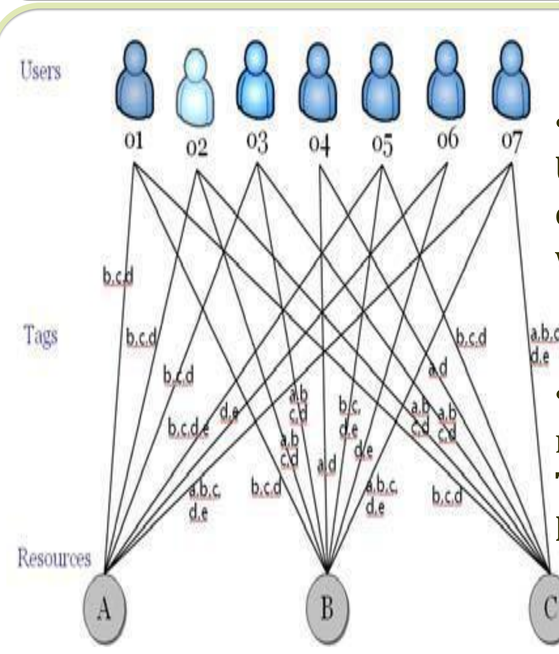
Detection of Overlapping Communities in Folksonomies

Abhijnan Chakraborty Saptarshi Ghosh Niloy Ganguly

Department of Computer Science & Engineering

Indian Institute of Technology Kharagpur, India

Background & Objectives



• Folksonomy: **Users** annotate online **Resources** with **Tags**

• Folksonomy is modeled as **Tripartite Hypergraph**

• Almost all existing community detection algorithms for folksonomies assign only a single community to each node

• Reality: Nodes belong to multiple overlapping communities
  • most users have multiple topics of interest
  • the same resource is often associated with semantically different tags by different users

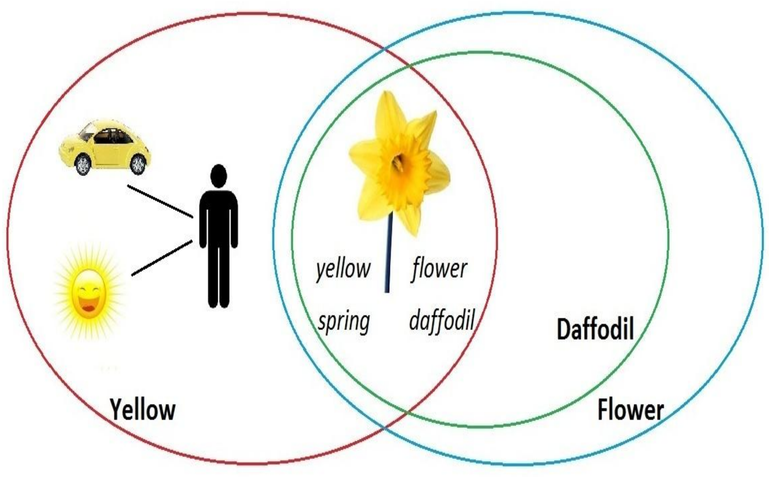
• Two prior approaches exist for overlapping community detection. Both work on projection of the tripartite hypergraph.

• Projections lose information and quality of communities is proved to be worse in projected network.

• Our Objective: Develop an algorithm to detect overlapping communities in folksonomies considering the complete tripartite hypergraph structure

Motivation

Why Overlapping Communities?



• Existing algorithms likely to put the daffodil image only into 'Daffodil' community based on majority tagging

• Algorithm for overlapping community detection
  • relate image with 'Yellow' community as well, can be recommended to users favoring yellow objects ⇒ better community-based recommendation

  • identify 'Daffodil' community as a subset of 'Flower' community ⇒ hierarchical organization of resources and tags into semantic categories

Algorithm

Idea: **Cluster links in stead of nodes**

i. Find similarities between all pairs of adjacent hyperedges using Algorithm 1.
ii. Construct the weighted line graph of the hypergraph. Hyperedges are nodes here and two such nodes are connected if they have non-zero similarity. Exact similarity score is represented as the edge-weight.
iii. Apply any community detection algorithm on that line graph. (We used Infomap algorithm [Rosvall et al., PNAS 2008])
iv. Each hyperedge gets placed into a single link-community.
v. A node inherits membership of all those communities into which the hyperedges connected with this node are placed.

Time Complexity = O($n.d^2$) where n = number of nodes and d = average degree

Algorithm 1 Compute Similarity between two Hyperedges

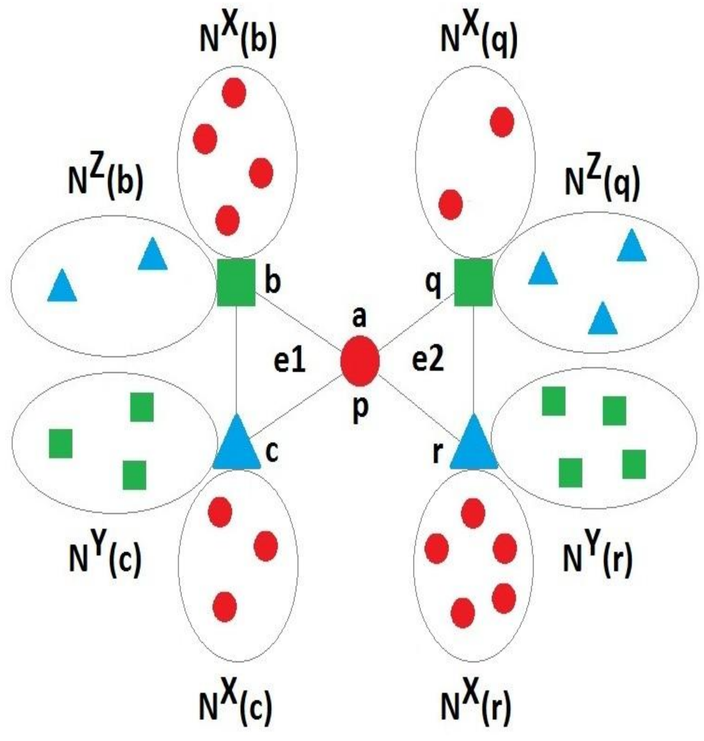
Input: hyperedges $e_1 = (a, b, c)$ and $e_2 = (p, q, r)$; $a, p \in V^X$; $b, q \in V^Y$; $c, r \in V^Z$
Output: sim, Similarity between e_1 and e_2

if $a \neq p$ AND $b \neq q$ AND $c \neq r$ then
  /* Hyperedges are non-adjacent */
  $sim \leftarrow 0$
else
  /* Without loss of generality, let $a = p$; Any of the other pairs may be common as well */

$$S_1 \leftarrow N^X(b) \bigcup N^X(c),\ S_2 \leftarrow N^Y(c),\ S_3 \leftarrow N^Z(b)$$
$$S_1' \leftarrow N^X(q) \bigcup N^X(r),\ S_2' \leftarrow N^Y(r),\ S_3' \leftarrow N^Z(q)$$

$$sim \leftarrow \frac{|S_1 \bigcap S_1'| + |S_2 \bigcap S_2'| + |S_3 \bigcap S_3'|}{|S_1 \bigcup S_1'| + |S_2 \bigcup S_2'| + |S_3 \bigcup S_3'|}$$

end if
return sim

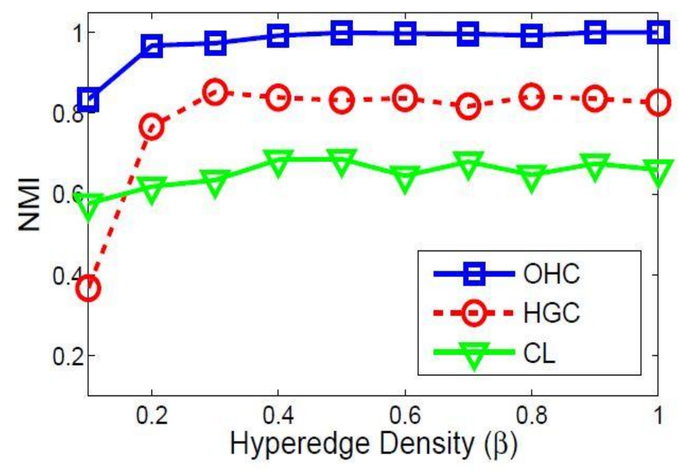


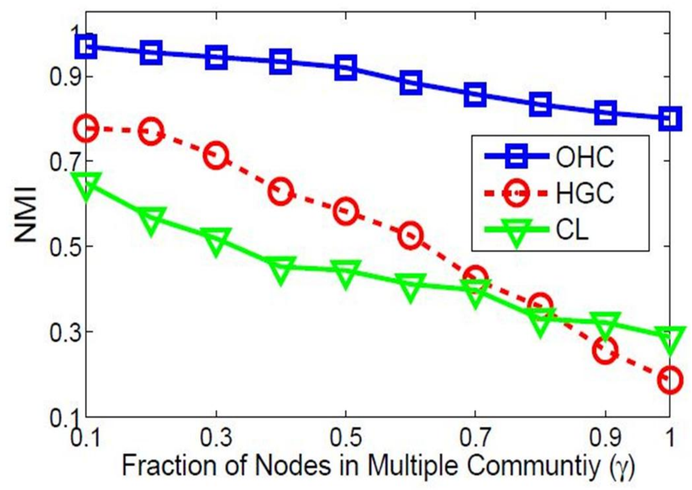
Experiments

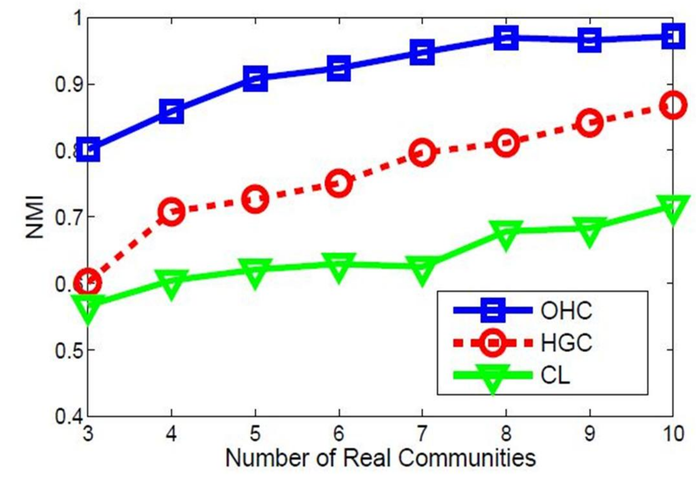
• Synthetic hypergraphs generated
  ▪ Each node assigned to one community, then β fraction of nodes assigned multiple communities
  ▪ Nodes in same community randomly connected with hyperedges
  ▪ Number of hyperedges is decided based on the specified density α

• Metric for Performance: **Normalized Mutual Information (NMI)**
  ▪ NMI is a measure of similarity between 'real' and 'detected' community structures.
  ▪ It falls in range [0,1]. Higher the NMI value, better the community detection algorithm.

• Compared the NMI performance of our algorithm (OHC) with the algorithms by
  ▪ Wang et al. [ICDM 2010] (CL) and
  ▪ Papadopoulos et al. [DWKDC 2010] (HGC)







Conclusion

- We proposed the first algorithm to detect overlapping communities considering the full tripartite hypergraph structure of folksonomies.
- It out-performs existing algorithms that consider projections of hypergaphs.
- The proposed algorithm can be used in recommending interesting resources and friends to users.

Contact

Abhijnan Chakraborty: abhijnan.cs@gmail.com
Saptarshi Ghosh: saptarshi.ghosh@gmail.com
Niloy Ganguly: niloy@cse.iitkgp.ernet.in

Complex Network Research Group (CNeRG)
Department of Computer Science & Engg
Indian Institute of Technology Kharagpur
http://cse.iitkgp.ac.in/resgrp/cnerg