A new approach of Community Detection Based on Seed Node

A thesis in partial fulfilment for the degree of

Masters of Computer Applications (MCA)

Under The Supervision of:

Dr. Prasenjit Choudhury

Assistant Professor NIT Durgapur

Submitted By : Bheem kumar (15/CA/639)

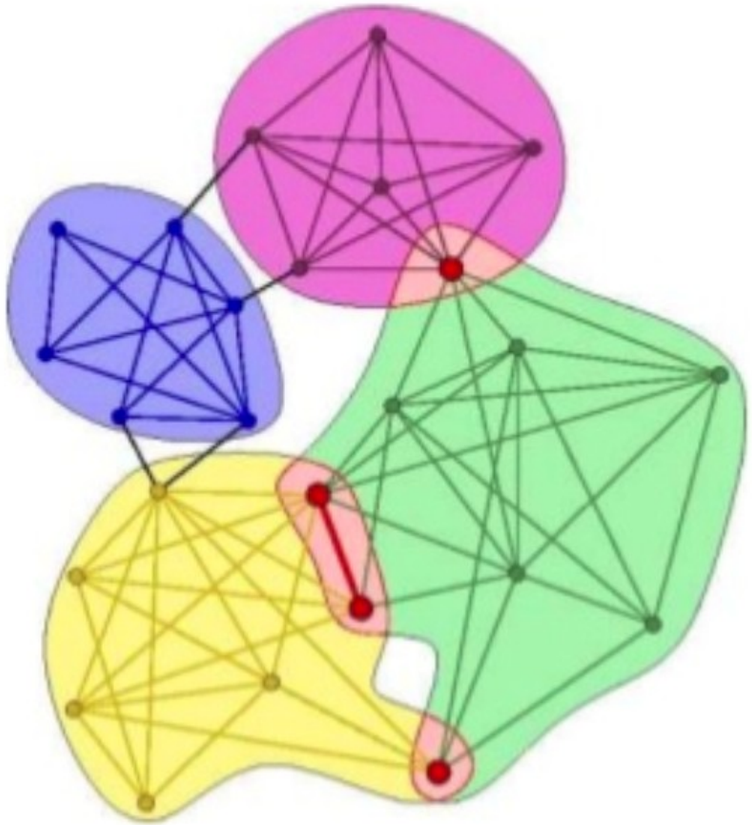
Outline

- Community
  - Types of community
  - Community detection
- Seed Node
- Clustering Coefficient
- Moudularity
- Basic Seed – centric Algorithm
- Our Algorithm
- Experimental results
- Comparision
- Conclusion
- References

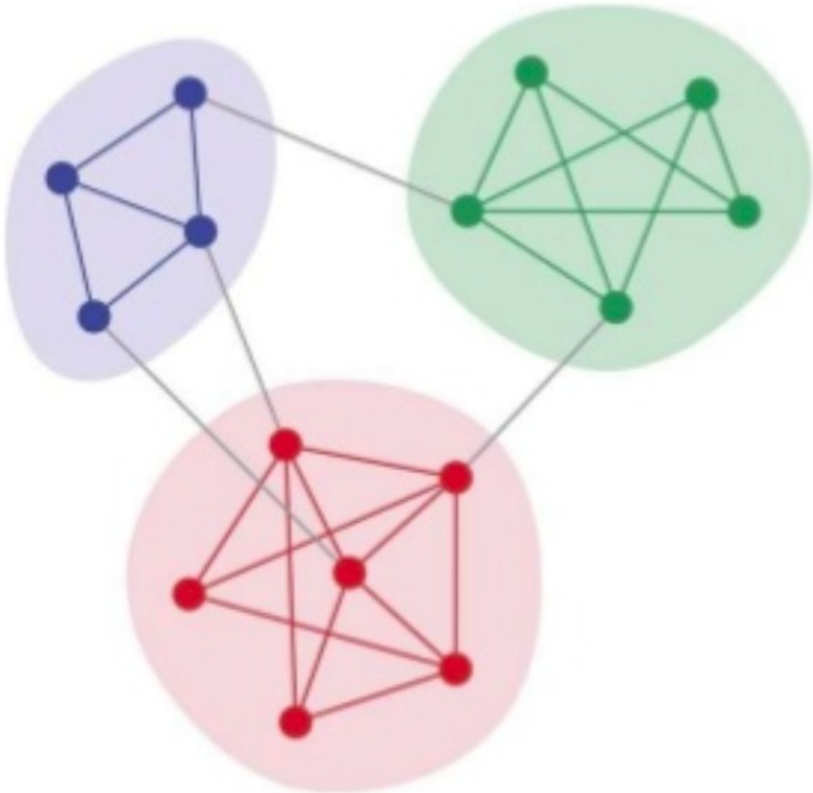
Community

A community is a small or large unit (a group of living or non- living things) who have something in common, such as religion, values, identity etc.

Types of Community



Overlapping Communities

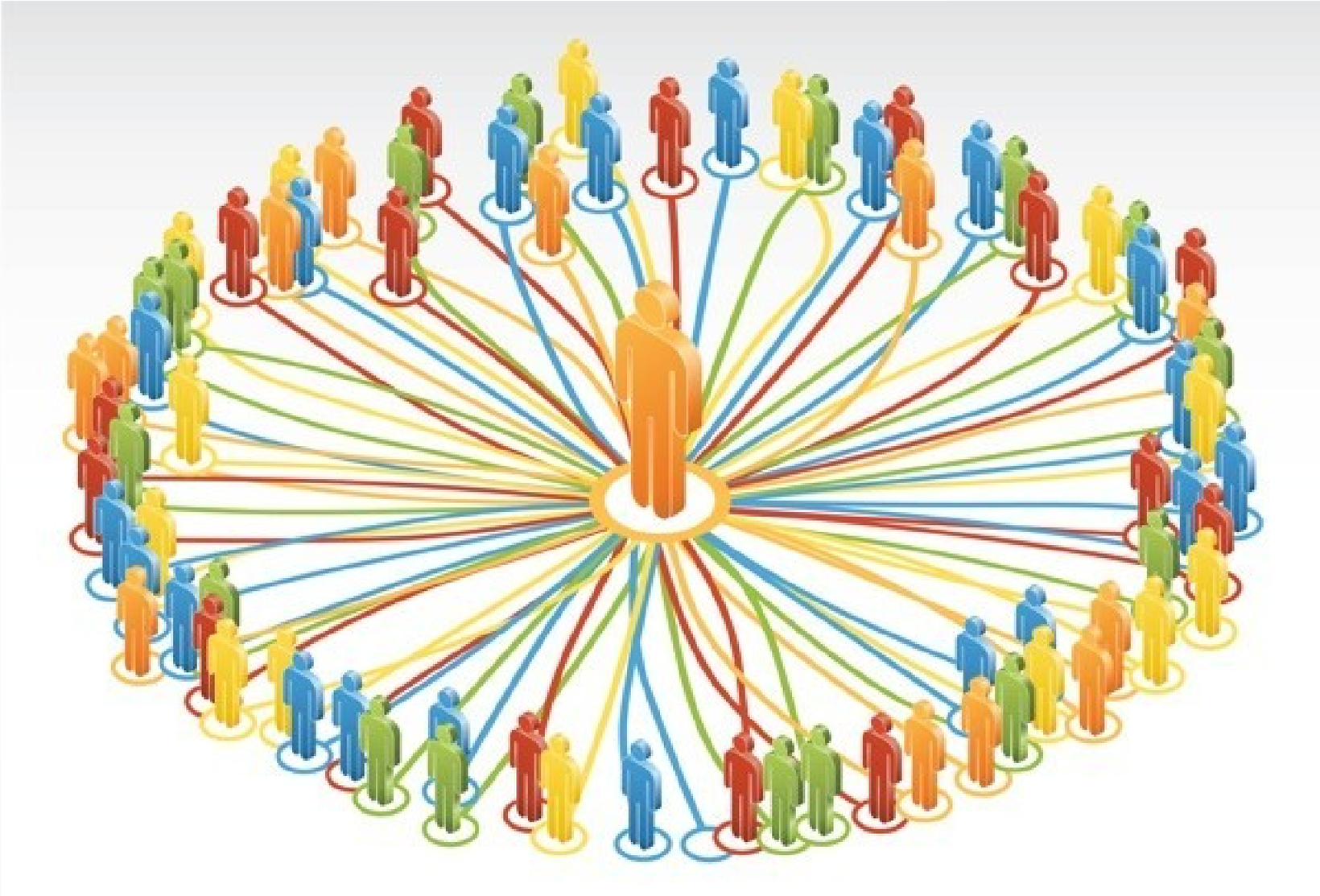


Disjoint Communities

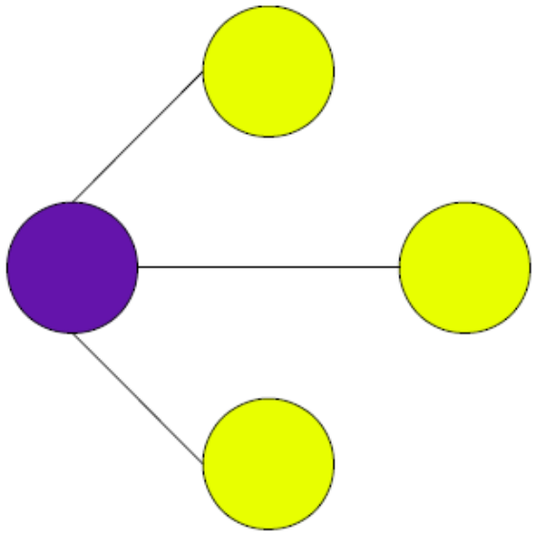
Community Detection

Discovering groups in a network where individual group or membership are not given explicitly.

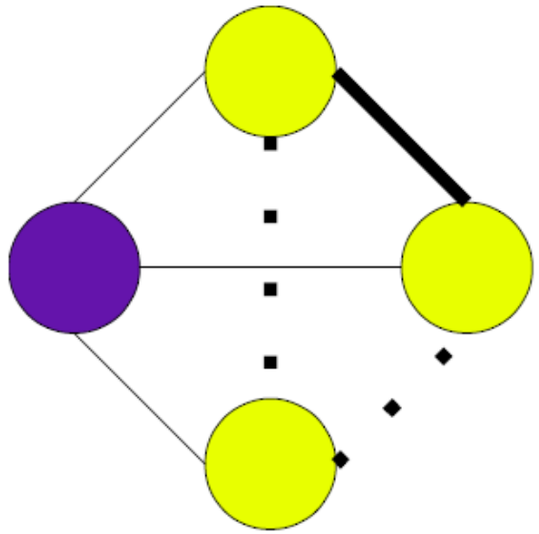
Seed Node



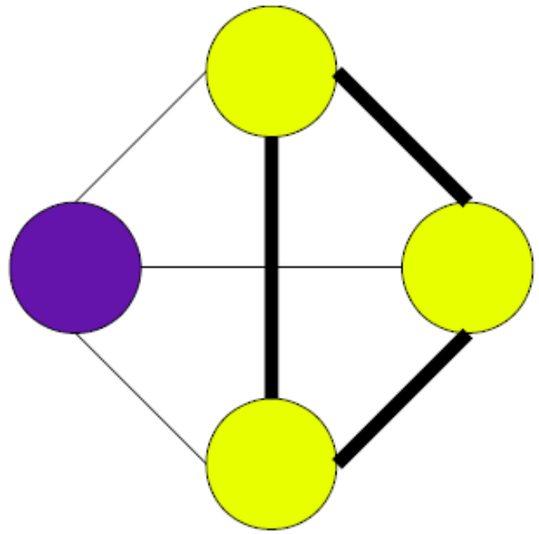
Clustering Coefficient



(a) No pairs formed among neighbors: $C = 0$



(b) One pair formed among neighbors: $C = 1 / 3$



(c) Three pairs formed among neighbors: $C = 3 / 3$

Modularity

- Modularity[1] is one of the measure structure of networks or graphs.
- Designed to measure the strength of division of a network into modules.
- The value of the modularity lies in the range [-1, 1].
- The network partition is as much good as Modularity of network is close to 1.

Basic seed-centric Algorithm

Algorithm 1 *General seed-centric community detection algorithm*

Require: $G =< V,E >$ *a connected graph,*
$C \leftarrow \emptyset$
$S \leftarrow$ *compute seeds(G)*
for $s \in S$ *do*
    $Cs \leftarrow$ *compute local com(s,G)*
    $C \leftarrow C + Cs$
end for
return *compute community(C)*

Our Seed-centric Community Detection

Notations:

CC - Clustering Coefficient of all the node available in the graph G.

Deg - Degree of all the nodes available in the graph G

CCMD - Clustering Coefficient Multiplied by Degree of corresponding node

LC - Local Community

LOC - List of Community

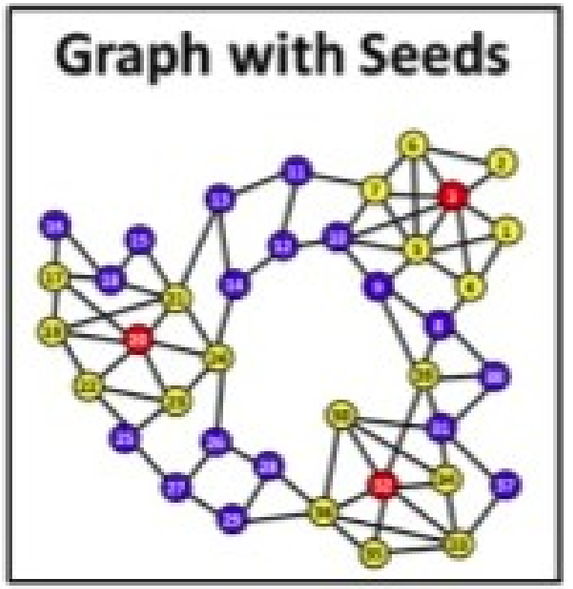
RN - Remaining Nodes

a - node id

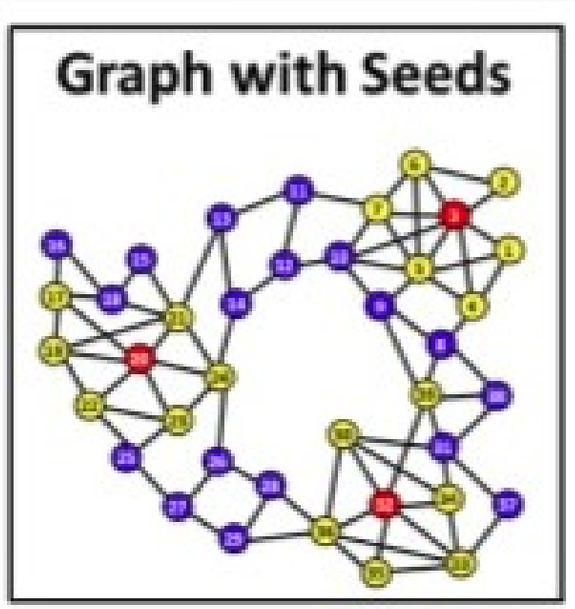
Our Algorithm

ALGORITHM 2 *Our algorithm for seed node and community detection*

Require: G =< V,E > a connected graph
output = seed node and communities
RN = 0
***while**(True)*
    CC = Clustering Coefficient of all the node in graph G
    Deg = degree of all the nodes in graph G
    *CCMD = CC * Deg*
    seed = max CCMD in CCMD list
    Deg = degree of seed in Deg list
    if**(neighbors of seed < threshold1 or CCMD of seed< threshold2) **then
        RN = remaining nodes in graph G
        break the loop
    else
        LC = seed + neighbors of seed
        LOC = LOC + LC
        G = G – LC
    end if
    if**(G is empty) **then
        break the loop
    end if
end while



Our Algorithm



ALGORITHM 3 *Our algorithm to Seed set expansion*

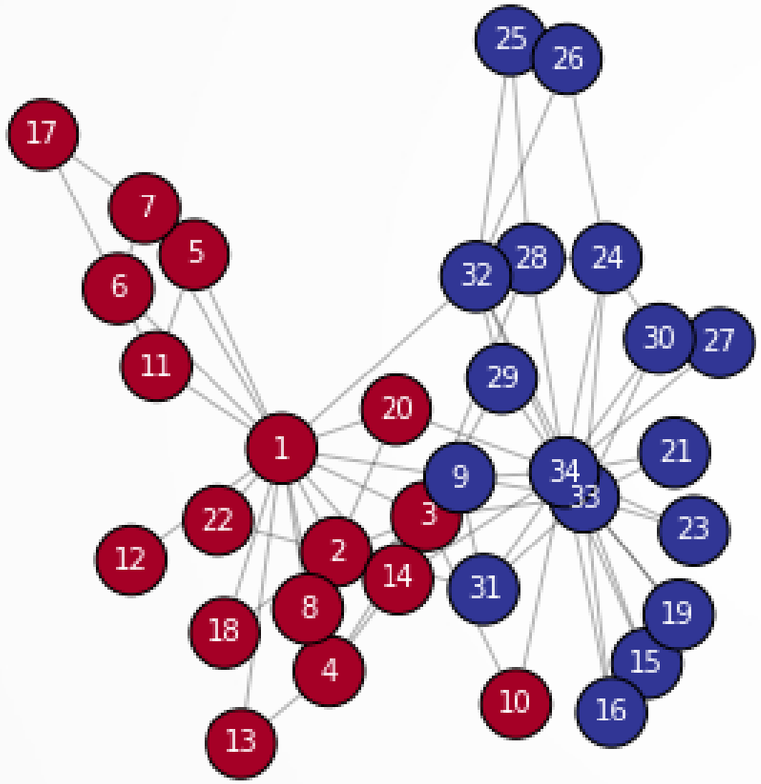
Input = RN
Output = communities
while*(length of RN > 0)*
    for *each node a ∈ RN*
        find the neighbors of a in each community(LOC)
        put the node a in that community, which have maximum number of neighbors of a
    end for
    RN = RN - a
end while

Experimental Results

Table 1. Real World Dataset

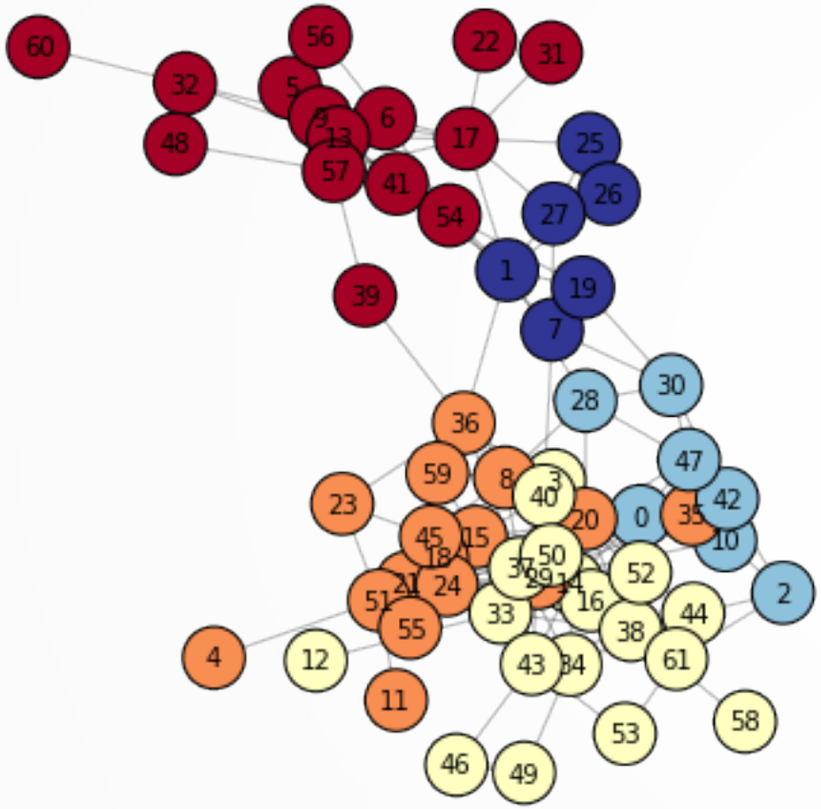
Networks	No. of nodes	No. of edges	References
Karate Club	34	78	[2]
Dolphins	62	159	[3]
Political Books	105	445	[4]

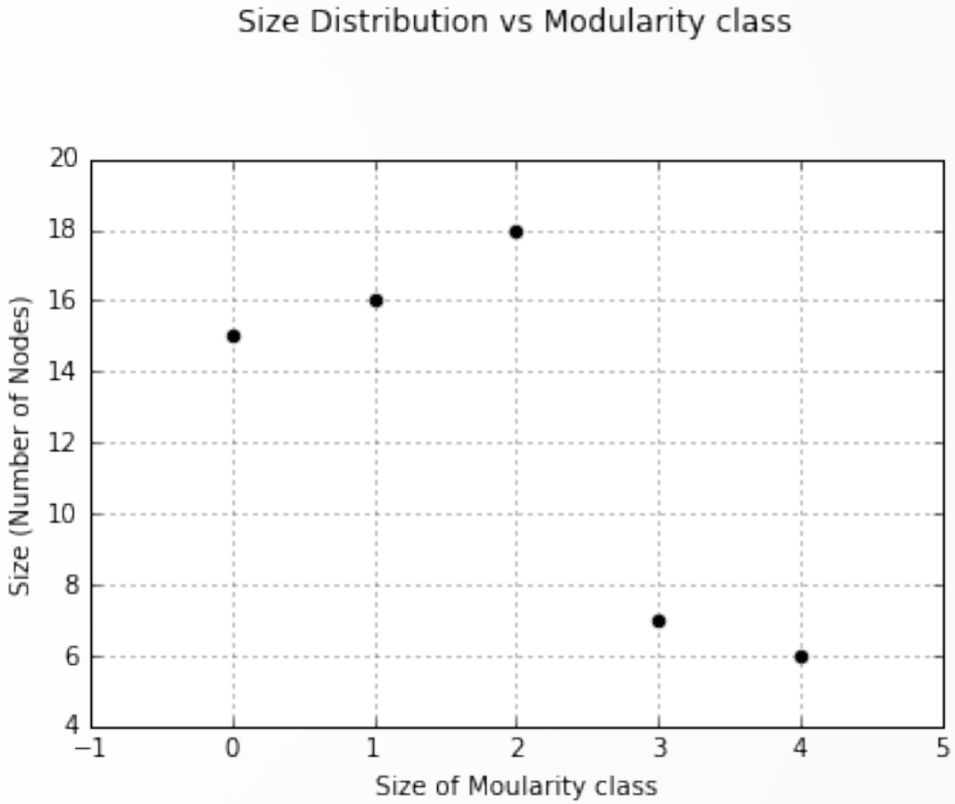
Experimental Results (contd.)



Zachary's karate club Network

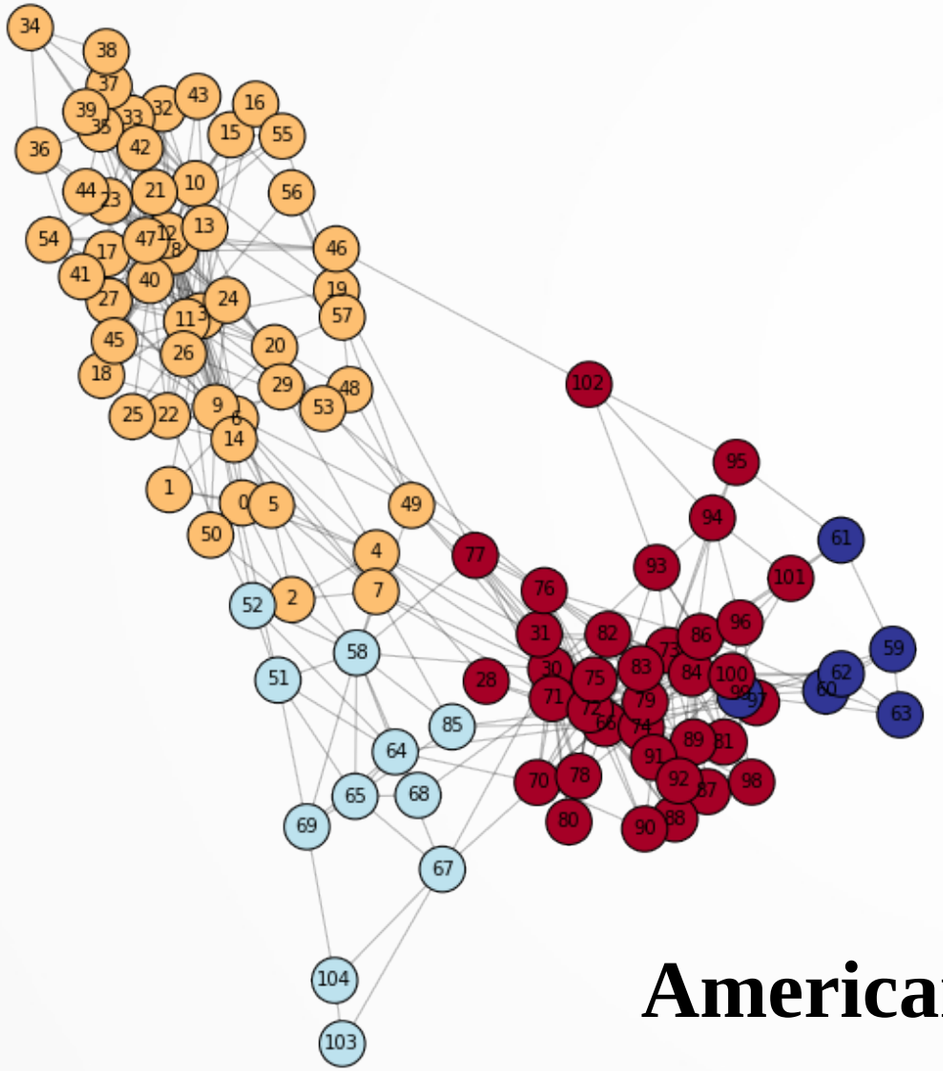
Experimental Results (contd.)





Dolphins social network

Experimental Results (contd.)



American political books

Experimental Results (contd.)

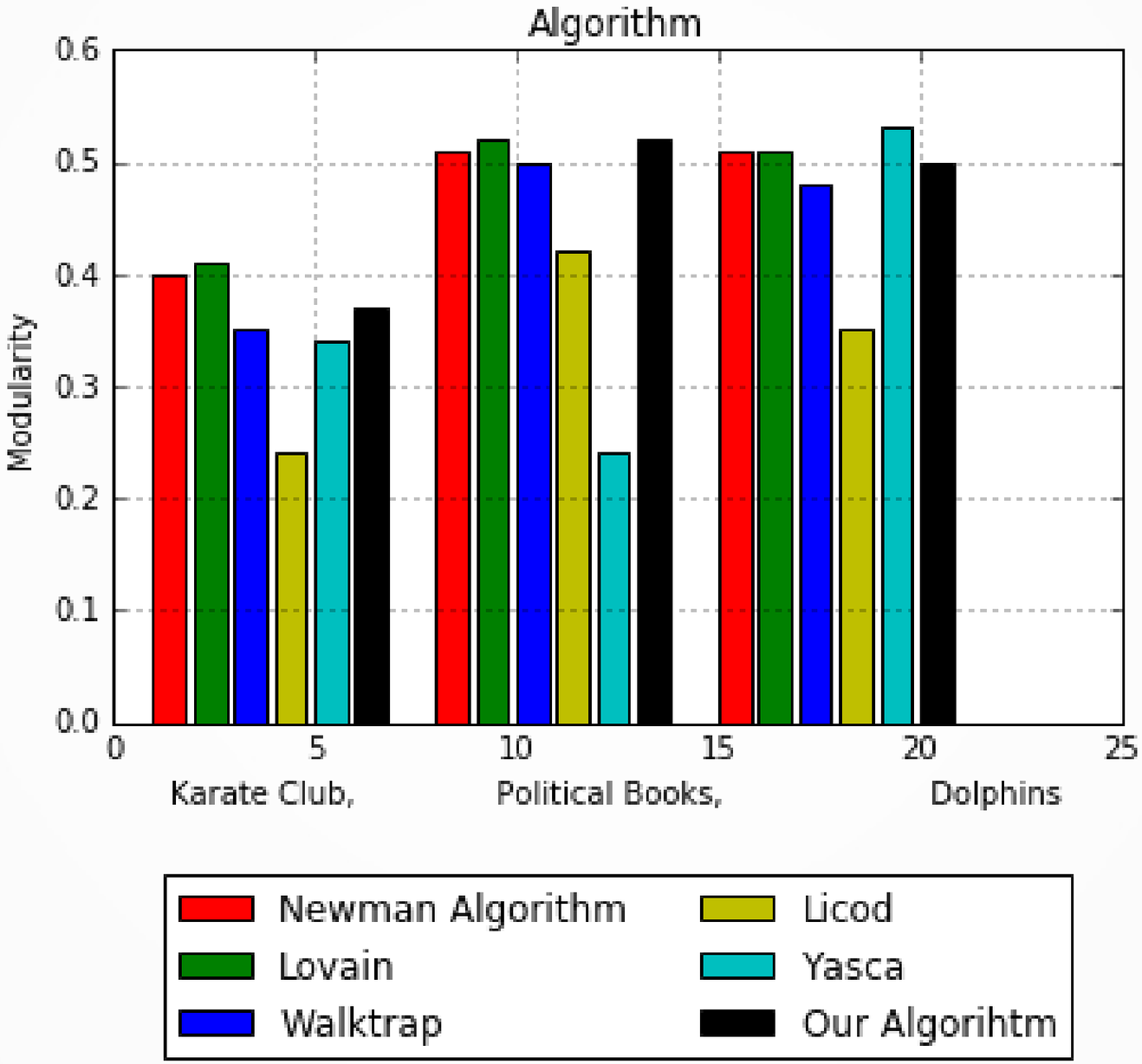
Table 2. Real World Dataset experiments result

Networks	#Nodes	#Edges	#Community	Modularity	References
Karate Club	34	78	2	0.371	[2]
Dolphins	62	159	5	0.505	[3]
Political Books	105	441	4	0.524	[4]

Name of Network	Algorithms	Community	Modularity
Karate Club	Newman[1]	5	0.40
	Lovain[5]	4	0.41
	Walktrap[6]	5	0.35
	Licod[7]	3	0.24
	Yasca[8]	2	034
	Our Algorithm	**2**	**0.37**
Dolphins	Newman	5	0.51
	Lovain	4	0.52
	Walktrap	4	0.50
	Licod	6	0.42
	Yasca	3	0.24
	Our Algorithm	**4**	**0.52**
Political Books	Newman	5	0.52
	Lovain	5	0.51
	Walktrap	4	0.51
	Licod	2	0.48
	Yasca	3	0.35
	Our Algorithm	**5**	**0.50**

Comparision between some popular existed community detection algorithm and **our algorithm**

Visual Comparision based on Modularity



Conclusion

- A new seed-centric community detection algorithm
- We compare the proposed algorithm with existing seed-centric community detection algorithm.
- Experimental results show that our proposed algorithm out performed the existing seed-centric community detection algorithm

References

[1] Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. PNAS 99(12), 7821–7826 (2002)

[2] Zachary, W.W.: An information flow model for conflict and fission in small groups.Journal of Anthropological Resea33, 452–473 (1977)

[3] Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.:The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. Behavioral Ecology and Sociobiology 54, 396–405 (2003)

[4] Krebs, V.: Political books network, http://www.orgnet.com/

[5] Blondel, V.D., Guillaume, J.I., Lefebvre, E.: Fast unfolding of communities in large networks, pp. 1–12 (2008)

[6] Pons, P., Latapy, M.: Computing communities in large networks using random walks. J. Graph Algorithms Appl. 10(2), 191–218 (2006)

[7] Kanawati, R.: Licod: Leaders identification for community detection in complex networks. In: SocialCom/PASSAT, pp. 577–582 (2011)

[8] Kanawati, R.: Yasca: A collective intelligence approach for community detection in complex networks. CoRR abs/1401.4472 (2014)

THANK YOU