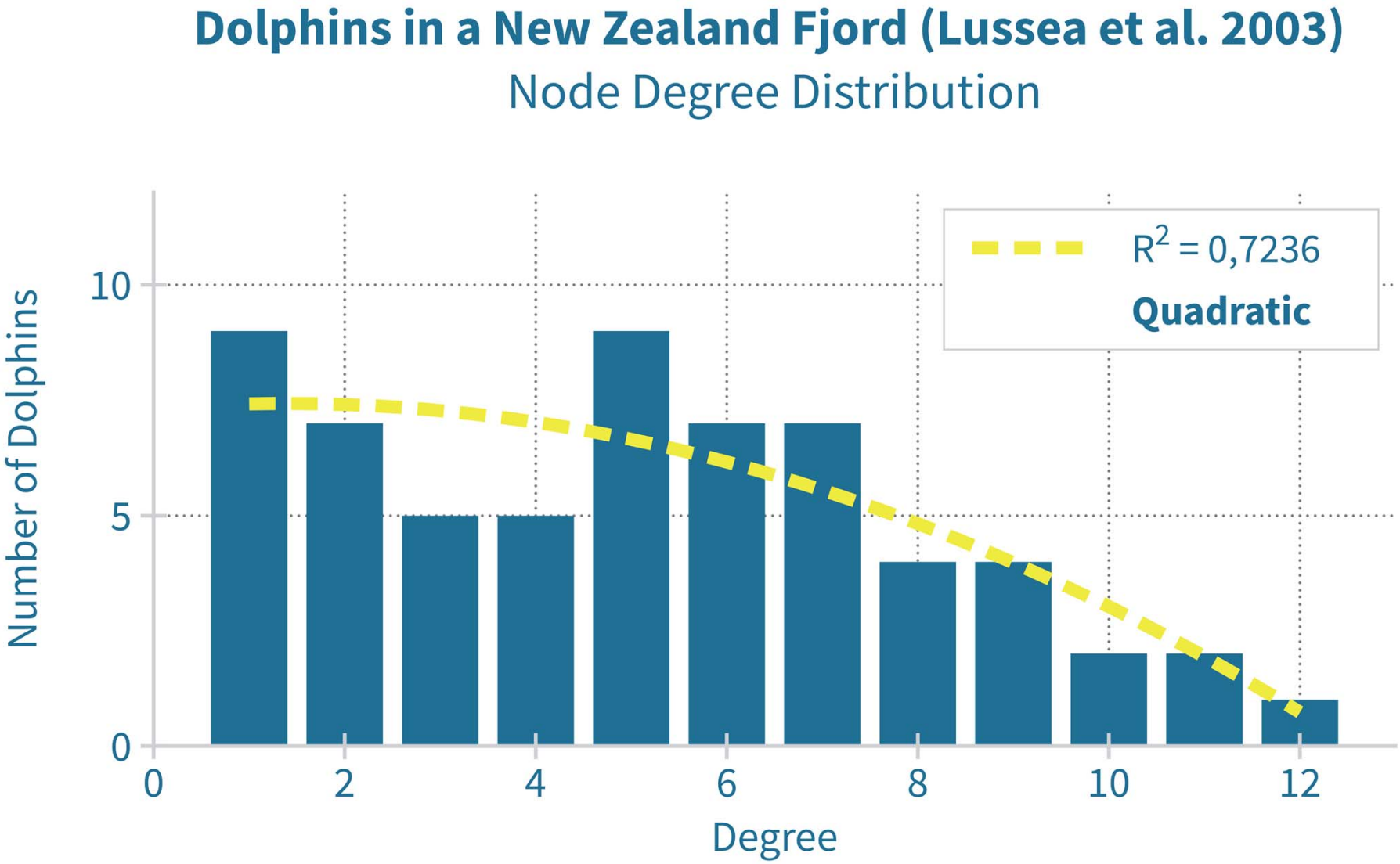

Wedekind: »Franziska« (1912)
Node Degree Distribution
Number of Characters
Degree
R^2 = 0,1436
Quadratic
0
2
4
5
10
15
20
25
30

Flights between US Airports in 2010 (Data cf. Opsahl 2011)

Node Degree Distribution




Windsurfer at Moreno Beach (Freeman et al. 1988)
Node Degree Distribution
Number of Surfers
0
2
4
5
10
15
20
25
30
Degree
R2 = 0,1436
Quadratic


Dolphins in a New Zealand Fjord (Lussea et al. 2003)
Node Degree Distribution
Number of Dolphins
10
5
0
0
2
4
6
8
10
12
Degree
R2 = 0,7236
Quadratic

Summary

- diversified discourse on quantitative dominance relations in drama
- **multi-dimensionality is key** – it makes no sense to rely on just one measure and declare that the killer measure
- power-law postulate has to be rejected
- our search for the protagonist also highlighted the role of middle characters as important part of the social fabric of a play
- research on dominance relations has to be as multidimensional as possible, only then it can be the basis for interpretive literary studies

Literature

- Albert & Barabási 2002 – Réka Albert & Albert-László Barabási: Statistical Mechanics of Complex Networks. In: Reviews of Modern Physics 74 (2002), 47–97.
- Algee-Hewitt 2017 – Mark Algee-Hewitt : Distributed Character: Quantitative Models of the English Stage, 1500-1920. In: Digital Humanities 2017. Conference Abstracts. McGill University & Université de Montréal 2017, 119–121.URL: https://dh2017.adho.org/abstracts/103/103.pdf
- Barabási 2018 – Albert-László Barabási: All You Need Is Love [Blogpost]. 6.3.2018. URL: https://www.barabasilab.com/post/love-is-all-you-need
- Broido & Clauset 2018 – Anna D. Broido & Alan Clauset: Scale-free Networks Are Rare [Preprint]. 9.1.2018. URL: https://arxiv.org/abs/1801.03400
- Fischer & Trilcke 2015 – Frank Fischer & Peer Trilcke: Introducing dlina Corpus 15.07 (Codename: Sydney) [Blogpost]. 20.6.2015. URL: https://dlina.github.io/Introducing-DLINA-Corpus-15-07-Codename-Sydney/
- Freeman et al. 1988 – Linton C. Freeman, Sue C. Freeman & Alaina G. Michaelson. On Human Social Intelligence. In: Journal of Social and Biological Structures 11.4 1988, 415–425.
- Kampkaspar, Fischer, Göbel & Trilcke 2015 – Dario Kampkaspar, Frank Fischer, Mathias Göbel & Peer Trilcke: The Biggest Chatterbox in German Literature [Blogpost]. 23.6.2015. URL: https://dlina.github.io/The-Biggest-Chatterbox-in-German-Literature/
- Kittel & Fischer 2017 – Christopher Kittel & Frank Fischer: dramavis (v0.4). September 2017. URL: https://github.com/lehkost/dramavis.
- Lussea et al. 2003 – D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten & S. M. Dawson: The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-Lasting Associations. In: Behavioral Ecology and Sociobiology 54 (2003), 396–405.
- Moretti 2011 – Franco Moretti: Network Theory, Plot Analysis. Literary Lab Pamphlet 2. May 2011. URL: http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf.
- Moretti 2013 – Franco Moretti: „Operationalizing“: or, the function of measurement in modern literary theory. Literary Lab Pamphlet 6. December 2013. URL: https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf
- Pfister [9]1997 – Manfred Pfister: Das Drama. Theorie und Analyse. 9. Aufl. München 1997.
- Trilcke, Fischer, Göbel, Kampkaspar 2016 – Peer Trilcke, Frank Fischer, Mathias Göbel & Dario Kampkaspar: Theatre Plays as ‚Small Worlds‘? Network Data on the History and Typology of German Drama, 1730–1930. In: Digital Humanities 2016. Conference Abstracts. Jagiellonian University & Pedagogical University, Kraków 2016, 385–387.
- Weingart 2012 – Scott Weingart: Power Law Rant [Blogpost]. 8.6.2012. URL: http://www.scottbot.net/HIAL/index.html@p=17824.html

Supporting Network Data

- Freeman et al. 1988 – http://konect.uni-koblenz.de/networks/moreno_beach
- Lussea et al. 2003 – http://konect.uni-koblenz.de/networks/dolphins
- Opsahl 2011 – http://konect.uni-koblenz.de/networks/opsahl-usairport