Complex Networks: Graph Theoretical Analysis of Structural and Functional Systems

Dr Smita S. Agrawal

Computer Science and Engineering Department
Nirma University

November 18, 2022

Outline

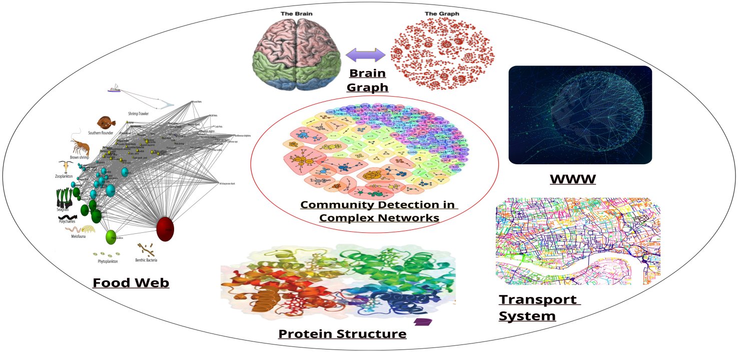
1. Introduction
2. Graph Theoretical Analysis
3. Case Study #1:A Collaborative Similarity Based Graph Clustering for Community Detection in Complex Networks
4. Case Study #1:Predicting overall Flight Delay caused by Origin Delay using Random Forest Algorithm and Path Finding
5. Case Study #3 :AI-enabled optimize routing for solid waste collection in Ahmedabad city
6. Research Opportunities
7. References

Introduction

Connections between data are as important as the data itself

Introduction

- The amount of data traveling across the internet today is not only massive in terms of size but is complex as well.
- Big data of complex networks include very large scale networks with structured, unstructured, or semi-structured data and a set of the graph [6, 29]



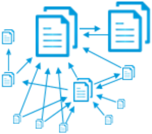
Applications of Community Detection in Complex Network - Brain graph [9], protein structure [12, 14, 21] food web, transportation system [4], World Wide Web [3]

Use of data connections has created industry leaders

altavista

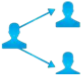
LYCOS
THE CATALOG OF THE INTERNET

Google



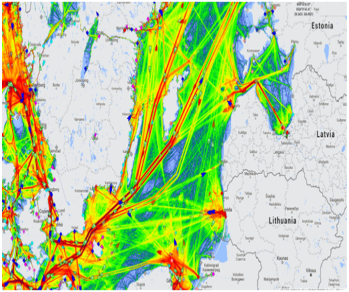
monster
Find better.™

LinkedIn

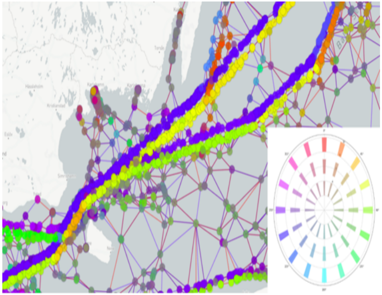


amazon

Use of data connections has created for Maritime



Maritime Traffic Data



Directed graph for the Baltic Sea generated on AIS data for an 8-week period

With regard to military operations,

- Social Network Analysis has been used to show that enemy networks and relationships can be

With regard to military operations,

- Social Network Analysis has been used to show that enemy networks and relationships can be
    - accurately represented using weighted layers

With regard to military operations,

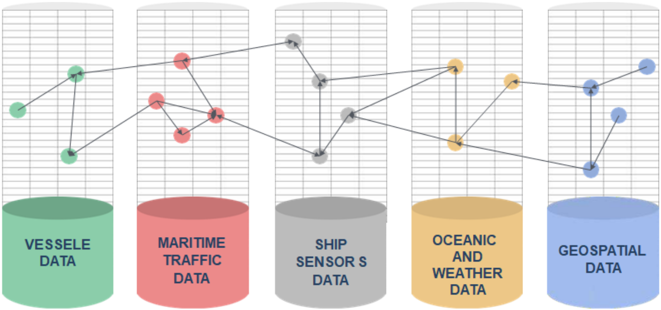
- Social Network Analysis has been used to show that enemy networks and relationships can be
  - accurately represented using weighted layers
  - with weighted relationships

With regard to military operations,

- Social Network Analysis has been used to show that enemy networks and relationships can be
  - accurately represented using weighted layers
  - with weighted relationships
  - in order to identify the key player(s) that must be influenced and/or removed so that

With regard to military operations,

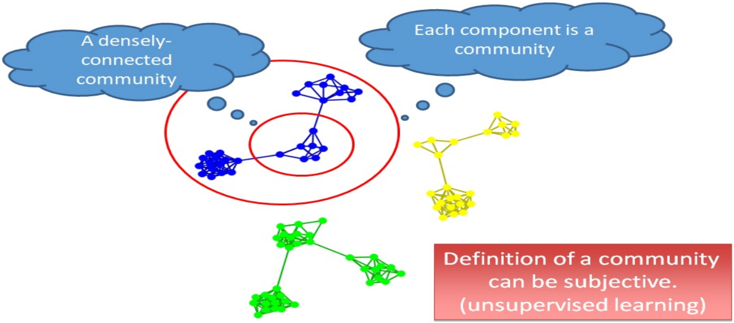
- Social Network Analysis has been used to show that enemy networks and relationships can be
  - accurately represented using weighted layers
  - with weighted relationships
  - in order to identify the key player(s) that must be influenced and/or removed so that
  - A particular effect on the enemy might be realized.



"The next wave of ***competitive advantage*** will be all about ***using connections*** to produce actionable insights."

Community Detection

- Community is formed by individuals such as those within a group interact with each other more frequently than with those outside the group. Such as, group, cluster, cohesive sub-group, module in different contexts.
- Community detection is discovering groups in a network where individuals' group memberships are not explicitly given.

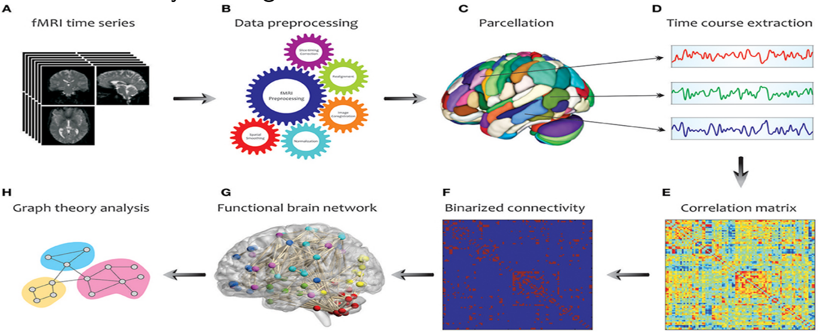


Graph Theoretical Analysis

- Most of the clustering methods on graphs predominantly emphasis on the topological structure without considering connectivity between vertices and not bearing in mind the vertex properties/attributes or similarity-based on indirectly connected vertices.
- Many scientific and commercial applications have complicated structures than sequential patterns and needs additional effort to analyze it.
- Community of brain graphs helps in the development of a new methodology to precisely parse and understand modularity in human brain function and structure [9].
- Communities of proteins networks identify functionally related to protein-protein interaction and also supports earlier stages of treatment of diseases and is related to drug discovery [11, 16, 18, 28].
- Modeling and generating a graph based on their structure, attributes, weight, and direction (whether directed or undirected) becomes an important task.
- Big Data Analytics refers to the process of collecting, organizing, and analyzing large data sets to discover different patterns and other useful information.

Graph Theoretical Analysis : for Brain Network

Schematic representation of brain network construction and graph theoretical analysis using fMRI data

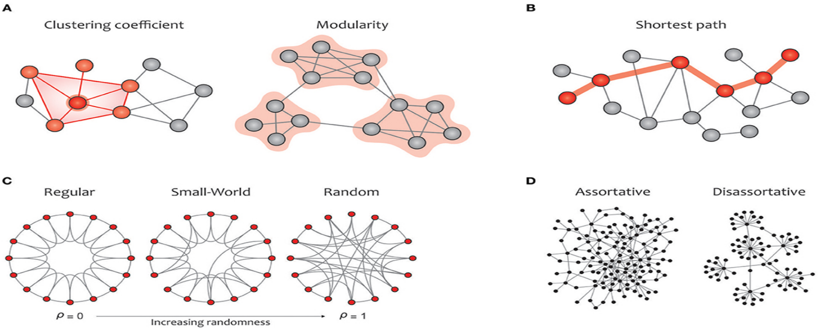


- After processing **(B)** the raw fMRI data **(A)** and division of the brain into different parcels **(C)**, several time courses are extracted from each region **(D)** so that they can create the correlation matrix **(E)**.
- To reduce the complexity and enhance the visual understanding, the binary correlation matrix **(F)**, and the corresponding functional brain network **(G)** are constructed, respectively.
- Eventually, by quantifying a set of topological measures, graph analysis is performed on the brain's connectivity network **(H)**.

Categorization of graph theory applied in neurological domain

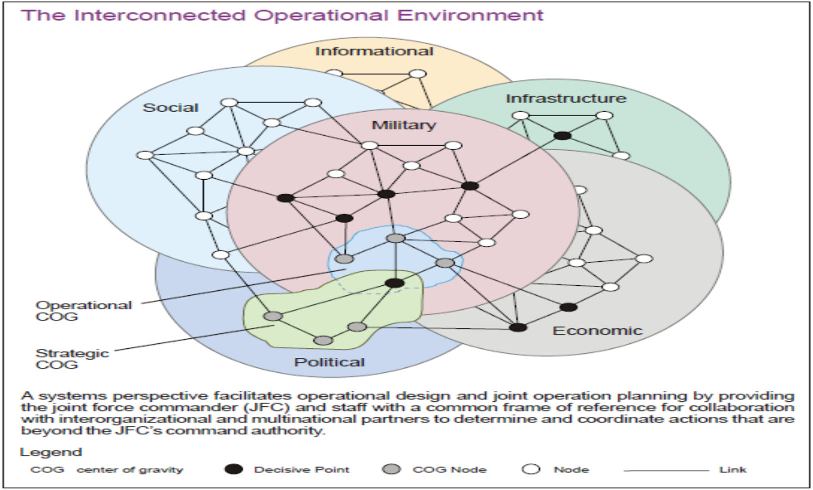
- Graph Theory applications in neurological disorder
    - Epilepsy
    - Alzheimer disease
    - Multiple sclerosis
    - Autism spectrum disorder
    - Attention-deficit/hyperactivity disorder
    - other mental disorder
- Graph Theory applications in human cognition
    - Human intelligence and the brain topology
    - Topological changes across the lifespan
    - Working memory performance and network efficiency
    - Effect of cognitive loads on modularity

Computation of Graph Measures



(A) Segregation measures include clustering coefficient, which quantify how much neighbors of a given node are interconnected and measures the local cliquishness ; modularity, which is related to clusters of nodes, called modules, that have dense interconnectivity within clusters but sparse connections between nodes in different clusters.
(B) Integration measure include characteristic path length, which measures the potential for information transmission, determined as the average shortest path length across all pairs of nodes.
(C) A regular network (left) displays a high clustering coefficient and a long average path length, while a random network (right) displays a low clustering coefficient and a short average path length.
(D) The assortativity index measures the extent to which a network can resist failures in its main components (i.e., its vertices and edges).

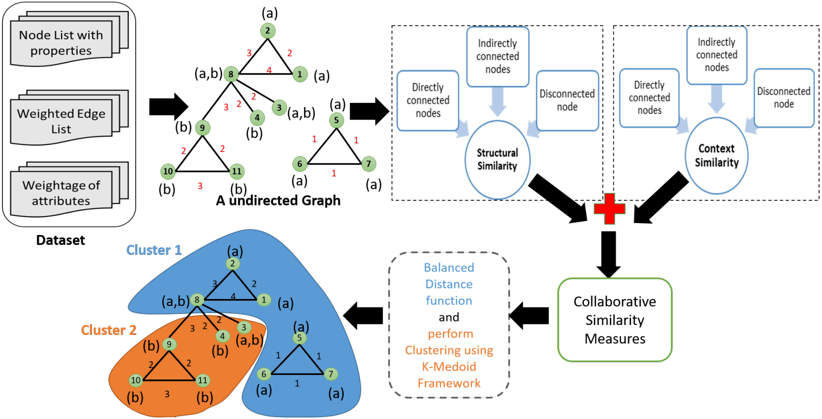
Graph Theoretical Analysis : for Military → The Interconnected Operational Environment



Categorization of Community Detection Criteria

Category	Methods
Node-Centric Community Detection	Complete Mutuality : cliques [26]
	Reachability of members :k-clique, k-clan, k-club [26]
	Nodal degrees :k-plex, k-core [26, 8]
Group-Centric Community Detection	Density-Based Groups [26, 27]
	Machine Learning Based Approaches [7]
Network-Centric Community Detection [26, 2, 13]	Clustering based on vertex similarity
	Latent space models (multi-dimensional scaling)
	Block model approximation
	Modularity maximization
	Spectral clustering
Hierarchy-Centric Community Detection [26, 5]	Divisive Hierarchical Clustering (top-down)
	Agglomerative Hierarchical clustering (bottom-up)

Case Study #1:A Collaborative Similarity Based Graph Clustering for Community Detection in Complex Networks

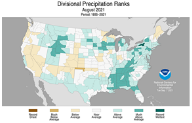


#2:Predicting overall Flight Delay caused by Origin Delay using Random Forest Algorithm and Path Finding

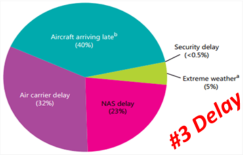
All the airports with their connections in the world

#2:Predicting overall Flight Delay caused by Origin Delay using Random Forest Algorithm and Path Finding

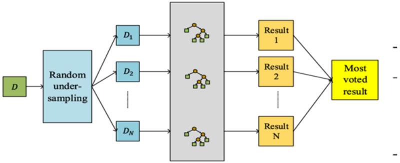
#1 Data Source



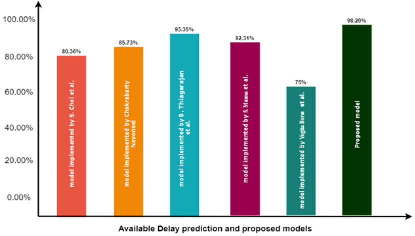
#2 Data Cleaning



#3 Delay Calculation



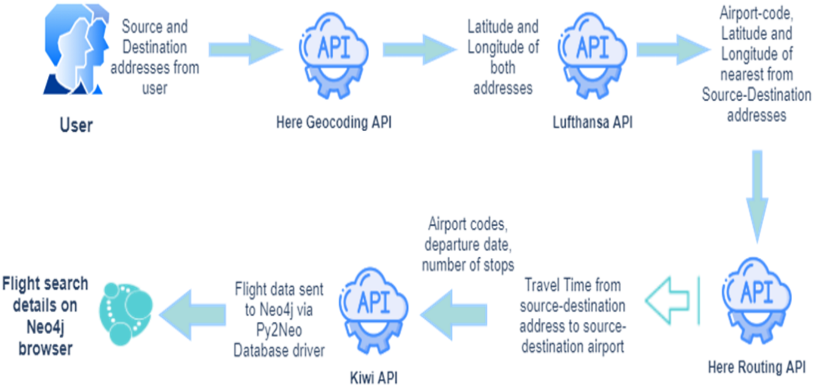
#4 Delay Prediction



#5 Result

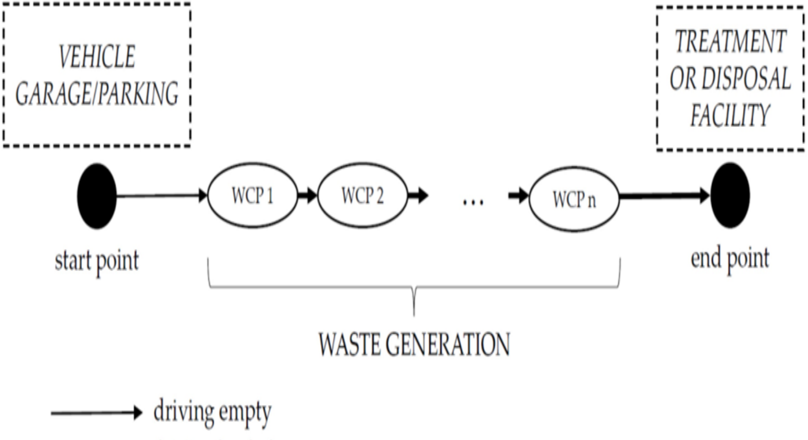
#2:Predicting overall Flight Delay caused by Origin Delay using Random Forest Algorithm and Path Finding

#Flight Search

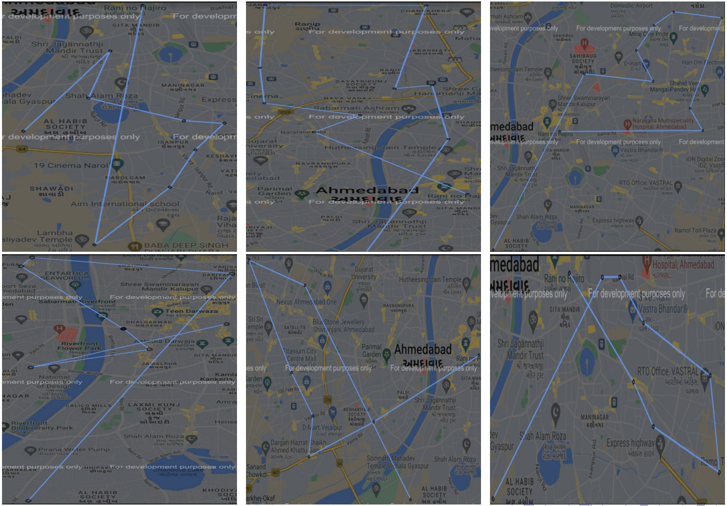


#3:AI-enabled optimize routing for solid waste collection in Ahmedabad city

Route finding for garbage collection is part of a waste management system that can be more cost-efficient with the help of technology.



#3:AI-enabled optimize routing for solid waste collection in Ahmedabad city



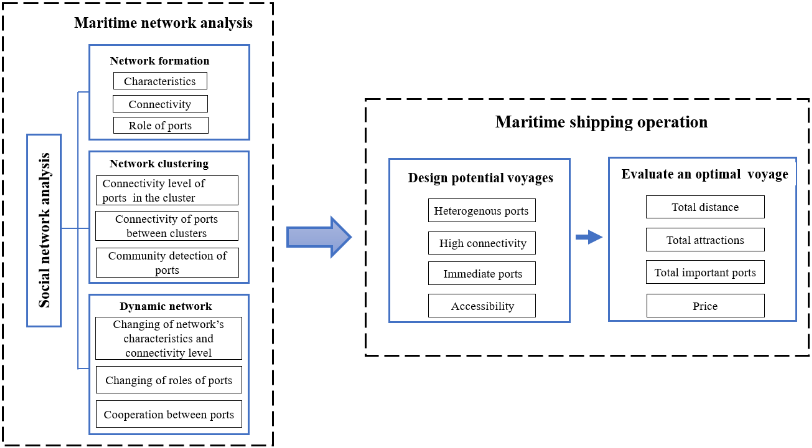
#3:AI-enabled optimize routing for solid waste collection in cities

- To find the optimized route for trucks to collect garbage from smart dustbins, reducing costs and making a more efficient and profitable system.
- Route finding is an NP- hard problem that can be solved by many algorithms, some are heuristics, meta heuristics, VRP, and genetic algorithms.
- Used clustering and genetic algorithm to find route optimization for the Ahmedabad city dataset

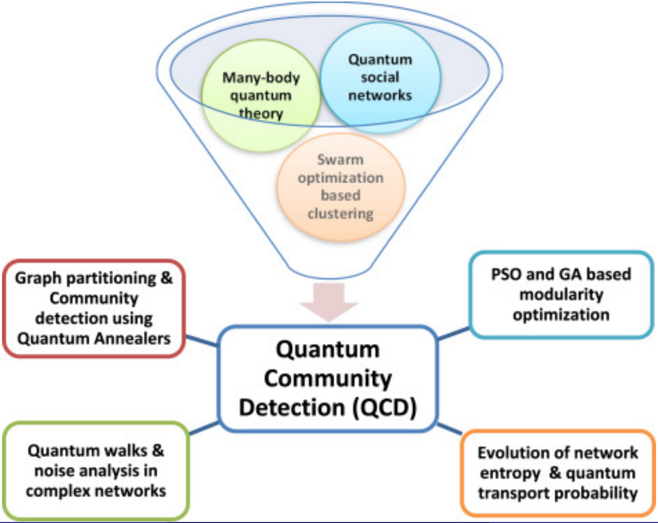
Research Opportunities

- The brain network topology is expected to be responsive to cognitive performance, behavioral variability, experimental task, and neurological disorders such as epilepsy, Alzheimer's disease, multiple sclerosis, autism, and attention-deficit/hyperactivity disorder.
- However, graph analysis in human neuroscience faces a number of issues that remain unaddressed, restricting its interpretation and application
- Some examples are heterogeneity of the results, sensitivity to parcellation strategy and node specification, statistical variability of brain graphs due to noise, lack of attention to the structure-function relationship, neglecting the variations in network density and connection strength, and dynamics of the brain network.
- Addressing any of these limitations in future studies will help advance our understanding of functional neural networks in the human brain.

Research opportunities in Maritime Network Analysis



Research opportunities in Quantum SNA paving way for QCD[1]



References I

[1] Sana Akbar and Sri Khetwat Saritha.
Towards quantum computing based community detection.
Computer Science Review, 38:100313, 2020.

[2] Neda Binesh and Mansoor Rezghi.
Fuzzy clustering in community detection based on nonnegative matrix factorization with two novel evaluation criteria.
Applied Soft Computing, 69:689–703, 2018.

[3] S Brin and L Page.
The anatomy of a large scale hypertextual Web search engine.
Computer Networks and ISDN Systems, 30(1/7):107–117, 1998.

[4] Joe Celko.
Graph Databases.
2014.

[5] R Lopez de Mantaras and L Saitia.
Comparing conceptual, divisive and agglomerative clustering for learning taxonomies from text.
In *16th european conference on artificial intelligence conference proceedings*, volume 110, page 435, 2004.

[6] Matthias Dehmer, Frank Emmert-Streib, Stefan Pickl, and Andreas Holzinger.
Big Data of Complex Networks.
Chapman & Hall/CRC, 2016.

[7] Xuhui Fan, Richard Yi Da Xu, Longbing Cao, and Yin Song.
Learning nonparametric relational models by conjugately incorporating node information in a network.
IEEE transactions on cybernetics, 47(3):589–599, 2016.

References II

[8] Rafael Ferreira, Luciano de Souza Cabral, Rafael Dueire Lins, Gabriel Pereira e Silva, Fred Freitas, George DC Cavalcanti, Rinaldo Lima, Steven J Simske, and Luciano Favaro.
Assessing sentence scoring techniques for extractive text summarization.
Expert systems with applications, 40(14):5755–5764, 2013.

[9] Javier O. Garcia, Arian Ashourvan, Sarah Muldoon, Jean M. Vettel, and Danielle S. Bassett.
Applications of Community Detection Techniques to Brain Graphs: Algorithmic Considerations and Implications for Neural Function.
Proceedings of the IEEE, 106(5):846–867, 2018.

[10] Yan Ge, Haiping Lu, and Pan Peng.
Mixed-Order Spectral Clustering for Networks.
1(c):1–12.

[11] Christos Giatsidis, Fragkiskos D. Malliaros, and Michalis Vazirgiannis.
Advanced graph mining for community evaluation in social networks and the web.
Proceedings of the sixth ACM international conference on Web search and data mining - WSDM '13, page 771, 2013.

[12] Jiawei Han, M Kamber, and J Pei.
Chapter 9: Graph Mining, Social Network Analysis, and Multirelational Data Mining.
Data mining: concepts and techniques, pages 535–589, 2006.

[13] Yue Huang and Xuedong Gao.
Clustering on heterogeneous networks.
Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(3):213–233, 2014.

References III

[14] Amir Kamil.
Graph Algorithms.
pages 1–7, 2003.

[15] Mingming Li and Jing Liu.
A link clustering based memetic algorithm for overlapping community detection.
Physica A: Statistical Mechanics and its Applications, 503:410–423, 2018.

[16] Grzegorz Malewicz, Matthew H Austern, Aart J C Bik, James C Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski.
Pregel: a system for large-scale graph processing.
Proceedings of the 2010 international conference on Management of data - SIGMOD '10, pages 135–146, 2010.

[17] Fahim Mohammadi.
A New Approach To Focused Crawling : Combination of Text summarizing With Neural Networks and Vector Space Model.
2(3):31–36, 2013.

[18] Ardashir Mohammadzadeh and Erkan Kayacan.
A non-singleton type-2 fuzzy neural network with adaptive secondary membership for high dimensional applications.
Neurocomputing, 338:63–71, 2019.

[19] Waqas Nawaz, Kifayat Khan, and Sungyoung Lee.
Intra graph clustering using collaborative similarity measure Intra Graph Clustering using Collaborative Similarity.
(January), 2015.

References IV

[20] M E J Newman.
Modularity and community structure in networks.
103(23):8577–8582, 2006.

[21] Antonio J Peña, Carlos Reaño, Federico Silla, Rafael Mayo, Enrique S Quintana-Ortí, and José Duato.
A complete and efficient CUDA-sharing solution for HPC clusters.
Parallel Computing, 40(10):574–588, 2014.

[22] Clara Pizzuti.
GA-Net : A Genetic Algorithm for Community Detection in Social Networks GA-Net : A Genetic Algorithm for Community Detection in Social Networks.
(January 2008), 2014.

[23] Matthew J Rattigan, Marc Maier, and David Jensen.
Graph clustering with network structure indices.
International Conference on Machine Learning, CI:783–790, 2007.

[24] Ronghua Shang, Jing Bai, Licheng Jiao, and Chao Jin.
Community detection based on modularity and an improved genetic algorithm.
Physica A, 392(5):1215–1231, 2013.

[25] Chuan Shi.
IN LARGE-SCALE COMPLEX NETWORKS.
13(1):3–17, 2010.

[26] Lei Tang and Huan Liu.
Community detection and mining in social media.
Synthesis lectures on data mining and knowledge discovery, 2(1):1–137, 2010.

References V

[27] Lei Tang and Huan Liu.
Graph mining applications to social network analysis.
In *Managing and Mining Graph Data*, pages 487–513. Springer, 2010.

[28] Chad Vicknair, M. Macias, Zhendong Zhao, Xiaofei Nan, Yixin Chen, and Dawn Wilkins.
A Comparison of a Graph Database and a Relational Database: A Data Provenance Perspective.
Proceedings of the 48th annual Southeast regional conference, pages 42–47, 2010.

[29] Yichuan Wang, Lee Ann Kung, and Terry Anthony Byrd.
Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.
Technological Forecasting and Social Change, 126:3–13, 2018.

[30] Sai Wu, Weichao Ren, Chengchao Yu, Gang Chen, Dongxiang Zhang, and Jingbo Zhu.
Personal Recommendation Using Deep Recurrent Neural Networks in NetEase.

[31] Xiaowei Xu and Thomas A J Schweiger.
SCAN : A Structural Clustering Algorithm for Networks.
pages 824–833, 2007.

[32] Chuantao Yin, Shuaibing Zhu, Hui Chen, Bingxue Zhang, and Bertrand David.
A Method for Community Detection of Complex Networks Based on Hierarchical Clustering.
2015, 2015.

[33] Hong Fang Zhou, Jin Li, Jun Huai Li, Fa Cun Zhang, and Ying An Cui.
A graph clustering method for community detection in complex networks.
Physica A: Statistical Mechanics and its Applications, 469:551–562, 2017.

[34] Xu Zhou, Yanheng Liu, Jian Wang, and Chun Li.
A density based link clustering algorithm for overlapping community detection in networks.
Physica A: Statistical Mechanics and its Applications, 486:65–78, 2017.

References VI

[35] Yang Zhou, Hong Cheng, and Jeffrey Xu Yu.
Graph Clustering Based on Structural / Attribute Similarities.
Vldb, 2(1):718–729, 2009.

Thank You

smita.agrawal@nirmauni.ac.in