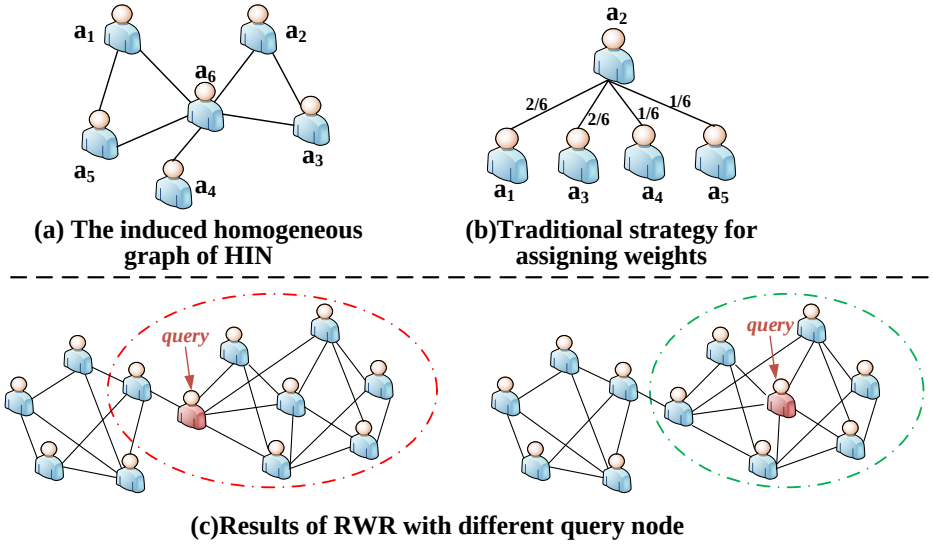
Community Search over Heterogeneous Information Networks via Weighting Strategy and Query Replacement

Fanyi YANG , Huifang MA ,Weiwei GAO, Zhixin LI

Frontiers of Computer Science, DOI: 10.1007/s11704-022-1329-9

Problems & Ideas

- Problems of Community Search over HINs:
  - When mapping HINs to the corresponding homogeneous information networks, they fail to properly utilize the intermediate nodes to assign weights on the edges of the induced homogeneous graph.
  - Existing methods struggle with the query-bias issue and especially perform unsatisfactorily if the query node stands at the boundary region.
- Ideas: Community search via weighted strategy and query Replacement over HINs model



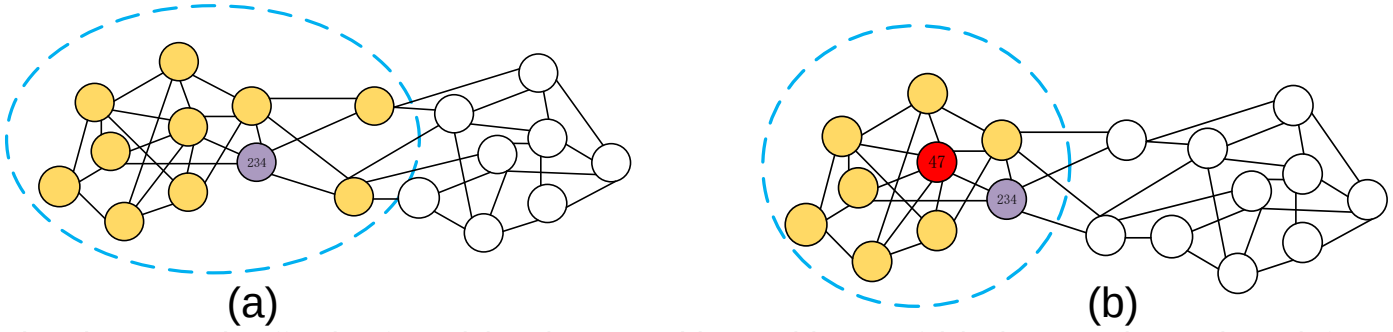
An illustrative example of an induced homogeneous graph and query-bias issue. (a) An induced homogeneous. (b) Traditional weight assignment strategy for v_2. (c) Local community detection results (with RWR) for different query node of a toy network.

Main Contributions

- Contributions:
  - a new weighting strategy to capture the rich semantic information in a meta-path by properly modeling the intermediate nodes of the meta path and integrating the information of multiple meta paths
  - a query node replacement strategy using CCLC score to avoid query bias in local community detection.

Table 1: Comparisons of overall performance between CSQR and baselines

Datesets	Metrics	Basic-core	CSQR-W	CSQR-R	CSQR-WR	CSQR
IMDB	*F1-score*	0.605	0.626	0.631	0.613	**0.642**
	NMI	0.594	0.611	0.629	0.603	**0.639**
DBLP	*F1-score*	0.542	0.561	0.572	0.559	**0.583**
	NMI	0.535	0.553	0.566	0.547	**0.579**
Last.fm	*F1-score*	0.563	0.573	0.584	0.554	**0.598**
	NMI	0.552	0.564	0.571	0.546	**0.587**



(a) (b)

seed replacement is effective for solving the query bias problem. Left(a): the experimental result for CSQR-R; Right(b): the experimental result for CSQR.