

Ontology-based Research Community Detection in UQU

Rahaf Al-Harbi - Maram Al-Sharif - Raghad Al-Mofarji - Rawan Al-Harbi - Lamia Al-Thagafi - Rasha Al-Harathi

Supervised by: Dr. Sarah Al-Shareef

Computer Science Department, Umm Al-Qura University, 2022

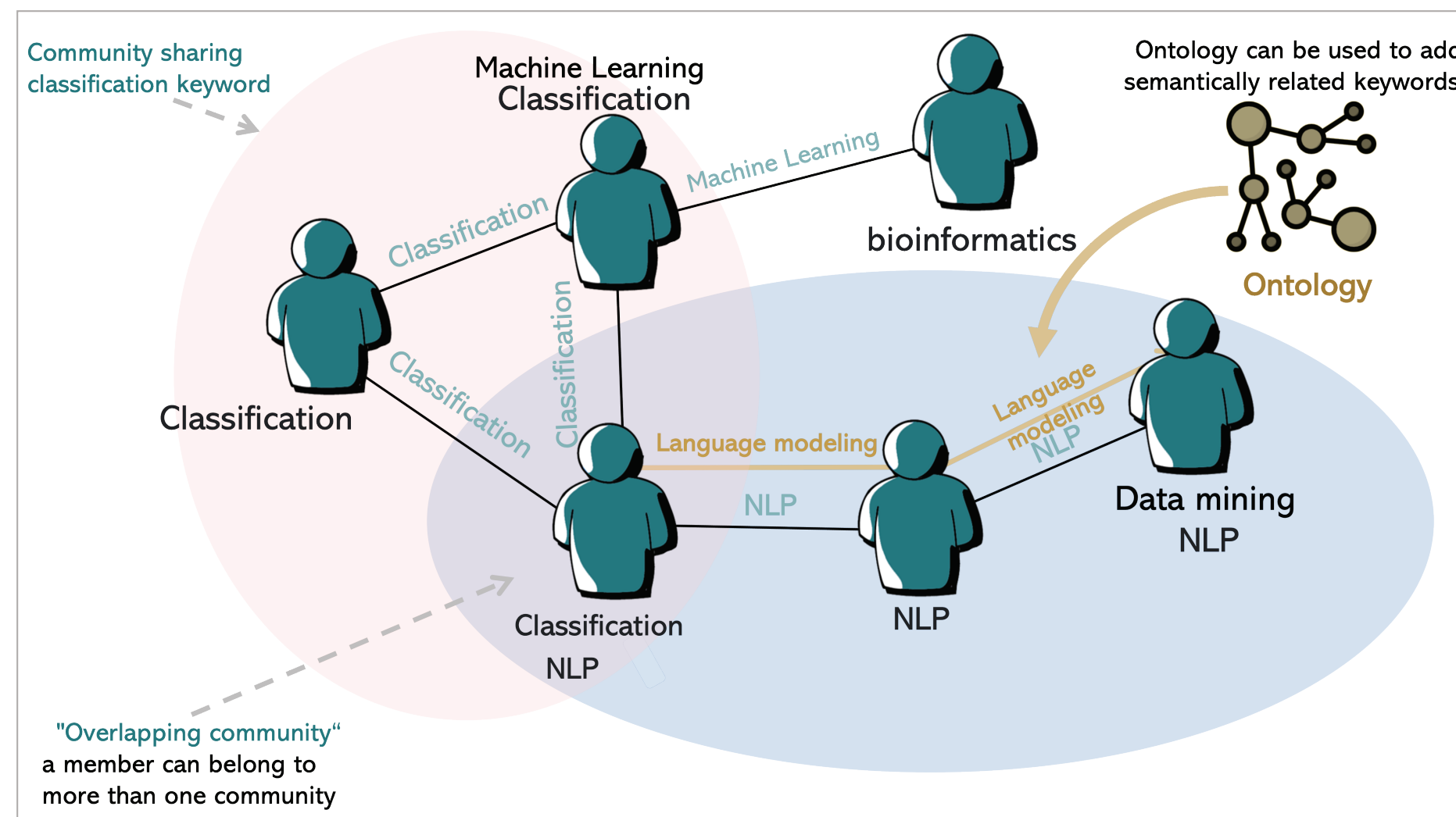


Abstract

A community refers to a group or a cluster that is densely connected within a complex network which is known as community detection. Clustering researchers in communities is an important task to support a range of techniques for analyzing the research environment and helps researchers find people in the same field of interest to collaborate. In computer science, ontology is capture knowledge about a particular area using relevant concepts and relations. **This study introduces a community detection framework that is employed on a multilayered Arabic scholarly network to expand the network using a cross-domain ontology and to detect communities based on co-authorship and keyword relations.** DBpedia is an existing cross-domain ontology that will be used to improve community detection. The evaluation was performed on three levels. First, the framework was applied to synthetic networks to determine the efficiency of the algorithms. Then, detect the communities within a real-data network based on **Umm-Al-Qura University (UQU)** faculty members from **three colleges with Arabic publications**. Finally, the UQU network was expanded by **adding semantically related keywords**. To our knowledge, **this is the first work that proposed detecting communities within the Arabic scholarly network** with and without the assistance of cross-domain ontology.

Introduction

- Example of a scholarly network of researchers sharing their publications' keywords and co-authorship



Community detection is useful for: and make a horizontal bulletin for three benefits: Research collaboration, Talent recruitment, and more.

Objectives

This research project aims to:

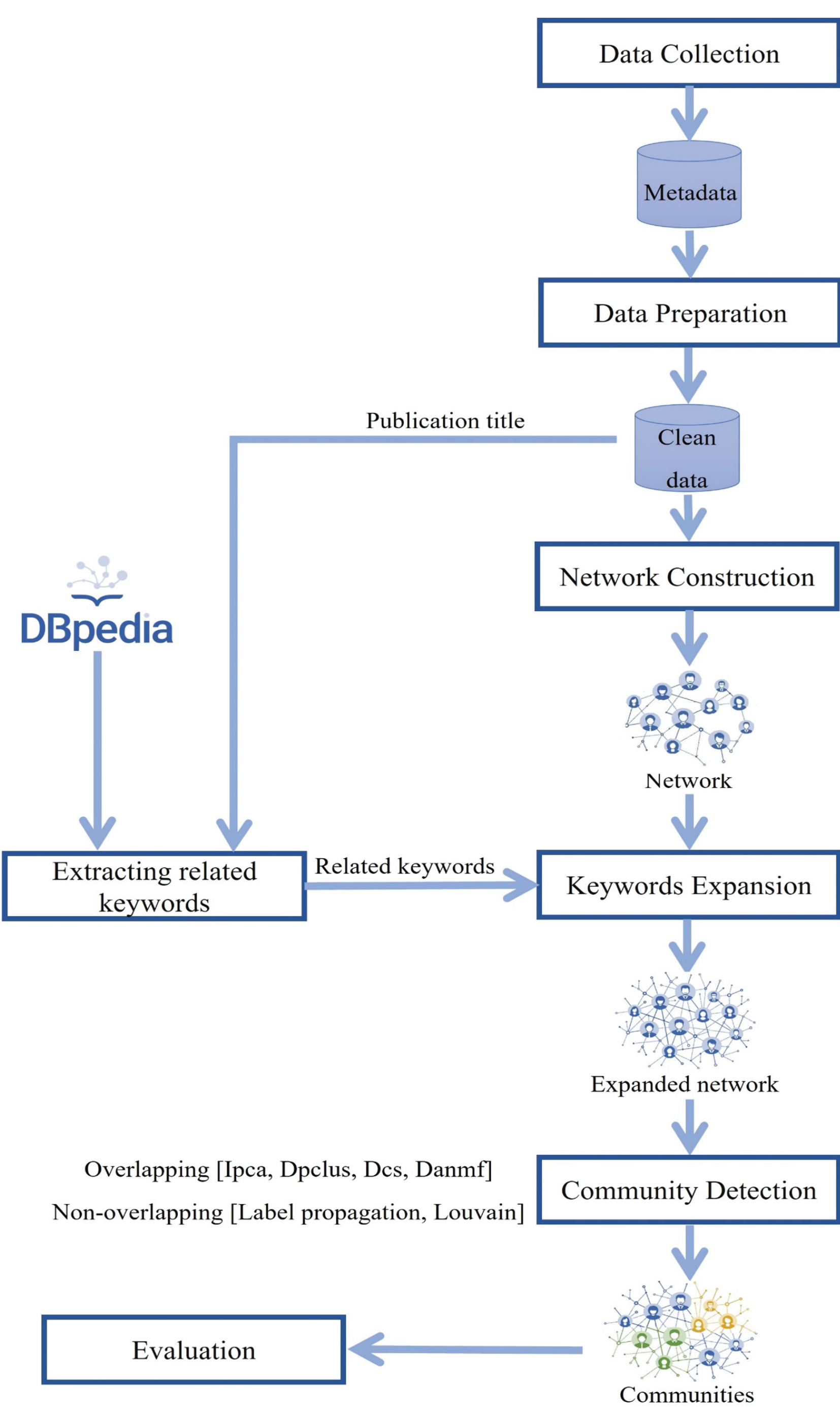
- **Detecting communities** within a given set of researchers based on their scientific papers published in UQU.
- **Using an existing ontology** as proof of improving community detection.

Related work

Table 1: Different system utilizes the ontology to enhance their outcomes, whether in hierarchical clustering, sentiment analysis, data mining, topic modeling, and network analysis.

Slabbekoorn et al. (2016)	Propose an ontology-assisted method to cluster Twitter users based on their interests. [1]
Ihab et al. (2019)	Suggested an ontology-based method for sentiment analysis and community detection on social media. [2]

Methodology



Experiment

1-Applying Community Detection on Synthetic Data

Algorithms	ONMI	
	Weighted	Un-Weighted
Ipca vs. ground truth	0.2442	0.2442
Dpclus vs. ground truth	0.1793	0.1793
Dcs vs. ground truth	0.2240	0.2380
Danmf vs. ground truth	0.2240	0.2380

Synthetic network:
50 nodes
136 edges
3 overlapping clusters for ground truth.

Algorithms	ONMI	
	Score	
Dcs	0.312	1
Dpclus	0.271	1
Ipca	0.271	1
Danmf	0.271	1

Identical outcomes

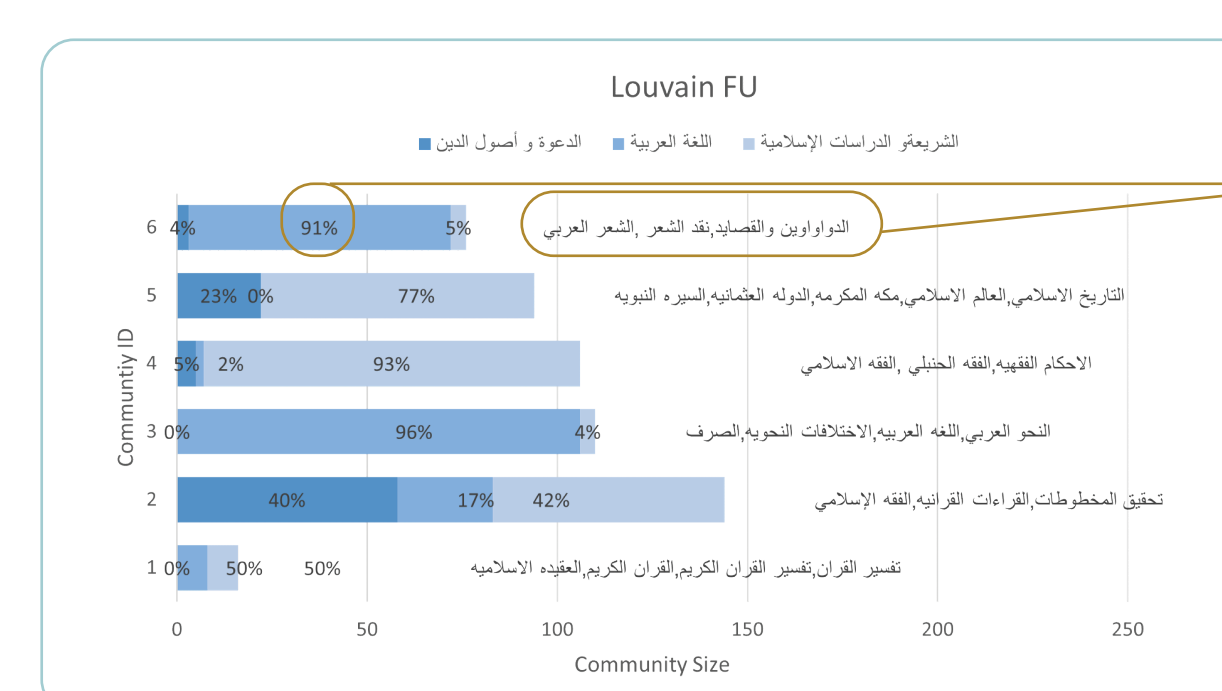
Finding: The unweighted and weighted results are not significantly different.

2. Applying Community Detection on UQU Data

Algorithms	Number of Communities		Average Number of Nodes per Community		Average Number of Communities per Node		ONMI	
	Weighted	Un-Weighted	Weighted	Un-Weighted	Weighted	Un-Weighted	Weighted	Un-Weighted
Ipca	48	48	94.771	94.771	6.612	6.612	0.1573	0.1573
Dpclus	46	45	18.696	19.067	1276	1277	0.0918	0.0823
Danmf	8	8	88.125	88.125	1	1	0.0470	0.0467
Dcs	3	3	561.333	561.333	2.382	2.382	0.2245	0.1900

High overlapping No overlapping Similar

Overlapping communities across colleges were detected in UQU Network and we found that Weights has no impact on the outcome.



The most three repeated keyword has a correlation with the majority of the affiliation

Finding: Weights does not play a significant role. Keyword list indicated the shared interest within a single community

3. Applying community detection on semantically expanded UQU data.

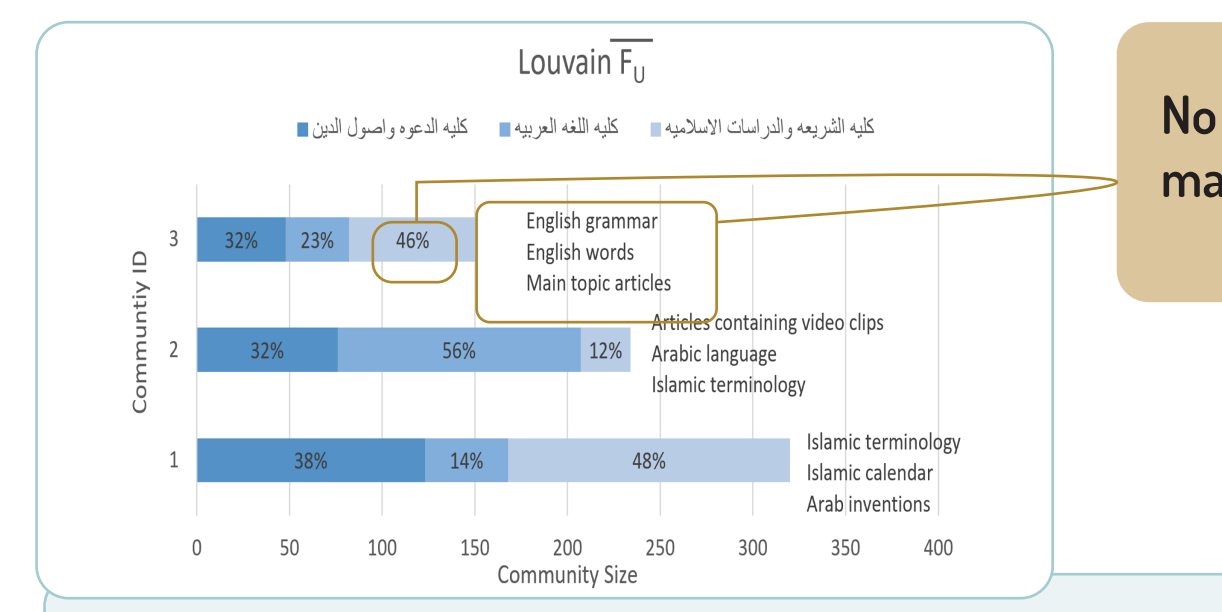
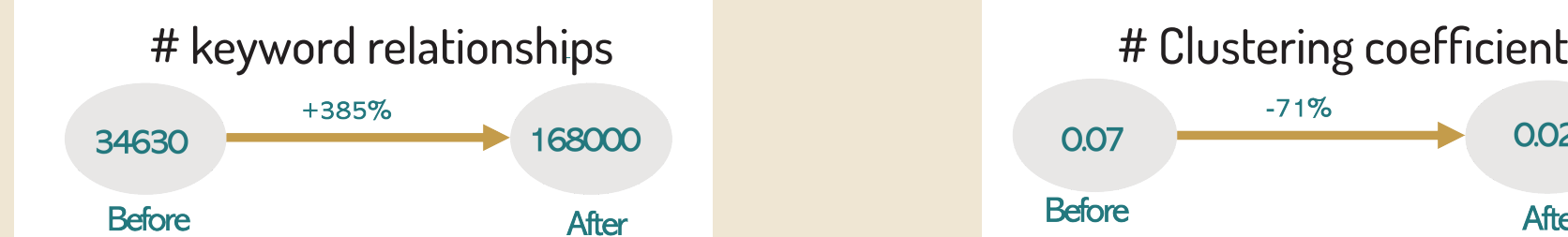
Algorithms	Number of Communities	Average Number of Nodes per Community	Algorithm vs. Colleges	ONMI	Change percentage
Dpclus	18	42.389	Dpclus vs. Colleges	0.0493	+14%
Dcs	1	705	Dcs vs. Colleges	0.0	0%
Danmf	7	100.714	Danmf vs. Colleges	0.0664	-18%

Fewer communities Larger communities Different from college affiliation

Why fewer but larger communities?

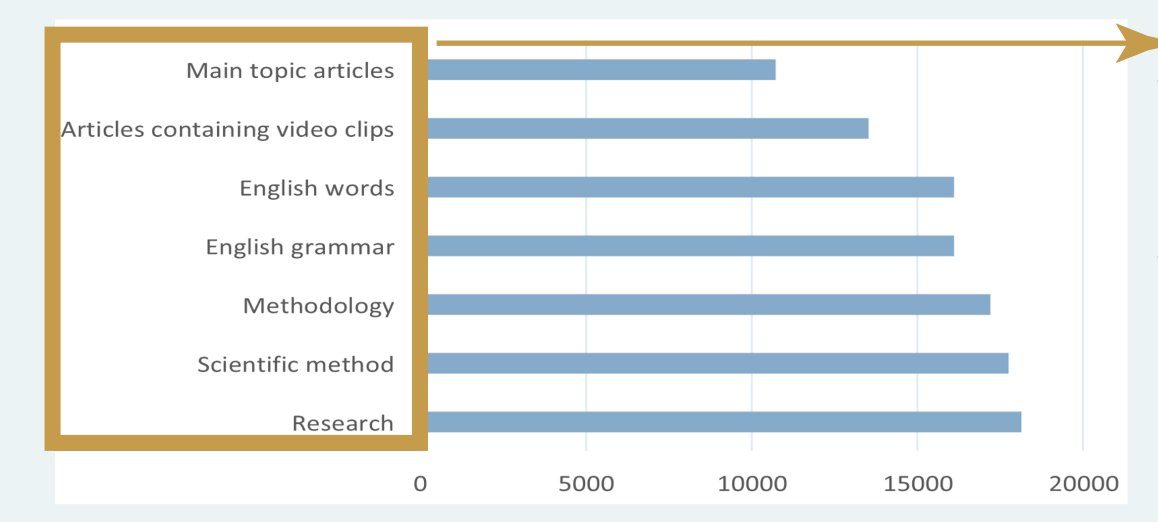
Finding: Fewer but larger communities were detected after keywords semantic expansion.

DBpedia semantic expansion



No correlation between keywords and majority of the affiliation

Fig: most frequent words in the network



DBpedia added generic keywords for the majority of the relationship make the network almost fully connected and harder to detect communities with real shared interest

4. Removing top 4 most frequent keywords and reapplying community detection

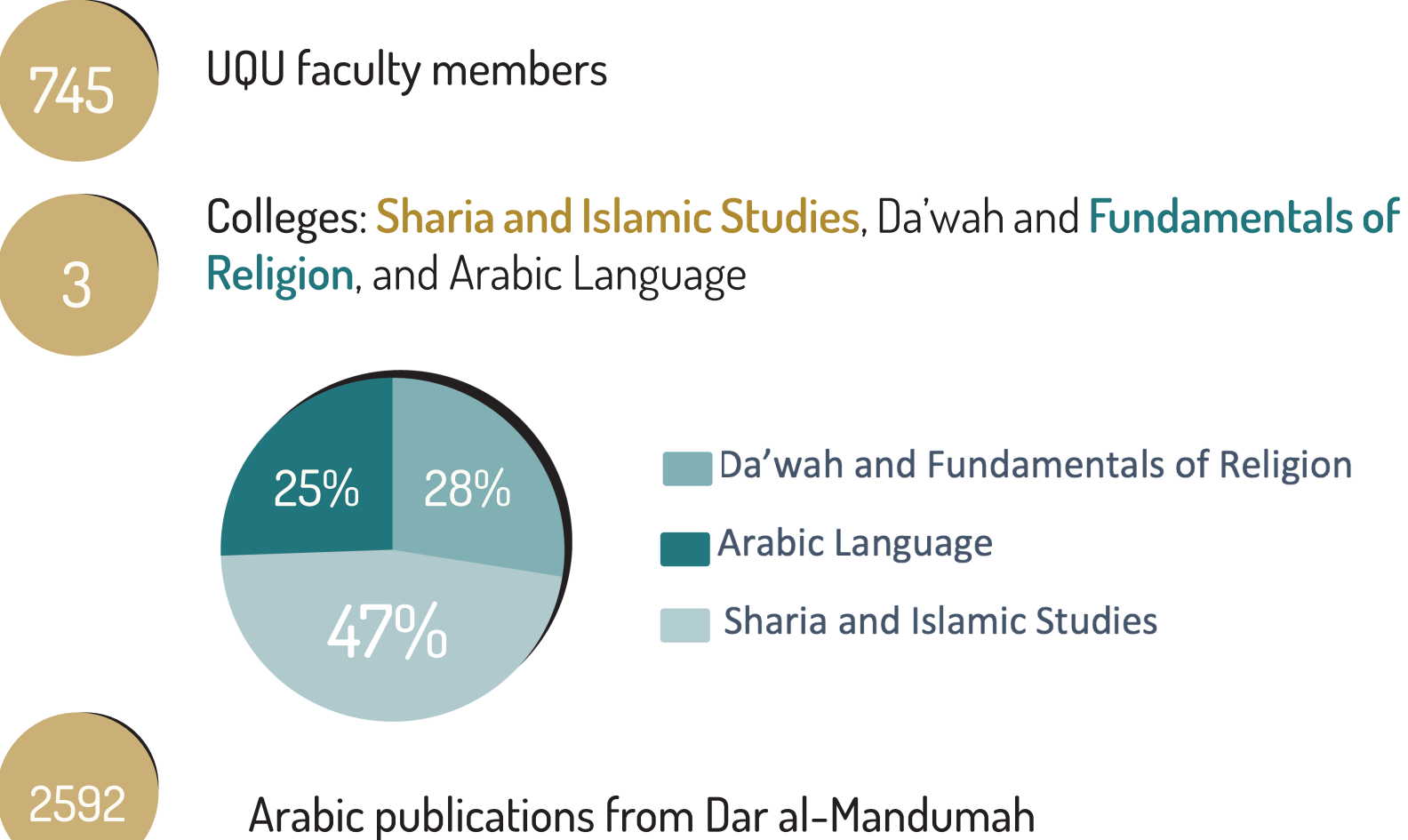
All the comparisons have increased except any comparison that include the algorithm Dcs as it is still failing to detect any communities with other comparison's increase between (+9%) and (+93%) and the decrease is in the range (-46%) to (-51%) where this decrease was only when applying on the un-weighted flattened network.

Algorithms	Number of Communities	Average Number of Nodes per Community	Algorithm vs. Colleges	ONMI	Change percentage
Dpclus	18	42.389	Dpclus vs. Colleges	0.0493	+14%
Dcs	1	705	Dcs vs. Colleges	0.0	0%
Danmf	7	100.714	Danmf vs. Colleges	0.0664	-18%

The decreased result The increased result

Finding: Better communities are extracted when removing generic keywords

Data



Networks	Relationship	Weights	Nodes and Edges
Authorship Layer	Co-authored a publication	# of shared publications	705,702
Keywords Layer	Shared keywords in their publications?	# of shared keywords	705,34630
Flattened Network	Either co-author a publication or shared a keyword	Normalized the sum of both	705,34645

References

- 1 Kristian Slabbekoorn, Tomoya Noro, and Takehiro Tokuda. Ontology-assisted discovery of hierarchical topic clusters on the social web. J. Web Eng., 15:361-396, 2016.
- 2 Zouaoui, Samia; Rezeg. Islamic inheritance calculation system based on arabic ontology (arafamonto). King Saud University - Computer and Information Sciences, 2018.

Conclusion

Using semantic expansion for detecting communities within UQU network is promising when used selectively

Future work

- 1 Generate domain ontology using publications' abstracts
- 2 Employ an intelligent method for keyword expansion such as word embeddings.