

# Curiositäten- und Memorabilien Lexicon von Wien

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

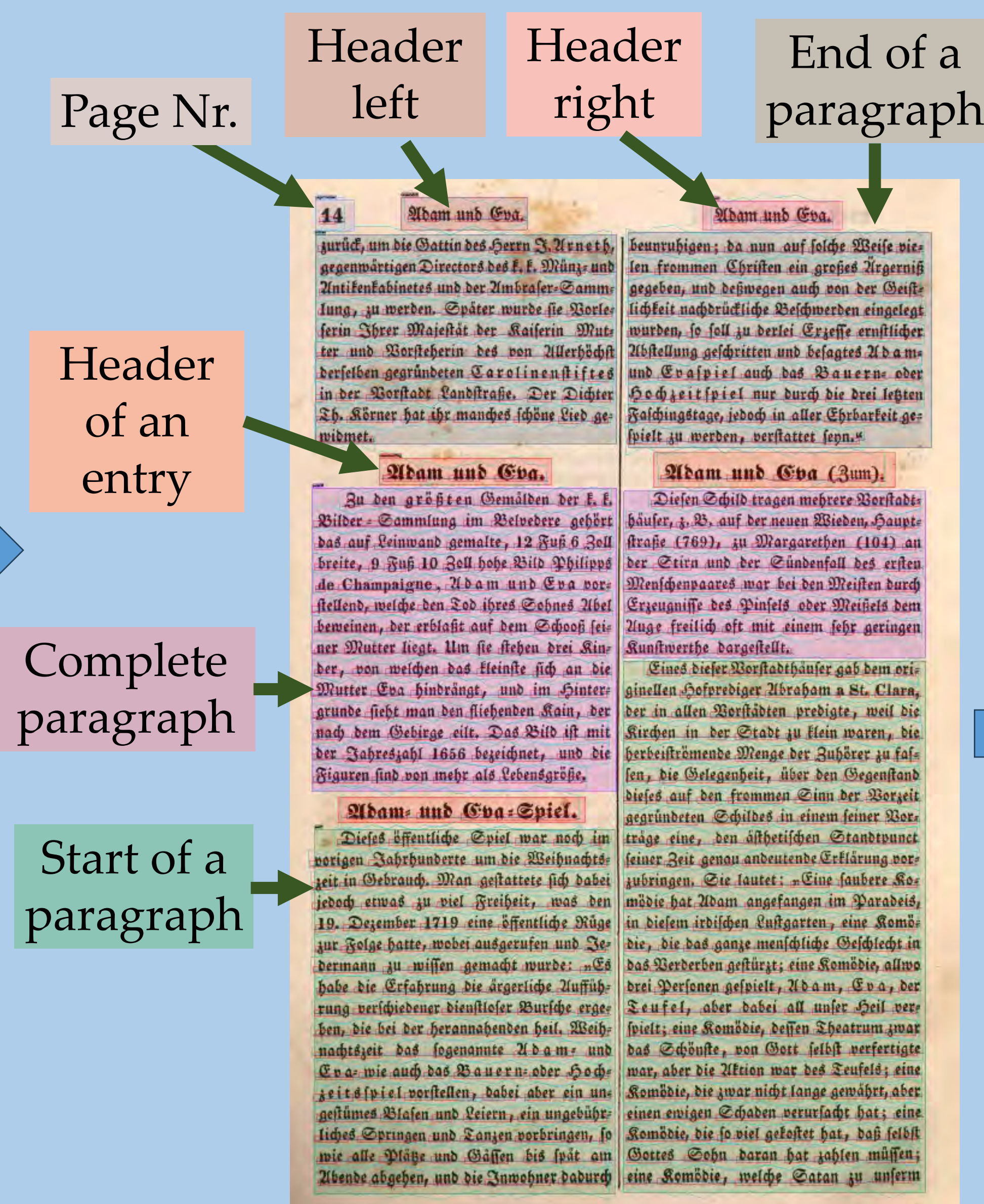
Nikola Krisztian Czindrity (University of Vienna) nikola.krisztian.czindrity@univie.ac.at

### 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.

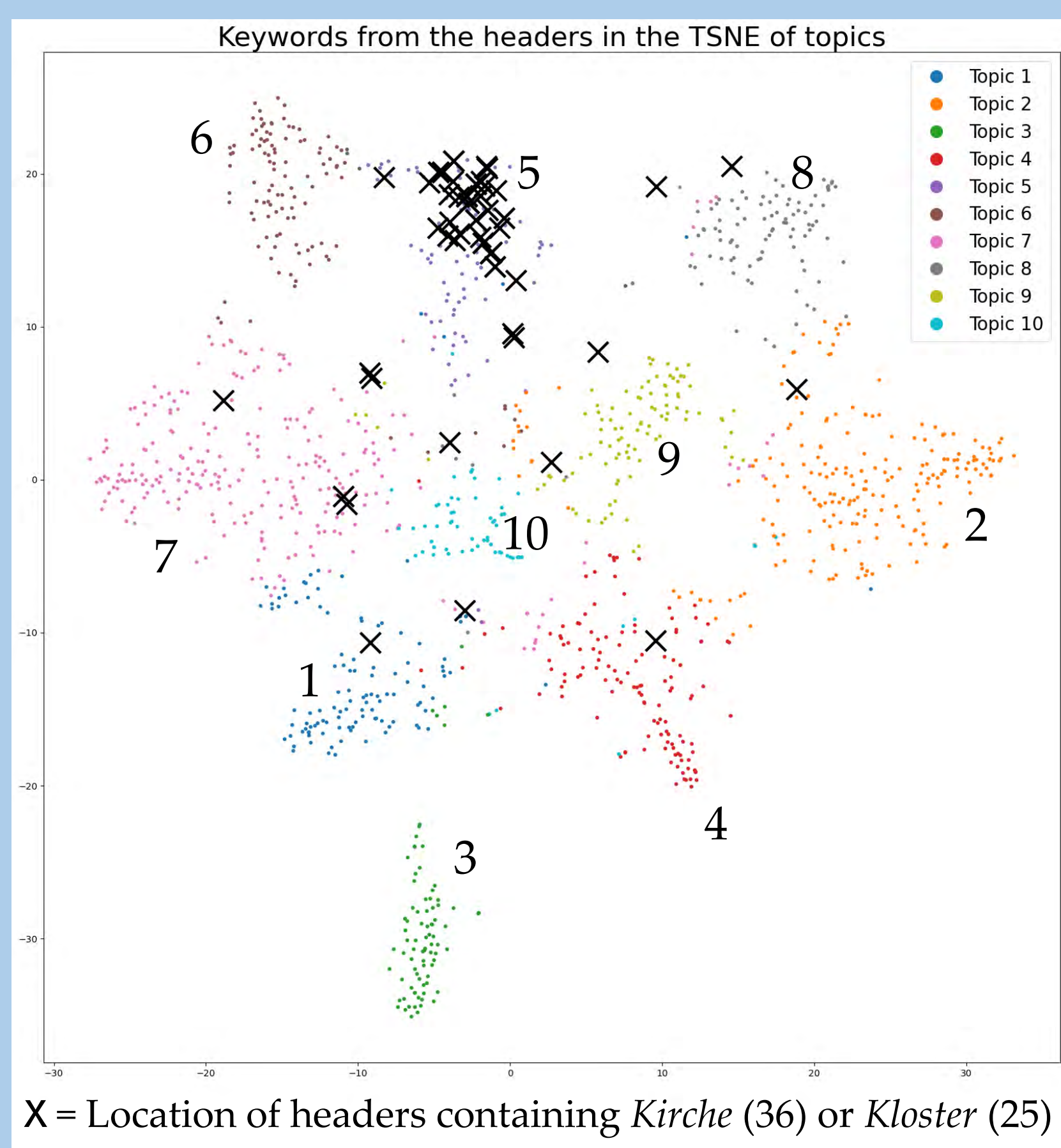
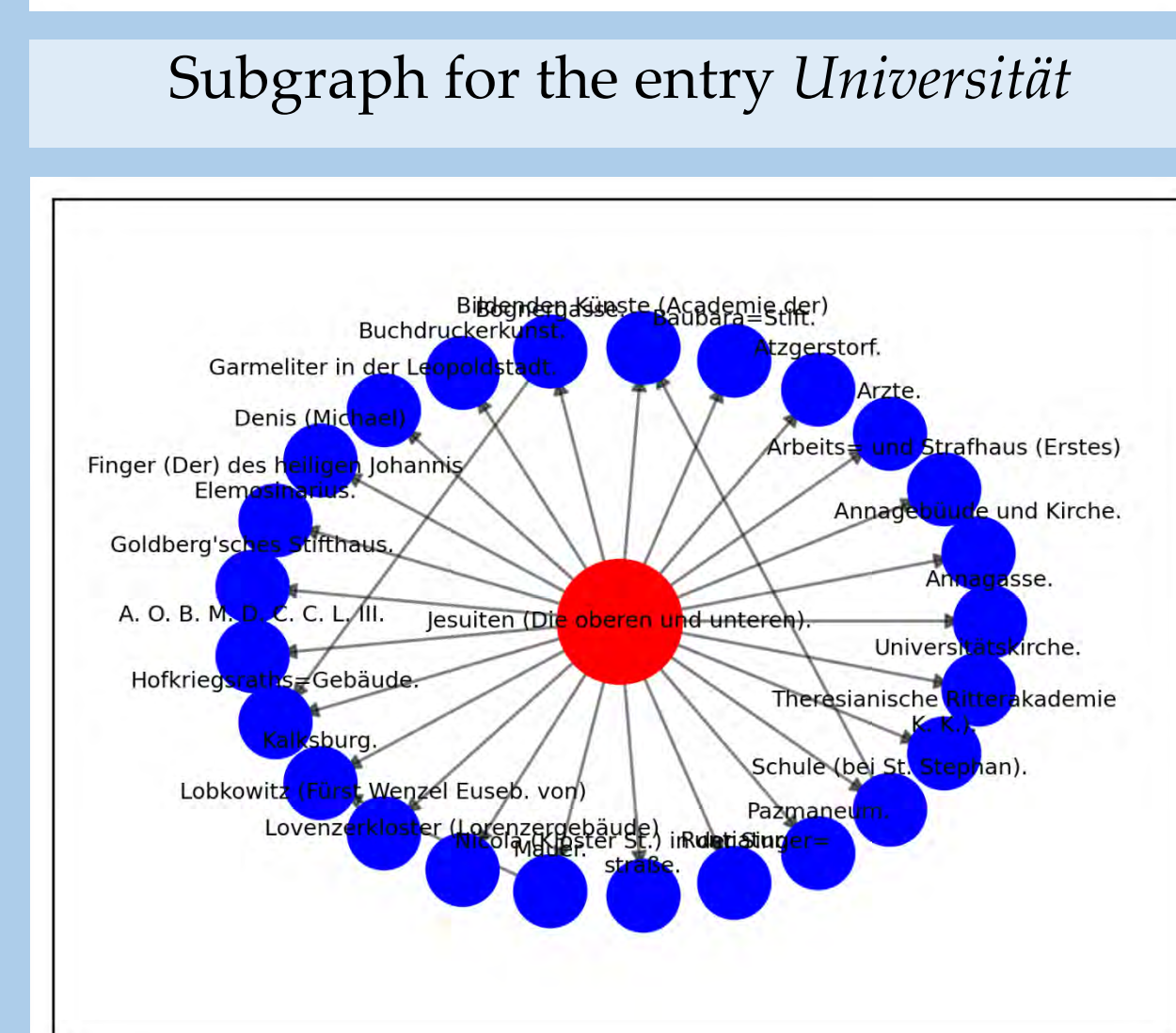
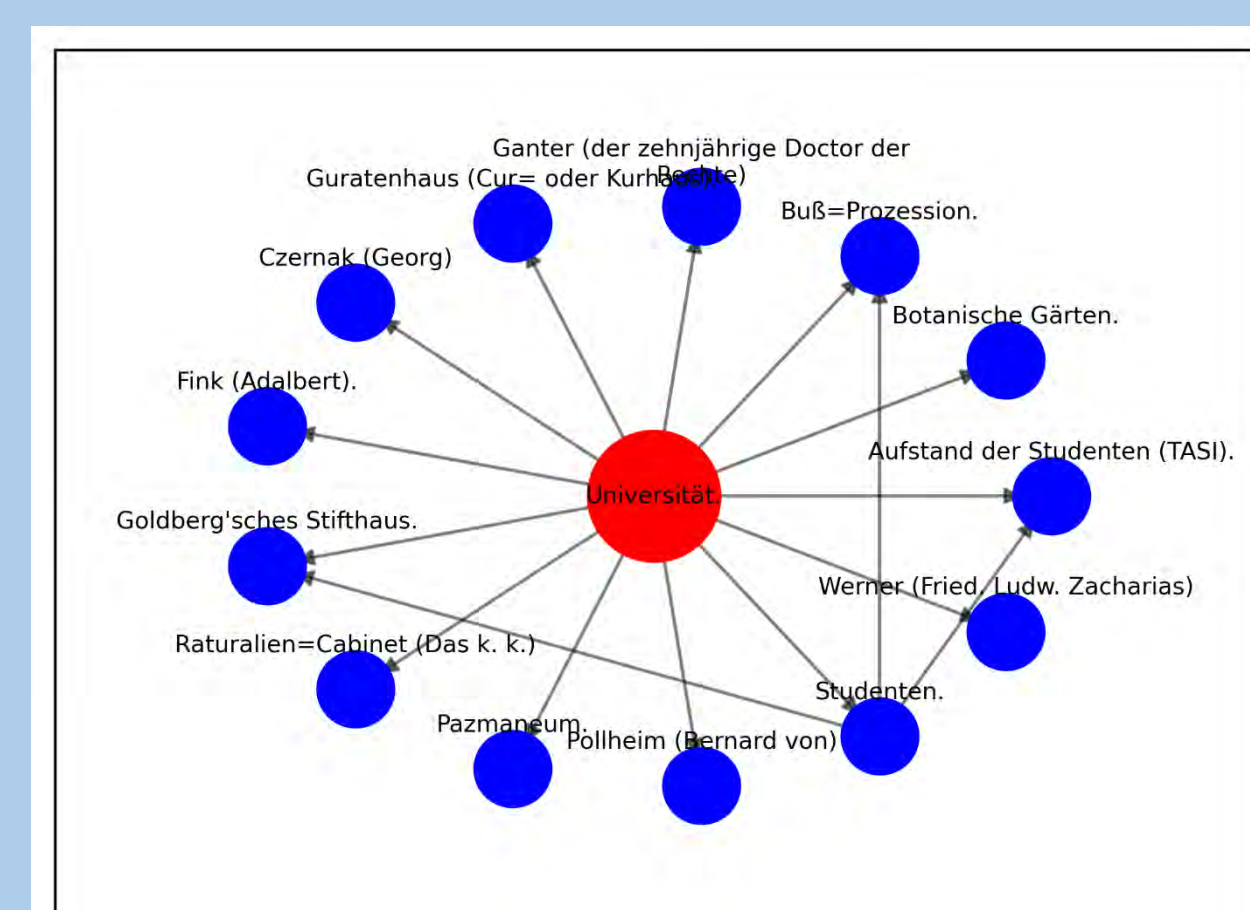
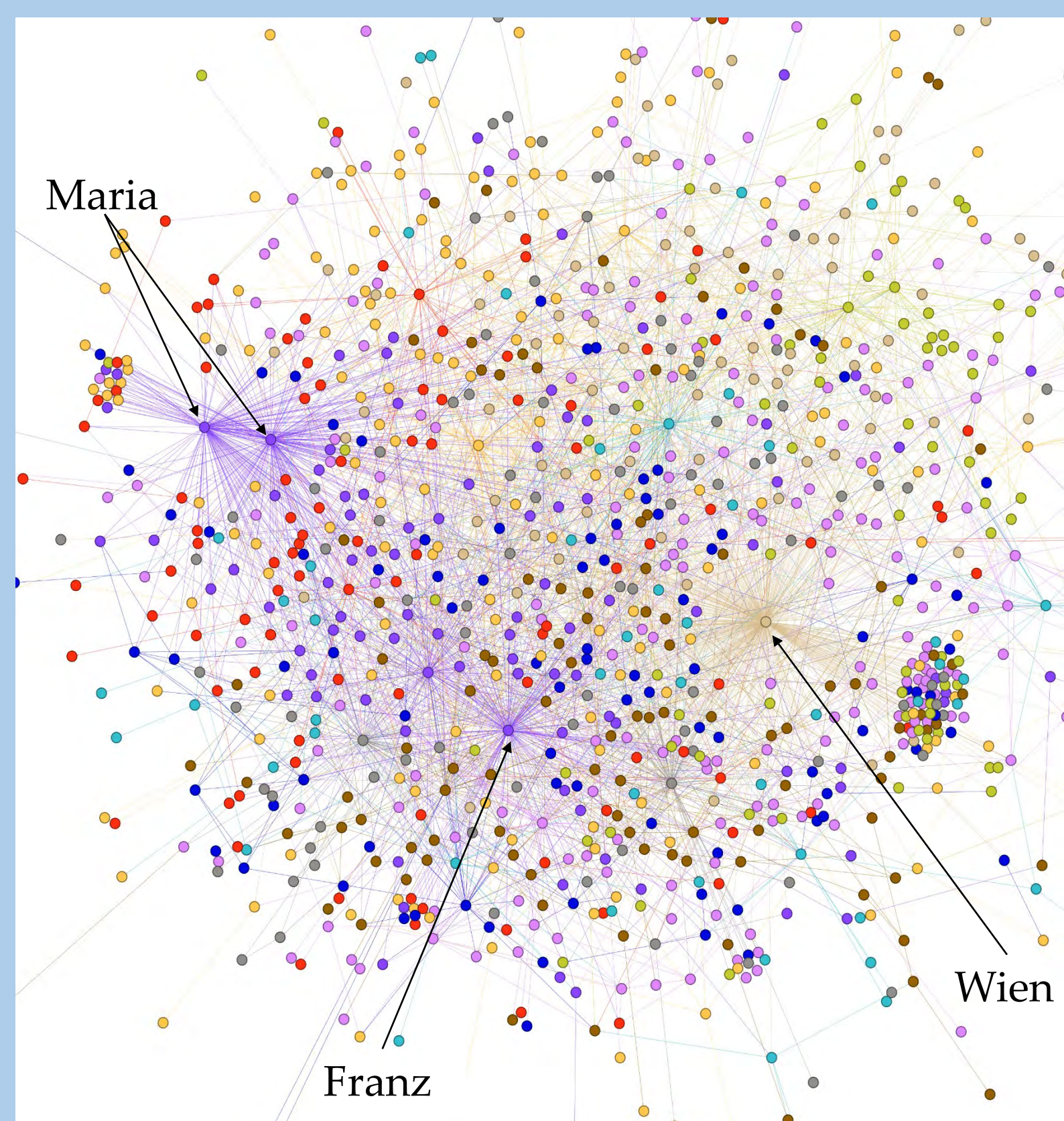
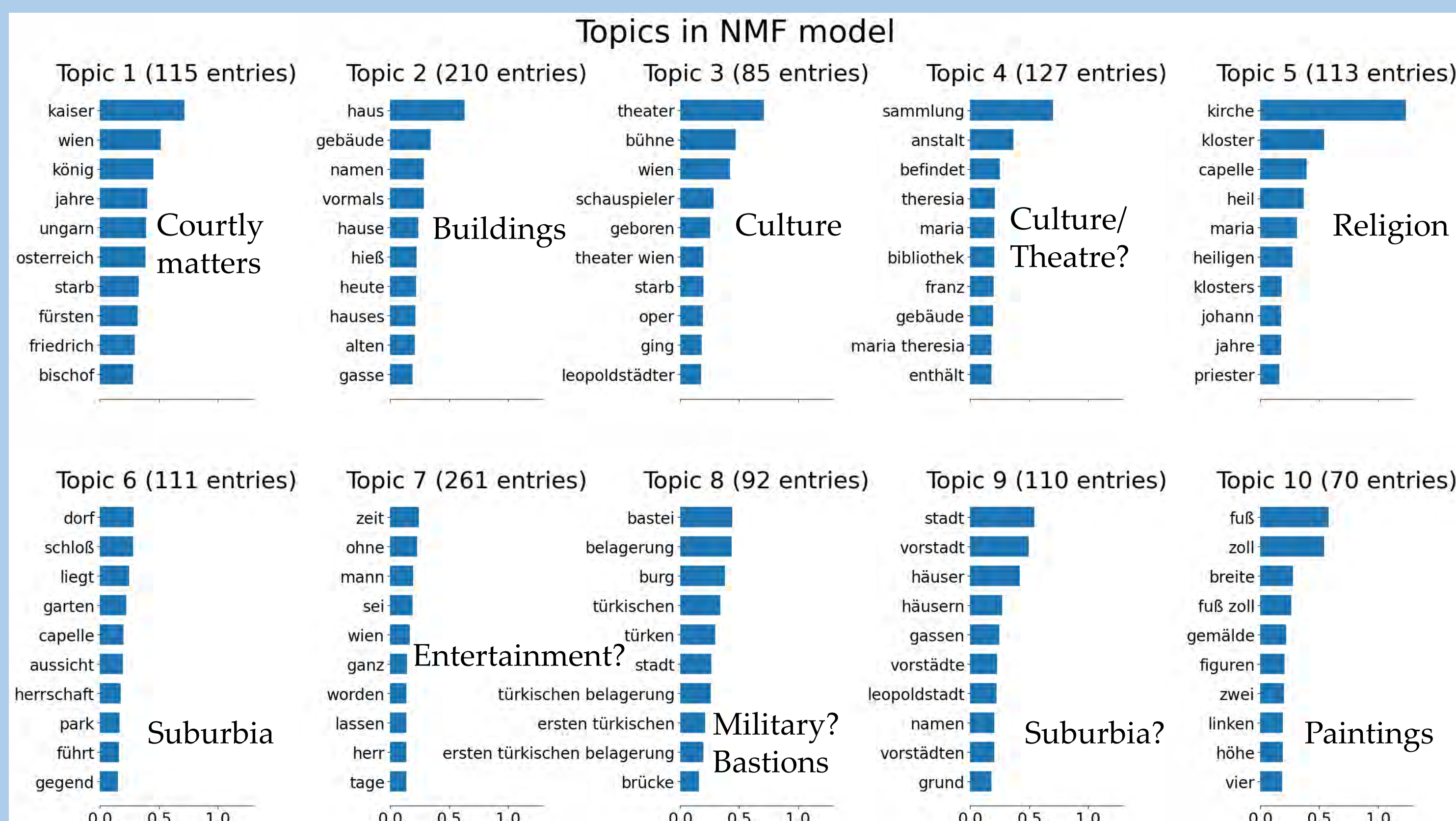
The field model underwent initial training with manual annotations from 30 pages. Subsequent iterations involved integrating annotations from previous models, which were manually corrected, progressively scaling up to 75 pages and then 102 pages.

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
1	book	header_id	book_page_start	book_page_end	header_string	entry_text	entry_id
2	CL_MID	CL_Facs_1_region	1699100518515_15	3	A	man man eine große Stadt vorder Ferne beschau, so lehrenm [Facs_1_region_170]	170
3	CL_MID	CL_Facs_2_region	1699100518796_50	4	A	Es gerücht dem gereden Sinne die Bewohner Wien gar [Facs_2_region_169]	169
4	CL_MID	CL_Facs_2_region	1699100518887_62	4	5	Abenm. Der Abenm sei ein Bett oder Tisch, vorworfomals, da die Kopf [Facs_2_region_169]	169
5	CL_MID	CL_Facs_3_region	1699100520652_41	5	7	7 Abenmge- und Trausliche (Gräblich) Diese in der terragege legendeitauer (240 und 242) gühon [Facs_3_region_169]	169
6	CL_MID	CL_Facs_5_region	1699100296676_47	7	8	8 Abenmge Haus (Das) In der Vorstadt Wien (187) ist ein an-erblickendes Monum [Facs_5_region_169]	169
7	CL_MID	CL_Facs_6_region	1699100287851_70	8	10	10 Abenmge des Mondes und Stornes vom 5 Jan 12. Jul 1688 um 11 Uhr Vor-mittag, began der Zepeldeck [Facs_6_region_169]	169
8	CL_MID	CL_Facs_7_region	1699100293637_39	9	10	10 Academie (Der) Wien zählt der wissenschaftlichen und Kunst-Anstalten, die als [Facs_7_region_169]	169
9	CL_MID	CL_Facs_8_region	1699100293483_110	10	10	10 Academie. Dieser Ort wird in alten Urkundenwache und Eick an genannt [Facs_8_region_169]	169
10	CL_MID	CL_Facs_8_region	1699100292561_113	10	11	11 Achth. Da Wien der Sitz sowohl aller deutschen Kaiser aus dem Haus [Facs_8_region_170]	170
11	CL_MID	CL_Facs_9_region	1699100297618_132	11	12	12 Achth. Hier von T. hatte den Entfall, dem Pfarren seines Kirchsplatz in ei [Facs_9_region_17]	17
12	CL_MID	CL_Facs_10_region	1699100313094_171	12	12	12 Achth. Adalbert (Abert), dritter Margrafen Oiters aus dem Hause ei [Facs_10_region_16]	16
13	CL_MID	CL_Facs_10_region	1699100313023_174	12	13	13 Achth. von Babenberg. Die erste ist die Tochter des hochfürstlichen Jodest, wurde [Facs_10_region_16]	16
14	CL_MID	CL_Facs_11_region	1699100329526_208	13	14	14 Adenberger (Maria Anna und Antoinette) Zu den größten Gemälden der k. k. Bilden-Sammlung im Bebelo [Facs_11_region_16]	16
15	CL_MID	CL_Facs_12_region	1699100324874_227	14	14	14 Adam und Eva.	

### 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.

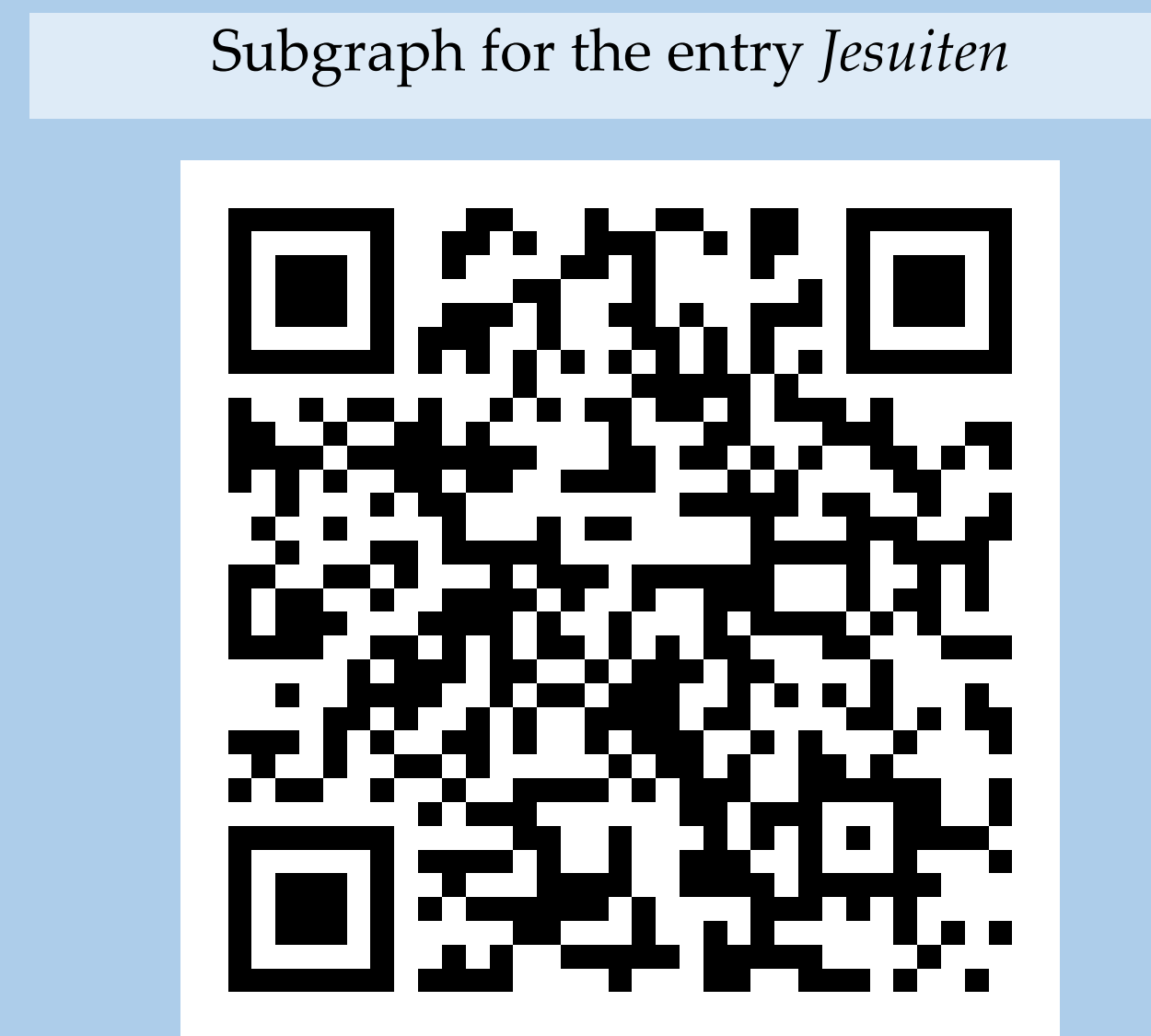
Further analysis of the reduced lexicon entry - topic matrix using t-SNE revealed that topics form distinct clusters, e.g., keywords such as *Kirche* and *Kloster* are prominent within Topic 5, which focuses on "Religion".

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.



Even more information is available at <https://nkc.z.githu.io/content/dsd2024/dsd2024.html>