Let data sing the Uyghur Twelve Muqam

A text mining on lyrics   ئۇيغۇر مۇقام

Limai Kai[1], Zekun Yang[1]*, Huiling Feng[1], Yuenan Liu[1], Jihong Liang[1], Huilin Yang[2], Jin Yan[2], Yanfen Huang[2], Kongwen Guan[2]

1 School of Information Resource Management, Research Center for Digital Humanities, Renmin University of China
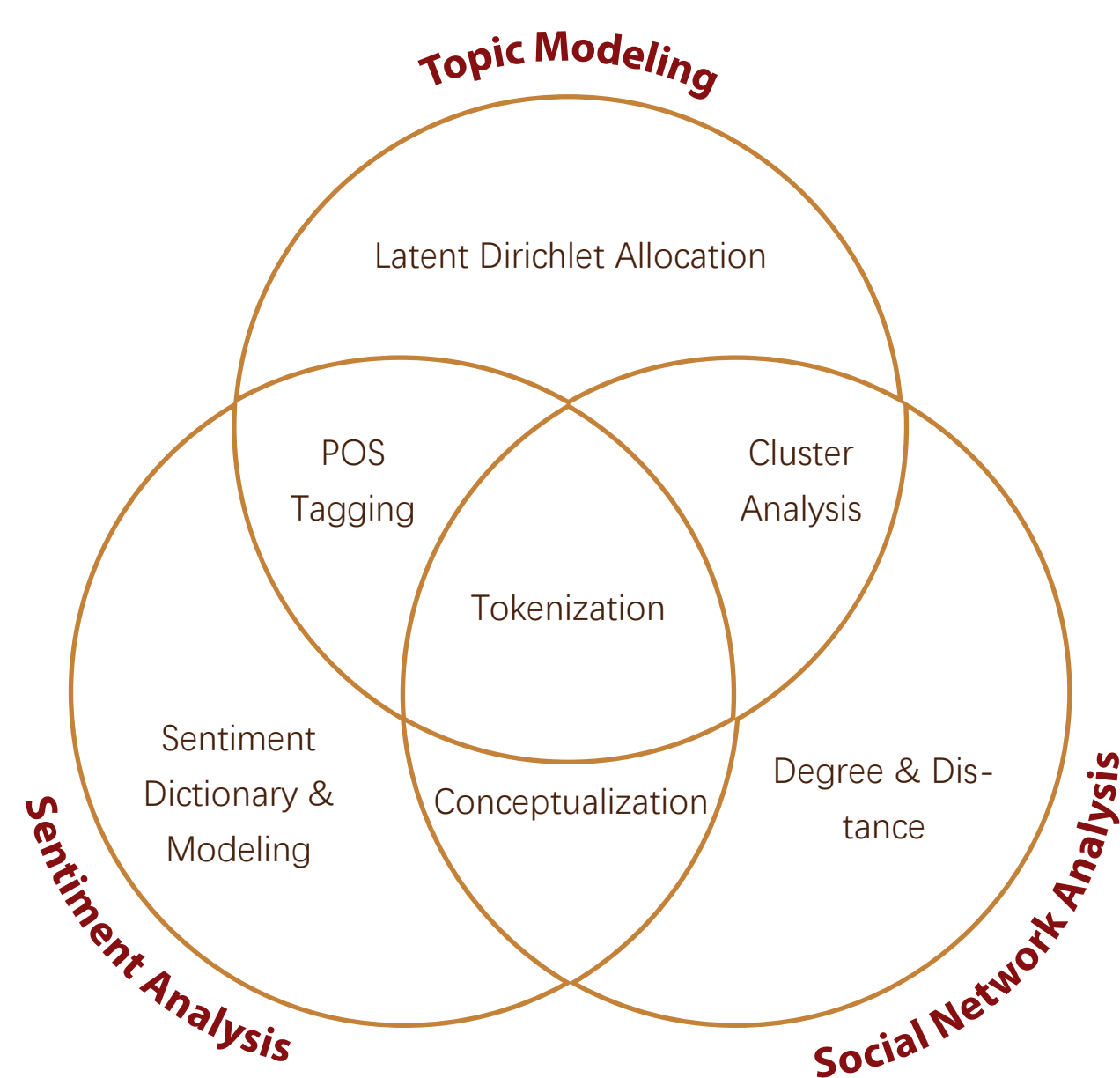2 Centre for European Studies/Centre for EU Studies, Renmin University of China
* Zekun Yang is the corresponding author and can be contacted at zekunyang@ruc.edu.cn

Research Center for Digital Humanities of Renmin University of China

DH 2023 Collaboration as an Opportunity

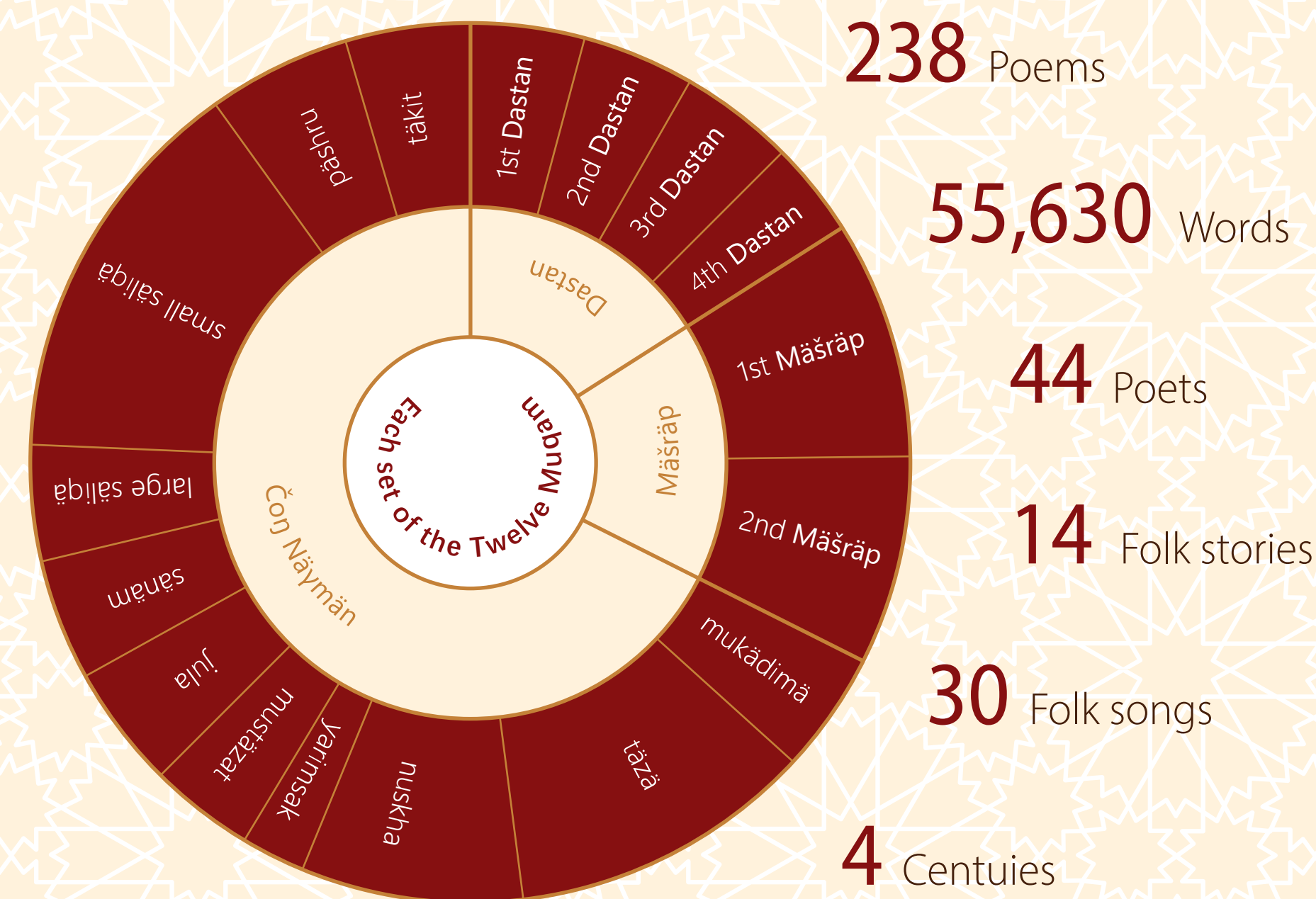ALLIANCE OF DIGITAL HUMANITIES ORGANIZATIONS

---

1 INTRODUCTION

Originated in the 4th century, Uyghur Twelve Muqam is an artistic treasure created by the Uyghur people, which integrates **songs, dances, and instrumental performances** *(Harris, 2008; Han & Xiao, 2017a; Tursun et al., 2019)*. The Uyghur Muqam of Xinjiang was designated as an Intangible Heritage of Humanity by UNESCO in 2005. Previous research on Uyghur Twelve Muqam mainly focuses on its history, origin, rhythm, tone, structure, notation, dance, lyrics, and characters *(Anderson, 2012; Han & Xiao, 2017a; Han & Xiao, 2017b; Imin et al., 2019; Light, 1998)*. However, digital humanities methodology has not been applied to study the Uyghur Twelve Muqam. Thus, we start the project "Let data sing the Uyghur Twelve Muqam", which applies text mining methods to gain deeper insights into Muqam lyrics. This paper demonstrates the preliminary results, which discover new insights beyond existing qualitative research on Muqam, including new Muqam poem topics, sentiment styles of individual poets, and topic similarity among poets.

2 MEHODS

Topic Modeling
- Latent Dirichlet Allocation
- POS Tagging
- Cluster Analysis
- Tokenization
- Sentiment Dictionary & Modeling
- Conceptualization
- Degree & Distance

Sentiment Analysis

Social Network Analysis

3 CORPUS

238 Poems

55,630 Words

44 Poets

14 Folk stories

30 Folk songs
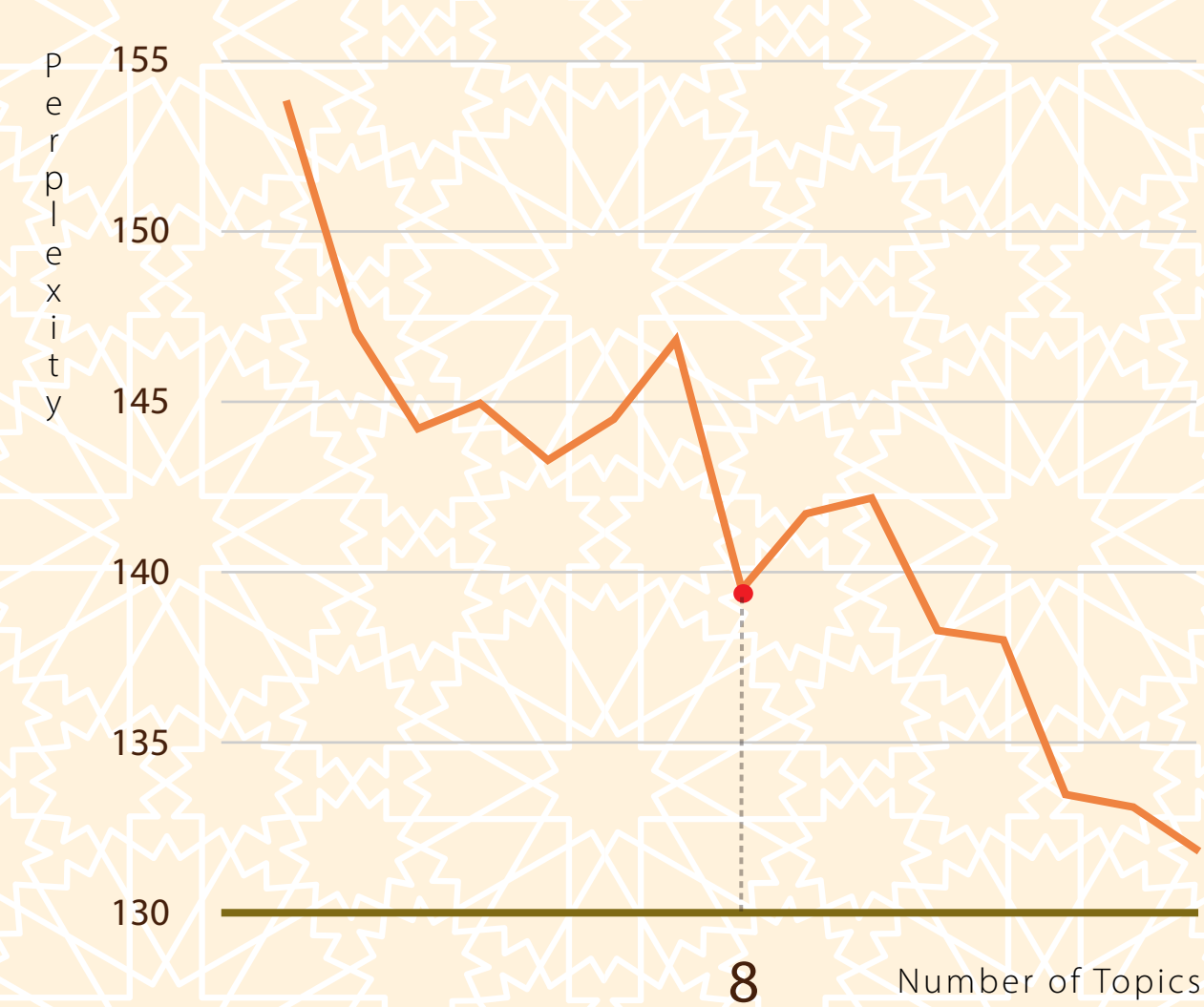
4 Centuies

4.1 TOPIC MODELING

We apply the statistical topic model **Latent Dirichlet Allocation (LDA)** to extract topics of semantically related words in Muqam poems, instead of manually identifying topics. We first preprocess the Muqam text using tokenization and POS tagging models in the Python package Jieba, and keep only the nouns and the adjectives in the poems for topic modeling. Next, we use the Python package scikit-learn to conduct LDA. By calculating the elbow point based on perplexity, we set the number of topics of LDA to eight. Among the eight topics, four are already identified by previous research, which are love, cry, disaster, and faith *(Zhou, 1995)*. The other four topics, including appearance, wine, kingdom, and garden, are newly discovered by our research.

Perplexity Value of LDA

Number of Topics: 8

8 Topics of twelve Muqam

DISASTER: disaster, world, sun, love, pain, destiny, society, suffer, enemy, friend, man's world, joy, people, infatuation, scar

APPEARANCE: cherry lips, face, girl, hair, eyes, visage, nightingale, veil, everybody, body, wish, haggard, sun, vellus hair, soul

LOVE: lover, pain, love, garden, nightingale, slender, promise, flowerbed, destiny, tryst, misery, wish, tender, wonderful, robber

WINE: lover, wine, beggar, pub, soul, thirst, Allah, winecup, secret, destiny, moth, world, tavern, wine glasses, believer

GARDEN: fairy, home, goddess, paradise, juniper, miss, valerian, cherry lips, musk, flowerbed, flower, hair, oasis, black eyes, garden

KINGDOM: Allah, flower, lover, monarch, slave, nightingale, beauty, beggar, king, binocular, heaven, love, alley, wealth, world

CRY: heart, hair, pain, lover, love, blood tears, body, tears, pearl, eyes, beauty, soul, eyelash, flower, lips

FAITH: Allah, nightingale, lover, rose, garden, faithful, face, precious, king, sorrow, lamentation, man's world, care, piety, sorrow
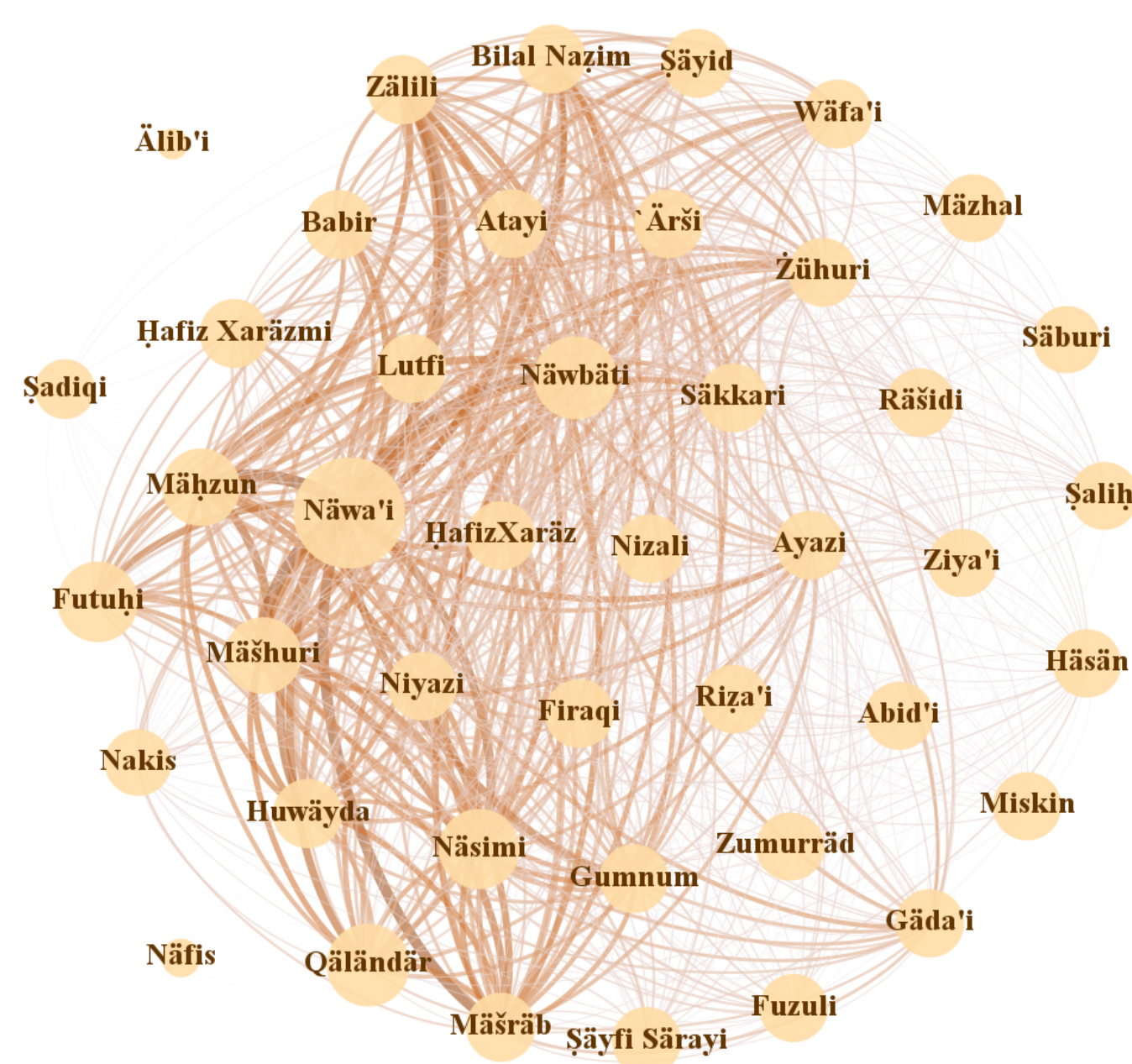
4.2 SENTIMENT ANALYSIS

Sentiment analysis tries to identify emotion from texts. Previous Muqam research studies the sentiment difference among different parts of Muqam, and the sentiment difference among different poets has not been investigated. Thus, we employ the Python package baidu-aip to directly calculate the sentiment of each poem and analyze the sentiment distribution of poets. The number and percentage of positive poems written by each of the Muqam poets are shown in Figure.

Poet Name	Number of Positive Poems	Percentage of Positive Poems	Poet Name	Number of Positive Poems	Percentage of Positive Poems
Näwa'i	33	80.49%	Säkkari	2	50.00%
Mäšräb	10	90.91%	Ḥafiz Xaräzmi	2	66.67%
Näwbäti	9	100.00%	Šäyfi Särayi	2	100.00%
Mäšhuri	7	77.78%	Niyazi*	2	100.00%
Näsimi	6	85.71%	Firaqi	2	66.67%
`Ärši	5	83.33%	Şadiqi*	2	100.00%
Huwäyda	4	80.00%	Atayi	2	66.67%
Zälili	4	100.00%	Fuzuli	1	100.00%
Futuḥi	4	100.00%	Zumurräd	1	50.00%
Bilal Naẓim	4	80.00%	Nizali*	1	100.00%
Mäḥzun	3	33.33%	Räšidi	1	50.00%
Gäda'i	3	100.00%	Lutfi	1	33.33%
Ayazi	3	75.00%	Riza'i	1	100.00%
Babir	3	75.00%	Ziya'i*	1	50.00%
Żuḥuri	3	60.00%	Miskin	1	100.00%
Gumnum	3	100.00%	Älbi'i*	1	100.00%
ḤafizXaräz	3	75.00%	Nakis*	1	100.00%
Qäländär	3	50.00%	Mäzhal*	1	100.00%
Wäfa'i	2	100.00%	Şalihi	0	0.00%
Šäyid*	2	100.00%	Näfis	0	0.00%
			Säburi	0	0.00%

*The asterisk mark indicates that the name is translated by the author according to Transliteration used from post-1978 Uyghur Arabic Script

---

4.3 SOCIAL NETWORK ANALYSIS

Social network analysis measures relationships between social actors. While previous research focuses on analyzing Muqam poets individually *(Zhou, 2006)*, social network analysis could discover the topic similarities between multiple poets. We select nouns that appear more than ten times in Muqam poems, calculate the **noun co-occurrence** between poets, and visualize the co-occurrence relationship using a social network via Gephi. Figure shows the noun co-occurrence network between Muqam poets. The thicker the lines in the graph, the greater number of nouns shared by the two poets. We could observe that many poets share similar nouns and topics in their poems.

REFERENCES

Anderson, Elise (2012): "The Construction of Ämännisä Khan as a Uyghur Musical Culture Hero", in: Asian Music 43,1: 64-90.
Dawamat, Tumur (Eds.) / Xinjiang Uygur Autonomous Region Twelve Muqam Research Society / Xinjiang Uygur Autonomous Region Classical Literature Research Society (1997): Uyghur Twelve Muqam. China Encyclopedia Publishing House.
Han, Yunxia / Xiao, Xuejun (2017a): "Summary of The Twelve Muqam study (1949-2015) (1) (Chinese)" in: Journal of Xinjiang Arts University 15, 1:8–25.
Han, Yunxia / Xiao, Xuejun (2017b): "Summary of The Twelve Muqam study (1949-2015) (2) (Chinese)" in: Journal of Xinjiang Arts University 15, 2:1–10.
Harris, Rachel (2008): The making of a musical canon in Chinese Central Asia: the Uyghur Twelve Muqam. Ashgate Publishing, Ltd.
Imin, Nurbiye / Tudaji, Rehemitulla / Juma, Osman (2019): "Study on China's Uyghur Twelve Muqam in the Past 40 Years of Reform and Opening up", in: Third International Conference on Art Studies: Science, Experience, Education, Moscow, Russia, October 2019 :78-80. DOI: 10.2991/icassee-19.2019.15
Li, Dongming / Meng, Xing (Eds.) (2008): Twelve Muqam Lyrics (1st ed.). Xinjiang People's Publishing House.
Light, Nathan (1998): Slippery paths: the performance and canonization of Turkic literature and Uyghur Muqam song in Islam and modernity. Ph.D. Thesis, Indiana University.
Tursun, Guzelnur / Tudaji, Rehemitulla / Juma, Osman (2019): "Overview of China's Uyghur Twelve Muqam in the Past 40 Years of Reform and Opening up", in: Third International Conference on Art Studies: Science, Experience, Education, Moscow, Russia, October 2019 :78-80. DOI: 10.2991/icassee-19.2019.35
Zhou, Ji (1995): "Ten topics in the Uyghur Twelve Muqam (Continued 1) (Chinese)", in: Journal of Xinjiang Normal University (Edition of Philosophy and Social Science), 2: 14–25.
Zhou, Ji (2006): "A review of Uyghur Muqam research in the 20th century (Chinese)". in: Journal of Xinjiang Arts University, 4, 11–17.