Detecting Communities in Social Network

Dr. V.S. Felix Enigo, DCSE, SSNCE

Introduction

- Importance of Detecting communities in social networks:

- Used for collaborative filtering in recommendation –members have similar tastes and preferences

- Understand the structures of given social networks – functions and properties of network

- Visualize large-scale social networks - information sharing and diffusions, growth

Definition of Community

- Community – sub-network with denser intra-community edges than inter-community edges
- Definitions of community can be classified as:
- Local definitions
- Global definitions
- Based on vertex similarity

Local Definitions

- Focused on vertices of subnetwork under investigation and its immediate neighborhood

- Self referring ones – subnetwork

  – Clique - a maximal subnetworks where each vertex is adjacent to all the others

  – n-clique - a maximal subnetwork such that the distance of each pair of vertices is not larger than n

  – k-plex – a maximal subnetwork such that each vertex is adjacent to all the others except at most k of them

Contd...

- Comparative ones - compares mutual connections of vertices of subnetwork with connections with external neighbors

- LS set - a subnetwork where each vertex has more neighbors inside than outside of the subnetwork

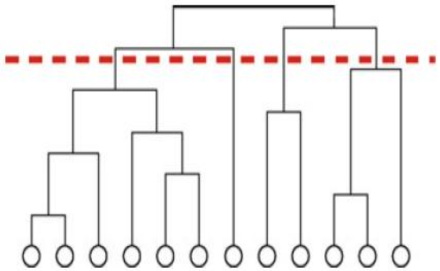
- Weak community - the total degrees of vertices inside community > number of edges lying between the community and the rest of network

Global Definitions

- It characterizes a subnetwork with respect to the network as a whole
- It starts from a null model (Newman and Girvan)
- Null model – a network that matches original network in some topological features, but no community structure
- To design a null model - introduce randomness in the distribution of edges among vertices of original network
- Link properties of subnetworks to the original network, If wide difference wrt subnetwork, then it is a community
- Null model is a way to evaluate goodness of partition of network into community (Modularity)

Definitions Based on Vertex Similarity

- Based on assumptions that communities are groups of vertices which are similar to each other
- Similarity between each pair of vertices done quantitatively
- In Hierarchical clustering - layers of communities composed of vertices similar to each other
- Ex. dendrogram, highly similar vertices found in lower part
- Subtrees got by cutting dendrogram with horizontal line correspond to communities



- Communities of different granularity got by changing position of horizontal line

Evaluating Communities

- Many ways to partition network into communities
- A quality function needed to evaluate goodness of a partition
- Modularity is the quality function proposed by Newman and Girivan:

$$Q = \sum_{s=1}^{n_m} \left[\frac{l_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right]$$

- n_m is the number of communities
- l_s is the total number of edges joining vertices of community s
- d_s is the sum of the degrees of the vertices of s
- upper term in each summand represents fraction of edges of network inside community
- lower term represents the expected fraction of edges in random network with same degree for each vertex (null model)
- i.e. comparison between real and expected edges

Contd...

- Newman and Girivan formula implies:
- Subnetwork is a community, if no.edges inside > expected number in modularity's null model (if true more tightly connected)
- Large +ve Q indicates good partitions
- Modularity of whole network, taken as single community, is zero
- Modularity is always smaller than one, but can be –ve
- Modularity optimization is a popular method for community detection
- Modularity optimization fails for communities smaller than scale (depends on size of network and resolution limit – degree of interconnectedness of community)

Methods for Community Detection

- Naive methods - graph partitioning, hierarchical clustering, and k-means clustering (no. of clusters or their size given in advance)

- Methods for detecting communities:
- Divisive algorithms
- Modularity optimization
- Spectral algorithms

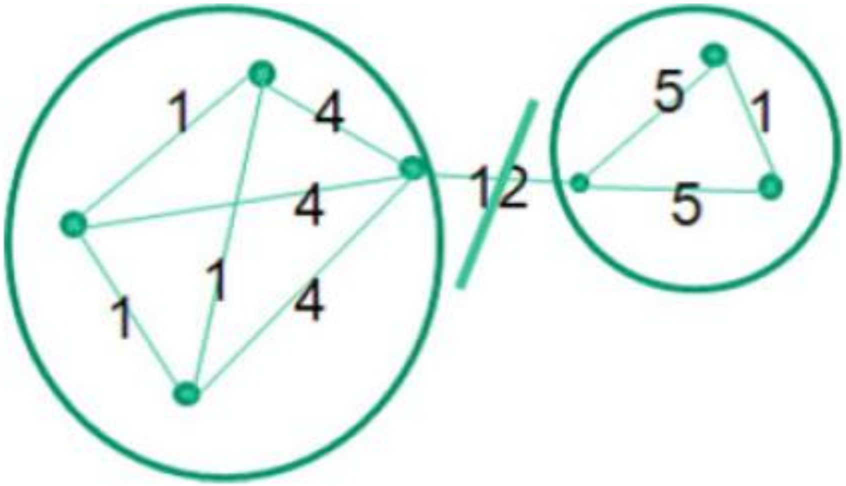
Divisive Algorithms

- To identify communities in a network - detect the edges that connect vertices of different communities and remove them
- Girvan and Newman proposed community detection algorithm based on edge betweenness centrality
- Edge betweenness centrality is defined as the number of the shortest paths that go through an edge in a network
- Steps in the algorithm
- (1) Computation of the centrality of all edges,
- (2) Removal of edge with largest centrality
- (3) Recalculation of centralities on the running network
- (4) Iteration of the cycle from step (2).

Contd...

- Intercommunity edges has larger edge betweenness value, as it serves as shortest path for many communities

- Ex. Edge 12 in the below figure



Modularity Optimization

- Numerous way partition a network can be done, so need best modularity optimization Q
- As NP hard problem, approximations algorithms that produce result in reasonable time is used
- Most popular modularity optimization is CNM algorithm
- Others greedy algorithms and simulated annealing

Spectral Algorithms

- Spectral algorithms cuts given network into pieces so that the number of edges to be cut will be minimized
- Basic algorithm is spectral graph bi-partitioning
- Laplacian matrix L of given network nxn symmetric matrix is used
- Laplacian matrix is defined as L=D-A
- where A is the adjacency matrix
- D is the diagonal degree matrix
- All eigen values of L are real and non-negative
- L has a full set of n real and orthogonal eigenvectors

- To minimize the above cut, vertices are partitioned based on the signs of the eigenvector that corresponds to the second smallest eigen value of L

- Community detection based on repetitive bi-partitioning is relatively fast.

Other Algorithms & Tools

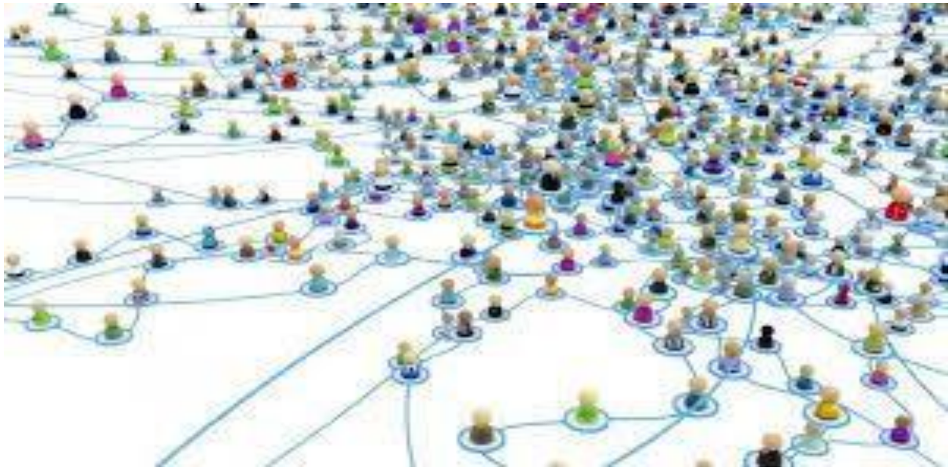
- Random walk, and the ones searching for overlapping cliques

- Tools for large scale networks:
  - CNM algorithm of community detection based on modularity optimization
  - Works for few million vertices

- Tools for Interactive Analysis:
- JUNG, Netminer, Pajek, igraph, SONIVIS, Commetrix, NetworkWorkbench, visone, Cfinder etc.

Social Network Analysis

Community Detection in SN



Course Instructor: Dr.V.S.Felix Enigo

Social Network Analysis

- Anetwork community mining problem (NCMP) is finding communities in a given network

Example:

- VLSI circuit board - processes frequently communicate with each other
- Image Segmentation - Segments of an image
- Web Page Clustering – Web pages related to common topics ………..

- Allows to discover hidden patterns
- Enables to understand the structural and/or functional characteristics of networks to efficiently utilize them

Community mining algorithms classified into two main categories:

- *Optimization based algorithms*
- *Heuristic based algorithms*

Optimization based algorithms – transforms NCMP to optimization problem and tries to find an optimal solution WRT a pre-defined objective function

Objective function can be:

- various cut criteria adopted by Spectral methods
- evaluation function introduced by the Kernighan–Lin algorithm
- network modularity

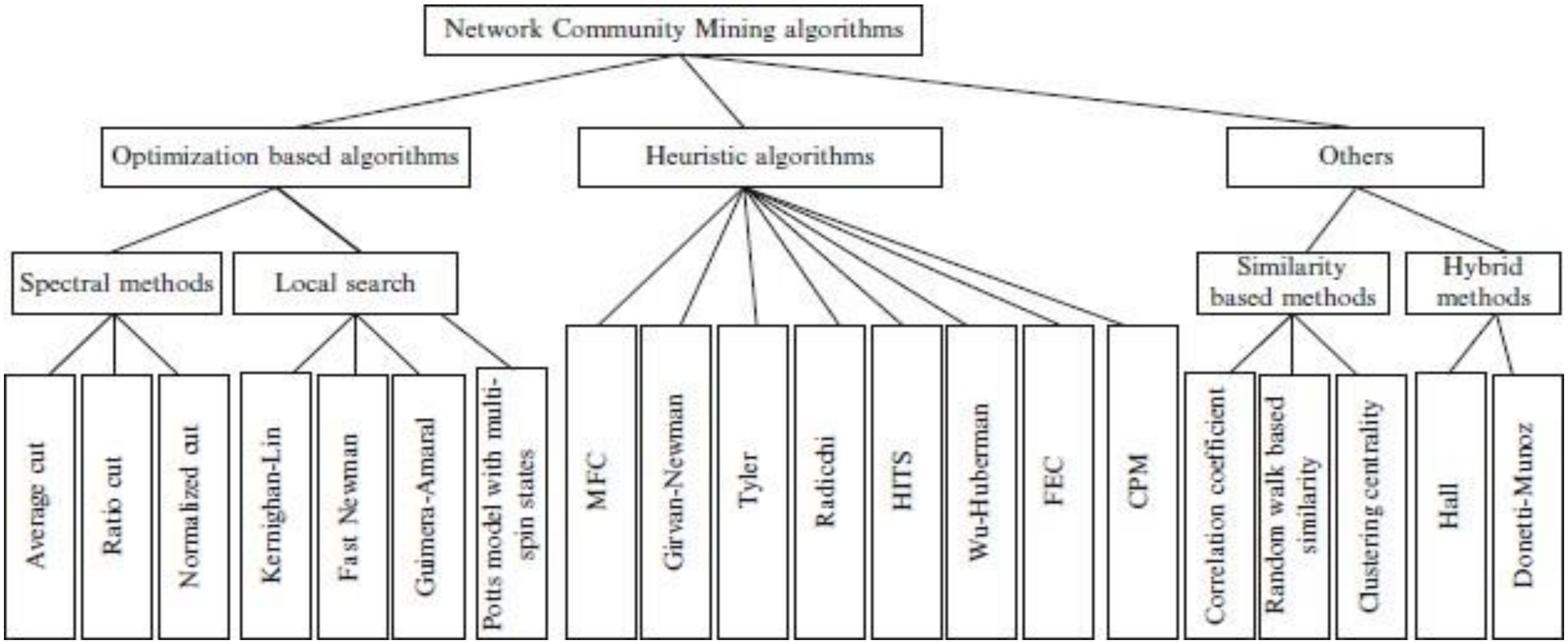
- Heuristic Algorithms solve an NCMP based on certain intuitive assumptions or heuristic rules.

Example:

- Maximum flow community (MFC) algorithm assumes that “flows” through inter-community links should be larger than those of intra-community links

- Heuristic rule in GN algorithm states that the ***edge betweenness*** of inter-community links should be larger than that of intra-community links

Social Network Analysis



Optimization Based Algorithms

Two Broad Classes: Spectral methods and local search based methods

Spectral methods optimize certain pre-defined cut criteria using the quadratic optimization technique

Cut of a bipartition of a network is number of inter-group links

An optimal bipartition of a network is the one with the minimum cut but leads to bias partition as it counts the number of inter-group link

In order to avoid this problem, other criteria such as *average cut*, *ratio cut*, *normalized cut*, and their variants are used to compute the density

Social Network Analysis

Problems of finding different optimal cuts have been proven to be NP-complete

Spectral methods finds approximately optimal cut by transforming the problem into a constraint quadratic optimization problem represented as $min(X^T MX) / (X^T X)$

Here X denotes the indicator vector of a bipartition

Minimizing the **Average cut**, M corresponds to the Laplacian matrix of a given graph

The **Ratio cut, Normalized cut, or others** a variant of Laplacian matrix

Optimal solution is found out by computing the second smallest eigenvector of M

Social Network Analysis

- Spectral graph theory, looks at the eigenvalues of the graph Laplacian, to say whether a graph is connected and also how well it's connected

- The graph Laplacian is the matrix $L = D - A$

where D is the diagonal matrix whose entries are the degrees of each node
A is the adjacency matrix

- The smallest eigenvalue of L, λ_1, is always 0, if the graph has two disconnected components
- The second smallest eigenvalue λ_2 tells you about how well a **graph is connected**.

- if λ_2 is **small**, this suggests the graph is **nearly** disconnected, that it has two components that are not very connected to each other

Social Network Analysis

- Spectral methods are bipartition methods that try to split a graph into two, with a balanced size and the minimum cut

- If a network contains multiple communities, one can find all of the communities with a hierarchical structure in a recursive way until a pre-defined stopping criterion is satisfied

Local search based optimization

- Kernighan–Lin algorithm
- Fast Newman algorithm
- Guimera–Amaral algorithm

- Adopt similar idea in finding a neighbour of the current solution in the problem space during each iteration

- Differs in optimization objectives and different strategies for regulating the local search

The Kernighan-Lin algorithm (or KL)

- Aims to minimize an evaluation function defined as the difference of the numbers of intra-community links and inter-community links
- In each iteration, KL moves or swaps nodes between communities in order to decrease the evaluation function
- Iterative process stops when the evaluation function remains unchanged

Social Network Analysis

- KL runs moderately fast with the time complexity of $O(n^2)$
- During the local search, KL only accepts better neighbor solutions and rejects all worse ones, finding local optimal than global
- KL needs prior knowledge such as the number and average size of, communities to generate initial partition.
- KL is also sensitive to initial partitions bad one causes slow convergence and poor solution

Social Network Analysis

Faster Newman algorithm (or FN)

- Proposed by Newman for detecting community structures with the time complexity of O(mn)
- FN is also a local search based optimization method.
- Starting from an initial state in which each community only contains a single node
- FN repeatedly joins communities together in pairs by choosing the best merge, until only one community is left
- In this bottom-up way, the *dendrogram* of community structure is constructed

Social Network Analysis

- To choose best merge in each iteration, a new metric *modularity* is used
- Modularity quantitatively measures how well-formed a community structure is
- The modularity of a given network in terms of a Q-function is defined as follows:

$$Q = \sum_i e_{ii} - a_i^2$$

- e ij is the weighted links that connects the nodes in Community ***i*** to nodes in Community *j and* $a_i = \sum_j e_{ij}$.
- Large Q-values is better partition

Guimera and Amaral Algorithm (or GA for short)

- Finds a partition of a network with the maximum modularity
- GA adopts simulated annealing (SA) to regulate the local search processing order to obtain a better solution
- Simulated Annealing is **a stochastic global search optimization algorithm**.
- It makes use of randomness as part of the search process
- This makes the algorithm appropriate for nonlinear objective functions where other local search algorithms do not operate well.
- Starting from an initial partition GA generates, evaluates, accepts or rejects a new neighbor partition from the current one in each iteration

Social Network Analysis

- To generate a new neighbor partition, GA moves or swaps nodes between groups, divides a group or merges two groups

- GA evaluates the new partition by calculating its modularity and decides whether or not to accept it by using the *metropolis* criterion based on the current system temperature

$$p = \begin{cases} 1, & C_{t+1} \le C_t \\ e^{-(C_{t+1}-C_t)/T}, & C_{t+1} > C_t \end{cases}$$

$C_t = -Q_t$ where p is the probability of accepting the solution at time t + 1

T is the system temperature at time t+1

Social Network Analysis

- GA offers good performance as it finds globally optimal solution
- Efficiency of GA depends by the convergent speed of SA which is slow and sensitive to initial parameters
- Parameters are: initial layout, the strategies of finding a neighbor solution, and the cooling system temperature
- GA outputs a partition of a network without a hierarchical structure, and does not require prior knowledge (no. of communities)

Potts Model

Proposed by Reichardt and Bornholdt

Network is considered as a multiple- state Potts model, in which each node is a spin with q values

Best network partition corresponds to the most stable state of the Potts model (minimum energy)

In the stable state the node that spins with the same values constitute one community

Distribution of spin values is found by minimizing a pre-defined energy function using a Monte Carlo optimization method + simulated annealing algorithm

Heuristic Methods

- Maximum flow community (MFC) algorithm
- Girvan–Newman algorithm (GN)
- Hyperlink Induced Topic Search algorithm (HITS)
- Wu–Huberman algorithm (WH)
- Clique Percolation Method (CPM)

Maximum Flow Community Algorithm

MFC algorithm was proposed by Flake et al

It is based on the Max Flow-Min Cut theorem in graph theory

Idea of MFC is maximum flow through a given network is decided only by the capacity of network ***bottle- necks***

Bottlenecks refers to the capacity of the Min-Cut sets and the sparse inter-community links

Inter-community links is discovered by calculating the Min-Cut sets

By iteratively removing ***bottle- necks*** links, involved communities will be gradually separated from each other

Girivan Newman Algorithm

- GN algorithm detects all communities by recursively breaking inter-community links
- The heuristic rule introduced in GN is that the inter-community links are those with the maximum ***edge betweenness***
- GN algorithm is a hierarchical method and can produce a dendrogram of community structure in a top-down fashion
- Its time complexity is $O(m^2n)$ which makes it not suitable for large-scale networks

Social Network Analysis

- To speed up the basic GN algorithm, several improvements have been proposed
- Tyler et al. introduced a statistical technique into the basic GN algorithm
- Monte Carlo method is used to estimate an approximate edge betweenness value for a selected link set
- Improvement in speed is gained obtained at the price of a reduction in accuracy

- Radicchi et al. defined a new metric, called *link clustering coefficient*, to replace edge betweenness as it is time-consuming
- Idea is inter-community links are unlikely to belong to a short loop, such as triangles and squares
- Heuristic rule defined is link clustering coefficient - the number of triangles or squares in which a link is involved
- In each iterative step, links with the minimum link clustering coefficient will be cut off
- The average time complexity for computing the link clustering coefficient of all links is $O(M^3 / M^2)$ $<$ for computing edge betweenness, which is $O(mn)$

Social Network Analysis

Hyperlink Induced Topic Search algorithm [**HITS]** (by Kleinberg)

- It aims to discover hyperlink- based Web communities (authority-hub communities from the Web)
- The basic assumption behind HITS is that there exist authorities and hubs on the Web
- Authorities are often pointed to by hubs that preferentially point to authorities
- Based on the mutually reinforcing relationship between authorities and hubs, an iterative method for inferring such authority-hub communities from the Web is developed
- HITS computes the principal eigenvectors of two special matrices in terms of the adjacency matrix of the Web
- Asearch engine based on HITS can return the most topic-related pages to users

WH algorithm

Proposed by Wu and Huberman

A network is modelled as an electrical circuit by allocating one unit resistor on each link

Selects two nodes from two distinct communities as the positive and negative poles

The idea is that the resistance within communities will be much less than that between communities as intra-community links are much denser than inter-community links

Social Network Analysis

- Heuristic is voltage difference of distinct communities should be more significant
- WH algorithm separates the group with a high voltage and the group with a low voltage from a network
- The node sequence sorted by their respective voltage values to find two maximum gaps
- Determines the final division by considering the co-occurrence of nodes in such separated groups
- Algorithm is very fast with a linear time in terms of the size of a network
- WH algorithm depends heavily on its prior knowledge, which is hard
- It needs to identify two “poles” belonging to different communities, needs the approximate size of each community in order to find multiple communities

CPM algorithm

- Proposed by Palla and his colleague to discover an overlapping community structure
- Network community is made of “adjacent” k cliques, which share at least k - 1 nodes with each other
- Heuristic is each clique uniquely belongs to one community, but cliques within different communities may share nodes
- CPM is able to find the overlaps of communities
- For a given K, CPM first locates all k cliques (k <= K) from a given network
- Build a clique-clique overlap matrix to find out communities in terms of different k.

Other Methods

Clustering in bottom-up approach repetitively joining pairs of current groups based on their similarities

Similarities are computed based on correlation coefficients and random walk similarities in terms of linkage relations

Proposed by Hall - Transforming an NCMP into a clustering problem in a vector space

Allocating coordinate to each node, and then cluster such spatial points using any typical spatial clustering algorithm, such as k-mean

Social Network Analysis

- Donetti and Munoz proposed a method for solving NCMPs based on quite a similar idea
- Mapping a network into a k-dimensional vector space using the k smallest eigenvectors of the Laplacian matrix before clustering spatial points

Summary

- Community detection allows to discover hidden patterns, understand structural and/or functional properties of a network

Community mining algorithms classified into two main categories:

- *Optimization based algorithms*
- *Heuristic based algorithms*

Optimization based algorithms – transforms NCMP to optimization problem and finds an optimal solution wrt pre-defined objective function

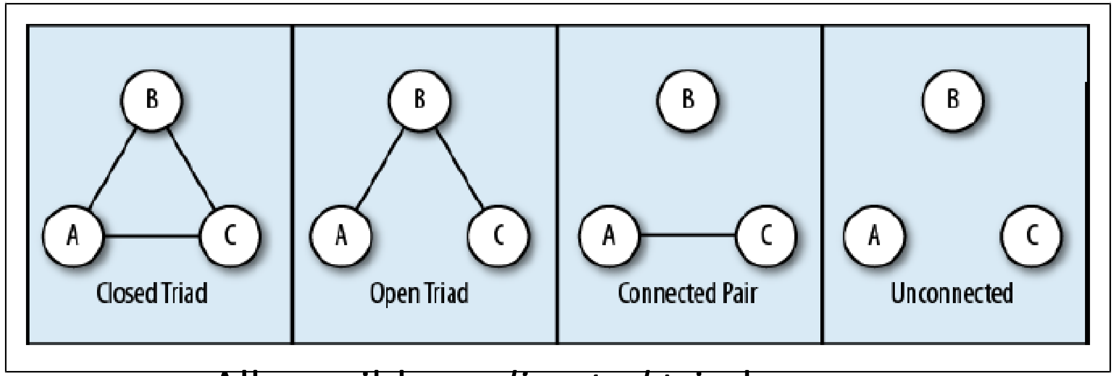
Heuristic Algorithms solve an NCMP based on certain intuitive assumptions or heuristic rules.

Triads

Dr.V.S.Felix Enigo

Introduction

- A triad is simply three nodes interlinked in some way.



All possible *undirected* triads

```
[alice, bob, carol]
[alice, bob, dave]
[carol, dave, alice]
[carol, dave, bob]
```

- Here closed triad represents a fully connected group: *A*, *B*, and *C* are connected to each other with equivalently strong ties.

Example

- A “nuclear family"—mother (Alice), father (Bob) and a child (Carol).
- These triads can overlap—for example, the same mother and father might have another child (Dan), in which case there is not one triad

Cond…

- In a *triad*, the third individual becomes at a source of balance (providing second opinions and calming nerves)

Findings:

- Asymmetric ties (e.g., “I like you more then you like me”) were the least stable
- Triadic structures were the most stable over time

Triads and Terrorists - Example

- Al Qaeda cells during training in safe houses form a triadic structure
- Everyone embedded in triads with everyone else
- All information from the outside world arrived highly filtered through the cell leader
- the groups generate their own cultural artifacts
- They define their identity as religious extremist and reinforcing their resolve to complete the attack
- Studying Hamburg Cell (responsible for 9/11 attack) shows that most common factor driving them was the social ties within their cell

Forbidden Triads or Structural Holes

- A person pays bank A 5% interest and bank B 7% interest
- Bank A and B never talk with each other but through middleman
- If they interacted, they could agree on *A* loaning money to *C* directly at a rate of 6% and realize that both could benefit by cutting out the middleman
- *Banker B* would be rather upset if that happens
- *Banker's* interests are to make sure that the two ends of their open-triad network never communicate directly
- Such open-ended triads are called Forbidden triads or structural holes or brokerage structure
- It is the number of structural holes to a person's ability to perform as an entrepreneur, a banker, a broker or a real-estate agent.

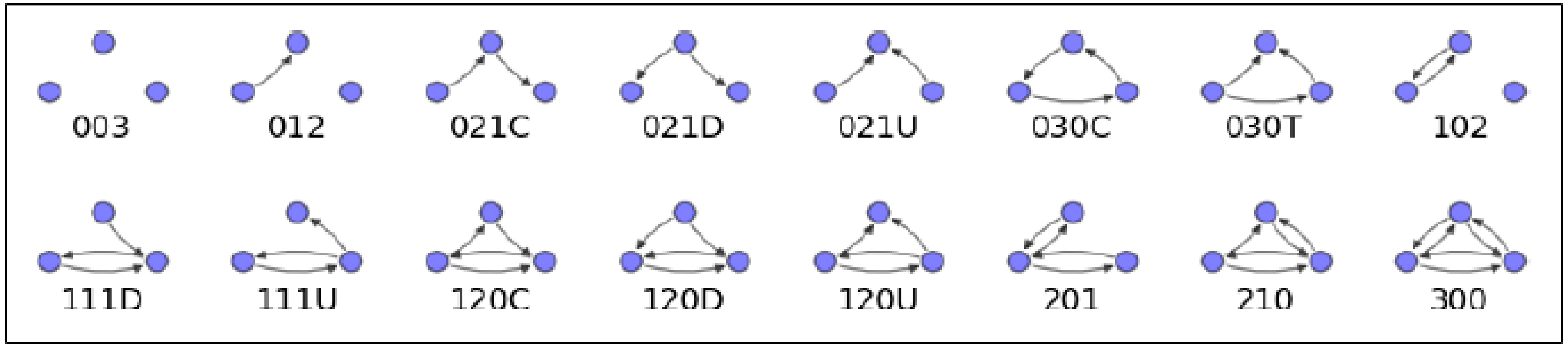
Structural Holes and Boundary Spanners

- Structural holes can span asymmetric information they can also bridge entire communities.

Example : community with scientist and community with musicians, here scientist and musician serve as a boundary spanners

Directed Triads

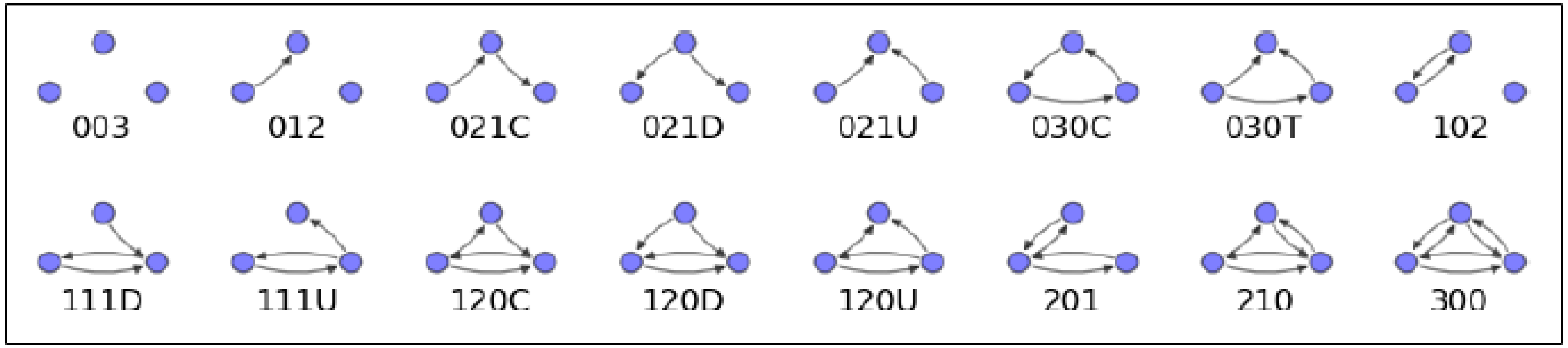
- In a directed triad, both one-way and bidirectional is considered
- So instead of 4 possible triads it has 16 variations



- The first number is the number of bidirectional edges.
- The second number is the number of single edges.
- The third number is the number of "non-existent" edges.
- A letter code to distinguish directed variations of the same triad—U for "up," D
- for "down," C for "circle," and T for "transitive" (i.e., having 2 paths that lead to
- the same endpoint).

Triadic Analysis

- The process of triadic analysis in a real network is called the *triad census*
- In this process, for every node we count occurrences of the 16 types of triads to determine the node's role in the network structure
- For example, a node with many occurrences of triads 4, 7, and 11 (i.e., rich in outgoing links and structural holes) is a *source* of information or possibly a group leader.



Example for Triads and census

- Triadic census lets one make high-level conclusions about the network structure in a macro form
- Triadic census on the 9/11 hijacker’s data shows 300 (all have bidirectional edges, with no non-existence edges) for Mohammed Atta
- Part of Hamburg cell that planned the September 11th attacks, and served as the hijacker-pilot of American Airlines Flight 11, crashing the plane into the North Tower of the World Trade Center

Clique

- Clique in a social network as a cohesive group of people that are tightly connected to each other
- A clique is defined as a *maximal complete subgraph* of a given graph
- Essentially, a clique consists of several overlapping closed triads, and inherits many of properties of closed triads.
- A clique must generate consensus or fall apart
- It is very easy to agree about conflict, and having a common enemy (or a group of common enemies) helps cliques unite

Clusters - Hierarchical Clustering

Algorithm:

1. Starting at the lowest level, every node is assigned its own cluster.
2. Using the distance table, find the closest pair of nodes and merge them into a cluster
3. Recompute the distance table, treating the newly merged cluster as a new node.
4. Repeat steps 2 and 3, until all nodes in the network have been merged into a single large cluster.
5. Choose a useful clustering threshold between the bottom and top levels

Cond…

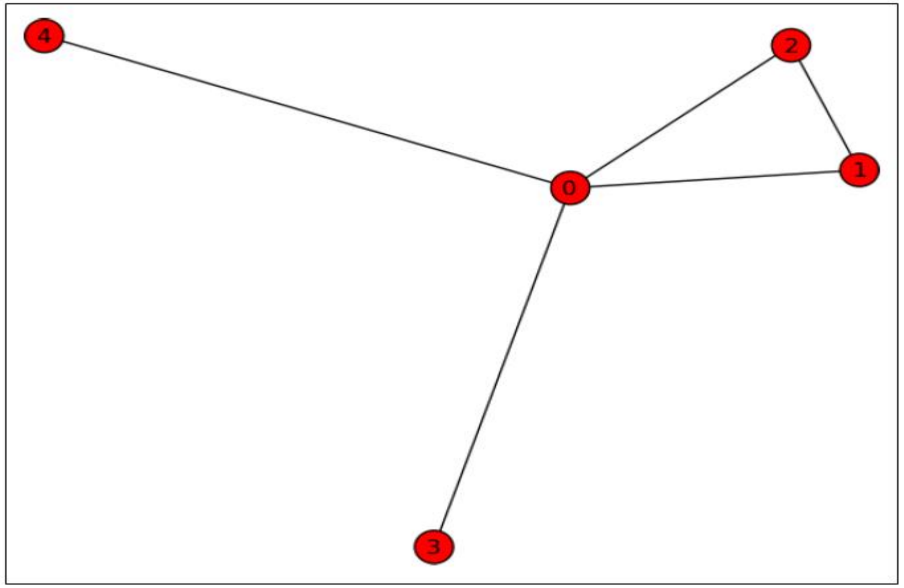
There are three common methods for merging two clusters:

- *Single-link*: merge two clusters with the smallest *minimum* pairwise distance
- *Average-link*: merge two clusters with the smallest *average* pairwise distance
- *Maximum-link* or *Complete-link*: merge the two clusters with the smallest *maximum* pairwise distance

- The *complete-link* method is considered most sensitive to outliers
- S*ingle-link* method tends to form long chains of
- *Average-link* method is considered a compromise between the two and the most frequently used.

Block Models

- A block model is a simplified network derived from the original network
- All nodes in a cluster are considered a single node, and all relationships between original nodes become aggregated into relationships between blocks
- The below block model shows relationships between the Russia-centric, Western-oriented, and Islamic-oriented clusters (0, 1, and 2)
- Clusters 3 and 4 are small republics that have significant ties with Russia, but almost no ties with anyone else



Application Areas of Community Detection: A Review

Conference Paper · December 2018

DOI: 10.1109/IBIGDELFT.2018.8625349

2 authors:

Arzum Karataş
Izmir Institute of Technology

Serap Şahin
Izmir Institute of Technology

Applications of Community Mining Algorithms

- Network Reduction
- Discovering Scientific Collaboration Groups from Social Networks
- Mining Communities from Distributed and Dynamic Networks

Application Areas of Community Detection: A Review

Arzum Karataş
Department of Computer Engineering
Izmir Institute of Technology
İzmir, Turkey
arzumkaratas@iyte.edu.tr

Serap Şahin
Department of Computer Engineering
Izmir Institute of Technology
İzmir, Turkey
serapsahin@iyte.edu.tr

***Abstract*— In the realm of today's real world, information systems are represented by complex networks. Complex networks contain a community structure inherently. Community is a set of members strongly connected within members and loosely connected with the rest of the network. Community detection is the task of revealing inherent community structure. Since the networks can be either static or dynamic, community detection can be done on both static and dynamic networks as well. In this study, we have talked about taxonomy of community detection methods with their shortages. Then we examine and categorize application areas of community detection in the realm of nature of complex networks (i.e., static or dynamic) by including sub areas of criminology such as fraud detection, criminal identification, criminal activity detection and bot detection. This paper provides a hot review and quick start for researchers and developers in community detection area.**

Index Terms*— Complex networks, community detection, application of community detection.

I. Introduction

With digital era, we are all intimate with information systems. In the realm of today's world, information systems are represented by complex networks. One of the characteristics 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. Disjoint community structure includes communities with no overlap, which is illustrated in Fig 1. (a). That is, the members of this type of communities can be assigned to only one group. Overlapping community structure represents a structure that a member of any communities can have one or more membership of other communities, which is seen in Fig 1. (b). That is to say, a person can be members of different interest groups on an online social network at the same time. Hierarchical community structure shows hierarchical grouping levels as seen in Fig. 1. (c). As for local communities, they show different structure from local view, but no structure from global perspective, which is illustrated in Fig. 1. (d).

Since domain dependency and diversity in nature of communities on the given network is unknown beforehand, community definition is an ill-defined concept [1]. Even so, a general adopted definition of community according to structure of the network is that within community members are highly connected and across community members are loosely connected [2]. Communities can come into existence according to not only their structural similarities but also functional similarities among the members of the network [3]. Therefore, detecting community structure provides us meaningful insights about the network structure and its organization principle.

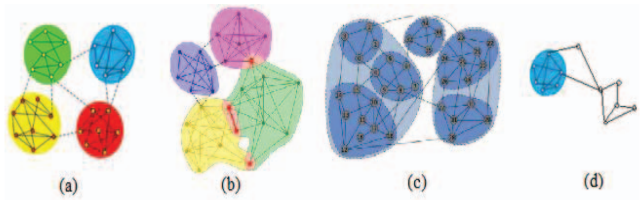


Figure 1. An example of illustrating different types of communities: (a) disjoint, (b) overlapping, (c) hierarchical and (d) local communities

Community detection is the task for revealing community structure of a given network at current time interval. It provides us a power to look from 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. For example, community detection is used in criminology by Pinheiro's work [4] by detecting anomalies on customer behaviors as possibly fraud. Another example for usage of community detection in criminology is done by Waskiewicz's work [5] to detect and reveal terrorist groups in terrorist social networks.

Since complex networks are modelled as either static or dynamic in its nature, community detection can be done for both. Static network may be considered as just frozen network for a specified time interval. However, communities in the network may growing or shrinking in size, even new communities may appear when some of them may disappear when 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 evolution of community structure over time.

In literature, there is a study [6] about community detection practical applications. However, it does neither give a taxonomy about the community detection methods and detailed explanation about them. It does not cover hot application areas as well. Additionally, it does not match application areas with 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

978-1-7281-0472-0/18/$31.00 ©2018 IEEE

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.

The rest of the paper is organized as follows. In Section II, we give preliminary information about community structure, static community detection and dynamic community detection within a taxonomy with their shortages. Next, we present application areas of community detection in Section III. In Section IV, we close the paper with our closing thoughts that includes application areas, sub areas and network applicability of community detection method (i.e., static, dynamic or both applicable) and possible effects of improvements in community detection methods for criminology area.

II. Preliminaries

In this part, we give a brief information about community structure, static and dynamic networks and static and dynamic community detection methods.

A. Community Structure

For a given network, represented by a graph G = V, E where V is the set of nodes and E the set of edges, the community detection task is interested in finding a community structure for a given graph (i.e., network). Community structure is the partition of the nodes in V of the form $C = C_1, \ldots, C_k$ such that each C_i , $1 \leq i \leq k$ exhibits the community structure that presents groups of nodes so called communities .

Let's assume that we are given a social network as seen in Figure 2. If a static community detection algorithm is run on it, we can obtain a community structure as seen in Figure 3. That is, the network seen in Fig. 2 is partitioned into five sub groups (i.e., communities as C_1, C_2 etc.) as seen in Figure 3. It is worth to note that each community detection method does not have to be provide same partitioning. That is, you may or may not see the same communities seen in Fig. 3.

Fig. 2. An illustration of a social network

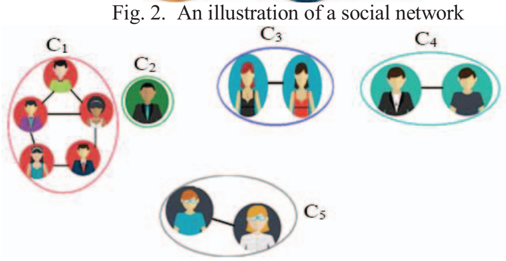


Figure 3. An illustration of a community detection algorithm application on the social network in Fig.2.

B. Static Community Detection

Graphs are good data structures to representing complex relationships. Therefore, community detection is done on complex network graphs that may be directed/undirected or weighted/unweighted or multiple-edges. In the context of this study, we examine all kind of graphs used by the works that we touch in the literature.

Static networks are regarded as snapshots of a network for a given time interval. As for static community detection, it tries to reveal underlying community structure by partitioning the network snapshot into different partitions according to friendship or relation properties or some properties of interests.

Many researchers have contributed in the field of static community detection on complex networks. Even if there are some categorizations in the literature, Fortunato [7] divides the traditional methods into four broad categories as spectral methods, methods based on statistical inference, methods based on optimization and methods based on dynamics.

(i) Spectral methods use spectral properties like eigen value spectrum of the matrix representations (e.g., adjacency matrix, Laplacian matrix, modularity matrix etc.) of networks to detect communities [1]. They generate a projection of vertices into a metric space by using eigen vectors as coordinates. For example, i^{th} entries of the eigen vector represent coordinates of vertex i in a m-dimensional Euclidean space where m is the number of eigen vectors used. Then, they apply a clustering algorithm like k-means to detect communities. However, they are not reliable when the network is very sparse because separation of eigenvalues is not sharp. They are computationally expensive as well. The works [8] [9] in can be given as examples of spectral methods.

(ii) Methods based on statistical inference generally adopts an ordinary approach to fit data to a generative network model. Most popular generative model with communities for networks is Stochastic Block Model (SBM). SBM depends on the maximization of log-likelihood of communities in the given graph. However, the methods using SBM need to know the number of communities in advance, which is unknown in real world networks in advance. The works in [10, 11] can be given as examples of statistical inference-based community detection methods.

(iii) Methods based on optimization try to find a maximum or minimum of a quality function that shows the quality of community structure. Most popular quality function is modularity proposed by Newman and Girvan [12]. Modularity is a concept depending on the maximization of difference between actual network and another form of actual network that have randomly destroyed community structure. Modularity maximization is an NP-hard problem [13]; therefore, approximation algorithms or heuristics are used. Modularity may be not detecting small size communities (i.e., resolution limit) as well. Additionally, they can suffer from instability problem (i.e., on each run they may detect different communities even on the same network). The methods in [14, 15] can be given as examples of methods based on optimization.

(iv) Methods based on dynamics uses running dynamics of the networks like diffusion, random walk and spin dynamics to detect community structure in a network. The methods using label propagation technique is based on

diffusion process. Most popular dynamic are random walk dynamics. The methods using random walks using homophily assumption. That is, vertices in similar nature is highly probable in same partition. Methods using spin dynamics first define a spin model on the network where spin variables are assigned to vertices of the network. Then, they assign an energy of the spin via Hamiltonian distance. They aim to find spin configurations that give minimum Hamiltonian distance. The works in [16, 17] can be given as examples for methods based on dynamics.

C. Dynamic Community Detection

We give a brief description for each community evolution approaches we have identified in the literature, framing them in the taxonomy in [18]. The taxonomy is a three-level classification of existing methods for tracking community evolution in dynamic social networks with respect to their network models (first level of tree structure), their functioning principles (second level) and algorithmic techniques (third level). According to their functioning principles, there are four fundamental approaches:

(i) Independent Community Detection and Matching contains methods that first detect community structure on each snapshot separately and match those communities across consecutive or non-consecutive time-stamps. All methods in this category target to track community evolution by identifying key community events (e.g., birth, death, growth etc.) through life cycles of communities. Core-based methods identify one or several specific nodes for each community called as core nodes. For example, the nodes have highest centrality value on the network can be core nodes. Then, the methods determine community events according to the core nodes. As for event-based methods, they consider all nodes to determine the community events. The works in [19-21] can be given as examples of dynamic community detection methods that use this approach.

(ii) Dependent Community Detection covers methods that detect community structure on snapshot at time t and past community information (i.e., on previous snapshot or some recent snapshots). In this approach, broadly there are two types of methods: evolutionary methods and cost-function methods. Evolutionary methods built on or modify The methods using this approach are based on coupling graphs or either optimizing metrics. In coupling graph-based methods, the basic idea is to build a coupling graph and detect communities on this graph. In metric optimization-based methods, the basic idea is designing a metric that can be directly optimized on all snapshots given. The works in [25, 26] can be given as examples of dynamic community detection that use this approach.

(iv) Dynamic Community Detection on Temporal Networks includes methods that detect community on only first snapshot network, then alter this community structure for each incoming update. The works in [27, 28] can be given as examples of dynamic community detection that use this approach.

basic community detection algorithms like Louvain [22]. They initialize community structure with this algorithms and re-run modified basic algorithm again. As for cost-function methods, they use a cost-function (i.e., any function to minimize the community changes between successive snapshots such as Modularity maximization between two consequent snapshots). The works in [23, 24] can be given as examples of dynamic community detection that use this approach.

(iii) Simultaneous Community Detection on All Snapshots encloses methods that first construct a single from all snapshots so-called coupling graph and detect community structure on the coupling (joint) graph.

Table 1. Community Detection methods and their shortages

Community Detection Methods	**Network Type**	**Shortage/Drawbacks**
Spectral methods	Static	-Not computationally efficient -Not reliable for sparse networks
Methods based on statistical inference	Static	-Selection of models -High time complexity -Specifying community numbers in advance.
Methods based on optimization	Static	-Resolution Problems -Instability Problems
Methods based on dynamics	Static	-They may need other clustering algorithms to work -They may suffer from poor stability
Independent community detection & matching	Dynamic	-High time complexity -Instability of traditional community detection methods
Dependent community detection	Dynamic	-No chance to parallelize community detection on snapshots -Traditional community detection methods are not directly applicable
Simultaneous community detection on all snapshots	Dynamic	-There is no chance to update the network structure. -Difficulty in handling some operations such as merge and split
Dynamic community detection on temporal networks	Dynamic	-Community drift can be occurred -Traditional community detection methods are not directly applicable

In Table 1, community detection methods and their shortages /drawbacks are seen.

III. Practical applications of community detection

Community detection is worked by many researchers from different disciplines so far. This paper represents such practical applications by categorizing them according to their domains rather than just chronological order.

A. Criminology

Community detection is used for identification of criminal user groups. Those groups can be built from either real person accounts or bot accounts. They can support or diffuse criminal ideas or terrorism-like activities. The authors in work [29] use community detection to detect communities in criminal networks, then they do some manual analysis. Pinheiro [4] make a study to identify fraud events on telecommunication networks by using community detection that helps to determine customer behaviors and examining outliers as possibly fraud. Similarly, Waskiewicz [5] provide a study to detect terrorist group activities on some online social networks by using community detection to detect such groups.

Additionally, usually bots(i.e., software robots) popularly used by attackers for impersonation for identity frauds, follower frauds and botnet attacks. The works for bot detection via community detection is introduced in detail in [30].

B. Public Health

In health domain, community detection is generally used for discovering dynamics of certain groups susceptible to an epidemic disease. Salathe and Jones[31] show the impact of community structure on disease dynamics.

Community detection is used detect a disease such as cancer and tumor types as well. Bechtel et al.[32] proposed a community-based lung cancer detection approach. Likewise, Haq and Wang [33] make a study on genomic datasets for detecting subgroups of twelve types of cancers and they look survival rates of those communities and distribution of tumor types over communities.

Additionally, community detection is used to organ detection. The works in [34, 35] are examples for organ or tissue detection.

C. Politics

In Politics, community detection is used for observation of influences of political ideologies or individual politicians on some social group. Later, it can be specialized to track evolution of this influence over time. This influence commonly is created by influence or astroturfer bots – bots make an attempt to create a fake impression on real grassroots to support a policy, individual, product campaign [36]. Community detection is used for detecting those types of bots as well [30].

D. Customer Segmentation, 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 solution if they know their customer groups intimately(i.e., customer segmentation). Then, they can do advertising and marketing for detected specific groups [37].

E. Recommendation Systems

Recommendation systems are another important services we use everyday whenever we wanto to buy a book from a website, watch a video or listen a music on a social media site etc. They try to recommend something that probably you would like to glad to meet. The task of community detection is akind of segregation of people which is like-mind. Although there are many works in the literature that use community detection in recommendation systems, we give only some example studies in [38, 39].

F. Social Network Analysis

Community detection can be a good means to understand communities on networking level and correlate them in real-life relations. For example, community detection on social networks and online social networks such as Facebook, Twitter, LinkedIn etc. are good examples for this area.

Social Network Analysis is one of the mostly used community detection; therefore, there are many works in the literature. Some of them are given in [40, 41].

G. Network Summarization and Privacy

Community detection provide a mesoscale view of networks. That is, it is provide a group-level point of view and summarize the network in group level. Therefore, it can help serving privacy in the network when sharing generalized features of them with another party.

On the other hand, community detection can be used break privacy of people on weak signal networks like Bitcoin. Authors of the work [42] used community detection to show how it can be efficiently used to re-identify multiple addresses that belong to same user. Fortunately, Kokkiligadda and Vatsavayi [43] and the authors of the work [44] show that anonymized social networks satisfactorily preserve the community structure of their original networks. That is, except weak signal networks, community detection does not lead to privacy leakage.

H. Link Prediction

Link prediction assess the possibility of future links between members of a network and it is used for determining fake links, missing links and future links. The underlying structure of the network is found via a community detection algorithm, then the possibility of being a link between two members is calculated. Valverde-Rebaza and Lopes [45] propose an approach based on community detection for link prediction. Similarly, Soundarajan and Hopcroft [46] show that community information obtained by community detection methods increase the accuracy of similarity-based link prediction.

İ. Community Evolution Prediction

Community evolution prediction regards the prediction of future form of a community given its past and present form in terms of community events such as growing, shrinking, merging, forming, solving etc. It is one of the hot topics in community analysis field. Naturally, there are some works in literature. Those address the predictability of community evolution issue as supervised learning task. Some of them like Brodka, Kazienko and Kołoszczyk[47] employ ordinary classifiers (e.g., Naïve Bayes, Bayes Net, Logistic Regression, SVM etc.) whereas others like the authors of [48] use both sequential (i.e., Conditional Random Fields with Linear

Chain and with Skip Chain) and ordinary classifiers for prediction. To achieve community evolution prediction, a dynamic community detection is needed to see evolution of the communities.

IV. Conclusion

From mentioned application areas just above, community detection has wide range of application domains. Some of them use static community detection, some other dynamic community detection or both. In Table 2, areas and usable community detection method categories are shown.

The number of applications presented here is neither complete nor exhaustive. However, we presented only a bunch of sample applications that indicate the usefulness of community detection. In future, it can be applied to smart cities or any other emerging fields that working on a group-level tasks; therefore, it should not be underestimated.

Let's focus the works in criminology area in Table 2. If there is an improvement on shortages mentioned in Table 1, there will be some effects on these works as well.

Table 2. Application areas, sub-areas and nature of correspondent community detection 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

For example, the work in [29] use a method based on optimization and it suffers from instability problem. If instability problem is solved for the method, it produces more stable community structure and analysis done on this structure will be more coherent and reliable. Similarly, the community structures detected by other works [4, 5] will be more dependable, and produce better results.

In this study, we introduce a taxonomy of community detection methods and share their shortages, which is lead to possible research areas. We discussed possible effects of possible improvements of instability of community detection algorithms on some case studies on criminology area. Then, we create a summary table that contains application areas and corresponding community detection method type. By using this table, researchers and developers can do a quick-start in the field.

Acknowledgment

This work is partially supported by YOK through its 100/2000 and TÜBİTAK through 2011-C PhD Scholarship Programmes to whom we are most grateful.

References

[1] S. Fortunato, "Community detection in graphs," Physics Rep., vol. 486, no. 3-5, pp. 75-174, 2010.
[2] M. Girvan, and M. E. Newman, "Community structure in social and biological networks," Proc. Nat. Academy of Sci., vol. 99, no. 12, pp. 7821-7826, 2002.
[3] M. E. Newman, "Detecting community structure in networks," The European Physical J. B., vol. 38, no. 2, pp. 321-330, 2004.
[4] C. A. R. Pinheiro, "Community detection to identify fraud events in telecommunications networks," *SAS SUGI proceedings: customer intelligence*, 2012.
[5] T. Waskiewicz, "Friend of a friend influence in terrorist social networks," in Proc Int. Conf. on Artificial Intelligence , 2012, p. 1.
[6] M. Ahuja, and J. Singh, Neha, "Practical applications of community detection," Int. J. Advanced Research in Comput. Sci. and Software Eng.vol. 6, no. 4, pp. 412-415,2016.
[7] S. Fortunato, and D. Hric, "Community detection in networks: A user guide," Physics Rep., vol. 659, pp. 1-44, 2016.
[8] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," Physical Review E, vol. 74, no. 3, pp. 036104, 2006.
[9] M. E. Newman, "Spectral methods for community detection and graph partitioning," Physical Review E, vol. 88, no. 4, pp. 042822, 2013.
[10] E. Côme, and P. Latouche, "Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood," Statistical Modelling, vol. 15, no. 6, pp. 564-589, 2015.
[11] M. E. Newman, and G. Reinert, "Estimating the number of communities in a network," Physical Review Lett., vol. 117, no. 7, pp. 078301, 2016.
[12] M. E. Newman, and M. Girvan, "Finding and evaluating community structure in networks," Physical Review E, vol. 69, no. 2, pp. 026113, 2004.
[13] M. Chen, K. Kuzmin, and B. K. Szymanski, "Community detection via maximization of modularity and its variants," IEEE Trans. Comput. Social Syst., vol. 1, no. 1, pp. 46-65, 2014.
[14] A. Clauset, "Finding local community structure in networks," Physical Review E, vol. 72, no. 2, pp. 026132, 2005.
[15] A. Lancichinetti, and S. Fortunato, "Limits of modularity maximization in community detection," Physical Review E, vol. 84, no. 6, pp. 066122, 2011.
[16] J. Reichardt, and S. Bornholdt, "Statistical mechanics of community detection," Physical Review E, vol. 74, no. 1, pp. 016110, 2006.
[17] M. Rosvall, and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," Proc. Nat. Academy of Sci., vol. 105, no. 4, pp. 1118-1123, 2008.
[18] N. Dakiche, F. B.-S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," Information Processing & Management, 2018 [Online] Available: https://www.sciencedirect.com/science/article/pii/S0306457317305551
[19] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in Proc. Int. Conf. on Advances in social networks analysis and mining, 2010, pp. 176-183.
[20] J. Hopcroft, O. Khan, B. Kulis, and B. Selman, "Tracking evolving communities in large linked networks," Proc. Nat. Academy of Sci., vol. 101, no. suppl 1, pp. 5249-5253, 2004.
[21] E. G. Tajeuna, M. Bouguessa, and S. Wang, "Tracking Communities over Time in Dynamic Social Network," Machine Learning and Data Mining in Pattern Recognition, , vol 9729, pp. 341-345, Perner P, Eds: Springer, 2016.
[22] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. statistical mechanics: theory and experiment, vol. 2008, no. 10, pp. P10008, 2008.

[23] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "Graphscope: parameter-free mining of large time-evolving graphs," in Proc. 13th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining, 2007, pp. 687-696.

[24] J. Zhu, J. Liu, X. Zhang, and Y. Zhao, "A reconstructed event-based framework for analyzing community evolution," in 2016 IEEE Int. Conf. on Big Data Analysis , 2016, pp. 1-4.

[25] B. Mitra, L. Tabourier, and C. Roth, "Intrinsically dynamic network communities," Comput. Networks, vol. 56, no. 3, pp. 1041-1053, 2012.

[26] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," Sci., vol. 328, no. 5980, pp. 876-878, 2010.

[27] G. Rossetti, L. Pappalardo, D. Pedreschi, and F. Giannotti, "Tiles: an online algorithm for community discovery in dynamic social networks," Machine Learning, vol. 106, no. 8, pp. 1213-1241, 2017.

[28] A. Zakrzewska, and D. A. Bader, "Tracking local communities in streaming graphs with a dynamic algorithm," Social Network Analysis and Mining, vol. 6, no. 1, pp. 65, 2016.

[29] H. Sarvari, E. Abozinadah, A. Mbaziira, and D. Mccoy, "Constructing and analyzing criminal networks," in 2014 IEEE Security and Privacy Workshops , 2014, pp. 84-91.

[30] A. Karataş, and S. Şahin, "A Review on Social Bot Detection Techniques and Research Directions," in Proc. Int. Security and Cryptology Conference Turkey, 2017, pp.156-161.

[31] M. Salathé, and J. H. Jones, "Dynamics and control of diseases in networks with community structure," PLoS Computational Biology, vol. 6, no. 4, pp. e1000736, 2010.

[32] J. J. Bechtel, W. A. Kelley, T. A. Coons, M. G. Klein, D. D. Slagel, and T. L. Petty, "Lung cancer detection in patients with airflow obstruction identified in a primary care outpatient practice," Chest, vol. 127, no. 4, pp. 1140-1145, 2005.

[33] N. Haq, and Z. J. Wang, "Community detection from genomic datasets across human cancers" 2016 IEEE Global Conf. on Signal and Infor. Process., 2016,pp. 1147-1150.

[34] F. Taya, J. de Souza, N. V. Thakor, and A. Bezerianos, "Comparison method for community detection on brain networks from neuroimaging data," Appl. Network Sci., vol. 1, no. 1, pp. 8, 2016.

[35] Y. Yang, P. G. Sun, X. Hu, and Z. J. Li, "Closed walks for community detection," Physica A: Statistical Mechanics and its Applications, vol. 397, pp. 129-143, 2014.

[36] A. Bienkov, "Astroturfing: what is it and why does it matter?," [Online]. Available: https://www.theguardian.com/commentisfree/2012/feb/08/what-is-astroturfing

[37] M. J. Mosadegh, and M. Behboudi, "Using social network paradigm for developing a conceptual framework in CRM," Australian J. Bus. and Manage. Research, vol. 1, no. 4, pp. 63, 2011.

[38] S. B. Abdrabbah, R. Ayachi, and N. B. Amor, "Collaborative filtering based on dynamic community detection," Dynamic Networks and Knowledge Discovery, vol. 85, 2014.

[39] D. Lalwani, D. V. Somayajulu, and P. R. Krishna, "A community driven social recommendation system," in Proc. 2015 IEEE Int. Conf. on Big Data, 2015, pp. 821-826.

[40] Y. Atay, I. Koc, I. Babaoglu, and H. Kodaz, "Community detection from biological and social networks: A comparative analysis of metaheuristic algorithms," Appl. Soft Computing, vol. 50, pp. 194-211, 2017.

[41] C. Wang, W. Tang, B. Sun, J. Fang, and Y. Wang, "Review on community detection algorithms in social networks," in Proc. IEEE Int. Conf. on Progress in Informatics and Computing,2015, pp. 551-555.

[42] C. Remy, B. Rym, and L. Matthieu, "Tracking bitcoin users activity using community detection on a network of weak signals," in Int. Workshop on Complex Networks and their Applications, 2017,pp. 166-177.

[43] N. Kokkiligadda, and V. K. Vatsavayi, "Community privacy preservation in dynamic social networks.", Int. J. Tech. Research and Applicat., vol. 4, no.6 , pp. 133-136, 2016.

[44] Campan, Y. Alufaisan, and T. M. Truta, "Community Detection in Anonymized Social Networks," in Proc. Workshops of the EDBT/ICDT 2014 Joint Conference, 2014, pp. 396-405.

[45] J. C. Valverde-Rebaza, and A. de Andrade Lopes, "Link prediction in complex networks based on cluster information," Proc. 21st Brazilian Conf. in Artificial Intelligence,2012, pp. 92-101.

[46] S. Soundarajan, and J. Hopcroft, "Using community information to improve the precision of link prediction methods," in Proc. 21st Int. Conf. on World Wide Web,2012, pp. 607-608.

[47] P. Bródka, P. Kazienko, and B. Kołoszczyk, "Predicting group evolution in the social network," in Int. Conf. on Social Informatics, 2012, pp. 54-67.

[48] G. Diakidis, D. Karna, D. Fasarakis-Hilliard, D. Vogiatzis, and G. Paliouras, "Predicting the evolution of communities in social networks," in Proc. 5th Int. Conf. on Web Intelligence, Mining and Semantics, 2015, p. 1.

Certainly, here are the top 5 applications of community detection explained in detail:

1. Criminology:

- **Identification of Criminal User Groups:** Community detection is used to identify and monitor groups of individuals engaged in criminal activities. These groups may include real person accounts or bot accounts that support or diffuse criminal ideas, terrorism-like activities, or fraud.
- **Terrorist Activity Detection:** In the realm of counterterrorism, community detection is applied to detect and track terrorist groups on online social networks. It helps identify the structure and interactions within these groups, aiding in threat assessment and prevention.
- **Fraud Detection:** Community detection assists in identifying fraudulent activities, such as anomalies in customer behaviors on telecommunication networks. It enables the detection of outlier behaviors that may indicate possible fraud.
- **Bot Detection:** Attackers often use software robots (bots) for impersonation, identity fraud, follower fraud, and botnet attacks. Community detection methods are employed to identify and differentiate these bots from genuine users.

2. Public Health:

- **Disease Dynamics:** Community detection helps in understanding the dynamics of groups susceptible to epidemic diseases. It aids in tracking the spread of diseases, identifying high-risk groups, and implementing targeted interventions.
- **Cancer Detection:** Community detection is used to identify subgroups related to different cancer and tumor types, enhancing cancer diagnosis and treatment planning.
- **Organ Detection:** The application of community detection can assist in the detection and categorization of organs or tissues within medical imaging and healthcare systems.

3. Link Prediction:

- **Future Link Assessment:** Link prediction assesses the likelihood of future connections between network members. It is used to determine potential fake links, missing links, and future links. Community detection plays a role in revealing the underlying network structure, which is crucial for link prediction algorithms.
- **Improving Link Prediction Accuracy:** Community information obtained through community detection methods enhances the accuracy of similarity-based link prediction models.

4. Social Network Analysis:

- **Understanding Human Interactions:** Community detection is a fundamental tool for analyzing social networks and online communities, offering insights into how individuals interact, form relationships, and share information.
- **Insight into Online Platforms:** Analyzing communities on social media platforms like Facebook, Twitter, and LinkedIn can reveal behavioral patterns, group dynamics, and user engagement. This knowledge is valuable for platform optimization and content delivery.

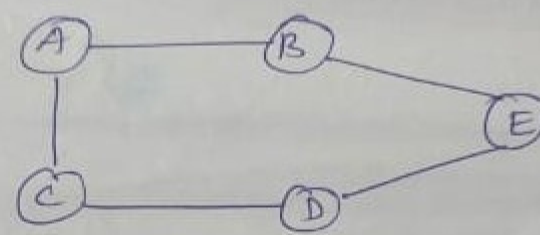
5. Customer Segmentation, Smart Advertising, and Targeted Marketing:

- **Customer Segmentation:** Community detection is used to segment customers based on their preferences, behaviors, and demographics. This segmentation helps businesses tailor their products, services, and marketing campaigns to specific customer groups.
- **Personalized Marketing:** With community detection, companies can offer personalized advertising and marketing strategies, improving customer engagement and increasing sales.
- **Enhanced Customer Experience:** By understanding customer segments, businesses can provide better service solutions and meet the unique needs of each group, ultimately enhancing the customer experience.

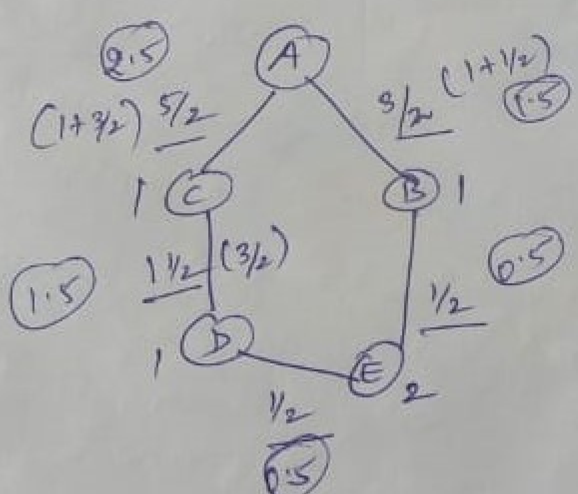
GIRVAN- NEWMAN COMMUNITY DETECTION ALGORITHM

1) Compute Edge betweeness Centrality
2) Remove Edge with Highest Centrality
3) Calculate edge betweeness for remaining edges
4) Until endup with Connected Component / Community repeat.

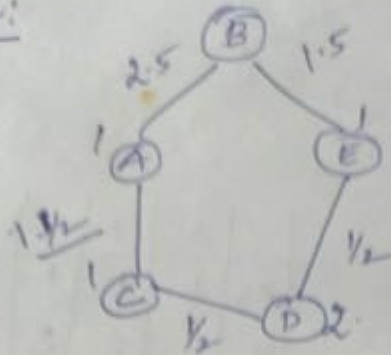
(1)



Step1: Find the Breadth first search of node (A)

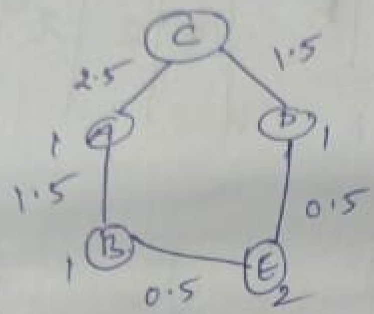


Step2:

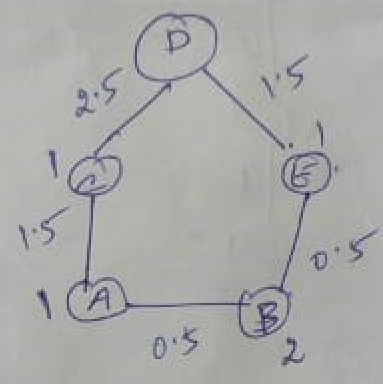


AB	1.5 + 2.5 + 1.5 + 0.5 + 0.5	6.5
AC	2.5 + 1.5 + 2.5 + 1.5 + 0.5	8.5
CD	1.5 + 0.5 + 1.5 + 2.5 + 1.5	7.5
BE	0.5 + 1.5 + 0.5 + 0.5 + 1.5	4.5
DE	0.5 + 0.5 + 1.5 + 1.5 + 2.5	6.5

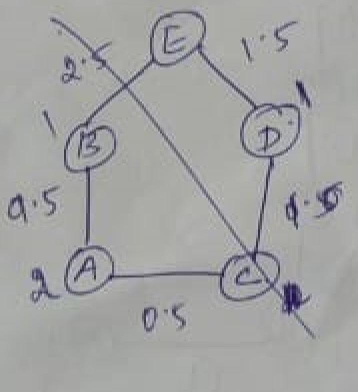
Step3:

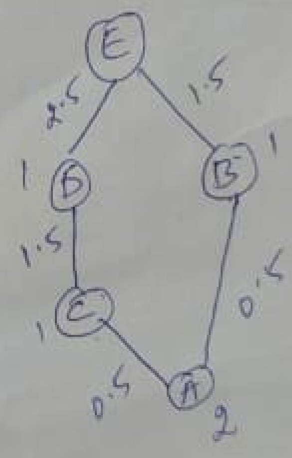


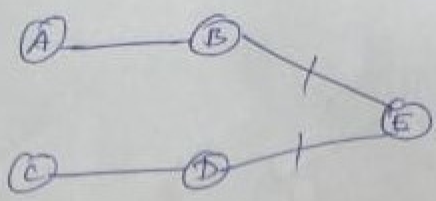
Step4:



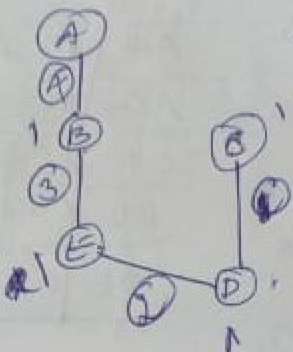
Step5:





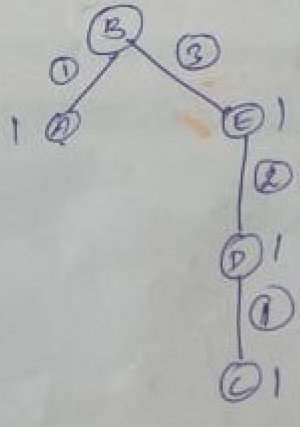


<u>Step1:</u>

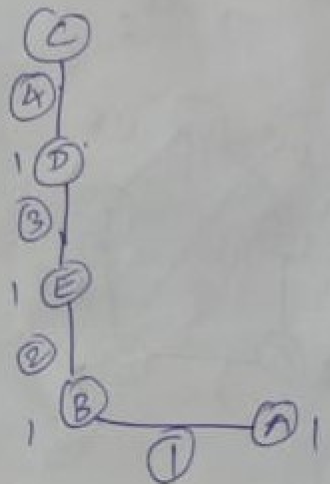


	A	B	C	D	E	Total
AB	4	1	1	1	1	8
BE	3	3	2	2	2	12
CD	1	1	4	1	1	8
DE	2	2	3	3	2	12

<u>Step2:</u>



<u>Step3</u>



Girvan-Newman Optimization Method

① Divide up the network

② Calculate Modularity (Q)

$$Q = \sum \left(\begin{matrix}\text{observed fraction of} \\ \text{links in group}\end{matrix} - \begin{matrix}\text{Expected fraction} \\ \text{of links in group}\end{matrix} \right)$$

Girvan-Newman Modularity $(Q) = \sum_{s=1}^{N_m} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$

N_m - no. of modules

l_s - no. of links between nodes in modules (s)

L - no. of links in the network

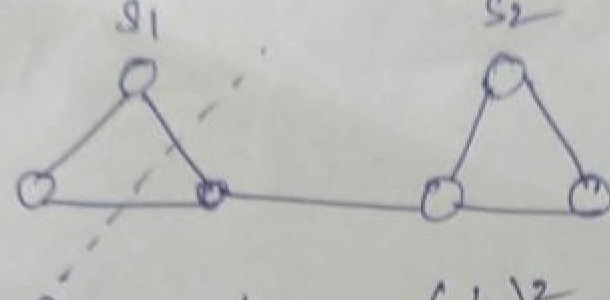
d_s - sum of the degrees of the nodes in modules 's'

③ Use this steps repeatedly until the solution is optimized (modularity is no longer increasing)

<u>Note:</u> for small scale network (Simulated Annealing) is accurately (brute-force approach).

Not applicable for large networks

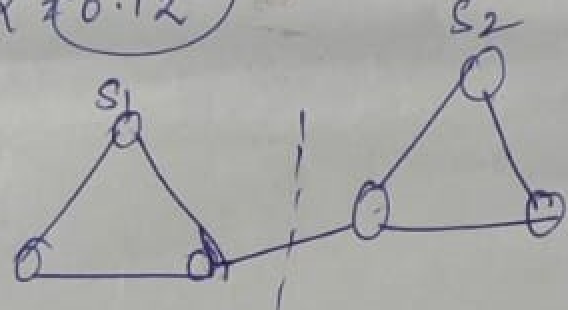
$$Q = \Sigma \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$



$$Q_{s_1} = \frac{1}{7} - \left(\frac{4}{14} \right)^2 = 0.06.$$

$$Q_{s_2} = \frac{4}{7} - \left(\frac{10}{14} \right)^2 = 0.06.$$

$$Q = 0.12$$



$$Q_{s_1} = \frac{3}{7} - \left(\frac{7}{14} \right)^2 = 0.18$$

$$Q_{s_2} = \frac{3}{7} - \left(\frac{7}{14} \right)^2 = 0.18$$

$$Q = Q_{s_1} + Q_{s_2} = 0.36$$

Repeating in different position with Simulated Annealing algorithm until modularity is optimized