Reformation of Co-Author Communities with Metagraph

Seema Gaur [1], Dr. Praveen Dhyani [2], Dr.Vishwambhar Pathak[3]
[1](Birla Institute Of Technology Mesra Ranchi.
Correspondind auther: Seema Gaur

Abstract : *Detecting Scientific Specialties For A Particular Area One Should Have Refined Information Related To That Field. For This Identification Of Author Groups Or Group Of People(Communities) That Are Related To That Particular Area Is Very Important . Detection Of These Communities Can Be Beneficial For Applications Areas, Like Finding A Common Research Area In Collaboration Networks. These Communities Are The Smallest Coherent "Groups" Of Researchers As They Are Actual Representative Of The Particular Area And With The Help Of Them We Know About The Structure Of A Given Scientific Discipline. We Represent These Communities In Form Of Community Metagraph. The Paper Focuses On The Association Of Group Effort Between Authors And Their Related Key Areas.*

Keywords - *Community, Metagraph, Collaboration Network, Community Metagraph .*

I. Introduction

A Metagraph Is A Graph That Shows Set To Set Relationship [1]. The Concept Of Metagraph Is Given By Basu And Blanning. Visualizations In Form Of Graph Cluster And Easily Discover Communities In The Graph Is Proposed By Fortunato[10]. Community Metagraph Is A Graph In Which Communities Represent As Node Of Metagraph For This Purpose Graph Is Split Into The Group Of Clusters, Where Each Cluster Represent A Community. Detection Of These Communities Can Be Beneficial For Applications Like Finding A Common Research Area In Collaboration Networks. These Communities Are The Smallest Coherent "Groups" Of Researchers. Communities Are Characterizing By Few Key Research Areas. It Helps In Analyzing And Understanding The Data.
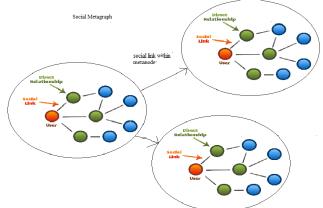

Fig.1. Community Metagraph

II. Community Detection:

Communities Create An Environment Where Members Can Share Their Thoughts, Opinions, And Experiences. Finding Communities Within The Network Is Clustering. Clusters Are Group Of Similar Node These Similar Nodes Create Communities. A Single Unit In A Community May Be Linked With Many Communities At The Same Time. Nodes Are Closely Related Within The Community And Share Some Common Attributes.

There Are Many Key Structural Measures And Notions For Community Detection. Besides This, Social Community Metagraph Have Critical Importance For The Development Of Methods For Social Network Analysis. So It Is Necessary To Do Metanode Analysis For The Metagraph. A Metanode Consist Of A Collection Of Individuals And These Individuals Are Densely Related To Each Other. Therefore, Exceptional Methods Are Needed For This Purpose.

Social Metagraph   Mainly Focuses On Set Of Actors And All The Relations That Could Be Defined On That Set. Related Work:

Survey Of Literature Revealed That There Are Many Types Of Social Networks That Can Be Studied. A Way To Categorizing Them Is Mode ; A Mode Is Defined As The Number Of Sets Of Entities On Which The Structural Variables Are Measured. Social Network   Data Can Be Represented Mathematically By Various Methods. Matrix Operations With Graph Theory Mainly Used As The Fundamentals Of Many Concepts In The Analysis Of Social Metagraph. Graph Theory Is Mainly Use For Identifying The Analysis Of Graphs. Social Metagraph Is Used To Identify The Most Important Actors/Persons In A Social Network. The Fundamental Measures For Analysis Of A Social Graph Are Based On Degree, Closeness, Betweenness. Some Common Methods Are-

III. Modularity Based Communities:

Newman And Girvan[4,5,6] Proposed A Method For Discovering Communities Based On Hierarchical Divisive Algorithms Where An Edge Is Removed Iteratively From The Network To Split It Into Communities. Modularity Is Based On The Given Function Here E Is The Total Number Of Edges , N Is The Number Of Nodes And Aij Is The Corresponding Element From The Adjacency Matrix ,Deg I And Deg J Is The Degree Of Vertex I And J δ (Ci,Cj) = 0 If I And J Are Did Not Belong To The Same Community Otherwise 1 .

$$Q = 1/2E \sum\nolimits_{i=1}^{N} \sum\nolimits_{j=1}^{N} (Aij - \frac{\deg i * \deg j}{2E})\delta(CiCj)$$

The Only Contributions To The Sum Come From Vertex Pairs Belonging To The Same Cluster: By Grouping Them Together The Sum.

A Modularity Optimization Finding Maximum Likelihood Method Is Also Explained By Newman[9].

Clique Based Communities: An Another Popular Method Known As Clique , It Is A Subgraph Of Graph. In An   Undirected Graph G= (V ,E). A Clique C, Is A Subset Of The Vertices, C ⊆ V, Such That Every Two Distinct Vertices Are Adjacent Or It Can Be Termed As Subgraph Of Graph .

N-Clique: A N Clique Is A Fully Connected N Nodes .The Two Nodes Belong To Same Community If They Are Connected Through N Adjacent Cliques. It Is A Fully Connected Graph Of N Nodes. A Community Is Formed By The Nodes That Are Connected Through Adjacent N-Cliques Defining Communities Using Clique Is Explained By Palla Et Al.,[11].

A Maximum Clique Of A Graph G Is A Clique If There Is No Possibility To Exist Cliques With More Vertices.

IV. Metagraph Coauthor Clustering Algorithm:

Coauthor Community Network Is Find Out By The Help Of Citation Dataset That Contains The Metadata Related To Author(S), Different Field(S) Of The Paper, Year Of Publication, Here Mostly Data Related To Current Year On Community Detection. Bipartite Graph Is Used To Represent Communities.

Pseudo Code Of Proposed Algorithm:

Input:

Author Dataset A = {A1, A2 ,… An} ,Set Of Related Area Or Topic K= {K1, K2 ,… Km}

Output: Set Of Author Clusters C = {Cmn1, Cmn2 ,… Cmnc},

   While !Eoauthor Dataset A

   For Each A∈ Author Group

   For Each A∈a

   For Each K∈ Topic Group

   For Each K∈k

    If Exist (A,K) Then

      G(A,K) =+1

Using Synthetic Data Set And A Data Set Created On Own References It Is An Attempt To Construct A Metagraph.

V. Experimental Results:

Experiments Were Done Initially On Synthetic Author Dataset Based On The Keyword Related To Area Of Computer Science. An Author Data Set Represented Mathematically By Its Adjacency Matrix A   In Which Aij = 1, If Node I And Node J Has A "Connection" Or Else Aij = 0 , If Node I And Node J Does Not

Has A "Connection". This "Connection" Here We Find Out The Authors That Are Related To Each Other For Writing Papers They Form Community Metagraph .

Using The Cliques Concept We Find Out The Author Metanodes And Similarly Same Cliques Method Is Used To Find Out Keywords Communities. Using These Communities An Affiliation Network Is Constructed. With The Help Of The Affiliation Network A Bipartite Graph Is Formed.

Creating Clique Base Affiliation Matrix:

Following Steps Are Require To Create Affiliation Matrix

1. An Author Adjacency Matrix Is Created To Find Similarity Between Different Authors.
2. A Keyword Adjacency Matrix Is Created To Find Similarity Between Different Related Areas.
3. Finding The Cliques On Each Of Them.
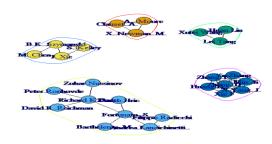4. Using These Cliques An Affiliation Matrix Is Made.

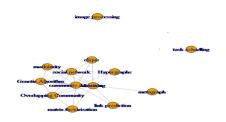**Fig. 2**. Show The Metanodes Of The Authors .   .
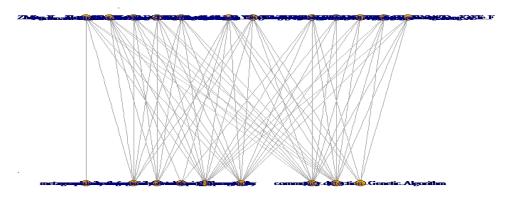

Fig. 3: Keywords Metanodes
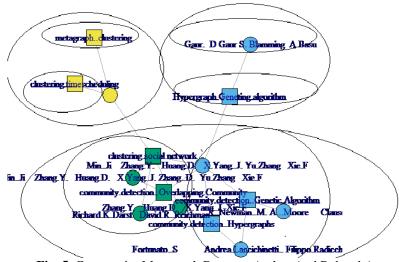

Fig. 4: Community Metagraph

Fig. 5. Community Metagraph Between Author And Related Areas

VI. Conculation

In The Era Of Big Data, Community Detection Becomes An Important Area Of Social Network Analysis. With The Help Of Coauthor Metagraph Communities Related To Particular Fields Can Easily Be Identified. The Data Analysis Shows The Currently Active Authors In Particular Field. This Can Be Used For Further Analysis Of Finding Recent Work Done In Any Area. Community Metagraph Provides A Way To Extract Meaningful Information Within A Compact Group And Find Out Internal Structure And Managerial Principal In A System. Knowing The Structure Of Network, Community Metagraph Also Provide A View To Identify Some Covered Area. This Work Can Further Be Extended In Identifying Dominating Communities.

References

[1]   Basu A. And Blanning R.W., "Metagraphs". Omega International Journal Of Management Sciences, Vol.23, 1995, Pp. 13-25.
[2]   Gaur D., Shastri A. And Biswas R., "Introducing 'Meta-Network': A New Concept In Network Technology". The Korean Institute Of Maritime Information And Commucation Sciences, Vol.6, 2008, Pp.470-474.
[3]   Heli Sun, Jiao Liu, Jianbin Huang, Guangtao Wang, Xiaolin Jia, Qinbao Song: Linklpa: A Link-Based Label Propagation Algorithm For Overlapping Community Detection In Networks. Computational Intelligence 33(2):2017,Pp.308-331.
[4]   Newman, MEJ (2001). The Structure Of Scientific Collaboration Networks. Proceedings Of The National Academy Of Sciences, 98(2):404–409. Doi: 10.1073/Pnas.98.2.404.
[5]   Newman, MEJ (2004b). Fast Algorithm For Detecting Community Structure In Networks. Phys. Rev. E, 69:066133. Doi: 10.1103/Physreve.69.066133.
[6]   Newman, MEJ And Girvan, M, Finding And Evaluating Community Structure In Networks. Phys. Rev. E, 69:026113. Doi: 10.1103/Phys- Reve.69.026113,2004.
[7]   Newman, MEJ (2004b). Fast Algorithm For Detecting Community Structure In Networks. Phys. Rev. E, 69:066133. Doi: 10.1103/Physreve.69.066133.
[8]   Newman, MEJ And Girvan, M , Finding And Evaluating Community Structure In Networks. Phys. Rev. E, 69:02611, 2004.
[9]   Newman, MEJ (2006). Finding Community Structure In Networks Using The Eigenvectors Of Matrices. Phys. Rev. E, 74:036104. Doi: 10.1103/Phys Reve.74.036104.
[10]   Santo Fortunato, Community Detec-On In Graphs, Physics Reports, Vol.486,2010,Pp.75—174.
[11]   Yilmaz Atay, Ismail Koc, Ismail Babaoglu, Halife Kodaz: Community Detection From Biological And Social Networks: A Comparative Analysis Of Metaheuristic Algorithms. Appl. Soft Comput. 50:,2017,Pp. 194-211 .