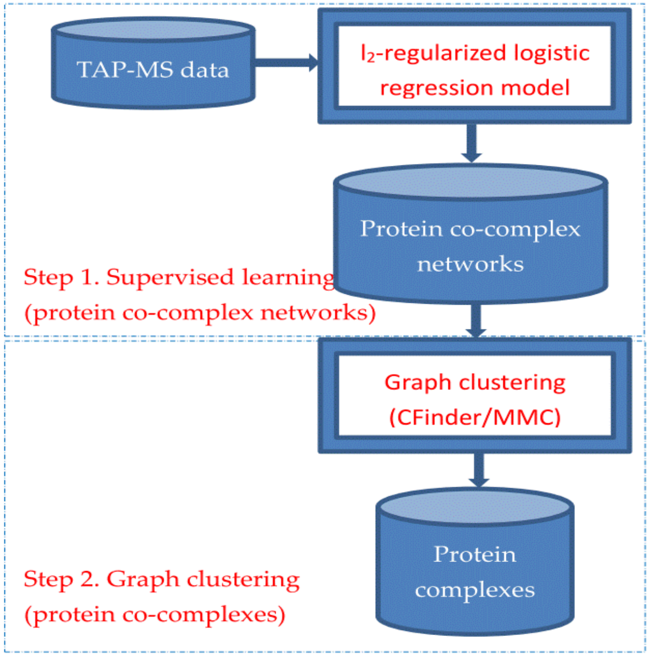
A framework combines supervised learning and dense subgraphs discovery to predict protein complexes

Suyu MEI

Frontiers of Computer Science, DOI: 10.1007/s11704-021-0476-8

Problems & Ideas

- Problems of existing methods
  - PPI networks contain a large number of interactions that do not form protein complexes, so that the inferred clusters are less biologically relevant;
  - The available TAP-MS data are not exploited to supervise the learning process (clustering) of protein complexes;
- Ideas: Combining TAP-MS supervised learning and graph clustering to infer protein complexes
  - Step 1. Supervised learning to predict protein co-complex networks;
  - Step 2. Graph clustering to infer protein co-complexes;


TAP-MS data

l_2-regularized logistic regression model

Protein co-complex networks

Step 1. Supervised learning (protein co-complex networks)

Graph clustering (CFinder/MMC)

Protein complexes

Step 2. Graph clustering (protein co-complexes)

Framework flowchart

Main Contributions

- **Performance comparison (step 1)**
- **Performance comparison (step 2)**