

## Performance Metrics for Chromatic Correlation Clustering for Social Network Analysis

Jaishri Gothania\*, Shashi Kant Rathore

Department of Computer Science & Engineering, School of Engineering & Technology, Career Point University, Kota 324005, Rajasthan, India

Corresponding Author Email: [jaishri.gothania2692@gmail.com](mailto:jaishri.gothania2692@gmail.com)

<https://doi.org/10.18280/ria.330507>

**Received:** 9 July 2019

**Accepted:** 16 September 2019

### **Keywords:**

*community detection, community discovery, chromatic correlation clustering, chromatic balls, performance metrics, social network analysis*

### **ABSTRACT**

The social network can be viewed as a chromatic graph of relations and entities. Thus, the community discovery in social network is essentially a problem of chromatic correlation clustering. This paper aims to develop metrics to measure the performance of community discovery algorithms in view of nonoverlapping strong communities. Three performance metrics, namely, chromatic density (CD), chromatic cut ratio (CCR) and chromatic conductance (CC), were proposed for thorough analysis on the output quality of chromatic clustering algorithms. In addition, synthetic graph generator was developed to generate sparse networks with few dense and strong communities. Five algorithms for chromatic correlation clustering, i.e. CB, ICB, LCB, OCB and RECB, were evaluated by the proposed metrics. The evaluation shows that the RECB is the most suitable algorithm for the discovery of strong communities in social network. The research results shed important new light on the detection of communities in social network.

## 1. INTRODUCTION

Increasing use and ease of availability of Internet has made social platforms like Facebook, Twitter, Google+ very popular. People prefer to connect to their acquaintances and even strangers with like mindedness through these social media. This popularity combined with commercial interests and trendiness has made social networks seemingly increasing day by day. From an analyst's point of view these networks are potential source of useful information for business and science alike. The technical term social network analysis (SNA) refers to mapping and measuring relationships and flow of data among people, groups, organizations, websites, machines and other entities that are connected via the social networks. The social network existing on any platform is modelled as a graph network where mostly nodes represent the entities and the interactions are denoted through edge, links and other "ties". A fundamental problem related to these social networks is the discovery of "clusters" or "communities". A community is a subset of data objects in a graph or a cluster of densely connected data nodes of a network. Other networks of objects could be transmission network and similarity networks. Karatas and Sahin [1] have listed all areas of application of community detection, like criminology, public health, politics, customer segmentation, marketing, recommendation systems, etc. Other applications can be found [2].

Any social network is represented as graph structure where the entities involve like users or subscribers are represented by nodes and the relationships between them by edges. Although the meanings of relations vary in different problems, there are two main types: two nodes are related if:

1. There is material transmission between them, and/or
2. They share some identical or similar properties.

A description of node and edge is what differentiates a network model from another. Thus, main purpose of a network

model is not modelling the data entities rather modelling the interaction of these entities. Though generally, a simple undirected graph suffices the need, there may be need of labels. The labels may be used either to describe the 'intensity' of a relationship between nodes or there may be many kinds of relationship existing. When multiple types of relationships exist, weight of the edge is rather a label pertaining to the type of edge. Weighted undirected graph models for social networks have been adopted [3, 4]. Bonchi et al. [5] have come up with a chromatic graph model where the color labels of edges represent a type of relationship in any object network.

Community discovery is a problem of finding groups of entities that are linked to each other more than to others in the network. It can be considered analogous to finding connected components of certain degree in a graph but the problem is much more complex. Neither the number of nodes, or edges or minimum degree of vertices in a community is known a priori. This makes discovering communities very difficult. The size of networks is too large and highly sparse making it computationally difficult. Dynamism of social networks makes the theoretical works to be implemented in real time scenario very challenging. Several algorithms have been proposed for community detection in past decade and each differs in the model considered and definition of community accepted according to the application for which it is designed. Some commonly accepted definitions can be concluded to define a community as a group of nodes that have a dense connection among themselves as compared to sparsely outside the group. Several formal discussions of a community structure can be found [6]. Besides these, there are strong communities and weak communities [7]. Algorithms for strong communities are more variedly applicable.

Scope of this paper is to develop metrics to measure performance of community discovery algorithms in view of strong communities that are non-overlapping. We also develop

a synthetic data generator that generates network structures with given sparsity and few strong communities.

## 2. RELATED WORK

A community detection algorithm is like any clustering algorithm that partitions the nodes of a network into few non-overlapping subsets. Let network be  $G(V, E)$  that is to be divided into  $K$  number of communities, ( $K$  is unknown before execution of algorithm and is known only when the algorithm has partitioned the network). The partition can be denoted as  $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$  such that each node participates in one and only one community, i.e.,  $|\omega_1| + |\omega_2| + \dots + |\omega_k| = |V|$ . The function  $f(\omega)$  also called a performance metric may characterize a quality of the community either on the basis of the connectivity of nodes in  $\omega$  or based on their differences. Yang and Leskovec [8] have summarized these scoring functions and grouped them into the following three broad classes. The metrics are collected from [6, 9-11]:

(A) Scoring functions based on internal connectivity:

**Internal density:** It is the internal edge density of the node in community  $\omega$  defined as

$$f(\omega) = \frac{|E_\omega^{in}|}{|\omega|(|\omega|-1)/2} \quad (1)$$

**Edge inside:** It is defined as simply the number of edges among the nodes in community, mathematically put as

$$f(\omega) = |E_\omega^{in}| \quad (2)$$

**Average degree:** It is the average number of edges that any node has within the community  $\omega$ , computed as

$$f(\omega) = 2|E_\omega^{in}|/|\omega| \quad (3)$$

**Fraction over median degree (FOMD):** Once the internal degree of all nodes in community have been measured, its median is computed as  $d_m$ . The number of nodes that have degree higher than median contribute to strength of community, hence FOMD is computed as

$$f(\omega) = \frac{|u:u \in \omega, \{v, \omega \in \omega, (u, v) \in E, (u, \omega) \in E, (v, \omega) \in E\} \neq \emptyset|}{|\omega|} \quad (4)$$

**Triangle Participation Ratio (TPR):** A triangle is smallest completely connected subgraph in any community and hence counting triangles gives a sense of strength of community. TPR is computed by counting number of nodes that are a part of triangle within the community, as

$$f(\omega) = \frac{|u:u \in \omega, \{v, \omega \in \omega, (u, v) \in E, (u, \omega) \in E, (v, \omega) \in E\} \neq \emptyset|}{|\omega|} \quad (5)$$

(B) Scoring functions based on external connectivity:

**Expansion** measures the number of edges per node that point outside the cluster:

$$f(\omega) = |E_\omega^{out}|/|\omega| \quad (6)$$

**Cut Ratio** is the fraction of existing edges (out of all possible edges) leaving the cluster

$$f(\omega) = |E_\omega^{out}|/|\omega|(N - |\omega|) \quad (7)$$

(C) Scoring functions that combine internal and external connectivity:

**Conductance:** It is the measure of fraction of edges pointing outside the community out of all edges related to a community. It is computed as

$$f(\omega) = \frac{|E_\omega^{out}|}{2|E_\omega^{in}| + |E_\omega^{out}|} \quad (8)$$

**Normalized Cut:** Computed as

$$f(\omega) = \frac{|E_\omega^{out}|}{2|E_\omega^{in}| + |E_\omega^{out}|} + \frac{|E_\omega^{out}|}{2(m - |E_\omega^{in}|) + |E_\omega^{out}|} \quad (9)$$

normalises the cut score.

**Maximum-ODF (Out Degree Fraction):**

$f(\omega) = \max_{u \in \omega} \frac{|(u, v) \in E: v \notin \omega|}{d(u)}$  is the maximum fraction of edges of a node in  $\omega$  that point outside  $\omega$

**Average-ODF:**  $f(\omega) = \frac{1}{|\omega|} \sum_{u \in \omega} \frac{|(u, v) \in E: v \notin \omega|}{d(u)}$  is the average fraction of edges of nodes in  $\omega$  that point out of  $\omega$ .

**Flake-ODF:**  $f(\omega) = \frac{|u:u \in \omega, \{v, \omega \in \omega, (u, v) \in E, d(v) < d(u)/2\}|}{|\omega|}$  is the fraction of nodes in  $\omega$  that have fewer edges pointing inside than to the outside of the cluster.

Besides these, there are metrics that are based on the model adopted by the algorithm. Most of the times it is the objective function of the partitioning problem, or a function associated to the objective function. Concept of modularity [12], density [10], cohesiveness [13] are all a measure of how compact a community is. It leads to measure how strong a community is with respect to itself. The metrics like separability [9, 10] and edges cut [13] are a measure of how strong a community is with respect to other communities. Reader is suggested to read [14] for detailed description of many metrics related to community detection. But majority of these metrics are for unlabelled graphs.

## 3. CHROMATIC CORRELATION CLUSTERING

Bonchi et al. [5] proposed the problem of chromatic correlation clustering. They suggested a model that extended Bansal et al's work [3] by introducing color labels on edges to represent categorical relations between objects in social media networks. Thus, objects at vertices of a network are joined through labelled edges and label is a color name that indicates the type of relation. For every absent edge in  $G$ , a special label  $l_0$  not belonging to the set of labels, is assigned to the edge. The formal definition of this model is stated below.

*A Chromatic Graph can be denoted as  $G = (V, E, L, l_0, \ell)$  where  $V$  is a set of vertices,  $E \subseteq V_2$  is a set of edges ( $V_2$  is the set of all unordered pairs of objects in  $V$ ),  $L$  is a set of labels,  $l_0 \notin L$  is a special label, and  $\ell: V_2 \rightarrow L \cup \{l_0\}$  is a labeling function that assigns a label to each unordered pair of vertices in  $V$  such that  $\ell(x, y) = l_0$  if and only if  $(x, y) \notin E$ .*

Thereafter, the problem of chromatic correlation clustering is defined as partitioning the set of vertices such that the vertices falling in same cluster (subset) should be connected to each other through edges of same color label. Mathematically through objective function it can be described as a clustering  $\mathcal{C}: V \rightarrow N$  and a clustering label  $cl: [V] \rightarrow L$  needed to

minimize the following cost

$$\begin{aligned} \text{cost}(G, \mathcal{C}, cl) = & \sum_{\substack{(x,y) \in V_2 \\ \mathcal{C}(x) = \mathcal{C}(y)}}} (1 - \mathbb{I}[\ell(x,y) = cl(\mathcal{C}(x))]) + \\ & \sum_{\substack{(x,y) \in V_2 \\ \mathcal{C}(x) \neq \mathcal{C}(y)}}} (1 - \mathbb{I}[\ell(x,y) = l_0]) \end{aligned} \quad (10)$$

where,  $\mathbb{I}[\cdot]$  denotes the indicator function. Eq. (10) contains two parts intra and inter-cluster costs respectively where the former is a measure of homogeneity of the labels on the intra-cluster edges and the latter is a cost incurred by the objective function only if vertices  $x$  and  $y$  are adjacent, regardless of the  $label(x,y)$  on the shared edge. Bonchi et al. [5] themselves have proposed two different algorithms to solve above problem. The Chromatic Balls (CB) algorithm randomly selects a pivot edge to begin a cluster and adds those points to the cluster which are connected to those already in cluster via edges of label same as cluster label. Thus, it gathers a ball or community out of entities that have similar relationship with more than one entity having that relationship among them. The lazy Chromatic Balls (LCB) algorithm appearing in same paper relaxes the condition of CB to include more entities in a ball. The LCB algorithm is heuristic and depends on the random selection of pivots based on probability proportional to the degree of vertex. Pivot edge is selected uniformly at random according to the probability of labels.

The performance guarantees of CB and LCB are not tight as the pivot selection at each iteration is done randomly [15]. The problem arises when the pivot edge chosen does not belong to the desired optimum solution. The Informed Chromatic Balls (ICB) by Gothania and Balabuksh [16] gives definition of good edge and bad edge in this respect. Good edge has more chances of being a part of an optimal solution and hence more probability of being chosen as pivot. Later Optimized Chromatic Balls (OCB) proposed [17] suggested how to pick a pivot so that the first community extracted pertains to a larger community. It counts the number of edges of each color occurring in the graph and thereby computes the probability of every edge to be chosen as pivot based on chromatic degree of its two vertices. The approach is to pick a pivot vertex first and the deciding which edge of this vertex to be taken as pivot.

Very recently, in our work [18] recursive extraction of chromatic balls (RCB) has been proposed that applies clique detection and extraction approach to a chromatic graph. Clique of certain labelled edges is extracted as a community. Strong connection with other vertices of any vertex in a ball is a requirement here. It assigns a chromatic label to each vertex and then assigns a cluster to each temporarily. These temporary clusters are evaluated on basis of value of cost function of chromatic correlation clustering. One with lowest cost is assumed to contain a strong community which can be extracted as a clique.

#### 4. PROPOSED PERFORMANCE MATRICS

The graph model we have assumed for our experiments and study are explained here. The graph model has to be related to the chromatic correlation clustering problem. Hence, we assume an undirected labelled chromatic graph. Entities in the interaction network are denoted as nodes of graph. The relationships or interactions between objects form edges between nodes. The interactions may be more than one type.

The labels on edges are such that each label corresponds to a colour/relationship. A special label denotes absence of edge between nodes.

Chromatic correlation clustering has an objective function associated with it which is a cost function. A lower value of cost function indicates better clustering. Yet, there is a need of other metrics too for judging quality of community structure discovered by any chromatic correlation clustering algorithm.

Picking ideas from above discussed metrics, we adapt them for the chromatic clustering problem. A metric for intrinsic property, a metric for extrinsic and one for combining both are required. The metrics proposed are:

**Chromatic density** – the fraction of edges in a cluster of same colour as the label of cluster out of all possible edges in the cluster is called chromatic density.

$$CD_\omega = \frac{| \{ (x,y) | \ell(x,y) = cl(\omega), x,y \in \omega \} |}{|\omega|(|\omega|-1)/2} \quad (11)$$

**Chromatic cut ratio** – the fraction of edges leaving the cluster and having same colour label as the label of cluster out of all possible edges.

$$CCR_\omega = \frac{| \{ (x,y) | \ell(x,y) = cl(\omega), x \in \omega, y \notin \omega \} |}{|\omega|(|V|-|\omega|)} \quad (12)$$

**Chromatic conductance** – the fraction of total edge volume that points outside the community/cluster.

$$CC_\omega = \frac{| \{ (x,y) | \ell(x,y) = cl(\omega), x \in \omega, y \notin \omega \} |}{2 * | \{ (x,y) | \ell(x,y) = cl(\omega), x,y \in \omega \} | + | \{ (x,y) | \ell(x,y) = cl(\omega), x \in \omega, y \notin \omega \} |} \quad (13)$$

Thus, total edges for a cluster are analyzed. The edges with both vertices inside the cluster and having same label as cluster contribute towards CD. The edges with a vertex inside the cluster and another outside the cluster but label same as the cluster contribute towards both CCR and CC, with different denominators. The metric CD takes into account the correctly clustered vertices. For any algorithm of community detection in a chromatic correlation clustering problem to be successful, the chromatic density of discovered community should be high; chromatic cut ratio should be low indicating that community is more cohesive and less adhesive towards other entities. Similarly, chromatic conductance should be low.

## 5. METHODOLOGY

### 5.1 Synthetic graph generator

Synthetic networks are generated by taking input number of nodes  $N$ , number of labels  $L$  and sparsity  $s$ . The method of generation is as follows:

Create a graph with  $N$  number of nodes and no edges.  $G = \{V, E\}, V = \{1, 2, \dots, N\}, E = \emptyset$

Compute total number of edges to be formed according to sparsity as

$$|E| = \left[ (1 - s) * \frac{n(n-1)}{2} \right] \quad (14)$$

Select  $L$  non-overlapping subsets of  $V$ , assign one label to each. The number of nodes per subset is computed as

$$N_s = \left\lceil \sqrt{\frac{1.8*|E|}{L}} \right\rceil \quad (15)$$

Join every vertex in a subset to every other vertex in it with edges labeled with label of subset.

Connect vertices in different subgroups and remaining vertices randomly through edges until required number of edges is formed.

This method produces sparse networks that have very dense subcomponents within them.

## 5.2 Using the proposed performance metrics

The metrics suggested are measured for individual communities discovered in the network. Often algorithms are designed to extract all possible communities from the network

and every community has a different measure of the metrics. Instead of comparing some combined metric value like sum or average of metric values of all discovered communities, it is better to compare the values for the largest community discovered. This is justified because several algorithms aim at discovering only one large community, if it exists in a sparse social network.

## 6. IMPLEMENTATION AND RESULTS

During experiments, the parameters were taken as L=4 and s=0.9 while value of N was varied from 1000 to 2000. The output of every algorithm is arranged in decreasing order of chromatic density and thereafter largest of top L communities is picked for recording the result.

**Table 1.** Results of experiment

N ↓	Algorithm	Runtime	No of Communities Discovered	For largest community discovered			
				Size	Chromatic Density	Chromatic Cut Ratio	Chromatic Conductance
1000	CB	1.06	5	173	1.484	0.00285	0.03611
	LCB	1.35	5	216	0.963	<b>0.00224</b>	0.15202
	ICB	9.67	4	171	1.534	0.00386	0.03945
	OCB	1.21	4	210	1.017	0.00226	0.14258
	RCB	10.03	4	149	<b>2</b>	0.00289	<b>0.03276</b>
1100	CB	1.16	8	184	1.531	0.00306	0.04567
	LCB	1.78	5	246	0.9	<b>0.00231</b>	0.03649
	ICB	13.83	5	186	1.554	0.00297	0.12786
	OCB	1.56	5	246	0.9	<b>0.00231</b>	0.12786
	RCB	11.82	6	164	<b>2</b>	0.00281	<b>0.03320</b>
1200	CB	1.79	7	202	1.553	0.00384	0.03901
	LCB	2.29	5	287	0.789	<b>0.00224</b>	0.18657
	ICB	18.67	4	206	1.51	0.00298	0.03695
	OCB	2.017	5	282	0.816	<b>0.00224</b>	0.18694
	RCB	16.55	6	179	<b>2</b>	0.00290	<b>0.03327</b>
1300	CB	2.56	7	218	1.585	0.00287	0.03646
	LCB	3.68	4	335	0.681	<b>0.00217</b>	0.22529
	ICB	46.07	5	218	1.584	0.00292	0.03673
	OCB	3.69	4	314	0.773	0.00230	0.18654
	RCB	26.58	6	194	<b>2</b>	0.00296	<b>0.03421</b>
1400	CB	2.95	8	237	1.555	0.00298	0.03650
	LCB	4.86	4	372	0.642	<b>0.00224</b>	0.21075
	ICB	59.036	6	227	<b>2</b>	0.00304	0.03335
	OCB	4.465	4	357	0.696	0.00233	0.19545
	RCB	30.67	5	209	<b>2</b>	0.00281	<b>0.03330</b>
1500	CB	3.54	10	248	1.617	0.00346	0.03827
	LCB	5.903	5	397	0.646	<b>0.00217</b>	0.21622
	ICB	74.53	6	245	<b>2</b>	0.00357	0.03598
	OCB	5.313	5	410	0.606	0.00228	0.23784
	RCB	37.654671	7	224	<b>2</b>	0.00279	<b>0.03380</b>
1600	CB	5.723001	8	259	1.675	0.00420	0.03994
	LCB	7.546703	5	467	0.533	0.00222	0.22479
	ICB	93.035557	4	260	1.69	0.00290	0.03435
	OCB	6.340645	4	477	0.511	<b>0.00215</b>	0.24833
	RCB	46.258443	5	239	<b>2</b>	0.00282	<b>0.03243</b>
1700	CB	5.131078	8	282	1.61	0.00340	0.03677
	LCB	7.759667	4	530	0.469	<b>0.00207</b>	0.27398
	ICB	106.141559	5	280	<b>2</b>	0.00373	0.03531
	OCB	6.566336	4	559	0.422	0.00224	0.28503
	RCB	54.090509	5	254	<b>2</b>	0.00276	<b>0.03209</b>
1800	CB	6.128113	13	293	1.661	0.00419	0.03882
	LCB	9.145659	4	555	0.479	<b>0.00211</b>	0.26255
	ICB	132.205452	8	287	<b>2</b>	0.00279	0.03209
	OCB	7.893057	5	587	0.428	0.00220	0.29404

	RCB	63.043635	6	269	<b>2</b>	0.00277	<b>0.03287</b>
	CB	7.284397	7	306	1.722	0.00286	0.03428
	LCB	9.734886	5	603	0.452	<b>0.00208</b>	0.27443
1900	ICB	151.371796	6	310	1.769	0.00279	0.03454
	OCB	8.651227	5	586	0.478	0.00212	0.27529
	RCB	73.092548	8	284	<b>2</b>	0.00280	<b>0.03180</b>
	CB	8.529425	10	327	1.65	0.00379	0.03810
	LCB	11.712076	4	656	0.424	<b>0.00208</b>	0.27559
2000	ICB	178.905398	4	324	1.703	0.00278	0.03374
	OCB	9.991534	4	715	0.358	<b>0.00208</b>	0.32311
	RCB	85.720609	5	299	<b>2</b>	0.00273	<b>0.03262</b>

The results can be interpreted into following observations and conclusions, as shown in Table 1:

**Runtime** - Runtime of ICB and RECB are too high as compared to others as it also includes the time required for pre-processing the input graph. Due to heuristic nature of CB, minimum runtime is achieved in every set of experiments.

**Number of communities discovered** – CB algorithm shows a general tendency of discovering a greater number of communities, followed by RECB which generates lesser communities than CB but higher than number of ground truth communities. This is due to emphasis on small strong community discovery strategy. The LCB and OCB algorithms tend to generate fewer communities.

**Size of community** - The size of community discovered by RECB is exactly equal to value of  $N_s$  that is ideal. Any larger or smaller community indicates that either algorithm is too relaxing or algorithm is heuristic and does not explore properly to find only sub-optimal results. ICB and RECB are good choice when strong communities which are rather small in size are to be discovered.

**CD of largest community** - Chromatic density of communities discovered by RECB is at ideal value of 2. Every other algorithm has a lower value.

**CCR of largest community** - Value of chromatic cut ratio is not consistent in performance. It is lower for LCB and OCB but not much higher in ICB and RECB too. Only CB gives high values of chromatic cut ratio because of its probabilistic nature. The communities discovered by CB are not isolated enough from other vertices.

**CC of largest community** - RECB consistently has lowest value of chromatic conductance in each set of experiments. It indicates presence of very strong communities in output of RECB.

**Other interpretations** - LCB and OCB are recommended when largest possible community from a network is to be extracted. CB is very probabilistic and does not guarantee quality but is quickest. It can be opted when quality can be compromised for just detection purpose in less time.

## 7. CONCLUSIONS

Community detection in social networks has several applications, which is aptly described by chromatic model of multiple relationships among entities. Major contributions of this paper are related to this model. Three performance metrics, namely Chromatic density, chromatic cut ratio and chromatic conductance are proposed for thorough analysis of quality of output of chromatic clustering algorithms. Secondly, a synthetic data generator that generates sparse networks with few dense and strong communities is suggested. Five algorithms (CB, ICB, LCB, OCB and RECB) for chromatic correlation clustering are evaluated based on proposed metrics.

The study finds RECB most suitable for discovering strong communities.

## FUTURE WORK

Efforts to combine the individual community metrics into a global metric for entire network that can be output by a community detection algorithm is an open problem right now. Extending the proposed chromatic correlation metrics to multichromatic clustering problem can also be addressed in future.

## ACKNOWLEDGMENT

The author acknowledges and expresses the gratitude for the support of research supervisor, Computer Science & Engineering Department of University and Dr. Sitesh Kumar Singh for his kind support.

## REFERENCES

- [1] Karataş, A., Şahin, S. (2018). Application areas of community detection: A review. In 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), pp. 65-70. <https://doi.org/10.1109/IBIGDELFT.2018.8625349>
- [2] Ahuja, M.S., Singh, N.J. (2016). Practical applications of community detection. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(4): 412-415.
- [3] Bansal, N., Blum, A., Chawla, S. (2004). Correlation clustering. *Machine Learning*, 56(1-3): 89. <https://doi.org/10.1023/B:MACH.0000033116.57574.95>
- [4] Liu, R.F., Feng, S., Shi, R.S., Guo, W.B. (2014). Weighted graph clustering for community detection of large social networks. *Procedia Computer Science*, 31: 85-94. <https://doi.org/10.1016/j.procs.2014.05.248>
- [5] Bonchi, F., Gionis, A., Gullo, F., Tsourakakis, C.E., Ukkonen, A. (2015). Chromatic correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(4). <https://doi.org/10.1145/2728170>
- [6] Lu, H.Y., Nayak, A. (2019). A scheme to design community detection algorithms in various networks. *Future Internet*, 11(2): 41. <https://doi.org/10.3390/fi11020041>
- [7] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, 101(9): 2658-2663.

- <https://doi.org/10.1073/pnas.0400054101>
- [8] Yang, J., Leskovec, J. (2015). Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1): 181-213. <https://doi.org/10.1007/s10115-013-0693-z>
- [9] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5): 75-174. <https://doi.org/10.1016/j.physrep.2009.11.002>
- [10] Shi, J., Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 888-905. <https://doi.org/10.1109/34.868688>
- [11] Flake, G.W., Lawrence, S., Giles, C.L. (2000). Efficient identification of web communities. *KDD '00 Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 150-160. <https://doi.org/10.1145/347090.347121>
- [12] Newman, M.E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23): 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- [13] Leskovec, J., Lang, K.J., Mahoney, M. (2010). Empirical comparison of algorithms for network community detection. *WWW '10 Proceedings of the 19th International Conference on World Wide Web*, pp. 631-640. <https://doi.org/10.1145/1772690.1772755>
- [14] Chakraborty, T., Dalmia, A., Mukherjee, A., Ganguly, N. (2017). Metrics for community analysis: A survey. *ACM Computing Surveys (CSUR)*, 50(4). <https://doi.org/10.1145/3091106>
- [15] Anava, Y., Avigdor-Elgrabli, N., Gamzu, I. (2015). Improved theoretical and practical guarantees for chromatic correlation clustering. In *Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee*, pp. 55-65. <http://dx.doi.org/10.1145/2736277.2741629>
- [16] Gothania, J., Buksh, B. (2016). A fast chromatic correlation clustering algorithm. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1870-1874.
- [17] Harikishan, MN., BalaBuksh. (2016). Community detection in social networks through chromatic correlation clustering. *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Jaipur, India.
- [18] Gothania, J., Rathore, S.K. (2019). Discovering strong communities in social networks through chromatic correlation clustering. *International Conference on Smart Computation and Technology, PIET. Jaipur, India.*