# Exploring Community detection in various areas of research

SRINIVAS.P, 2021CSM1014
M.PAVAN NAIK, 2021AIM1011

## 1 ABSTRACT

According to today's world, everything is formed into a complex network. These complex networks include various types of community structures inherently. A community is nothing but a group of people connected strongly within the people and loosely connected with the remaining networks. The task of Community detection is to detect the people who are strongly connected and form a community structure. These networks may be both static and dynamic, but community detection works on both. The detected communities can be used furthermore to examine and categorize application areas of community detection such as plagiarism check, social network analysis, and also subareas of criminology like fraud detection, and criminal identification.

## 2 INTRODUCTION

Nowadays everyone is closely related to information systems, where information systems are represented by complex networks. The main characteristic of complex networks is that they inherently contain a community structure. Community structures observed in complex networks can be different in their natures such as disjoint, overlapping, hierarchical, and local communities.

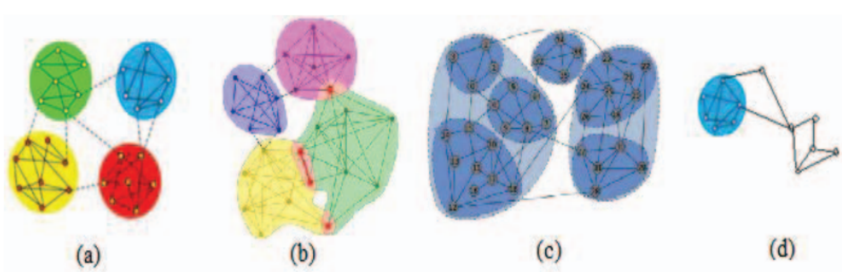



Fig. 1. An example of illustrating different types of communities:

a) The disjoint communities, include communities with no overlap.
b) Overlapping communities, show that a member of any community can have one or more members of other communities.
c) Hierarchical communities, show hierarchical grouping levels.
d) Local communities, show different structures from a local view, but no structure from a global perspective.

All the communities can come into existence by using functional similarities among the members of the network but not only using their structural similarities [3]. Therefore, detecting community structure provides us with meaningful insights into the network structure and its organizing principle.

Authors' addresses: Srinivas.P, 2021CSM1014; M.Pavan Naik, 2021AIM1011.

## 3 COMMUNITY DETECTION

It is the task of revealing the community structure of a given network at the current time interval. It provides us the opportunity to look from a mesoscale (group-level) perspective. Therefore, it has many application areas where group-level tasks are done. It is used for market segmentation, criminal detection, recommendation systems, and many more.

Since complex networks are modelled as either static or dynamic, community detection can be done for both. A static network may be considered as just a frozen network for a specified time interval. However, communities in the network may grow or shrink in size, even new communities may appear when some of them may disappear as time goes on. Dynamic community detection helps to detect and handle gracefully with this dynamicity. In summary, static community detection is interested in finding actual community structure as dynamic community detection is interested in detecting and tracking the evolution of community structure over time.

## 4 LITERATURE SURVEY

In literature, there is a study [2] about community detection practical applications. However, it does neither give a taxonomy about the community detection methods nor a detailed explanation of them. It does not cover hot application areas as well. Additionally, it does not match application areas with a network nature. To overcome this shortage, in this study, we (i) give a taxonomy of community detection methods, (ii) examine and categorize practical application areas of community detection according to their working nature (i.e., either static or dynamic) and (iii) talk about possible effects of some improvements on shortage of community detection methods on some case studies in criminology especially fraud detection, criminal identification, criminal activities detection and bot detection as well. This paper provides a hot review and quick start for researchers and developers in community detection.

## 5 ALGORITHM

```
REPEAT
   LET n BE number of edges in the graph
   FOR i=0 to n-1
      LET B[i] BE betweenness centrality of
edge i
      IF B[i] > max_B THEN
         max_B = B[i]
         max_B_edge = i
      ENDIF
   ENDFOR
   remove edge i from graph
UNTIL number of edges in graph is 0
```

## 6 RESULTS

For this project we have taken Facebook dataset for evaluation which has around 4039 nodes and 3234 edges. We have used the inbuilt Girvan Newman for community which detects the edge betweenness of the each edge in the graph and later with highest edge betweenness edge is been removed from the graph. This process is done untill all the edge in the graph are divide into separate communities.

Here for the purpose of presentation we have taken two data sets i) sample dataset with 12 nodes and 17 edges ii) a open source facebook dataset from standford. The figure 2 represents the graph of the small datase, after calculating the edge betweenness of the edges in the graph edge (2,7) as the maximum edge betweenness among all the edges in the graph, which is removed first from the graph. Later, edge (9,12) is removed followed by edge (1,4). Here, only three edges edges are removed because we have considered only 3 iterations. The final output of the graph after removing edges is shown in the step 4 of the figure 3. Here all communities are been assigned with different colors for easy identification and classification of communities.

The output for large data is show in table 1 where around 26 edges are been removed and it finally got divide into two communities or connected components. The graph of the large dataset is shown in figure 4.

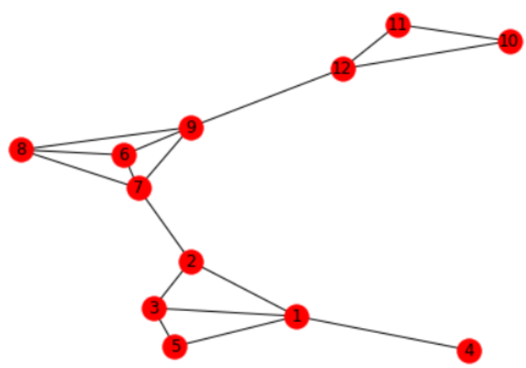



Fig. 2. Graphical representation of Small data set

## 7 APPLICATIONS OF COMMUNITY DETECTION

### 7.1 Plagiarism Check:

Consider a course that is enrolled by many students of a university, then in that case it's difficult for the instructor to check the plagiarism for all the students in that course. It takes a lot of effort and time for an instructor to check for every student.

To save instructors time and effort, they can detect the community among all the students who have enrolled in the course. After finding the communities, they need to check the plagiarism among the detected community first, as the similarity of code or assignment can be seen among the same community people rather than the students out of the community. It's highly seen that a group of students who are close to each other gets caught in plagiarism checks as they mostly do their assignments or exams together.

### 7.2 Criminology:

Community detection is used to identify the criminal user groups. Those groups can be formed from their real person accounts or fake accounts. If someone identifies such criminal user groups then, they can support or diffuse criminal ideas or criminal activities. Mostly this community detection method in criminology is used to eradicate criminal activities rather than supporting them.

To find a criminal group in a community, must focus questions are:

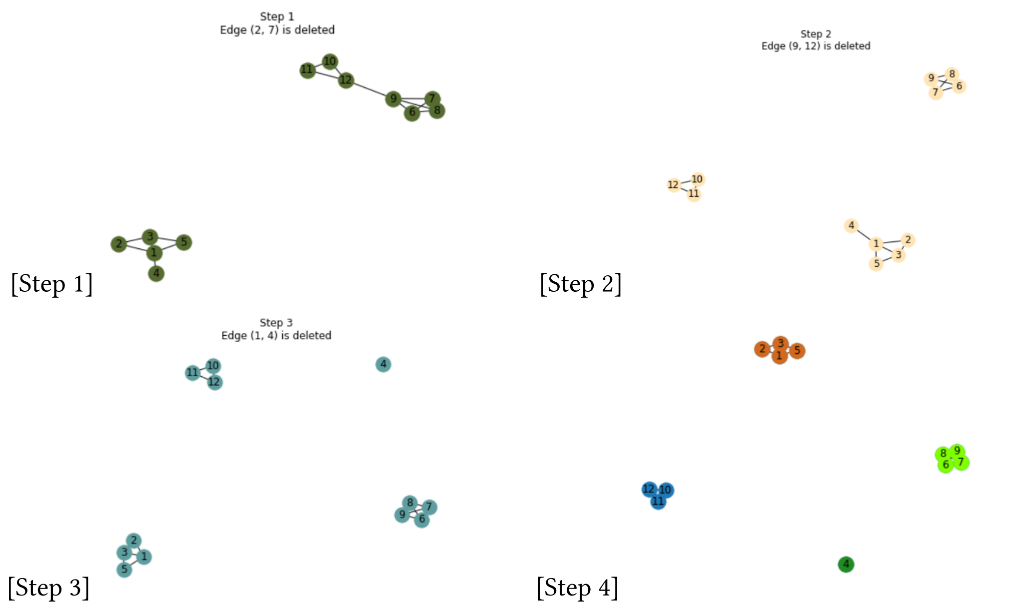



Fig. 3. Process of Community detection

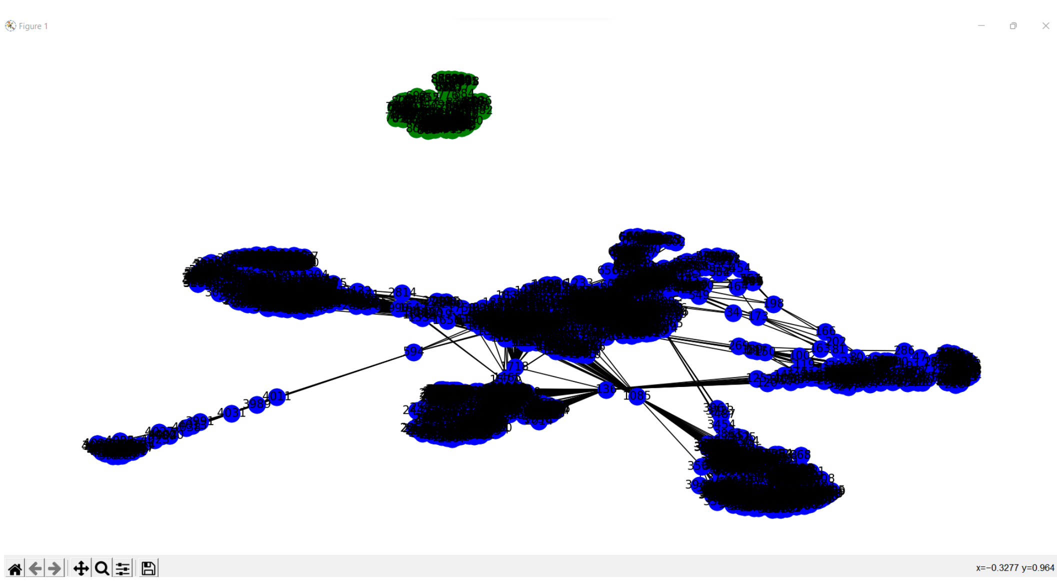



Fig. 4. Output graph of large data set

- What ranking of a criminal is based on his influence and importance in a network?
- Which subgroups and communities were found?
- Which criminal has a central position?

To answer the above questions lies in calculating graphs centrality, page rank and using community detection techniques.

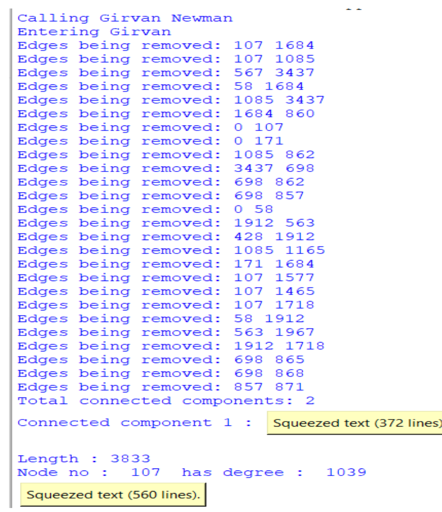
Table 1. Outputs of large data set

```
Connected component 2 : {'696', '750', '758', '885', '881', '891', '702', '835',
 '821', '836', '730', '698', '704', '751', '812', '873', '705', '825', '842', '8
79', '779', '791', '748', '762', '814', '871', '802', '892', '726', '773', '747'
, '805', '866', '894', '740', '849', '734', '801', '829', '843', '831', '797', '
720', '759', '701', '744', '850', '728', '852', '763', '864', '788', '872', '736
', '817', '732', '828', '764', '837', '717', '707', '694', '854', '787', '851',
'767', '755', '782', '798', '826', '737', '714', '760', '858', '686', '691', '78
6', '853', '832', '695', '735', '830', '778', '781', '777', '855', '794', '718',
 '727', '839', '743', '689', '712', '690', '770', '687', '752', '772', '799', '7
89', '792', '808', '824', '844', '845', '833', '863', '692', '703', '699', '796'
, '708', '880', '838', '895', '884', '848', '874', '753', '795', '769', '875', '
723', '765', '738', '815', '816', '724', '774', '790', '768', '733', '807', '803
', '887', '793', '846', '819', '813', '771', '739', '688', '883', '761', '722',
'841', '820', '783', '757', '754', '741', '785', '721', '882', '742', '870', '77
5', '716', '806', '893', '709', '818', '706', '823', '710', '784', '876', '811',
 '766', '800', '809', '886', '700', '725', '856', '859', '869', '804', '810', '7
13', '861', '834', '719', '731', '878', '822', '729', '888', '877', '693', '867'
, '697', '840', '711', '780', '827', '715', '860', '745', '890', '889', '749', '
776', '847', '746', '756'}
```

(b) list of nodes in connected component 2



(c) Dictionary of nodes with their degrees



```
Calling Girvan Newman
Entering Girvan
Edges being removed: 107 1684
Edges being removed: 107 1085
Edges being removed: 567 3437
Edges being removed: 58 1684
Edges being removed: 1085 3437
Edges being removed: 1684 860
Edges being removed: 0 107
Edges being removed: 0 171
Edges being removed: 1085 862
Edges being removed: 3437 698
Edges being removed: 698 862
Edges being removed: 698 857
Edges being removed: 0 58
Edges being removed: 1912 563
Edges being removed: 428 1912
Edges being removed: 1085 1165
Edges being removed: 171 1684
Edges being removed: 107 1577
Edges being removed: 107 1465
Edges being removed: 107 1718
Edges being removed: 58 1912
Edges being removed: 563 1967
Edges being removed: 1912 1718
Edges being removed: 698 865
Edges being removed: 698 868
Edges being removed: 857 871
Total connected components: 2

Connected component 1 :  Squeezed text (372 lines).

Length : 3833
Node no :  107  has degree :  1039
Squeezed text (560 lines).
```

(a) Removed edges and connected components 1 nodes

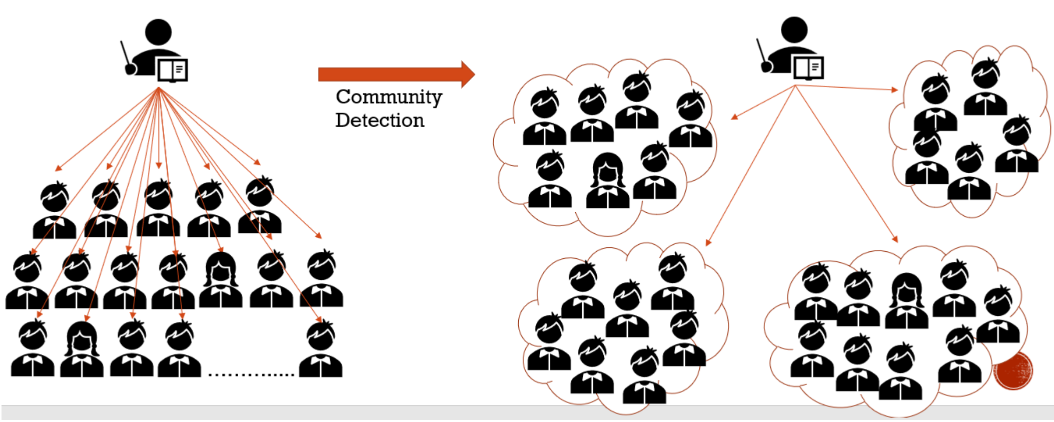




Fig. 5. Process of plagarism check

I)Centrality Measures:

- *Degree Centrality:* The first measure of distinguishing an important node is the number of its neighbours. It is believed that the node that has the most neighbours has the most activity and influence in its local neighbourhood and hence is a key member.
- *Betweenness Centrality:* The betweenness of a node is the number of shortest paths in the graph which pass through that node. A node with high betweenness has a key role in flowing information. It usually connects two densely connected parts of the graph and acts as the broker of messages between those communities. Removal of such nodes can lead to major shortcomings in message passing and communications in the network.

- *Closeness Centrality:* Distance(farness) of a node from other nodes in a graph is defined as the sum of the shortest paths between that node and all other nodes in the graph. A node with a low distance from other nodes can reach other nodes easier and faster [27]. • Eigenvector Centrality: Determines to what extent a node is connected to other well-connected nodes.

II)PageRank:

PageRank [20] can also be used as a ranking among nodes of a graph, giving us a chance to compare the relative" importance" of the nodes. The reason behind choosing PageRank is that there is a clear similarity between web pages and links between them and social networks. PageRank is designed to produce a global" importance" for web pages and we are trying to find the overall importance of the criminal actors based on their graph position. PageRank is introduced in this section so that we can make a comparison between centrality measures and PageRank.

### 7.3 Community-based Recommender Systems:

This type of system recommends items based on the preferences of the users friends. This technique follows the epigram "Tell me who your friends are, and I will tell you who you are" [1]. Evidence suggests that people tend to rely more on recommendations from their friends than on recommendations from similar but anonymous individuals [5].

This observation, combined with the growing popularity of open social networks, is generating a rising interest in community-based systems or, as or as they usually referred to, social recommender systems [4]. This type of RSs models and acquires information about the social relations of the users and the preferences of the users friends. The recommendation is based on ratings that were provided by the users friends. In fact these RSs are following the rise of social-networks and enable a simple and comprehensive acquisition of data related to the social relations of the users. The research in this area is still in its early phase and results about the systems performance are mixed.

For example, overall social-network based recommendations are no more accurate than those derived from traditional CF approaches, except in special cases, such as when user ratings of a specific item are highly varied (i.e. controversial items) or for cold-start situations, i.e., where the users did not provide enough ratings to compute similarity to other users. Others have showed that in some cases social-network data yields better recommendations than profile similarity data and that adding social network data to traditional CF improves recommendation results [4].

### 7.4 Smart Advertising and Targeted Marketing:

Community detection is used for customer segmentation directly, smart advertising and targeted marketing indirectly by companies as well. Companies can provide better service solutions if they know their customer groups intimately.

Targeted content is an opportunity to reach audiences with a custom experience that's personalized to their interests. Community detection is applied on a set of user's and detect the communities and their interests. Based on their interests the company/brand suggests the similar products and most likely the user selects/buy's that are suggested to him.

This is how Community detection is used and helps to segregate the products as per the customers interests and suggests. This technique can be widely use in all aspects of marketing.

## 8 CONCLUSION

From the above-mentioned application areas, it has been observed that there is a wide range of application domains in community detection for both static and dynamic networks. We presented only a few sample applications that indicate the usefulness of community detection. Still, few of the approaches suffer from instability problems. If the instability problem is solved for the method,

| Application Area | Sub-Area | C.D. Method |
|---|---|---|
| **Criminology** | Criminal Identification | Static |
| | Fraud Detection | Static |
| | Criminal Activities Detection | Static |
| | Bot Detection | Static |
| **Public Health** | Dynamics of Epidemic spreading | Dynamic |
| | Cancer/Tumor Detection | Static |
| | Tissue/Organ Detection | Static |
| **Politics** | Evolution of Influence | Dynamic |
| | Astroturfing | Static |
| **Smart Advertising** | Customer Segmentation | Static |
| **Targeted Marketing** | Customer Segmentation | Static |
| **Recommendation Systems** | Customer Segmentation | Static |
| **Social Network Analysis** | Community Detection | Both |
| **Network Summarization** | Member Grouping | Static |
| **Privacy** | Group Segmentation | Static |
| **Link Prediction** | Link Prediction | Both |
| **Community Evolution Prediction** | Community Evolution | Dynamic |

Fig. 6. Some more applications are listed above

it produces a more stable community structure and analysis done on this structure will be more coherent and reliable.

In the future, it can be applied to smart cities or any other emerging fields that work on group-level tasks; therefore, it can never be underestimated.

## REFERENCES


[1] David Ben-Shimon, Alexander Tsikinovsky, Lior Rokach, Amnon Meisles, Guy Shani, and Lihi Naamani. Recommender system from personal social networks. In *Advances in intelligent web mastering*, pages 47–55. Springer, 2007.

[2] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics reports*, 659:1–44, 2016.

[3] Mark EJ Newman. Detecting community structure in networks. *The European physical journal B*, 38(2):321–330, 2004.

[4] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

[5] Alan Smeaton and Jamie Callan. Joint delos-nsf workshop on personalisation and recommender systems in digital libraries. In *ACM SIGIR Forum*, volume 35, pages 7–11. ACM New York, NY, USA, 2001.