

Curiositäten- und Memorabilien Lexicon von Wien

Topics and Co-occurrence Network of a mid-19th Century Lexicon

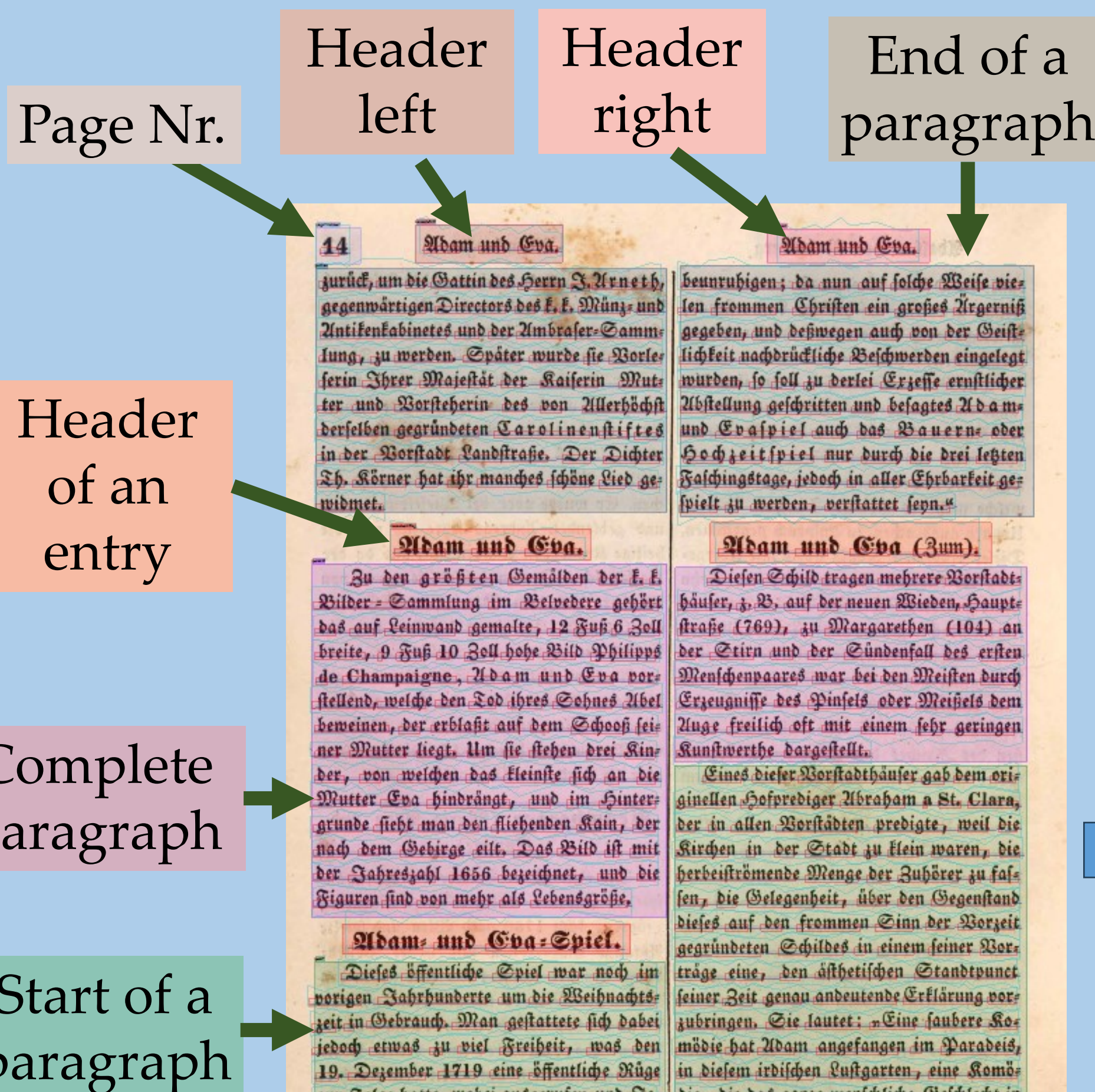
Nikola Krisztian Czindrity (University of Vienna) nikola.krisztian.czindrity@univie.ac.at

1. Back to the Sources



Author: Realis (Pseudonym for Walter Ritter von Cöckelberghe-Dützele [1786-1857])
Year of publication: 1846
Size: 2 volumes with 962 pages of content with 1295 entries (490K Tokens with 85K Types).
Source: Münchener Digitalisierungs Zentrum
While some of the entries provide factual information, others include anecdotal and comical content.

2. Field Model and Transcription



The software *Transkribus* was utilised for field modeling and transcription tasks.

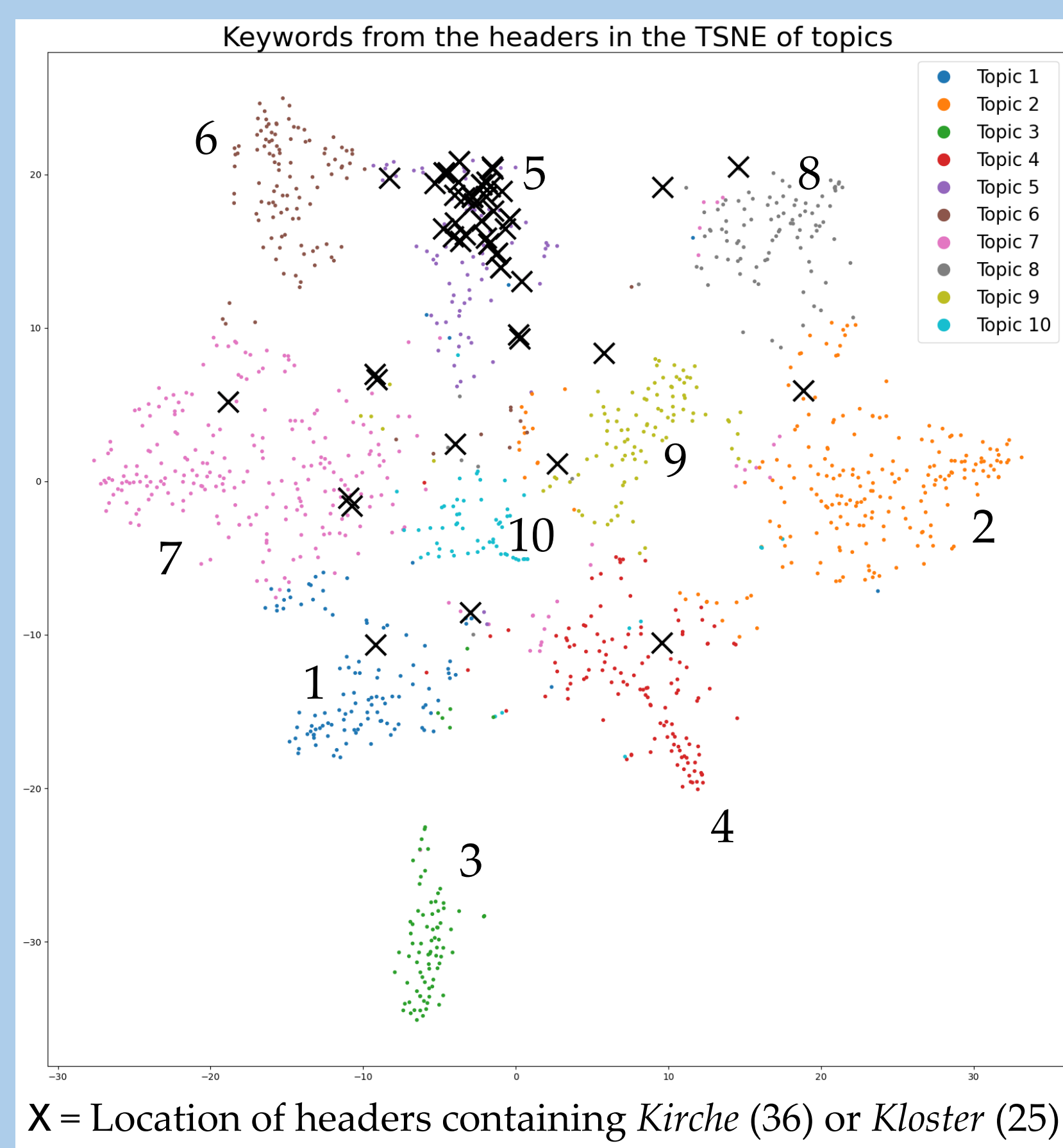
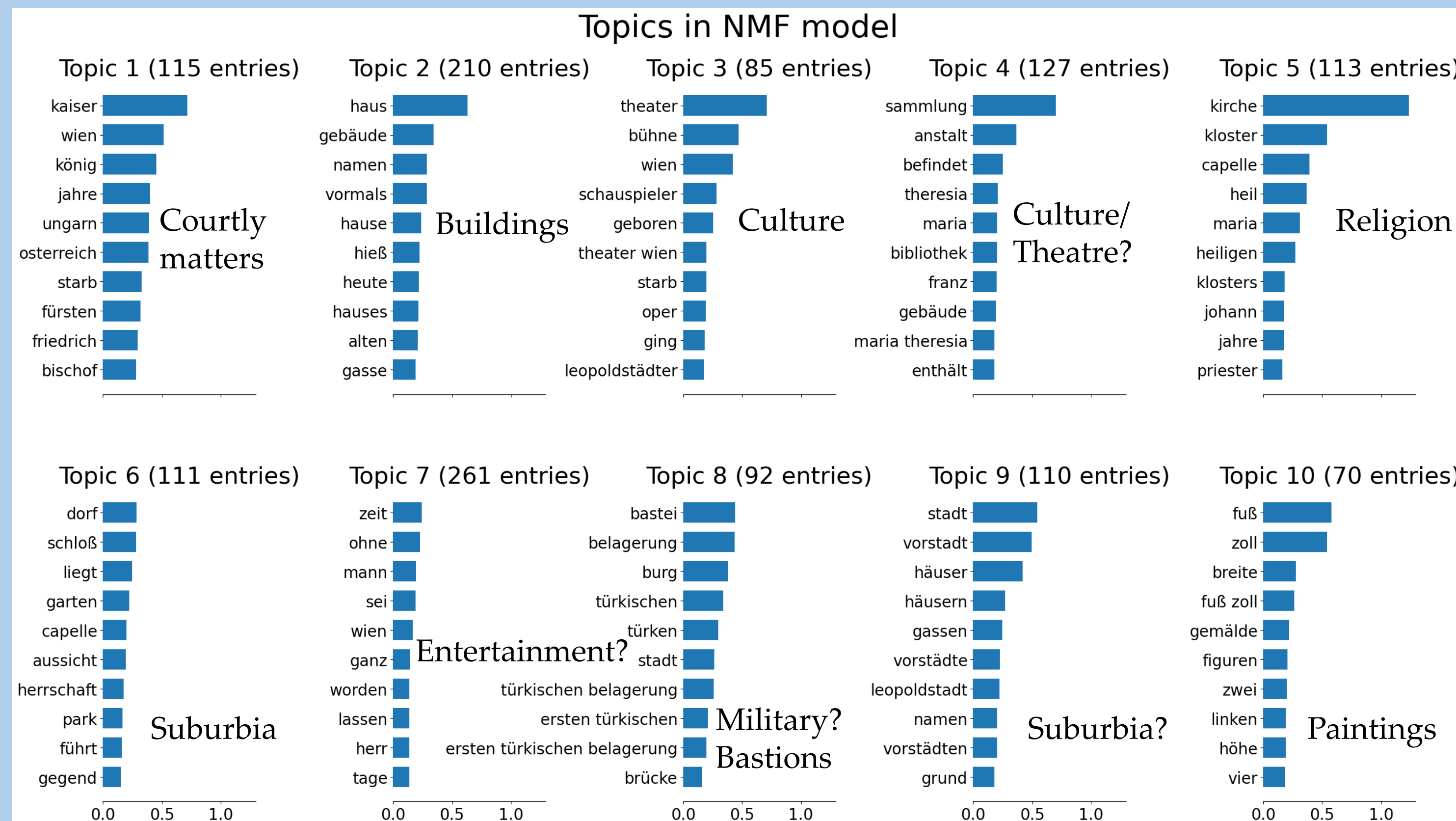
The field model underwent initial training with manual annotations from 30 pages. Subsequent iterations involved integrating annotations from previous models, which were manually corrected, progressively scaling up to 75 pages and then 102 pages.

For transcription, the “German Fraktur 18th Century” model (WrDiarium_M9, ID 47114; ACDH-CH) was utilized.

After ensuring that both the field and transcription models yielded favorable results, the data was extracted in TEI-XML format and tabulated using Python.

1	A	B	C	D	E	F	G
	book	header_id	book_page_start	book_page_end	header_entry	entry_text	entry_id
1	CL MDZ	CL_facs_1_region	1699100518315_15	3	4 A	von man eine große Stadt vorder Lenna beschau, so lernest du	170
2	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Es gereicht dem gaudium, wenn der Benachtheilte von der Stadt	170
3	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Der Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
4	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
5	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
6	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
7	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
8	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
9	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
10	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
11	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
12	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
13	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
14	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169
15	CL MDZ	CL_facs_2_region	1699100518315_15	4	4 A	Die Abschied ist ein Brief oder Tisch, voranformals, die Kopf	169

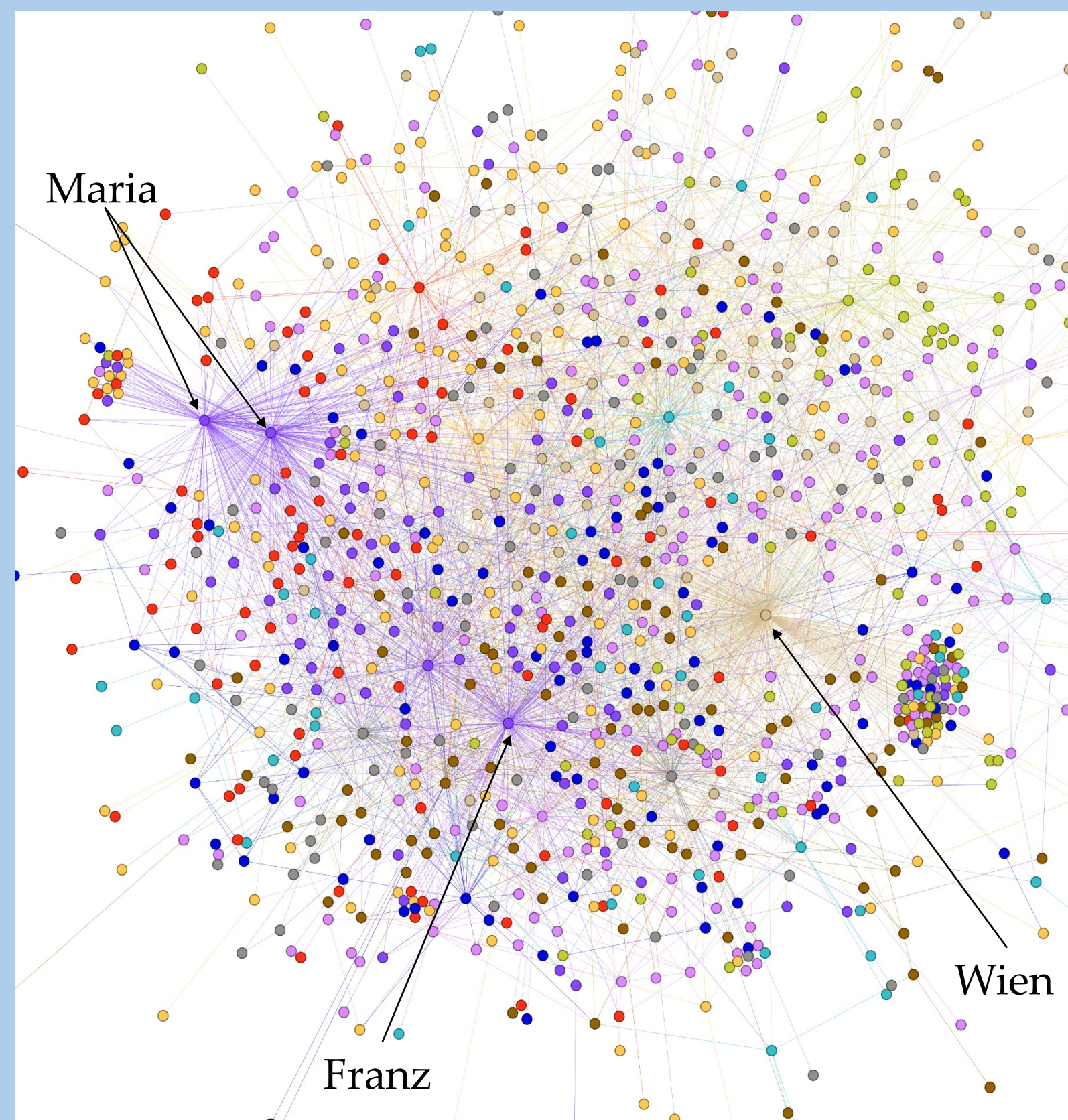
3. Topic Modeling and Co-occurrence Network



Negative Matrix Factorization (NMF) was employed for topic modeling.

Upon inspection of the maximal weight for each topic per entry and associated keywords, it is evident that NMF produces meaningful results, with the ten topics identified being well-distributed and coherent.

Further analysis of the reduced lexicon entry - topic matrix using t-SNE revealed that topics form distinct clusters, e.g., keywords such as *Kirche* and *Kloster* are prominent within Topic 5, which focuses on "Religion".

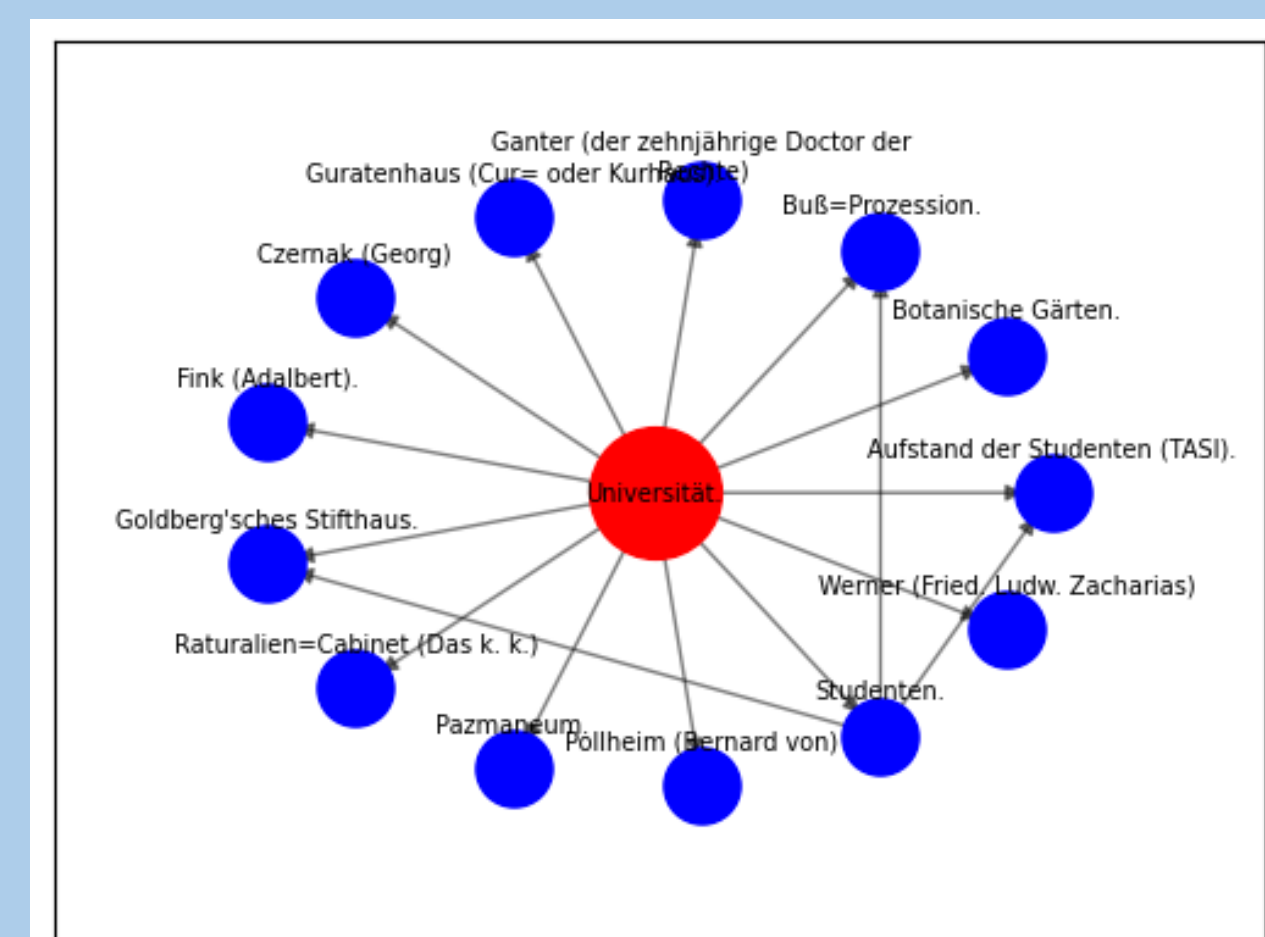


A network based on the co-occurrence of entry headers within the entries consisted of 1102 nodes and 3417 edges. Both nodes and edges are colored according to their association with topics identified by NMF.

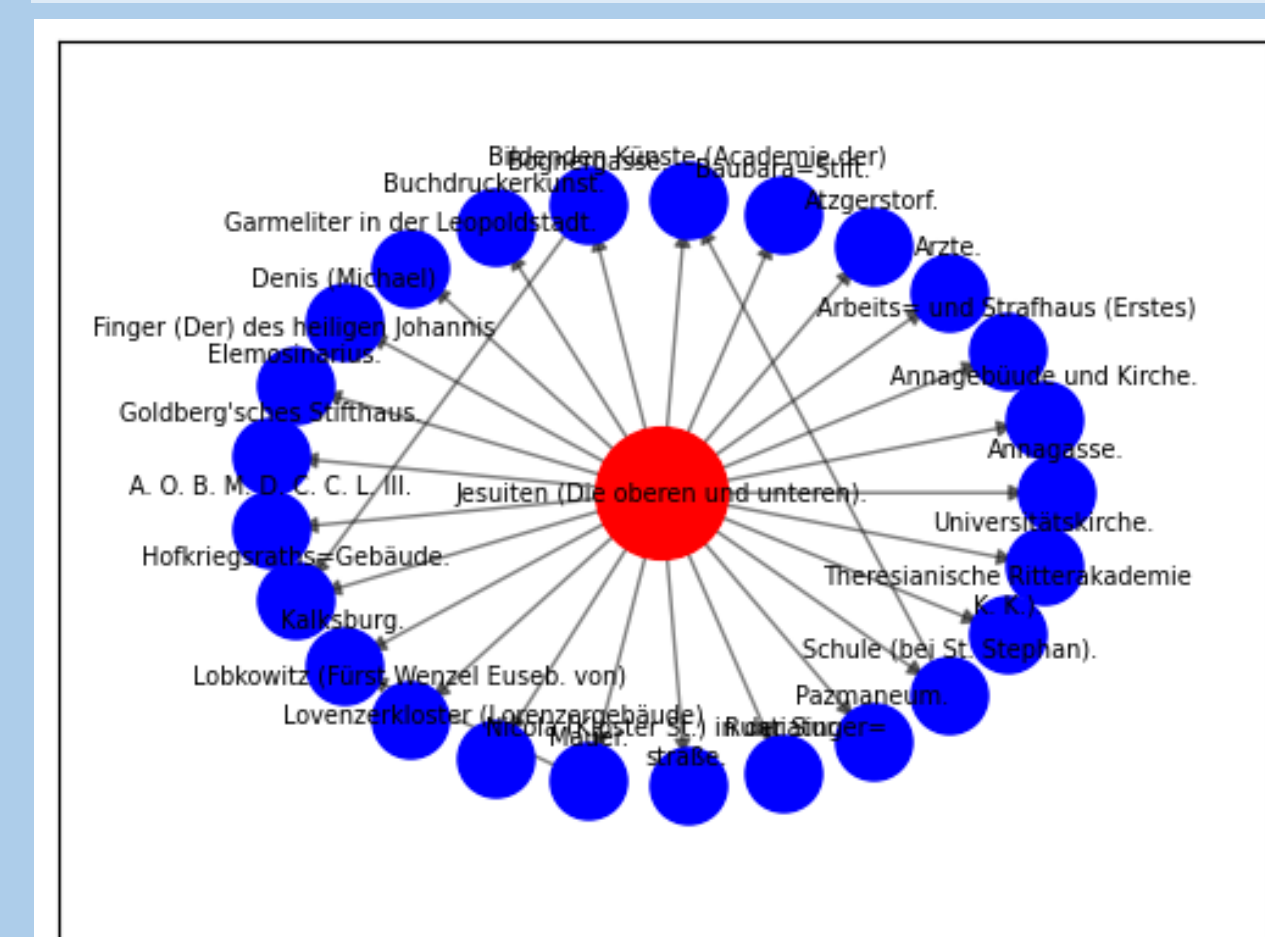
Despite efforts, it was not feasible to identify clusters using betweenness centrality or to produce meaningful interpretations between sets of connections.

However, upon inspecting subgraphs, it becomes apparent that the connections are meaningful. For instance, the entry *Universität* is connected to *Studenten* and *Botanischer Garten*.

Further investigation or different approaches are necessary to disentangle the connections.



Subgraph for the entry *Universität*



Subgraph for the entry *Jesuiten*



Basilisk from the entry „Basiliskenhaus“