

# Vergleich von Topic Modeling Verfahren

# Fragestellungen

- \* Gibt es Besonderheiten von Online Partizipationsverfahren?
- \* Wie werden die Daten am besten verarbeitet?
- \* Welche Topic Modeling Verfahren funktionieren besonders gut?

# Vorverarbeitung

Named Entities  
zu einem Token



Verschiedene Schreibweisen auf  
gleiches Lemma bringen

Wenn eines dieser  
Wörter am Ende

steht:

- \* strasse
- \* straÙe
- \* str
- \* str.
- \* platz
- \* gasse
- \*allee
- \* ufer
- \* weg

Hier nur:

- \* str. -> straÙe
- \* str -> straÙe
- \* strasse -> straÙe
- \* Bei einem Leerzeichen  
zusammenfügen
- \* Bei mehr, mit Bindestrichen  
verbinden
- \* Großschreibung korrigieren

# Vorverarbeitung

## Beispiele

- (1) ,endenicher', 'str.'
- (2) ,endenicher str.'
- (3) ,Endenicherstraße'

- (1) ,endenicher', 'strasse'
- (2) ,endenicher strasse'
- (3) ,Endenicherstraße'

- (1) ,clemens', 'august', 'strasse'
- (2) ,clemens august straße'
- (3) ,Clemens-August-Straße'

- (1) ,clemens-august-str'
- (2) ,Clemens-August-Straße'

# Probleme

## Bei der Vorverarbeitung

- \* Beruht auf spaCy's Erkennung von Named Entities
- \* False Positive z.B. **Der-Kleine-Mann-Von-Der-Straße**
- \* Es wird nicht nur ‚Straße‘ abgekürzt
- \* Kreuzungen (Straßennamen mit / verbunden)
- \* Mögliche Lösung: SpaCy NER mit Straßennamen trainieren
- \* Rechtschreibfehler

# Evaluation

- \* Precision kann leicht ermittelt werden, auch auf großem Datensatz
- \* Um Recall zu berechnen muss Datensatz annotiert werden
- \* Betrachtung nach Anzahl des Auftretens insgesamt
- \* Fehlerhafte Token treten häufig nur einmal auf

# Evaluation

	<i>Raddialog Bonn</i>	<i>Raddialog Köln-Ehrenfeld</i>	<i>Bonn 2011</i>
<i>Ersetzungen insgesamt</i>	1318	430	242
<i>Precision insgesamt</i>	0.88	0.94	0.83

Falsch sind nur Tokens, deren Semantik durch die Bearbeitung verfälscht wurden

z.B.: **Kaugummis-Von-Der-Straße**

Nicht aber: **Blu-Ray-Disc**

# Evaluation

<i>Raddialog Bonn</i>	
Token	# Ersetzungen
Endenicherstraße	103
Poppelsdorferallee	86
Königswintererstraße	74
Bornheimerstraße	54
Bertha-Von-Suttner-Platz	35
Siegburgerstraße	32
Niederkasselerstraße	31
Graurheindorferstraße	30
Sankt-Augustiner-Straße	29

# Evaluation

<i>Raddialog Köln-Ehrenfeld</i>	
Token	# Ersetzungen
Venloerstraße	177
Subbelratherstraße	67
Oskar-Jäger-Straße	15
Widdersdorferstraße	12
Bartholomäus-Schink-Straße	4
Wilhelm-Mauser-Straße	4
Alte-Escher-Straße	4
Hugo-Eckener-Straße	4
Mengenicherstraße	3

# Evaluation

*Bonn 2011*

Token	# Ersetzungen
Bertha-Von-Suttner-Platz	8
Poppelsdorferallee	8
Koblenzerstraße	6
Königswintererstraße	5
Siegburgerstraße	4
Pariserstraße	4
Alte-Escher-Straße	3
Blu-Ray-Discs	3
Niederkasselerstraße	3

# Filtering

Nach Häufigkeit in Dokumenten

- \* Häufige Begriffe, d.h. kommen in vielen Dokumenten vor
- \* Seltene Begriffe, d.h. nur in wenigen Dokumenten
  - \* Für ein Thema mindestens 2
- \* Betrachtung nach Stopwortfilterung durch spaCy

# Häufigste Begriffe

Raddialog Bonn	
Token	
bonn	0.89856
radfahrer	0.50365
radweg	0.45946
fahren	0.41098
die	0.39339
das	0.33333
richtung	0.33033
auto	0.29387
der	0.26684

# Häufigste Begriffe

<i>Raddialog Köln-Ehrenfeld</i>	
Token	
köln	0.89153
ehrenfeld	0.61640
radweg	0.41270
die	0.39683
radfahrer	0.35714
fahren	0.29894
auto	0.28836
das	0.27513
dank	0.25926

# Häufigste Begriffe

Bonn 2011	
Token	
die	0.82266
stadt	0.77931
vorschlag	0.72315
bonn	0.70345
verwaltung	0.69754
das	0.64433
ich	0.58424
es	0.56946
sie	0.56749

# Topic Modeling

- \* Non-negative Matrix Factorization
- \* Mit Probabilistic Latent Semantic Indexing als Spezialfall
- \* Latent Dirichlet Allocation
- \* Scikit-learn und Gensim

# Topic Modeling

- \* NMF benutzt TF-IDF Vektoren
- \* LDA TF Vektoren
- \* Jeweils 1000 Features. Bei mehr als 1000 Tokens nur die 1000 häufigsten

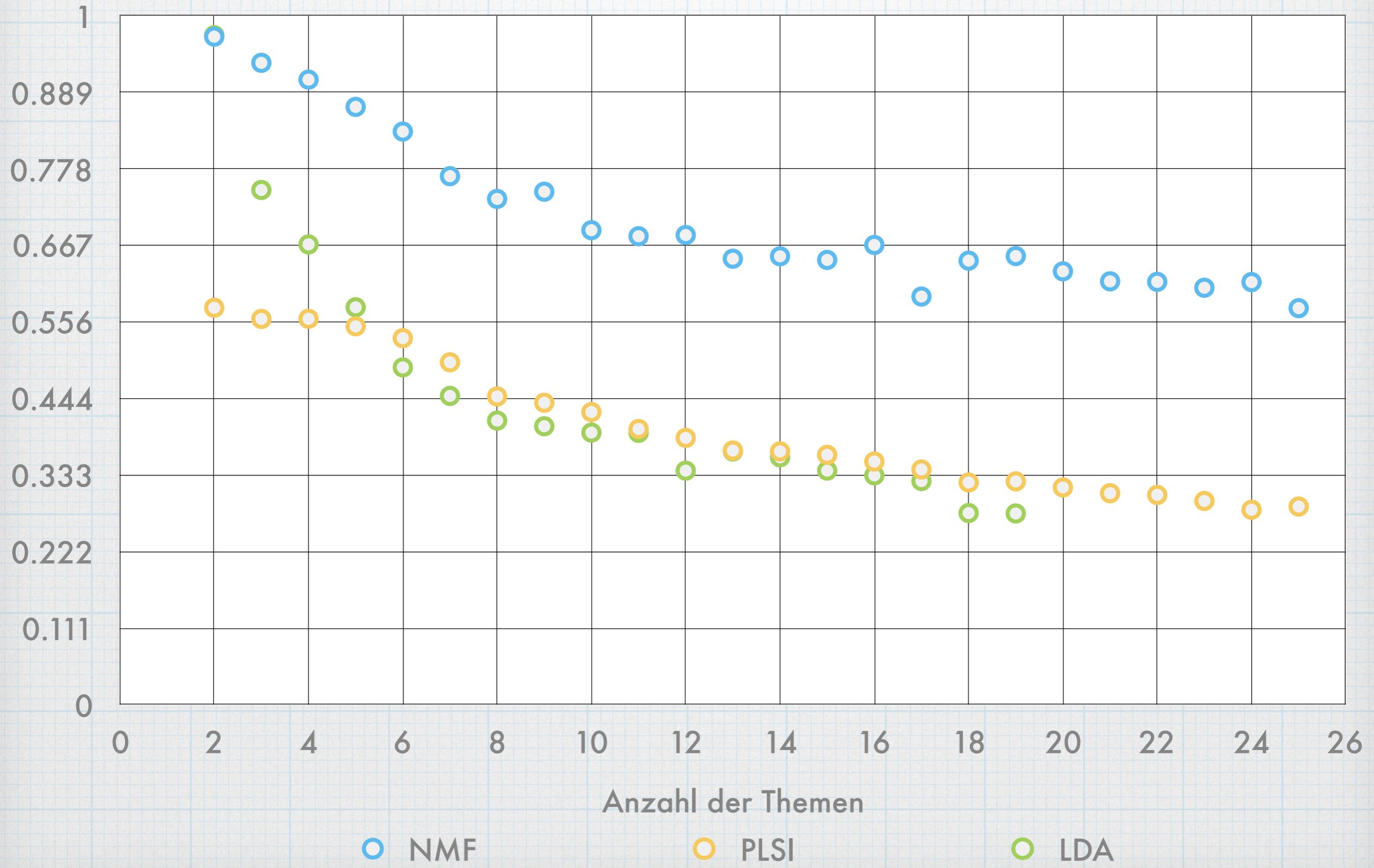
# Finden der optimalen Themen Anzahl

- \* Erstellung von 100 Samples, die je 80 % des Korpus enthalten
- \* Vergleich von Modell mit jedem Sample
- \* Stabilität ist durchschnittlicher Agreement Score über alle Vergleiche
- \* Je besser man die gefundenen Themen in den Samples wiederfindet, desto stabiler
- \* GREENE, Derek; O'CALLAGHAN, Derek; CUNNINGHAM, Pádraig. How many topics? stability analysis for topic models. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2014. S. 498-513.

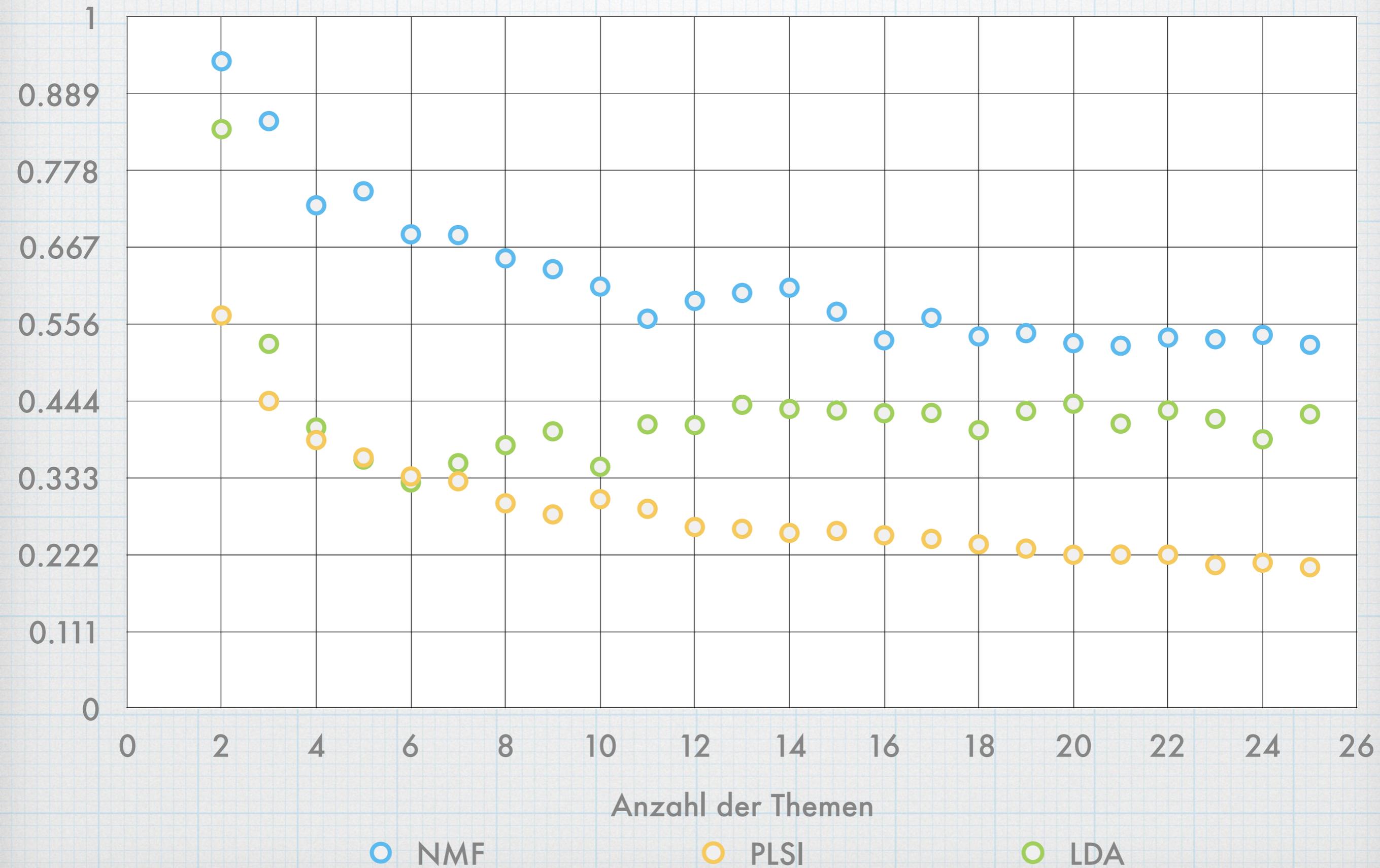
# Agreement Score

- \* Average Jaccard: Durchschnittliche Jaccard Index über die Top n (hier 20) Begriffe
- \* Finde beste Permutation der Themen
- \* Durchschnitt über die Average Jaccard Werte der Permutation

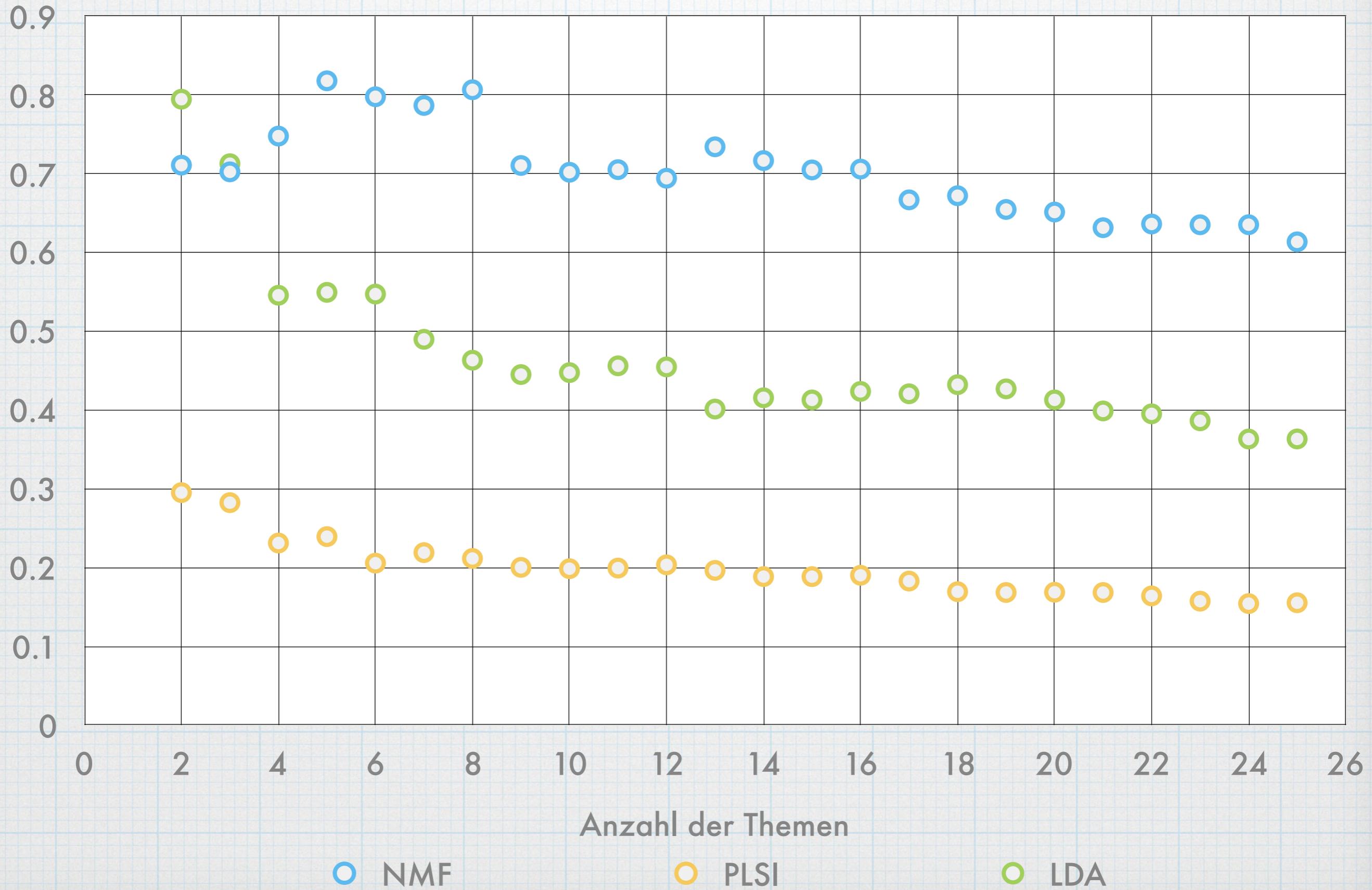
# Raddialog Bonn



# Raddialog Köln-Ehrenfeld



# Bonn 2011



# Manueller Vergleich

- \* NMF, PLSI und LDA
- \* 1000 Features
- \* Token maximal in 90%, aber mindestens 2 Dokumenten
- \* 15 Top Begriffe pro Thema
- \* Stopwortfilterung und Parsen der Straßennamen

# Raddialog Bonn - NMF

Topic #0	kategorie beitrag zuordnen verständnis vorschlag herzlich dank sie ihnen angesprochen auftauchen themenbereich sicherstellen thema richtig
Topic #1	auto schutzstreifen parkend parken radstreifen straße parkplatz überholen abstand tür schmal poppelsdorf parkende knapp autotür
Topic #2	radweg schmal bonn kaiserstraße entlang richtung fußgänger schlecht königswintererstraße zugeparkt regelmäßig straße wurzel fußweg gronau
Topic #3	ampel grün rot kreuzung ampelschaltung warten schalten radfahrer überqueren autofahrer pfeil grünphase fußgängerampel kreisverkehr stehen
Topic #4	ort dank rathaus herzlich beitrag handeln betrachten separat konkret spät einfach da moderation anliegen vorschlag
Topic #5	duisdorf fahrradstraße vorfahrt vorrang fahrradstraßen schlagloch richtung ändern feld einzig bordsteinkante verbinding schön einrichten entfernen
Topic #6	name beitrag entfernt [ firma ] werbung unangemessen betreffend vorbeugen vorwurf umstand missverständnis verständlich dennoch
Topic #7	radfahrer fahren richtung straße autofahrer bonn ich gefährlich kreuzung stelle fahrrad links es fußgänger fahrbahn
Topic #8	weg schmal fußgänger beleuchtung belag rheinaue fahrradweg bonn rhein schlagloch schlecht venusberg zustand nutzen friesdorf

# Raddialog Bonn - NMF

Topic #0	kategorie beitrag zuordnen verständnis vorschlag herzlich dank sie ihnen angesprochen auftauchen themenbereich sicherstellen thema richtig
Topic #1	auto schutzstreifen parkend parken radstreifen straße parkplatz überholen abstand tür schmal poppelsdorf parkende knapp autotür
Topic #2	radweg schmal bonn kaiserstraße entlang richtung fußgänger schlecht königswintererstraße zugeparkt regelmäßig straße wurzel fußweg gronau
Topic #3	ampel grün rot kreuzung ampelschaltung warten schalten radfahrer überqueren autofahrer pfeil grünphase fußgängerampel kreisverkehr stehen
Topic #4	ort dank rathaus herzlich beitrag handeln betrachten separat konkret spät einfach da moderation anliegen vorschlag
Topic #5	duisdorf fahrradstraße vorfahrt vorrang fahrradstraßen schlagloch richtung ändern feld einzig bordsteinkante verbinding schön einrichten entfernen
Topic #6	name beitrag entfernt [ firma ] werbung unangemessen betreffend vorbeugen vorwurf umstand missverständnis verständlich dennoch
Topic #7	radfahrer fahren richtung straße autofahrer bonn ich gefährlich kreuzung stelle fahrrad links es fußgänger fahrbahn
Topic #8	weg schmal fußgänger beleuchtung belag rheinaue fahrradweg bonn rhein schlagloch schlecht venusberg zustand nutzen friesdorf

# Raddialog Bonn - PLSI

Topic #0	vorschlag sie wir moderation dank zuordnen verständnis beitrag bitten um herzlich richtig thema ihnen kategorie
Topic #1	bonn straße auto fahren radfahrer wenn stehen richtung radweg sehen seite gefährlich parkend hier ich
Topic #2	radweg bonn schmal schlecht weg hochkreuz zustand rhein fahrradwegentlang stelle regelmäßig wurzel gronau nutzen
Topic #3	ampel richtung kreuzung überqueren grün radfahrer fußgänger wartenrot radweg gronau kommend zebrastreifen fußgängerampel schalten
Topic #4	bonn bonn-zentrum es stadt rad fahrrad sicher unterführung ich fehlen in schaffen radfahrer einfach mal
Topic #5	richtung fahrradstraße radfahrer straße freigeben duisdorf radverkehr sinnvoll vorfahrt einbahnstraße verbindung fahren kessenich strecke fahrradfahrer
Topic #6	sicht poller zugeparkt godesberg bad stehen entfernt verstehen parken bleiben sicherlich sinn vorbeugen plittersdorf regelmäßig
Topic #7	radfahrer richtung fahren kommend radweg links rechts gefährlich straße hier nordstadt Autofahrer kreuzung situation schild
Topic #8	weg radfahrer fahren fußgänger stelle endenich schmal gefährlich hier schlagloch poppelsdorf schnell venusberg überholen mal

# Raddialog Bonn - PLSI

Topic #0	vorschlag sie wir moderation dank zuordnen verständnis beitrag bitten um herzlich richtig thema ihnen kategorie
Topic #1	bonn straße auto fahren radfahrer wenn stehen richtung radweg sehen seite gefährlich parkend hier ich
Topic #2	radweg bonn schmal schlecht weg hochkreuz zustand rhein fahrradwegentlang stelle regelmäßig wurzel gronau nutzen
Topic #3	ampel richtung kreuzung überqueren grün radfahrer fußgänger wartenrot radweg gronau kommend zebrastreifen fußgängerampel schalten
Topic #4	bonn bonn-zentrum es stadt rad fahrrad sicher unterführung ich fehlen in schaffen radfahrer einfach mal
Topic #5	richtung fahrradstraße radfahrer straße freigeben duisdorf radverkehr sinnvoll vorfahrt einbahnstraße verbindung fahren kessenich strecke fahrradfahrer
Topic #6	sicht poller zugeparkt godesberg bad stehen entfernt verstehen parken bleiben sicherlich sinn vorbeugen plittersdorf regelmäßig
Topic #7	radfahrer richtung fahren kommend radweg links rechts gefährlich straße hier nordstadt autofahrer kreuzung situation schild
Topic #8	weg radfahrer fahren fußgänger stelle endenich schmal gefährlich hier schlagloch poppelsdorf schnell venusberg überholen mal

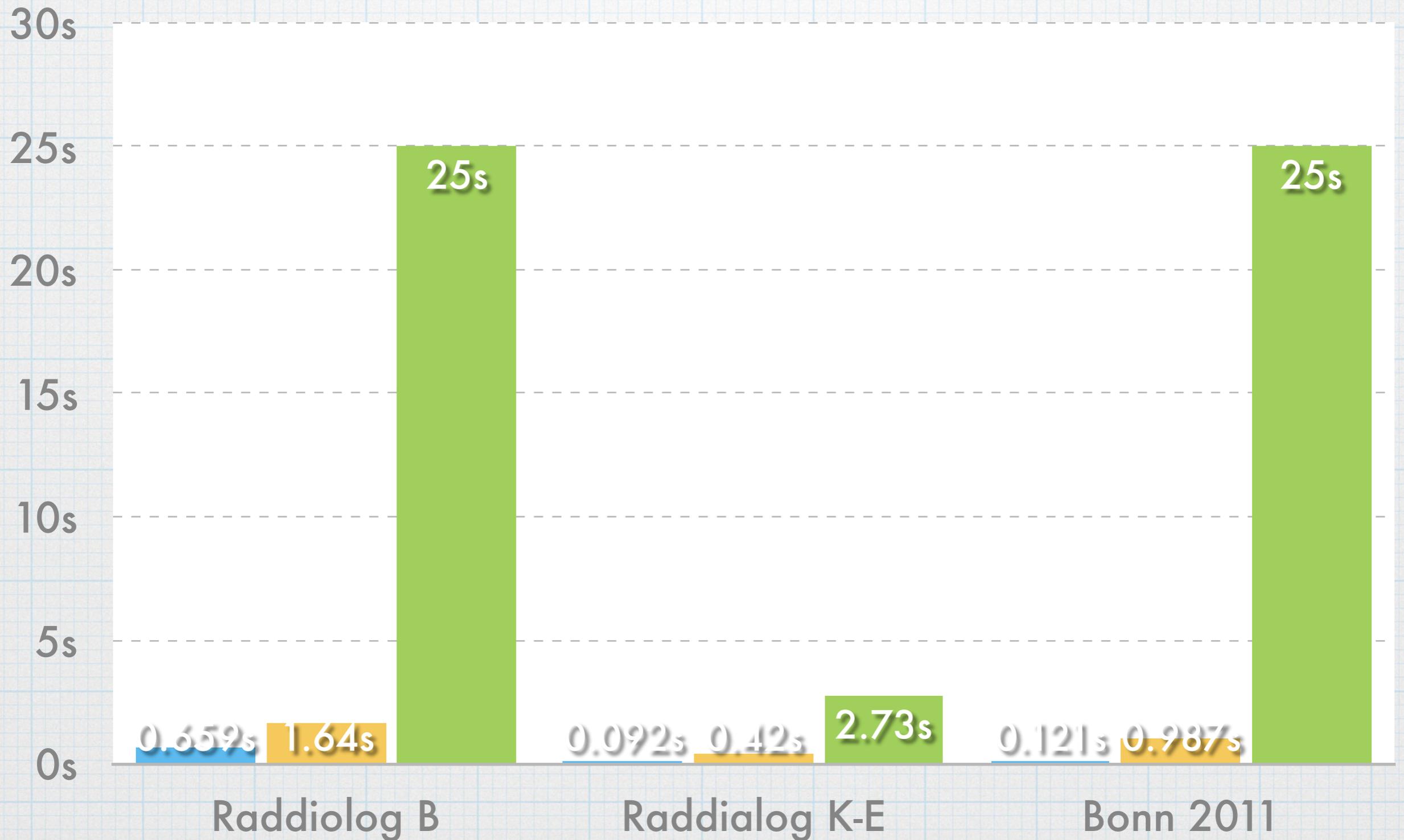
# Raddialog Bonn - LDA

Topic #0	kreisel bushaltestelle nordstadt altstadt westlich landgrabenweg heerstraße fehlende telekom aufstellen schutzstreifen bonn vernünftig gehweg radverkehrsführung
Topic #1	fahrradstraße bonn einbahnstraße fahrradstraßen richtung duisdorf freigeben straße radfahrer schild fahrradfahrer frei vilich radverkehr schwarzerheindorf/vilich-rheindorf
Topic #2	poppelsdorf bonn venusberg endenich übergang straße weg gehweg gefährlich siegburgerstraße ippendorf hügel auffahrt fahrradweg robert-koch-straße
Topic #3	bonn radfahrer bonn-zentrum unterführung stadt fahrrad brücke rad einfach innenstadt mal hauptbahnhof viktoriabrücke platz bonner
Topic #4	beitrag kategorie herzlich vorschlag verständnis zuordnen moderation auswertung bitten thema aufheben angesprochen sicherstellen auftauchen themenbereich
Topic #5	auto radweg fahren straße bonn radfahrer autofahrer schutzstreifenparkend gefährlich parken fahrbahn parkplatz poller überholen
Topic #6	fehlen bordstein bordsteinkante absenken absenkung kaiser-karl-ring zum teil vorhanden und/oder westlich unfallgefahr kreuzungsbereich mittlerweile lengsdorf angesichts
Topic #7	radfahrer ampel kreuzung richtung fahren bonn autofahrer kommend grün abbiegen links fußgänger überqueren rot auto
Topic #8	radweg bonn radfahrer fahren weg fußgänger richtung schmal straße stelle gefährlich fahrradweg fahrbahn entlang nutzen

# Raddialog Bonn - LDA

Topic #0	kreisel bushaltestelle nordstadt altstadt westlich landgrabenweg heerstraße fehlende telekom aufstellen schutzstreifen bonn vernünftig gehweg radverkehrsführung
Topic #1	fahrradstraße bonn einbahnstraße fahrradstraßen richtung duisdorf freigeben straße radfahrer schild fahrradfahrer frei vilich radverkehr schwarzerheindorf/vilich-rheindorf
Topic #2	poppelsdorf bonn venusberg endenich übergang straße weg gehweg gefährlich siegburgerstraße ippendorf hügel auffahrt fahrradweg robert-koch-straße
Topic #3	bonn radfahrer bonn-zentrum unterführung stadt fahrrad brücke rad einfach innenstadt mal hauptbahnhof viktoriabrücke platz bonner
Topic #4	beitrag kategorie herzlich vorschlag verständnis zuordnen moderation auswertung bitten thema aufheben angesprochen sicherstellen auftauchen themenbereich
Topic #5	auto radweg fahren straße bonn radfahrer autofahrer schutzstreifenparkend gefährlich parken fahrbahn parkplatz poller überholen
Topic #6	fehlen bordstein bordsteinkante absenken absenkung kaiser-karl-ring zum teil vorhanden und/oder westlich unfallgefahr kreuzungsbereich mittlerweile lengsdorf angesichts
Topic #7	radfahrer ampel kreuzung richtung fahren bonn autofahrer kommend grün abbiegen links fußgänger überqueren rot auto
Topic #8	radweg bonn radfahrer fahren weg fußgänger richtung schmal straße stelle gefährlich fahrradweg fahrbahn entlang nutzen

# Laufzeiten



# Raddialog Bonn - NMF mit weniger Vorverarbeitung

Topic #0	sein ich und werden einen nicht zu auch sich mit für in können dieradfahrer
Topic #1	kategorie ihr beitrag zuordnen verständnis ihren herzlich dank vorschlag mein um sie haben in auswertung
Topic #2	radweg auf schmal sein und schlecht bonn sehr zu am straße weg fußgänger an viel
Topic #3	man auf fahren straße links abbiegen aus kommend kreuzung radfahrer richtung kommen von rechts dann
Topic #4	auto schutzstreifen parkend parken auf radstreifen 53115 straße poppelsdorf im parkplatz tür neben überholen abstand
Topic #5	richtung für in einbahnstraße beid radfahrer freigeben bonn sein werden einen sollen kölnstraße fahrradstraße kessenich
Topic #6	duisdorf 53123 fahrradstraße vorfahrt vorrang schlagloch feld einzig ändern die bordsteinkante zur zufahrt radschnellweg 50
Topic #7	beitrag ihr name [ entfernt herzlich dank firma ] sicht anliegen unangemessen werbung betreffend vorbeugen
Topic #8	ampel grün rot ampelschaltung schalten an warten kreuzung für radfahrer pfeil lang grünphase überqueren kreisverkehr

# Spezifische Stoppwörter

[	sicherstellen	
werbung	sieihnen	
anliegen	ort	
verständnis	missverständnis	vorbeugen
dennoch	themenbereich	auswertung
auftauchen	name	aufheben
konkret	beitrag	bitten
herzlich	unangemessen	gast
betreffend	vorwurf	lieb
verständlich	handeln	nowzamani
zuordnen	richtig	wähler
spät	einfach	hallo
thema	umstand	radwegqualität
separat	vorschlag	vorschlagen
da	firma	sonstiges
kategorie	rathaus	thematisieren
entfernt	moderation	
betrachten	dank	
]	angesprochen	

# Raddialog Bonn - NMF mit spezifischer Stopwörtern

Topic #0	radfahrer fahren richtung straße Autofahrer Bonn geben gefährlich Stelle Kreuzung fahrrad links sehen fahrbahn kommend
Topic #1	radweg schmal Bonn kaiserstraße entlang richtung fußgänger straße Königswintererstraße zugeparkt Stelle regelmäßig Gronau enden Zustand
Topic #2	auto schutzstreifen parkend parken parkplatz straße radstreifen überholen abstand tür Poppelsdorf fahren parkende schmal knapp
Topic #3	unterführung poppelsdorferallee königstraße Bonn-Zentrum Hauptbahnhof Bonn kaiserplatz tunnel rampe radfahrer beuel-mitte beueler poppelsdorfer spiegel schieben
Topic #4	fahrradstraße fahrradstraßen kaiserstraße adelheidsstraße radverkehrsführung Bonn vilich straße florentiusgraben sinnvoll Vorfahrt anlieger einrichten tempo einrichtung
Topic #5	ampel grün rot kreuzung ampelschaltung warten schalten radfahrer überqueren pfeil Autofahrer grünphase fußgängerampel kreisverkehr abbiegen
Topic #6	godesberg bad alt-godesberg fehlen villenviertel plittersdorf fahrradstände blau bonnerstraße fahrtrichtung vorhanden fahrrad beschilderung bemängeln richtung
Topic #7	duisdorf maarweg vorfahrt vorrang schlagloch feld richtung einzig kommend berg bordstein Zufahrt bordsteinkante radschnellweg verbindung
Topic #8	weg fußgänger schmal beleuchtung rheinaue belag Bonn rhein fahrradweg venusberg schlagloch friesdorf nutzen Zustand dunkel

# Raddialog Bonn - NMF mit 18 Themen

Topic #0	radfahrer fahren straße Autofahrer gefährlich fahrbahn stelle geben richtung sehen gehweg bonn situation mal stadt
Topic #1	radweg schmal bonn richtung entlang königswintererstraße kaiserstraße zugeparkt straße enden regelmäßig zustand stelle gronau häufig
Topic #2	auto parkend parken parkplatz straße radstreifen parkende falschparker endenich zugeparkt häufig reihe poppelsdorf hindernis bonn
Topic #3	unterführung poppelsdorferallee königstraße bonn kaiserplatz rampetunnel bonn-zentrum beuel-mitte radfahrer beueler hauptbahnhof poppelsdorfer spiegel hindernis
Topic #4	fahrradstraße fahrradstraßen kaiserstraße adelheidisstraße bonn radverkehrsführung vilich florentiusgraben sinnvoll straße einrichtenanlieger vorfahrt tempo einrichtung
Topic #5	ampel grün rot warten ampelschaltung schalten pfeil radfahrer überqueren grünphase Autofahrer stehen kreisverkehr fußgängerampel rechtsabbieger
Topic #6	fahrradständer fahrrad bonn-zentrum rad bonn fehlen geben parkplatz stellplatz sicher abstellen hauptbahnhof innenstadt stadt anschließen
Topic #7	duisdorf maarweg vorrang vorfahrt schlagloch feld richtung einzig kommend bordstein berg zufahrt bordsteinkante radschnellweg verbindung
Topic #8	weg schmal beleuchtung belag bonn rheinaue venusberg rhein schlagloch friesdorf zustand dunkel nutzen kottenforst geben
Topic #9	schutzstreifen parkend streife pkw abstand tür überholen linie sicherheitsabstand schmal direkt autotür breit knapp poppelsdorf
Topic #10	einbahnstraße freigeben richtung gegenrichtung radfahrer kessenichradverkehr fahrradfahrer frei parkplatz bonn entgegen öffnen nordstadt freigabe
Topic #11	poller entfernen stehen unfallgefahr durchfahrt einfahren b9 drängelgitter dunkelheit godesbergerallee hindernis endenich behindern eng bonn
Topic #12	godesberg bad alt-godesberg villenviertel fehlen plittersdorf blaubonnerstraße fahrtrichtung bemängeln richtung fahrbahn vorhanden behindern selten
Topic #13	kreuzung links abbiegen kommend rechts fahrradfahrer richtung gefährlich linksabbiegen geradeaus überqueren verkehrsführung Autofahrer kreuzungsbereich linksabbieger
Topic #14	fußgänger radfahrer bonn-zentrum heck fußweg schmal zebrastreifen fußgängerzone überqueren fußgängerweg grünphase konflikt friedrichstraße ampelschaltung rhein
Topic #15	schild beschilderung vorfahrt fehlen blau aufstellen frei kennzeichnen einfahrt hinweis nehmen verbieten radverkehrsführung stehen radfahrer
Topic #16	fahrradweg kaiserstraße fahrradfahrer kaiserstr verbreitern gronaubushaltestelle neu bonn adenauallee radverkehrsführung eng stelle entlang richtung
Topic #17	kölnstraße bonn-castell rosental nordstadt kaiser-karl-ring heerstraße bordstein richtung lievelingsweg absenken bonn-zentrum bonn geben stadteinwärts kommend

# Agreement Scores

	NMF	PLSI	LDA
NMF	1.0		
PLSI	0.36281015	1.0	
LDA	0.29462090	0.31503636	1.0
NMF weniger Vorverarbeitung	0.04579244		
NMF mit Stoppwörter	0.05913829		

# Fazit

- \* NMF > LDA > PLSI
- \* Stabilität eher nicht geeignet
- \* Adresse bei ortsbezogenen Beiträgen wichtig
- \* Beteiligung mit einbeziehen