Author Set Identification via Quasi-Clique Discovery

Discovery

A presentation for DAA 2022/2023

Manuel Gomes

06/01/2023

Table of Contents

Table of Contents

Author Identification

The problem of author identification
has been extensively studied, which
aims to learn a model to rank
potential authors for an anonymous
paper based on public information.

Author Identification

The problem of author identification has been extensively studied, which aims to learn a model to rank potential authors for an anonymous paper based on public information.
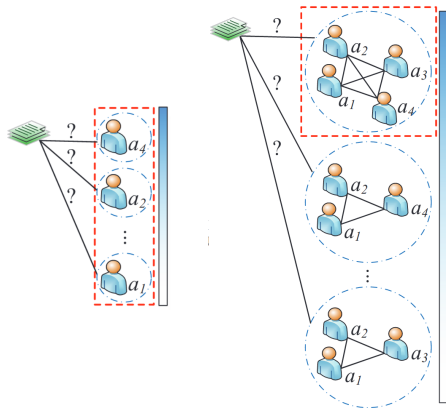
Possible Uses

- Author identification
- Relevance search
- Personalized recommendations
- Reviewer recommendation
- Collaboration discovery
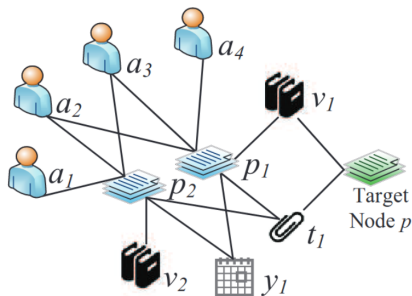
Problem formulation

Current approaches:

- Focus on single author identification
- Do not account for the optimal number of writers
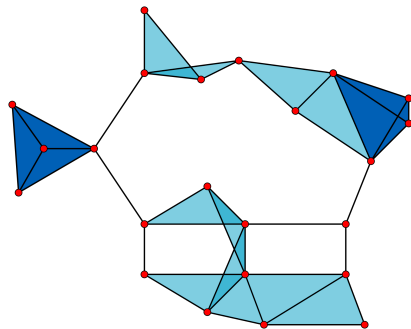- Do not take into account the relation between possible authors

Heterogeneous Information Network (HIN)

- Directed graph
- Nodes and edges have types
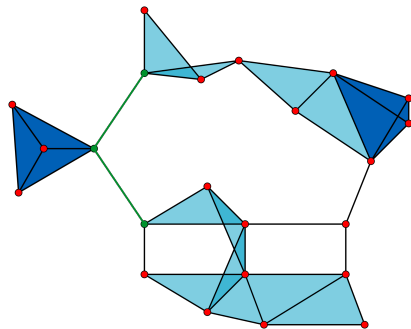
Clique (Graph Theory)

- Undirected graph
- Subset of vertices such that every two distinct vertices are adjacent

Quasi-Clique (Graph Theory)

A set of nodes S is an α-quasi-clique
if the edge density of the subgraph
induced by S exceeds a threshold
parameter $\alpha \in (0, 1)$
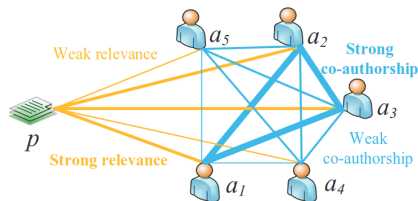
Table of Contents

Heterogeneous Information Network (HIN)

- HINs are complex structures
- Complex structures lead to complex algorithms
- How to simply this HIN without removing relevant information?
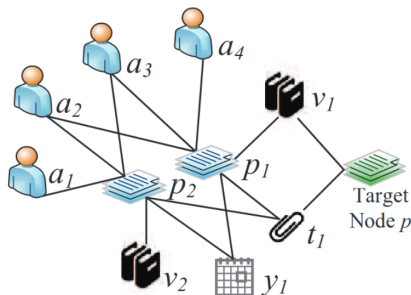
Weighted Paper-Ego-Network

- Extraction of a task-guided embedding to learn the low-dimensional representation of network
- Creation of a simpler weighted graph

Optimal Author Subset

- To find an optimal author subset that is related to the anonymous paper is a NP-hard problem
- Discovering the largest clique is inapproachable and the clique concept is in practice too strict to miss a single edge in an otherwise dense subgraph
- Quasi-clique has been significantly used to discover dense subgraphs
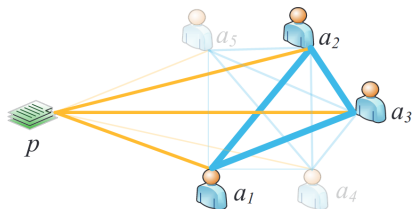
Algorithm description

- Optimization problem
- Local search algorithm

Algorithm 1: OQCCE

Input : Weighted paper ego network $G_p = (V, E, W)$;
maximum number of iterations I_{max}; the
constrained node p

Output: A subset of nodes $S \subseteq V$ and $p \in S$

$S \leftarrow p$, $b_1 \leftarrow$ TRUE, $i \leftarrow 1$;

while b_1 and $i \leq I_{max}$ **do**

    $b_2 \leftarrow$ TRUE;

    while b_2 **do**

        if *there exists* $u \in V \setminus S$ *and* $g_{\alpha,\beta}(S \cup \{u\}) \geq g_{\alpha,\beta}(S)$

        then

          | $S \leftarrow S \cup \{u\}$;

        else

          | $b_2 \leftarrow$ FALSE;

    if *there exists* $u \in S$ *and* $u \neq p$ *and* $g_{\alpha,\beta}(S \setminus \{u\}) \geq g_{\alpha,\beta}(S)$

    then

      | $S \leftarrow S \setminus \{u\}$;

    else

      | $b_1 \leftarrow$ FALSE;

    $i \leftarrow i + 1$;

Table of Contents

Sensor Calibration

Multi-sensor or multi-modal
perceptual systems operate under
the assumption that accurate
geometrical transformations between
the sensors that collect the data are
known

ATOM Calibration Framework

- A general sensor calibration
  framework
- Based on an optimization
  algorithm
- Potential for enhancement