

# Discourse Network Analysis

Philip Leifeld

Max Planck Institute for Research on Collective Goods, Bonn

Compon Workshop, IHDP Open Meeting, Bonn, 26 April 2009



## Analysis of political discourse

- actor-centered approaches: advocacy coalitions, discourse coalitions, epistemic communities, etc.
- content-related approaches: critical discourse analysis, qualitative data analysis (QDA), semantic network analysis, etc.
- missing link: relations between actors and the contents of the discourse
- idea: discourse networks
- combining content analysis and social network analysis

The logo for Discourse Network Analyzer features a stylized 'D' and 'N' formed by a grid of colored circles. The 'D' is composed of blue circles, and the 'N' is composed of orange circles. 

# Discourse Network Analyzer

- name: *Discourse Network Analyzer (DNA)*
- download: <http://www.philipleifeld.de>
- operating system: any (platform-independent!)
- requirements: Java 6 (*DNA* will not work with Java 4 or 5!)
- purpose:
  - ① assign tags to text data
  - ② convert these structured data into networks

## Encoding statements

- 1 Collect text material, e.g. newspaper articles
- 2 Go manually through the text
- 3 Whenever an actor makes a statement, add four tags to this text portion:
  - person
  - organisation
  - category
  - agreement

A statement is a part of the text where an actor expresses his beliefs or solution concepts for the problem.

## Coding examples

**Actor 1:** *“The pension gap should be closed by inviting migrants to settle down in our country.”*

- category: immigration
- agreement: yes

**Actor 2:** *“The number of immigrants does not affect the old-age ratio at all; increasing women’s share in the labour market should be our primary goal.”*

- category: immigration
- agreement: no
- category: increase women’s labour force participation
- agreement: yes

**Actor 3:** *“Gender equality is good, but we don’t have enough jobs anyway, and if more women work, fertility will go down and make the problem even worse.”*

- category: increase women’s labour force participation
- agreement: no
- category: increase fertility
- agreement: yes

## Data format

DNA is based on an XML file format:

```
<discourse>
  <article day="" month="" year="" title="">
    <statement person="" organization="" category=""
      agreement=""> ... </statement>
  </article>
</discourse>
```

title	day	month	year
Renten-Gerede	26	08	1993
Verlängerung der Lebensarbeitszeit bleibt Thema	27	08	1993
Rentendebatte mit allen alten Argumenten	01	09	1993
Der Wirtschaftsstandort Deutschland soll gestärkt werden	01	09	1993
Scharfe Kritik am Bonner Standortpapier	02	09	1993
Bundeskanzler Kohl: Deutschland steht vor großen gesellschaftlichen Aufgaben	04	09	1993
Des Kanzlers Wahrheit	04	09	1993
Blüm: Wirtschaft ist für Rentenproblem verantwortlich	06	09	1993
DIE HAUSHALTS-DEBATTE IM BUNDESTAG	09	09	1993

Finanzierung der betrieblichen Sozialpläne für vorzeitig ausscheidende Beschäftigte herangezogen werden.</statement> <statement person="Ruland, Franz" organization="VDR" category="Deutsche Einheit über RV finanzieren" agreement="no">Die Rentenversicherung solle nicht für die Folgen des SED-Unrechts aufkommen müssen.</statement> <statement person="Ruland, Franz" organization="VDR" category="Beiträge zur Rentenversicherung anheben" agreement="yes">Kritisch bewertet Ruland die Tatsache, daß 1992 und 1993 gegen den Rat der Rentenversicherung zweimal die Beitragssätze gesenkt worden sind. Der Beitragsanteil, den die Rentenversicherung für die Pflegeversicherung von 1994 an zu zahlen habe, trage zu dem Anstieg des Beitragssatzes auf 19,2 Prozent bei.</statement>

In einem Fernsehgespräch hat Bundeskanzler Kohl nicht mehr davon gesprochen, daß die Renten tabu seien. Niemand hätte den heute Dreißigjährigen sagen, wie im Jahr 2030 die Welt und seine Alters werden. <statement person="Ost, Friedhelm" organization="CDU" category="Grundrente" agreement="yes">Der CDU-Abgeordnete Ost sprach sich für eine variable Grundsicherung aus.</statement>, <statement person="Ost, Friedhelm" organization="CDU" category="private Zusatzvorsorge" agreement="yes">durch die die Umlagefinanzierung mit einem Kapitalansammlungsverfahren verknüpft werden könnte.</statement>.

<statement person="Klose, Hans-Ulrich" organization="SPD-60p" category="Rentenproblem durch Zuwanderung lösen" agreement="yes">Nach Ansicht des Vorsitzenden der SPD-Bundestagsfraktion, Klose, sind die Renten mit einem Bündel von Maßnahmen zu sichern. Dazu gehöre, daß nach dem Jahr 2000 jährlich etwa 300 000 Einwanderer ins Land geholt werden müßten.</statement> <statement person="Klose, Hans-Ulrich" organization="SPD-60p" category="Frauenanteil an Erwerbstätigen erhöhen" agreement="yes">Auch sei die Beschäftigungsquote für Frauen zu erhöhen.</statement>

Current file: /home/philip/Desktop/kodierung neu.dna

# Inserting a statement

A dialog box titled "Person" with four dropdown menus and two buttons. The dropdown menus are labeled "Person", "Organization", "Category", and "Agreement". The selected values are "Ost, Friedhelm", "CDU", "private Zusatzvorsorge", and "yes". The buttons are "OK" and "Abbrechen".

<b>Person</b>
Ost, Friedhelm
<b>Organization</b>
CDU
<b>Category</b>
private Zusatzvorsorge
<b>Agreement</b>
yes
<input type="button" value="OK"/>
<input type="button" value="Abbrechen"/>

## Advantages

- lists are updated on the fly
- no mistakes → no overestimation of ties

## Categories and sources

### Possible categories

- solution concepts
- the use of language (symbols, metaphors)
- other kinds of beliefs
- other actors
- groups, committees, interlocking directorates

### Sources

- newspaper articles (sampling depends on the algorithm)
- interview transcripts
- position papers
- parliamentary protocols

# Components of the model

## Components

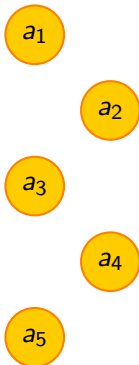
- 1 actors
- 2 concepts
- 3 relations: agreement or disagreement
- 4 discrete time points

## Data storage

- imagine an actor  $\times$  concept matrix
- there is one such matrix for agreement and one for disagreement
- this can be repeated several times at different time steps
- think of it as an  $\mathbf{X}^{a \times c \times r \times t}$  incidence array

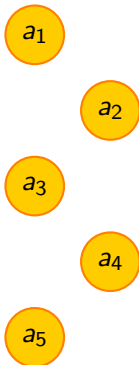
# Illustration of the basic cross-sectional model

actors

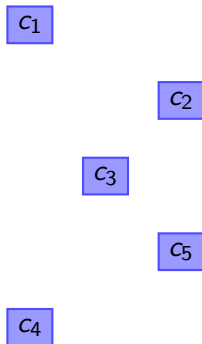


# Illustration of the basic cross-sectional model

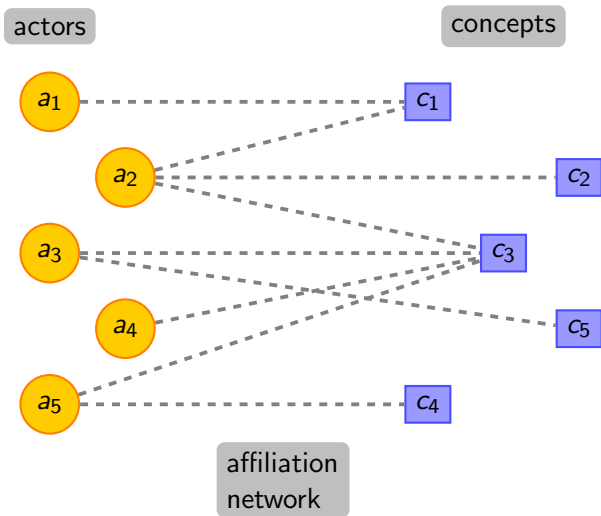
actors



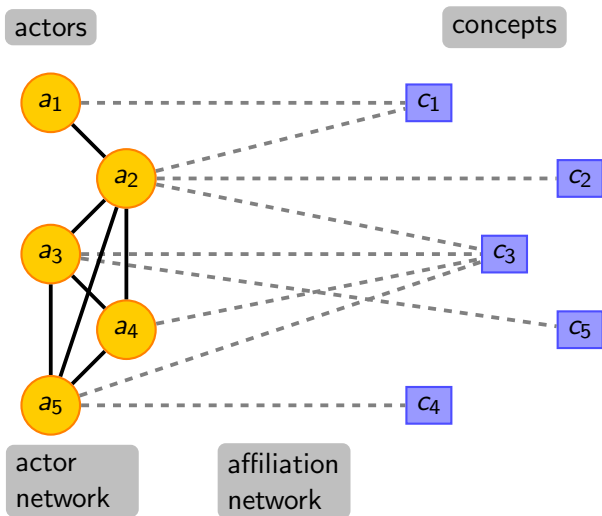
concepts



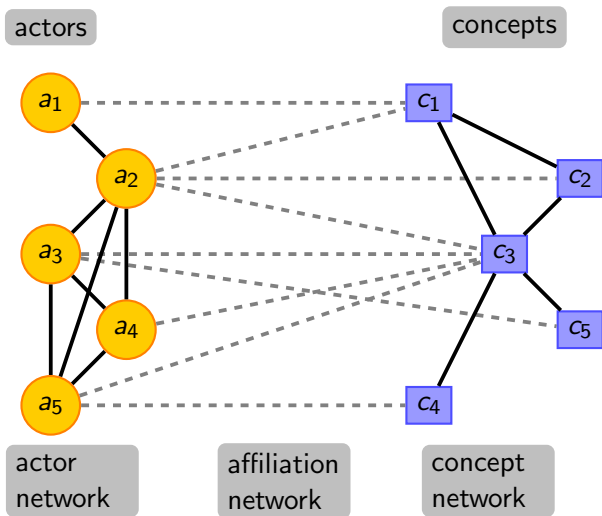
# Illustration of the basic cross-sectional model



# Illustration of the basic cross-sectional model

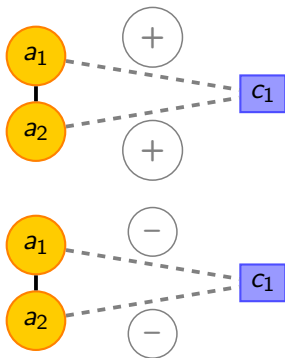


# Illustration of the basic cross-sectional model

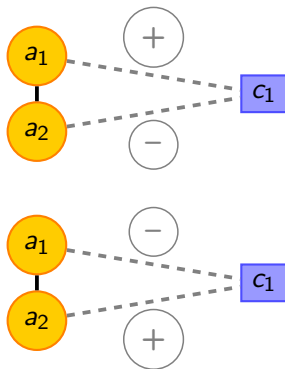


## Extension: agreement and disagreement

congruence networks



conflict networks



## Mathematical notation

- Actors:  $A = \{a_1, a_2 \cdots a_m\}$
- Concepts:  $C = \{c_1, c_2 \cdots c_n\}$
- Discrete time points:  $T = \{t_1, t_2 \cdots t_k\}$
- Relations:  $R = \{r_1 \cdots r_l\}$  where  $l = 2$
- $\forall r : G_{rt}^{\text{aff}} = (A, C, E_r^{\text{aff}})$  with  $\{a, a'\} \notin E_r^{\text{aff}} \wedge \{c, c'\} \notin E_r^{\text{aff}}$
- $(a, c) \neq (c, a) \wedge (c, a) \notin E_r^{\text{aff}}$
- $\forall r : G_{rt}^a = (A, E_r^a)$  with  $\{a, a'\} \in E_r^a$  iff  
 $(a, c) \in E_r^{\text{aff}} \wedge (a', c) \in E_r^{\text{aff}}$
- $\forall r : G_{rt}^c = (C, E_r^c)$  with  $\{c, c'\} \in E_r^c$  iff  
 $(a, c) \in E_r^{\text{aff}} \wedge (a, c') \in E_r^{\text{aff}}$
- $G_t^a = (A, E)$  with  $(a, a') \in E$  iff  $(a, a') \in E_{r_1}^a \vee (a, a') \in E_{r_2}^a$
- $G_t^c = (C, E)$  with  $(c, c') \in E$  iff  $(c, c') \in E_{r_1}^c \vee (c, c') \in E_{r_2}^c$

# Types of discourse networks

- ① affiliation network
- ② (static) co-occurrence network
  - ① congruence network
  - ② concept network
  - ③ conflict network
- ③ dynamic network
  - ① sequential network
  - ② aggregated network
  - ③ referral network

**1 Export format**

CSV  DL edgelist  graphML

**2 Network type**

Organizations ▼ by Organizations ▼ via Categories ▼

**3 Algorithm**

Number of co-occurrences  Time window  Attenuation

**4 Agreement**

yes  no  combined  conflict

**5 Normalization**

yes  no

**6 Time period**

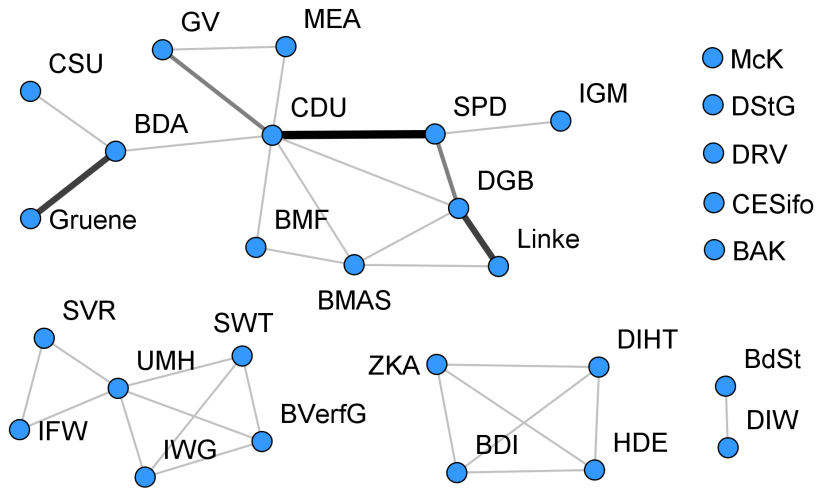
start: 02.01.1993  stop: 29.05.1996

**7 Chaining parameter**

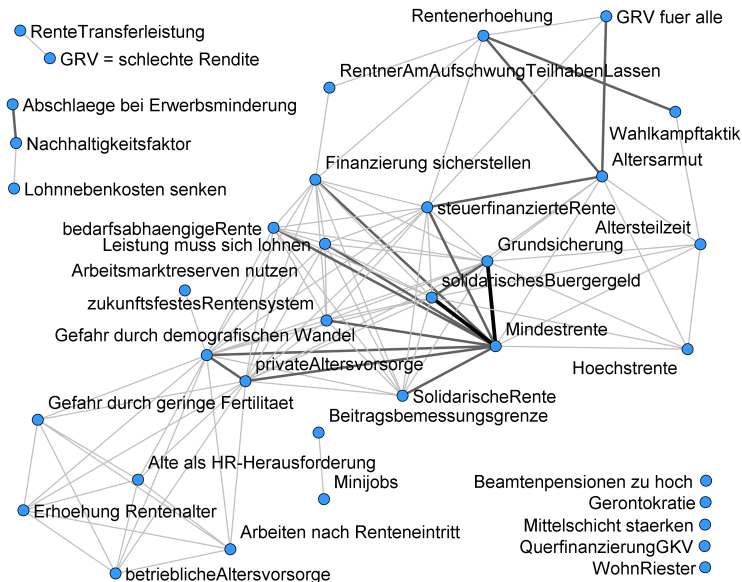
moving time window of  days which is shifted by  days

Calculating...

## Example: Congruence network, May 2008



# Example: Concept network, May 2008



# Sequential networks: Including the time dimension

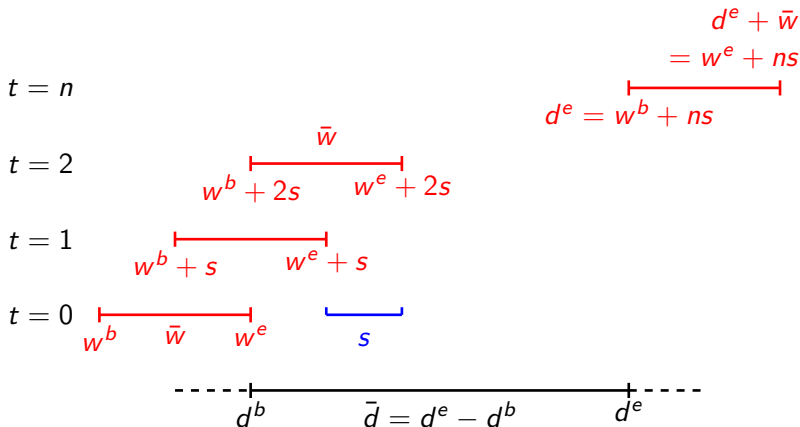
## How do we measure change?

- simple answer: split the duration into equal parts and calculate co-occurrence network for every time step
- compare the time steps and try to identify interesting changes
- use dynamic visualization techniques (SONIA, dynamic visone, Commetrix, Nevada, etc.)

## Problem

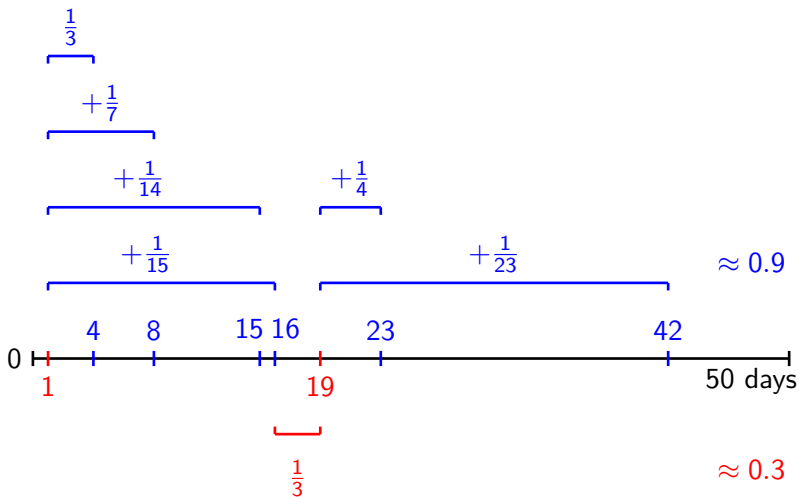
- meaning and context may change over time
- results in overestimation of network ties
- random variation increases this problem

## Refinement: Chaining algorithm



- now this avoids an overestimation of ties
- but actors interact, i.e. they refer to one another

## Refinement: Attenuation algorithm





## Advantages of this approach

- allows to analyse the link between actors and contents
- more systematic than a purely hermeneutic approach
- hermeneutic approaches allow to look at the link between actors and beliefs, but the second step, analysing similarities between actors or between concepts, is a mental challenge
- can be replicated
- you can use the resulting network data with your other network data, e.g. regressing cooperation on belief similarity
- a picture says more than a thousand words
- cross-country comparisons or longitudinal analyses are possible

## Planned features

- new algorithm which connects concepts if they are both cited within  $n$  days
- make *DNA* more user-friendly
- improve normalization algorithms
- search, replace and recode functionality for statements
- window where you can select the actors and categories to include for export

## The software package



what *Discourse Network Analyzer (DNA)*

where <http://www.philipleifeld.de>

who [Leifeld@coll.mpg.de](mailto:Leifeld@coll.mpg.de)

Thanks for your attention!