

## **Comparative Analysis of Community Detection Algorithms in Biological Networks**

KARTHIK, Vempati Sai, VARDHAN, Gorla Pavan Sai Vishnu, REDDY,  
Kandula Lohith Ranganadha and TANGIRALA, Jaya Lakshmi  
<<http://orcid.org/0000-0003-0183-4093>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/34628/>

---

This document is the Accepted Version [AM]

### **Citation:**

KARTHIK, Vempati Sai, VARDHAN, Gorla Pavan Sai Vishnu, REDDY, Kandula Lohith Ranganadha and TANGIRALA, Jaya Lakshmi (2024). Comparative Analysis of Community Detection Algorithms in Biological Networks. In: 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS). IEEE. [Book Section]

---

### **Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

# Comparative Analysis of Community Detection Algorithms in Biological Networks

Vempati Sai Karthik  
School of Computer Science  
Engineering  
SRM University AP  
Guntur, India  
saikarthik\_vempati@srmmap.edu.in

Gorla Pavan Sai Vishnu Vardhan  
School of Computer Science  
Engineering  
SRM University AP  
Guntur, India  
pavansaivishnuvardhan\_g@srmmap.edu.in

Kandula Lohith Ranganadha Reddy  
School of Computer Science  
Engineering  
SRM University AP  
Guntur, India  
lohith\_ranganadhareddy@srmmap.edu.in

T. Jaya Lakshmi  
SRM University AP  
Guntur, India and  
Sheffield Hallam University  
Sheffield, United Kingdom  
jaya.phd.hcu@gmail.com

**Abstract**— Community detection in biological networks is crucial for understanding complex interactions among biological entities. This research focuses on performing community detection using several algorithms such as Kernighan lin bisection algorithm, Louvain algorithm, Girvan Newman algorithm, Fast Greedy algorithm, and Asynchronous fluid community algorithm on various biological datasets. We evaluated the modularity and partition quality for all the communities using all these algorithms separately and did a comparative analysis on the results. Using those results we were able to identify which algorithm is more efficient and scalable in performing the community detection for biological networks.

**Keywords**— nodes, edges, community, network, modularity, partition quality

## I. INTRODUCTION

The study of biological networks has become pivotal in unraveling the intricate relationships among molecular components. These networks, representing interactions within biological systems, present a complex tapestry of relationships that underlie cellular functions. Community detection, an essential aspect of network analysis, facilitates the identification of cohesive modules within these intricate structures. This research endeavors to advance community detection algorithms tailored to biological networks, aiming to reveal underlying organizational principles. By deciphering modular patterns, this investigation contributes to a deeper understanding of cellular processes, offering insights into disease mechanisms and potential therapeutic targets. The exploration of biological networks stands as a key frontier in contemporary life sciences.

In this paper, we present the following.

- Performing Community Detection using various algorithms.
- Evaluating the performance of each algorithm using quality functions like modularity and partition quality.
- Performed a Comparative analysis of each algorithm using various biological datasets and predicted the better algorithm.

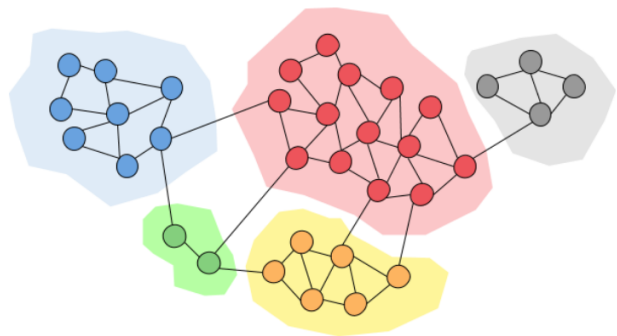


Fig.1 An example of a network having different communities.

The above Fig.1 is a basic example of a network having different communities in it. [4]

- All the numbered circles in the above network are represented as different nodes, and the connections between the nodes are referred to as edges (connections).
- Nodes colored with the blue color belongs to a certain type of (biological domain) (example) and are interconnected to each other to form a community.
- Nodes colored with the green color belongs to a certain type of (ecological domain) (example) and are interconnected to each other to form a community.
- Nodes colored with the yellow color belongs to a certain type of (scientific domain) (example) and are interconnected to each other to form a community.

In the above example blue color community relates to red color and green color, all other communities are also connected.

## Applications of community discovery

Community discovery has vast applications in the following domains. More details are given in (Table I).

Table I. Applications of community discovery in various domains

Network domain	Node	Edge	Communities
<b>Biological</b>	Protein heads present at different parts of body.	Physical interactions. protein	Population of animals living in a particular habitat. Microbial community, such as the microbiome of the human gut.
<b>Ecological</b>	Genes, individuals, populations, species, etc.	It may depict interactions of different kind, or movement measured in numerous ways.	A forest ecosystem may include trees, shrubs, and plants that provide food and shelter for animals such as birds, insects, and mammals.
<b>Engineering</b>	Different types of engineering branches, variety of students choosing variety of branches.	Connection between each branch like common subjects or different students studying common subjects.	communities in engineering can be used for the job searching processes and calculation of literacy rate.
<b>linguistic</b>	Nodes are typically represented by words or phrases	Relationships can be based on a variety of factors, such as semantic similarity, syntactic structure, or cooccurrence in a text corpus.	The communities in linguistic domain include, In general, Words or phrases that are semantically similar, syntactically related, or cooccur frequently in the corpus are more likely to belong to the same community

## II. LITERATURE SURVEY

Community detection within complex networks remains a pivotal endeavor for understanding system dynamics and organization beyond local constituents. A multitude of algorithms have been proposed, yet comprehensive evaluations of their efficacy are sparse [5]. Prior assessments often relied on small networks or simplified artificial graphs, failing to capture the complexities inherent in real-world systems [6]. Recent studies, however, have undertaken rigorous evaluations against benchmark graphs featuring heterogeneous degree and community size distributions, shedding light on algorithmic performance [7]. Notably, algorithms introduced by Rosvall and Bergstrom, Blondel et al., and Ronhovde and Nussinov exhibit promising results, showcasing both effectiveness and computational efficiency for large-scale analyses [8]. Community detection transcends disciplinary boundaries, permeating sociology, biology, and computer science, where networks serve as quintessential representations [6]. Despite concerted interdisciplinary efforts, challenges persist, necessitating thorough methodological expositions and discussions on testing methodologies and real-world applications [6]. Moreover, the significance of community detection extends beyond theoretical exploration, finding practical utility in domains like social media analysis and biological network modeling [7]. In parallel, advancements in computational methods have facilitated drug repositioning efforts, with novel approaches like HINGRL leveraging heterogeneous information networks for precise drug-disease association predictions [9]. Additionally, the burgeoning field of systems biology faces challenges in visualizing and interpreting complex biological networks, spurring the development of user-friendly visualization tools with enhanced

data integration capabilities [10]. Furthermore, high-throughput technologies have revolutionized data generation in biological research, necessitating effective storage, analysis, and comparison algorithms for pathway and interactomics data [11]. The identification of open problems, particularly in network alignment algorithms, underscores the ongoing quest for robust methodologies in biological network analysis [14]. Alongside algorithmic discoveries like the Louvain, Kernighan Lin, Girvan Newman, Fast Greedy, and Asynchronous Fluid algorithms, the field of community detection continues to evolve, offering new avenues for exploration and application [12][13][15]. In this context, our research contributes to the field by performing community detection using several algorithms such as Kernighan Lin bisection algorithm, Louvain algorithm, Girvan Newman algorithm, Fast Greedy algorithm, and Asynchronous Fluid community algorithm on various biological datasets. We evaluated the modularity and partition quality for all the communities using all these algorithms separately and conducted a comparative analysis on the results. Our findings allowed us to identify which algorithm is more efficient and scalable in performing community detection for biological networks [1-3], contributing to the ongoing advancement in this area of research.

## III. METHODOLOGY

### A. QUALITY FUNCTIONS

In our research we used various quality functions like modularity and partition quality to evaluate the performance of various algorithms on several biological datasets. A variety of functions are introduced or proposed for better division of networks and to maintain the goodness of them.

## 1. MODULARITY

Modularity, also known as community detection, is a valuable measure of a network's structure. Originally designed to evaluate the degree of module separation in a network, high modularity networks exhibit strong interconnections between nodes within modules, but fewer connections between nodes in different modules. Although a diversity of community detection algorithms has been proposed, the quality of community detection is usually measured by modularity and some 10 benchmark graphs. Hence, we can say that higher the modularity, better the division of communities. The modularity  $Q$  of a graph division into  $k$  clusters  $V_1, \dots, V_k$  is given by:

$$Q = \sum_{c=1}^k \left[ \frac{A(V_i, V_i)}{m} - \left( \frac{\text{degree}(V_i)}{2m} \right)^2 \right] \quad (1)$$

Formula for calculation of modularity is given in the above Eq.1.

## 2. PARTITION QUALITY

The ratio of intra-community edges to total edges in the graph represents the coverage of a partition. To determine the effectiveness of a partition, one can measure the number of potential edges against the combined number of intracommunity and intercommunity non-edges. This calculation provides valuable insight into the partition's performance. It is a set of two values and these two values typically represent the precision and recall of the community detection algorithm, respectively. Hence, we can say that higher the partition quality, better the division of communities.

## B. COMMUNITY DETECTION ALGORITHMS

In our research we used various algorithms to perform community detection on several biological datasets and performed a comparative analysis.

### 1. KERNIGHAN-LIN-BISECTION ALGORITHM

The Kernighan-Lin (KL) bisection algorithm is a graph partitioning algorithm that is used to divide a graph into two smaller sub-graphs while minimizing the connections between those sub-graphs. The algorithm works by iteratively moving nodes from one subgraph to the other to reduce the number of connections between the two sub-graphs.

### 2. GIRVAN NEWMAN ALGORITHM:

The Girvan-Newman algorithm is a powerful tool for detecting clusters or communities within networks. Combining the principles of divisive hierarchical clustering, it initially considers the entire network as a single cluster before partitioning it into increasingly smaller sub-clusters through iterative processes.

### 3. LOUVAIN ALGORITHM

Louvain Community Detection Algorithm is a straightforward approach for determining a network's community structure. This is a heuristic strategy for optimizing modularity. To

optimize community formation, the algorithm follows a two-step process. First, each node is initially placed in its own community. Then, for each node, there is a concerted effort to increase positive modularity by relocating the node to neighboring communities. If no improvements are achieved, the node stays in its original community.

### 4. FAST GREEDY ALGORITHM

The Fast Greedy algorithm is a hierarchical clustering algorithm used for community detection in networks or graphs. It is particularly popular for identifying communities in larger networks. The algorithm iteratively merges nodes into communities based on a greedy approach that optimizes a modularity function. The quality of a network's partition into communities is measured by modularity.

### 5. ASYNCHRONOUS FLUID COMMUNITY ALGORITHM

The core concept behind the algorithm is inspired by the behavior of fluids, where they expand and exert force on each other within a particular region. Its initialization is random, therefore the communities detected may differ between executions.

## C. DATASET

In our research, we examine the area of biological networks and utilize them as main datasets for a comprehensive study of community discovery. To this end, we apply five different algorithms, and in doing so, we analyze and compare their effectiveness in revealing the underlying structures of biological networks. We quantify each algorithm's modularity and partition quality to determine which performs best at defining meaningful communities inside these intricate networks. This comparison analysis provides important insights into the performance and application of each algorithm in the context of biological network research, in addition to highlighting the benefits and drawbacks of each approach. More details are given in (Table II).

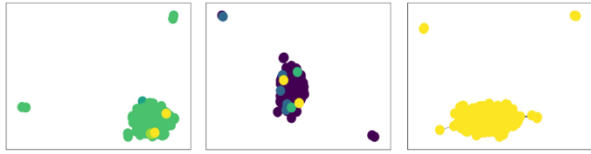
Table II. Representing the graph properties of all the 5 datasets.

Dataset	No of Nodes	No of Edges	Graph Density	Avg Clustering Coefficient
ChCh-Miner_durg_bank-chem-chem	1514	48514	0.0424	0.3040
ChG-Miner_mine_r-chem-gene	7341	15138	0.0006	0.0
bio-yeast-protein-inter	1870	2277	0.0014	0.0672
bio-CE-LC	1387	1648	0.0018	0.0758
bio_DM_L_C	658	1129	0.0053	0.1048

## IV. EXPERIMENTAL RESULTS

Community Graphs for all the 5 datasets using,

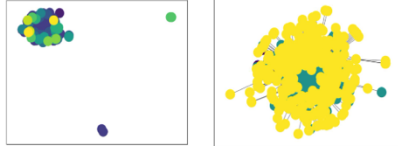
Dataset\_1: ChCh-Miner\_durgbank-chem-chem



**Louvain Algorithm**  
Modularity: 0.39112510636413506  
Partition Quality: (0.6217174423877644, 0.8245343526513065)

**Girvan-Newman Algorithm**  
Modularity: 0.00024731367820234436  
Partition Quality: (0.9999381621799893, 0.054199578990012584)

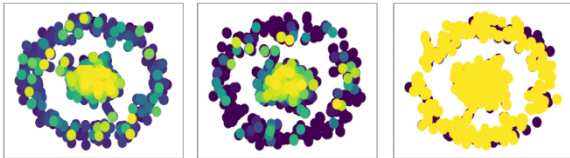
**Kernighan lin Bisection Algorithm**  
Modularity: 0.4144712118941707  
Partition Quality: (0.8501051242940182, 0.5299897585060614)



**Fast Greedy Algorithm**  
Modularity: 0.33947199850760534  
Partition Quality: (0.7108875788432205, 0.686912456639551)

**Asynchronous Fluid Algorithm**  
Modularity: 0.28189631569580637  
Partition Quality: (0.7367867744063324, 0.6525079105938322)

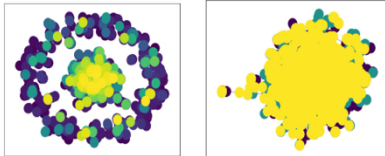
Dataset\_2: ChG-Miner\_miner-chem-gene



**Louvain Algorithm**  
Modularity: 0.7504815367444155  
Partition Quality: (0.8448275862068966, 0.9625953223784746)

**Girvan-Newman Algorithm**  
Modularity: 0.08795286299064903  
Partition Quality: (0.9996036464526358, 0.214307348485439)

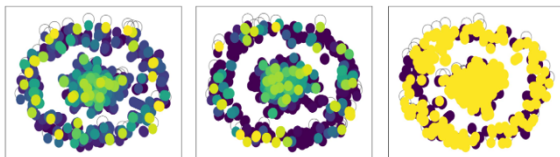
**Kernighan lin Bisection Algorithm**  
Modularity: 0.3738434429845152  
Partition Quality: (0.9367155502708416, 0.5005588781903882)



**Fast Greedy Algorithm**  
Modularity: 0.7134911703506721  
Partition Quality: (0.8770643413925221, 0.9495017161275907)

**Asynchronous Fluid Algorithm**  
Modularity: 0.492096777507946  
Partition Quality: (0.910774295315822, 0.6667789615664887)

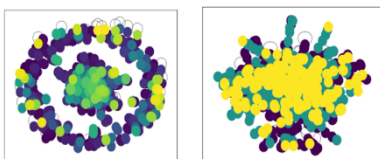
Dataset\_3: bio-yeast-protein-inter



**Louvain Algorithm**  
Modularity: 0.8477234779291241  
Partition Quality: (0.8796662274923145, 0.976341833975674)

**Girvan-Newman Algorithm**  
Modularity: 0.30186765788529996  
Partition Quality: (0.9964866051822574, 0.4637030293078914)

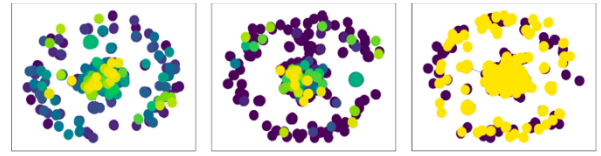
**Kernighan lin Bisection Algorithm**  
Modularity: 0.41433631343123245  
Partition Quality: (0.9143610013175231, 0.50134734179678)



**Fast Greedy Algorithm**  
Modularity: 0.8491429349537821  
Partition Quality: (0.8884497145366711, 0.9730571697524771)

**Asynchronous Fluid Algorithm**  
Modularity: 0.5115210059090409  
Partition Quality: (0.9031610637230306, 0.6139294433099562)

Dataset\_4: bio-CE-LC



**Louvain Algorithm**  
Modularity: 0.9067416045963256  
Partition Quality: (0.9375, 0.9683299156983367)

**Girvan-Newman Algorithm**  
Modularity: 0.390603680452261  
Partition Quality: (0.9993932038834952, 0.5470431995305823)

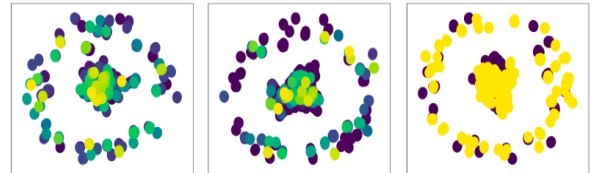
**Kernighan lin Bisection Algorithm**  
Modularity: 0.43248402954619075  
Partition Quality: (0.9162621359223301, 0.5017878860705104)



**Fast Greedy Algorithm**  
Modularity: 0.9059746935435984  
Partition Quality: (0.9423543689320388, 0.9659141627418484)

**Asynchronous Fluid Algorithm**  
Modularity: 0.5553046631135101  
Partition Quality: (0.9546153846153846, 0.6095206769970438)

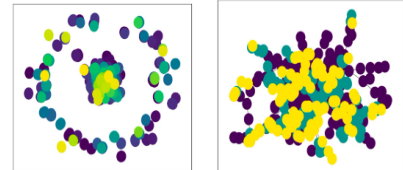
Dataset\_5: bio\_DM\_LC



**Louvain Algorithm**  
Modularity: 0.8142312227521392  
Partition Quality: (0.9229406554472985, 0.9592094488626112)

**Girvan-Newman Algorithm**  
Modularity: 0.5064433828819264  
Partition Quality: (0.9902568644818424, 0.7145818008540247)

**Kernighan lin Bisection Algorithm**  
Modularity: 0.4144712118941707  
Partition Quality: (0.9627989371124889, 0.5055955735058038)



**Fast Greedy Algorithm**  
Modularity: 0.7892908669970676  
Partition Quality: (0.912311780336581, 0.9523855787335822)

**Asynchronous Fluid Algorithm**  
Modularity: 0.5617942091067589  
Partition Quality: (0.9327983951855566, 0.6512632835923473)

RECORDED VALUES

The values of partition quality and Modularity for all the 5 datasets as determined by the 5 respective algorithms are shown in the (Table III, IV) below.

Table III. modularity of 5 datasets

Data Set	Kernighan	Louvain	Girvan Newman	Fast greedy	Asynchronous fluid community
ChChMiner durgbank-chem-chem	0.4145	0.3912	0.0003	0.3395	0.2819
ChGMiner miner-chem-gene	0.3739	0.7505	0.0880	0.7135	0.4921
Bio-yeast-protein-inter	0.4144	0.8478	0.3019	0.8492	0.5116
Bio-CE-LC	0.4325	0.9068	0.3907	0.9060	0.5554
Bio_DM_LC	0.4145	0.8143	0.5065	0.7893	0.5618

Table IV. partition quality of 5 datasets

Data Set	Kernighan	Louvain	Girvan Newman	Fast greedy	Asyn fluid community
ChChMiner_durgbank-chem-chem	(0.851, 0.530)	(0.622, 0.825)	(0.100, 0.055)	(0.711, 0.688)	(0.737, 0.653)
ChGMiner_miner-chem-gene	(0.937, 0.501)	(0.845, 0.963)	(0.910, 0.215)	(0.878, 0.950)	(0.911, 0.667)
bio-yeast-protein-inter	(0.915, 0.502)	(0.880, 0.977)	(0.997, 0.464)	(0.889, 0.974)	(0.904, 0.615)
bio-CE-LC	(0.917, 0.502)	(0.938, 0.969)	(0.910, 0.548)	(0.943, 0.967)	(0.955, 0.610)
bio_DM_LC	(0.963, 0.506)	(0.923, 0.960)	(0.991, 0.715)	(0.913, 0.953)	(0.933, 0.652)

## V. DISCUSSION

For ChCh-Miner\_durgbank-chem-chem dataset, Louvain algorithm has obtained better modularity and partition quality values on average, indicating it performs better for this dataset. For ChG-Miner\_miner-chem-gene dataset, Louvain algorithm has obtained better modularity and partition quality values on average, indicating it performs better for this dataset. For bio-yeast-protein-inter dataset, Fast Greedy Algorithm dominates with the highest modularity and strong partition quality values. For bio-CE-LC dataset, Fast Greedy Algorithm achieves the highest modularity and strong partition quality values, indicating superior performance. For bio\_DM\_LC dataset, Louvain Algorithm once achieves the highest modularity and strong partition quality, indicating its effectiveness for this dataset. In summary, based on the average overall performance, the Louvain Algorithm appears to be the best-performing algorithm across the five datasets, considering both modularity and partition quality.

## VI. CONCLUSION

Community detection in biological networks, akin to its applications in diverse domains, employs mathematical models and optimization algorithms to unveil the modular structure inherent in complex biological interactions. By leveraging objective functions like modularity, this approach provides valuable insights into cellular organization. The combination of computational tools and biological insights not only improve our understanding of complex network dynamics, but it also holds promise for breakthroughs in drug development, disease understanding, and personalized therapy in the field of systems biology. In the analysis of biological networks, community detection algorithms were applied to datasets using Kernighan Lin Bisection, Girvan Newman, Fast Greedy, Asynchronous Fluid and Louvain.

Across diverse biological datasets, the Louvain Algorithm consistently exhibited superior performance, demonstrating higher modularity and robust partition quality. Its efficiency and scalability make it a reliable choice for detecting communities in biological networks.

## REFERENCES

- [1] R. Rossi and N. Ahmed. "The Network Data Repository with Interactive Graph Analytics and Visualization." AAAI. - <https://networkrepository.com> - Accessed November. 17, 2023.
- [2] "Drug-drug interaction network." SNAP. <https://snap.stanford.edu/biodata/datasets/10001/10001-ChCh-Miner.html> - Accessed November. 17, 2023.
- [3] "Drug-target interaction network." SNAP. <https://snap.stanford.edu/biodata/datasets/10002/10002-ChG-Miner.html> - Accessed November. 17, 2023.
- [4] Jayawickrama, T. D. (2021b, December 29). Community Detection Algorithms - towards Data science. Medium - Retrieved November 17, 2023, from - <https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae>
- [5] Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: a comparative analysis. *Physical review E*, 80(5), 056117.
- [6] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3-5), 75-174.
- [7] Bedi, P., & Sharma, C. (2016). Community detection in social networks. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, 6(3), 115-135.
- [8] Zhao, B. W., Hu, L., You, Z. H., Wang, L., & Su, X. R. (2022). HINGRL: predicting drug-disease associations with graph representation learning on heterogeneous information networks. *Briefings in bioinformatics*, 23(1), bbab515.
- [9] Pavlopoulos, G. A., Wegener, A. L., & Schneider, R. (2008). A survey of visualization tools for biological network analysis. *Biodata mining*, 1, 1-11.
- [10] Milano, M., Agapito, G., & Cannataro, M. (2022). Challenges and limitations of biological network analysis. *BioTech*, 11(3), 24.
- [11] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
- [12] Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12), 7821-7826.
- [13] Clauset, A. (2005). Finding local community structure in networks. *Physical review E*, 72(2), 026132.
- [14] Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, 49(2), 291-307.
- [15] Parés, F., Gasulla, D. G., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., ... & Suzumura, T. (2018). Fluid communities: A competitive, scalable and diverse community detection algorithm. In *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)* (pp. 229-240). Springer International Publishing.