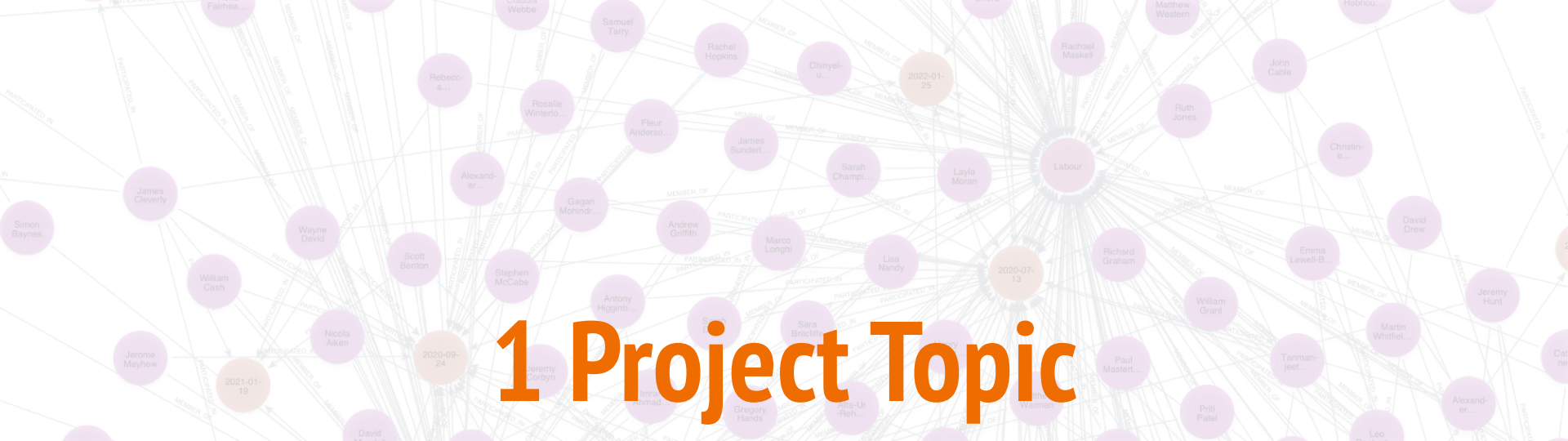
Bipartite Graph for Political Debates

Julian Wagner, Jana Lennartz

Text Technology SS-2025

Dr. Kerstin Jung

Table of Contents



1 Project Topic

1 Project Topic

Goal: Extract speaker data from British parliamentary debates on Yemen and represent it in a Neo4j graph database.

Key elements:

- Who spoke?
- How often?
- Their party and chamber affiliation?


Samuel Tarry
Rachel Hopkins
Rachael Maskell
Matthew Western
John Cable
Ruth Jones
Rosalie Winterto...
Fleur Anderso...
James Sunderl...
Sarah Champi...
Layla Moran
Labour
Christin-e...
James Cleverly
Simon Baynes
Wayne David
Alexand-er...
Gagan Mohindr...
Andrew Griffith
Marco Longhi
Lisa Nandy
Richard Graham
David Drew
Emma Lewell-B...
Scott Benton
Stephen McCabe
William Cash
2020-07-13
Jeremy Hunt
Antony Higginb...
William Grant
Nicola Aiken
2020-09-24
Jerome Mayhew
Jeremy Corbyn
Paul Masterl...
Martin Whitfiel...
Tanman-jeet...
2021-01-19
Matthew Warman
Priti Patel
Alexand-er...
Gregory Hands
David

2 Collect

ParlaMint-GB

- Linguistically annotated corpus of British parliamentary speeches
- Coverage: each parliamentary session between 2015 and 2022
- TEI-XML format with three different file types
  - Debate files

```
<div type="debateSection">
    <head>Yemen</head>
    <u who="#JeremyCorbyn">...</u>
    <u who="#AlanDuncan">...</u>
    ...
</div>
<meeting n="2025-01-21" corresp="#parliament.HC"/>
```

ParlaMint-GB Metadata

Speaker Metadata File

```
<person xml:id="TobiasEllwood">
   <persName>
      <forename>Tobias</forename>
      <forename>Martin</forename>
      <surname>Ellwood</surname>
   </persName>
   <sex value="M"/>
   <affiliation ref="#party.CON"/>
   <affiliation ref="#parliament.HC"/>
   ...
</person>
```

Party Metadata File

```
<org xml:id="party.LAB">
   <orgName full="yes">Labour</orgName>
</org>
```

3 Prepare

XSL Transformation

- Filter and group by speaker

```
<xsl:key name="speechesByWho"
         match="tei:u[ancestor::tei:div[@type='debateSection']/tei:head[contains(., 'Yemen')]]"
         use="@who"/>
```

- Look up of speaker metadata in external files

```
<xsl:variable name="person"
              select="document($listPersonDoc)//tei:person[@xml:id=$whoID]"/>

<xsl:variable name="partyOrg"
              select="document($listOrgDoc)//tei:org[@xml:id=$partyID]"/>
```

Output

```
<debate>
   <date>2015-01-21</date>
   <chamber>House of Commons</chamber>

   <speaker>
      <id>#KeithVaz</id>
      <name>Keith Vaz</name>
      <sex>M</sex>
      <parliament>House of Commons</parliament>
      <party>Labour</party>
      <contributions>2</contributions>
   </speaker>
   ...
</debate>
```

4 Access

Import Debates in Node4j

One (long) query per debate XML file

```
1   MERGE (d:Debate {date: '2015-01-21', chamber: 'House of Commons'})
2   MERGE (p0:Party {name: 'Labour'})
3   MERGE (p1:Party {name: 'Conservative'})
4   MERGE (p2:Party {name: 'Liberal Democrat'})
5   MERGE (p3:Party {name: 'Democratic Unionist Party'})
6
7   MERGE (s0:Speaker {id: '#KeithVaz'})
8   ON CREATE SET s0.name = 'Keith Vaz', s0.sex = 'M',
9       s0.parliament = 'House of Commons', s0.contributions = 2
10  MERGE (s0)-[:MEMBER_OF]->(p0)
11  MERGE (s0)-[:PARTICIPATED_IN]->(d)
12
13  MERGE (s1:Speaker {id: '#TobiasEllwood'})
14  ON CREATE SET s1.name = 'Tobias Ellwood', s1.sex = 'M',
15      s1.parliament = 'House of Commons', s1.contributions = 13
16  MERGE (s1)-[:MEMBER_OF]->(p1)
17  MERGE (s1)-[:PARTICIPATED_IN]->(d)
18
19  ...
```

Create the Debate & Party nodes if they don't exist already

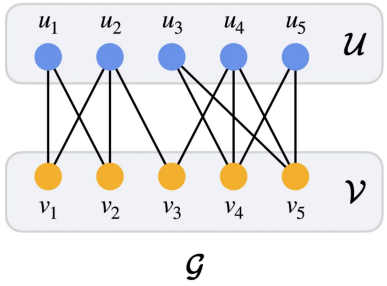
Create a Speaker node and add edges to the corresponding Party and Debate node

Continue with the next Speaker

Graph Schema

- Node Types
  - **Speaker**
  - **Debate**
  - **Party**
- Edge Types
  - **MEMBER_OF**: **Speaker → Party**
  - **PARTICIPATED_IN**: **Speaker → Debate**
- Total
  - 415 Nodes
  - 1306 Edges

Bipartite Requirement



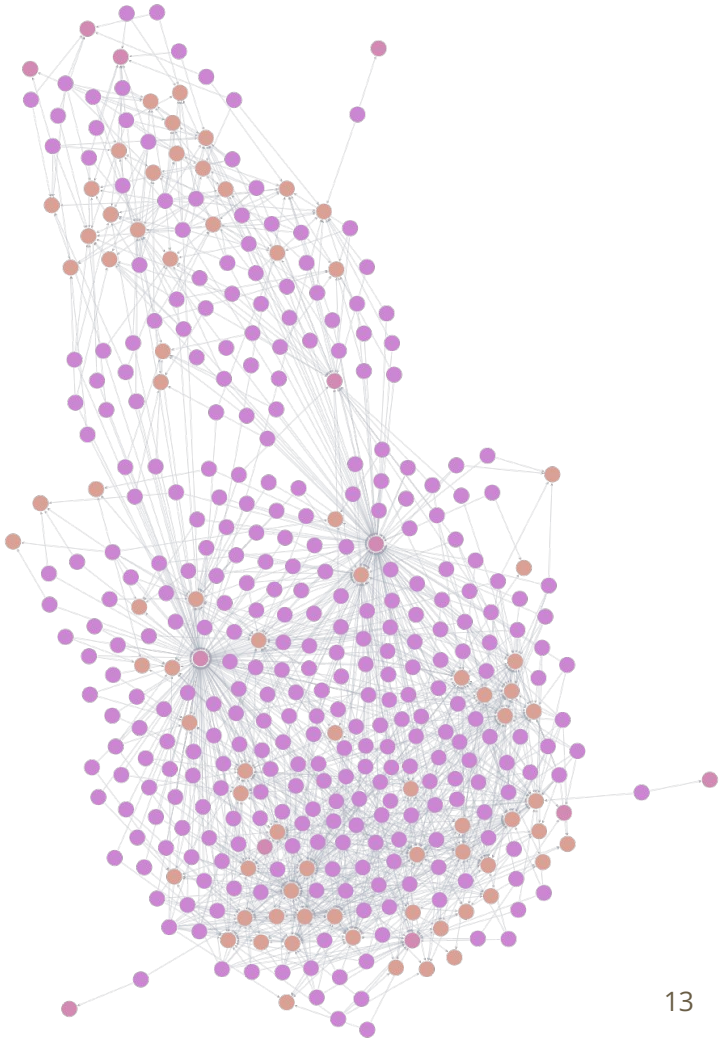
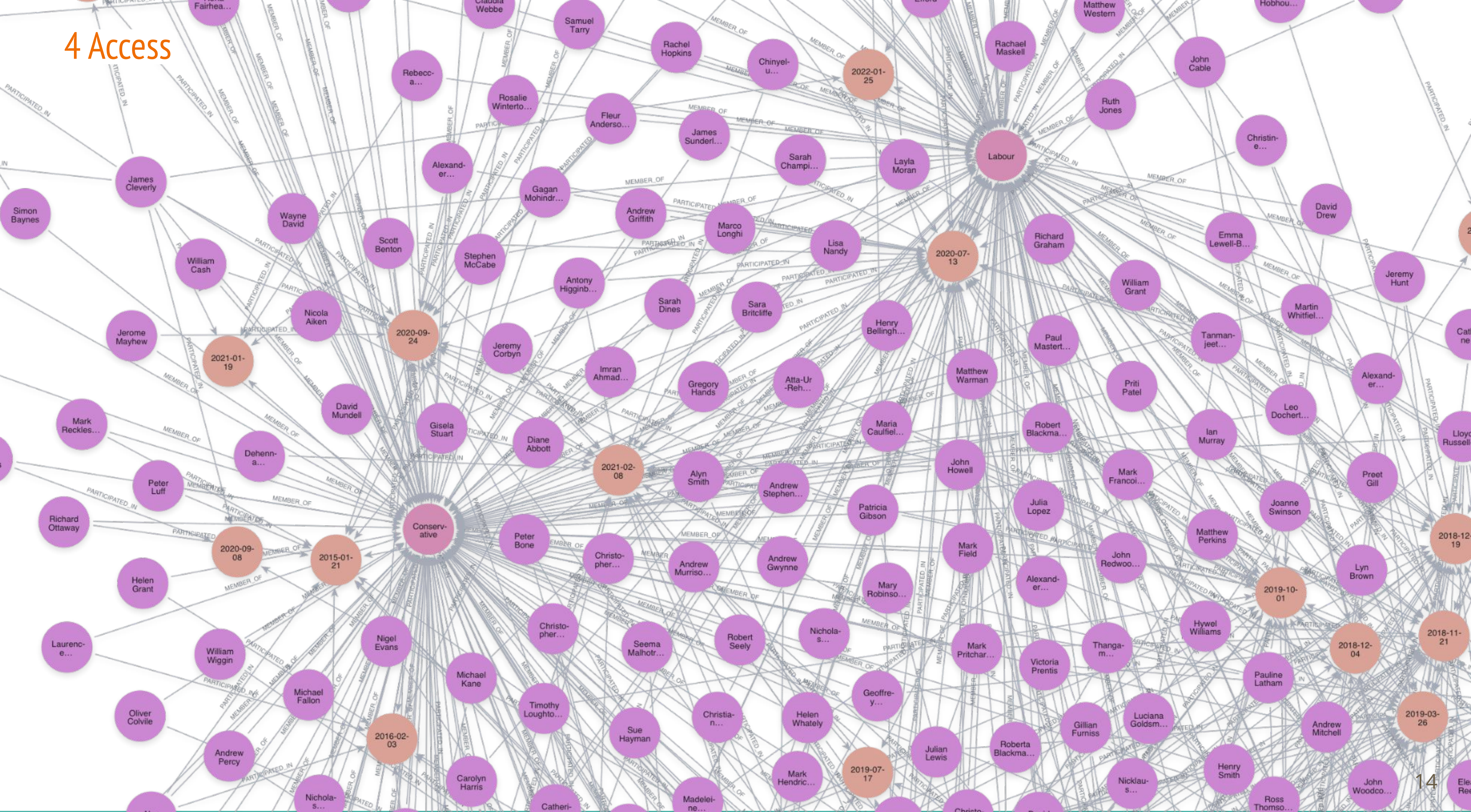


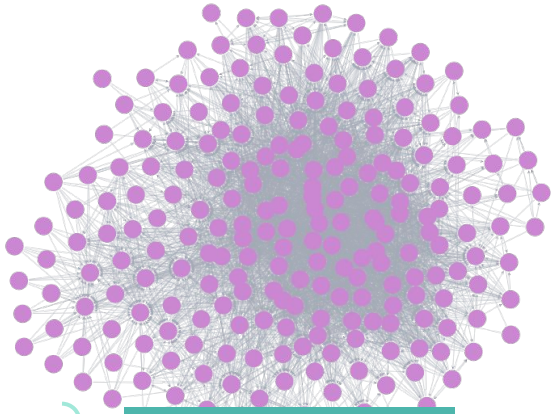
Image from Guan et al. (2023)

4 Access


Fiona Fairhea…
Claudia Webbe
Samuel Tarry
Rachel Hopkins
Chinyel-u…
2022-01-25
Rachael Maskell
Matthew Western
Hobhou…
John Cable
Rebecc-a…
Rosalie Winterto…
Fleur Anderso…
James Sunderl…
Sarah Champi…
Layla Moran
Labour
Ruth Jones
Christin-e…
James Cleverly
Alexand-er…
Gagan Mohindr…
Andrew Griffith
Marco Longhi
Lisa Nandy
2020-07-13
Richard Graham
Emma Lewell-B…
David Drew
Simon Baynes
Wayne David
Scott Benton
Stephen McCabe
Antony Higginb…
Sarah Dines
Sara Britcliffe
Henry Bellingh…
William Grant
Jeremy Hunt
William Cash
Nicola Aiken
2020-09-24
Jeremy Corbyn
Martin Whitfiel…
Tanman-jeet…
Paul Mastert…
Jerome Mayhew
2021-01-19
Imran Ahmad…
Gregory Hands
Atta-Ur -Reh…
Matthew Warman
Priti Patel
Alexand-er…
Leo Dochert…
David Mundell
Maria Caulfiel…
Robert Blackma…
Ian Murray
Mark Reckles…
Gisela Stuart
Diane Abbott
2021-02-08
Alyn Smith
Andrew Stephen…
John Howell
Mark Francoi…
Preet Gill
Lloyd Russell
Dehenn-a…
Peter Luff
Patricia Gibson
Julia Lopez
Joanne Swinson
Richard Ottaway
Conserv-ative
Peter Bone
Christo-pher…
Andrew Murriso…
Andrew Gwynne
Mark Field
John Redwoo…
Matthew Perkins
2018-12-19
2020-09-08
2015-01-21
Helen Grant
Mary Robinso…
Alexand-er…
2019-10-01
Lyn Brown
Laurenc-e…
Nigel Evans
Christo-pher…
Seema Malhotr…
Robert Seely
Nichola-s…
Mark Pritchar…
Victoria Prentis
Thanga-m…
Hywel Williams
2018-12-04
2018-11-21
William Wiggin
Michael Fallon
Michael Kane
Timothy Loughto…
Geoffre-y…
Pauline Latham
Oliver Colvile
2016-02-03
Sue Hayman
Christia-n…
Helen Whately
Julian Lewis
Roberta Blackma…
Gillian Furniss
Luciana Goldsm…
Andrew Mitchell
2019-03-26
Andrew Percy
Carolyn Harris
Mark Hendric…
2019-07-17
Henry Smith
Nicklau-s…
John Woodco…
Alan Duncan
Nichola-s…
Catheri-ne…
Madelei-ne…
Stephen
Christo-pher…
David Doogan
Ross Thomso…
MEMBER_OF
PARTICIPATED_IN

Speaker-to-Speaker Projection

Create a **CO_DEBATED_WITH** relation



```
20  MATCH (s1:Speaker)-[:PARTICIPATED_IN]->(d:Debate)<-[:PARTICIPATED_IN]-(s2:Speaker)
21  WHERE elementId(s1) < elementId(s2)
22
23  WITH s1, s2, count(d) AS overlap
24  MERGE (s1)-[r:CO_DEBATED_WITH]-(s2)
25  SET   r.sharedDebates = overlap;
26
27  MATCH (s1:Speaker)-[r:CO_DEBATED_WITH]->(s2:Speaker)
28  RETURN s1, r, s2;
```

Select all Speaker & Debate nodes

Add an edge between every speakers that share a debate, add the shared debate count as property

Query all speaker nodes with the new co-debate edge

5 Extensions

5 Extensions

Cross-document lookup using `document()` in XSLT:

- Retrieving speaker and party data from external TEI files.

Parameterization of XSLT:

- Used `<xsl:param>` to make the stylesheet reusable with external files.

Data Creation in Neo4j with Cypher:

- MERGE clause matches existing pattern in the graph or, if not present, creates new nodes/edges
- SET clause applies labels and properties on nodes/edges
- ON CREATE SET clause applies labels and properties only when MERGE just created the node/edge
- WITH clause binds matches into new keyword that can be used in the query
- COUNT clause counts number of rows returned by the subquery

6 Difficulties

6 Difficulties

- Figuring out how to use keys and document lookup
- Filtering the right elements
  - `<date>2024-05-14</date>`
  - `<date>2015-01-21</date>`
- Apostrophe in properties not allowed
  - `MERGE (s30:Speaker {id: '#BrendanO'Hara'})`

References

Corpus

- ParlaMint-GB Corpus: https://www.clarin.si/repository/xmlui/handle/11356/1912

Tools & Frameworks

- Neo4j 2025.05 Enterprise
- Cypher 5
- Python
- XSL Transformations (XSLT) Version 1.0
- TEI Namespace

Resources

- Guan, Y., Lu, R., Zheng, Y., Zhang, S., Shao, J., & Wei, G. (2023). *Achieving efficient and privacy-preserving (α, β)-core query over bipartite graphs in cloud*. *IEEE Transactions on Dependable and Secure Computing, 20*(3), 1979–1993. https://doi.org/10.1109/TDSC.2022.3169386