Pranav Nerurkar

Personal Data

Place and Date of Birth:	Rennes, France | 24 April 1991
Phone:	+91 9619997797 / +91 7021743410
email:	panerurkar_p16@ce.vjti.ac.in
Google Scholar:	Citations: 64, H-index: 3, i10-index: 2
Scopus:	Citations: 42, H-index: 3, i10-index: 2
Web of Science:	Citations: 17, H-index: 2, i10-index: 1

Research Publications in Journals (Publications-7)

1. Securing Logistics System and Supply Chain using Blockchain (Communicated)

   Applied Stochastic Models in Business and Industry(ASMBI)

   Abstract Past international trade practices have been associated with opaque information flows that have hindered traceability and created hurdles in hassle-free trade. Blockchain and allied technologies have been investigated as a panacea for the problems faced by supply chain and logistics industry. However, earlier works of literature have focused on limited aspects of a typical supply chain such as monitoring assets, securing traceability, data integrity negligence, and data access. To overcome such drawbacks, the current paper proposes permissioned Blockchain with relevant processes and functions to obtain a holistic framework for securing the supply chain and logistic operations. The efficacy of the proposed framework was demonstrated in the case study. It was found that critical loopholes in a current supply chain can be overcome using the proposed framework. Additionally, several outlines for future research are outlined.

2. Towards cough sound analysis using the internet of things and deep learning for pulmonary disease prediction (Communicated)

   Wiley Transactions on Emerging Telecommunications Technologies

   Abstracts Cough is a symptom in over a hundred respiratory diseases. The audio features in cough signals contain erudition about the predicament of the respiratory system. Using deep learning or signal processing, these features can be used to build an effective disease prediction system. However, cough analysis remains an area that has received scant attention from machine learning researchers. This can be attributed to several factors such as inefficient ancillary systems, high expenses in obtaining datasets, or difficulty in building classifiers. This paper categorized and reviewed the current progress on cough audio analysis for the classification of pulmonary diseases. It also explored potential future issues in research. Additionally, it proposed a model for the classification of ten serious pulmonary ailments commonly seen in Indian adolescents. The proposed model is evaluated against four existing state of the art techniques in the literature.

3. Investigations of Residual Graph Convolutional Network for Representation Learning (Communicated)

   Neurocomputing-https://www.journals.elsevier.com/neurocomputing

   Abstract A large number of critical real-world or artificial systems possesses network constitution. There is no straight forward method to analyze network data as knowl-

edge models are intended for simple euclidean data or grids. Network representations also suffered from other issues such as lack of independent data-points, computationally costly calculations for graph statistics, in-applicability of parallel or distributed algorithms and the curse of dimensionality. To address such challenges, the focus of this dissertation is on Latent Space Representations (L.S.R.) of networks, i.e., learning low dimension vector representations of network data. L.S.R. techniques have a data-driven mechanism for learning vector embedding of network data structures. Learning vector space embedding of graph-structured is comparable to mapping multifaceted data into low-dimensional geometries. The embedding procedure minimizes pitfalls linked with graph-structured data. Deep learning has found applications in L.S.R.. A drawback of deep learning architectures is the complexity involved in building deep graph neural network models. The backbone of deep graph neural networks is the graph convolution operation. G.C.N. based architectures have a small receptive field. Hence, to build deeper models with many layers of neighborhood aggregation without hampering training time and accuracy, the current paper investigates the effects of addition of a residual connection to the Graph Convolutional Network architecture. In the experimental setting, the performance of Residual G.C.N. compared to a standard G.C.N. for the parameter transitivity showed an improvement of $12 - 42\%$ on the data-sets.

4. Internet of Things based Pulmonary disease prediction through cough spectrograms (Communicated)

   Multimedia Tools and Applications-https://www.springer.com/journal/11042

   Abstract Cough is a manifestation of over a hundred respiratory infections. The audio features in cough signals and the spectrograms contain information about the state of the respiratory system. Using deep learning or signal processing, these features can be used to build an effective disease prediction system. However, cough analysis remains an area that has received scant attention from machine learning researchers. This can be attributed to several factors such as inefficient ancillary systems, high expenses in obtaining datasets, or difficulty in building classifiers. The current paper categorizes and reviews the progress on cough audio analysis, machine learning models, and the data collection techniques through IoT for the classification of pulmonary diseases. Furthermore, it proposes a Multi-layered Convolutional Neural Network (DCNN) for the classification of eight pulmonary diseases. The DCNN uses spectral features, cepstral coefficients, chroma features, and spectrograms from cough audio for training. To test the efficacy of the model, a comparative study with four baseline models was conducted on a dataset of 112 patients collected from a pediatric facility in India through a cloud server and wearable sensors. Results showed that the proposed model achieved a precision of 0.4 on the test partition, which was analogous with recent models proposed in the literature surveyed.

5. Supervised Learning model for Identifying illegal activities in Bitcoin (Communicated)

   Applied Intelligence- https://www.springer.com/journal/10489

   Abstract Since its inception in 2009, Bitcoin has been mired in controversies for providing a haven for illegal activities. Several types of illicit users hide behind the blanket of anonymity. Uncovering these entities is key for forensic investigations. Current methods utilize machine learning for identifying these illicit entities. However, the existing approaches only focus on a limited category of illicit users. The current paper proposes to address the issue by implementing an ensemble of decision trees for supervised learning. More parameters allow the ensemble model to learn discriminating features that can categorize multiple groups of illicit users from licit users. To evaluate the model, a dataset of 1216 real-life entities on Bitcoin was extracted from the Blockchain. Nine Features were engineered to train the model for segregating 16

different licit-illicit categories of users. The proposed model provided a reliable tool for forensic study. Empirical evaluation of the proposed model vis-a-vis three existing benchmark models was performed to highlight its efficacy. Experiments showed that the specificity and sensitivity of the proposed model were comparable to other models. CPU and RAM utilization were also monitored to demonstrate the usefulness of the proposed work for real-world deployment.

6. Dissecting bitcoin blockchain: Empirical Analysis of Bitcoin network (2009-2020) (Communicated)

   Journal of Network and Computer Applications

   Abstract Bitcoin system (or Bitcoin) is a peer-to-peer and decentralized payment system that uses cryptocurrency named bitcoins (BTCs) and was released as open-source software in 2009. Unlike fiat currencies, there is no centralized authority or any statutory recognition, backing, or regulation for Bitcoin. All transactions are confirmed for validity by a network of volunteer nodes (miners) and after collective agreement is subsequently recorded into a distributed ledger "Blockchain". Bitcoin platform has attracted both social and anti-social elements. On the one hand, it is social as it ensures the exchange of value, maintaining trust in a cooperative, community-driven manner without the need for a trusted third party. At the same time, it is anti-social as it creates hurdles for law enforcement to trace suspicious transactions due to anonymity and privacy. To understand how the social and anti-social tendencies in the user base of Bitcoin affect its evolution, there is a need to analyze the Bitcoin system as a network. The current paper aims to explore the local topology and geometry of the Bitcoin network during its first decade of existence. Bitcoin transaction data from 03 Jan 2009 12:45:05 GMT to 08 May 2020 13:21:33 GMT was processed for this purpose to build a Bitcoin user graph. The characteristics, local and global network properties of the user's graph were analyzed at ten intervals between 2009-2020 with a gap of one year. Small diameter, skewed distribution of transactions, power-law distributed in and out degrees, disconnected graph, and presence of large connected components were the observations from network analysis. Thus, it could be inferred that despite anti-social tendencies, Bitcoin network shared similarities with other complex networks.

7. Multilabel classification of remote sensed satellite imagery. Wiley Emerging Telecommunication Transactions (2020) (Accepted)

   DOI:https://doi.org/10.1002/ett.3988

   Scopus Citescore 2020:1.56

   Abstract Multilabel scene classification has emerged as a critical research area in the domain of remote sensing. Contemporary classification models primarily emphasis on a single object or multiobject scene classification of satellite remote sensed images. These classification models rely on feature engineering from images, deep learning, or transfer learning. Comparatively, multilabel scene classification of very high resolution (V.H.R.) images is a fairly unexplored domain of research. Models trained for single label scene classification are unsuitable for the application of recognizing multiple objects in a single remotely sensed V.H.R. satellite image. To overcome this research gap, the current inquiry proposes to fine tune the state of the art convolutional neural network (C.N.N.) architectures for multilabel scene classification. The proposed approach pre trains C.N.N on the ImageNet dataset and further fine tunes them to the task of detecting multiple objects in V.H.R. images. To understand the efficacy of this approach, the final models are applied on a V.H.R. dataset: the U.C.M.E.R.C.E.D. image dataset containing 21 different terrestrial land use categories with a submeter

resolution. The performance of the final models is compared with graph convolutional network based model by Khan et al. From the results on performance metrics, it was observed that proposed models achieve comparable results in significantly fewer epochs.

8. Survey of network embedding techniques for social networks. Turkish Journal of Electrical Engineering & Computer Sciences 27.6 (2019).

   DOI:doi:10.3906/elk-1807-333

   Scopus Citescore 2020:0.86

   Abstract High dimensionality of data is a challenging scenario in the current era as the digital transformation of the society is in process. This problem is particularly complex in social networks as in such systems, it is coupled with other challenges such as interdependency of data points and heterogeneity of data sources. To overcome such disadvantages and aid in creation of downstream applications for social network analysis, network embedding techniques have been proposed. These techniques, in themselves, are not important but are the backbone of various network-based applications. Due to the scientific interest in this domain there has been a mushrooming of embedding techniques. It has therefore become crucial to learn the intuitions behind these techniques in order to compare and contrast them. The current analytical study is drawn with the following broad objectives: providing practitioners with understanding of network representative learning mathematical study of state-of-the-art techniques and highlighting the evolution of the literature in this field.

9. Exploring convolutional auto-encoders for representation learning on networks. Computer Science 20.3 (2019).

   DOI:https://doi.org/10.7494/csci.2019.20.3.3167

   Scopus Citescore 2020:0.34

   Abstract A multitude of important real-world or synthetic systems possess network structure. Extending learning techniques such as neural networks to process such non-euclidean data is therefore an important direction for machine learning research. However, till very recently this domain has received comparatively low levels of attention. There is no straight forward application of machine learning to network data as machine learning tools are designed for iid data, simple euclidean data or grids. To address this challenge the technical focus of this dissertation is on use of graph neural networks for Network Representation Learning (NRL) ie learning vector representations of nodes in networks. Learning vector embeddings of graph-structured data is similar to embedding complex data into low-dimensional geometries. After the embedding process is completed, drawbacks associated with graph structured data are overcome. The current inquiry proposes two deep learning auto-encoder based approaches for generating node embeddings. The drawbacks in existing auto-encoder approaches such as shallow architectures and excessive parameters are tackled in the proposed architectures using fully convolutional layers. Extensive experiments are performed on publicly available benchmark network data-sets to highlight the validity of this approach.

10. Understanding attribute and social circle correlation in social networks. Turkish Journal of Electrical Engineering & Computer Sciences 27.2 (2019).

Scopus Citescore 2020:0.86

Abstract Social circles, groups, lists, etc. are functionalities that allow users of online social network (OSN) platforms to manually organize their social media contacts. However, this facility provided by OSNs has not received appreciation from users due to the tedious nature of the task of organizing the ones that are only contacted periodically. In view of the numerous benefits of this functionality, it may be advantageous to investigate measures that lead to enhancements in its efficacy by allowing for automatic creation of customized groups of users (social circles, groups, lists, etc). The field of study for this purpose, i.e. creating coarse-grained descriptions from data, consists of two families of techniques, community discovery and clustering. These approaches are infeasible for the purpose of automation of social circle creation as they fail on social networks. A reason for this failure could be lack of knowledge of the global structure of the social network or the sparsity that exists in data from social networking websites. As individuals do in real life, OSN clients dependably attempt to broaden their groups of contacts in order to fulfill different social demands. This means that 'homophily' would exist among OSN users and prove useful in the task of social circle detection. Based on this intuition, the current inquiry is focused on understanding 'homophily' and its role in the process of social circle formation. Extensive experiments are performed on egocentric networks (ego is user, alters are friends) extracted from prominent OSNs like Facebook, Twitter, and Google+. The results of these experiments are used to propose a unified framework: feature extraction for social circles discovery (FESC). FESC detects social circles by jointly modeling ego-net topology and attributes of alters. The performance of FESC is compared with standard benchmark frameworks using metrics like edit distance, modularity, and running time to highlight its efficacy.

11. Measurement of network-based and random meetings in social networks. Turkish Journal of Electrical Engineering & Computer Sciences 27.2 (2019).

DOI:10.3906/elk-1806-103

Scopus Citescore 2020:0.86

Abstract Social networks are created by the underlying behavior of the actors involved in them. Each actor has interactions with other actors in the network and these interactions decide whether a social relationship should develop between them. Such interactions may occur due to meeting processes such as chance-based meetings or network-based (choice) meetings. Depending upon which of these two types of interactions plays a greater role in creation of links, a social network shall evolve accordingly. This evolution shall result in the social network obtaining a suitable structure and certain unique features. The aim of this work is to determine the relative ratio of the meeting processes that exist between different actors in a social network and their importance in understanding the procedure of network formation. This is achieved by selecting a suitable network genesis model. For this purpose, different models for network genesis are discussed in detail and their differences are highlighted through experimental results. Network genesis models are compared and contrasted with other approaches available in the literature, such as simulation-based models and block models. Performance measures to compare the results of the network genesis models with baselines are statistics of networks recreated using the models. The socially generated networks studied here belong to various domains like e-commerce, electoral processes, social networking websites, peer to peer file-sharing websites, and Internet graphs. The insights obtained after analyzing these datasets by network genesis models are used for prescribing measures that could ensure continuous growth of these social networks and improve the benefits for the actors involved in them.

12. Understanding structure and behavior of systems: a network perspective. International Journal of Information Technology (2019). (Springer)

DOI:https://doi.org/10.1007/s41870-019-00354-2

Abstract Networks are interesting representation models for analysis of systems. The entities of the systems under review can be denoted as the nodes of the networks and the relationships between these entities as the edges connecting them. Such a representation has advantages in analysis as network theory has a rich collection of well defined concepts and methods. These concepts of can be applied on such networks to draw inferences about the systems. As digitization has penetrated almost all aspects of mankind, a wide variety of systems from diverse domains such as computer science, transportation, social science have become available in the form of networks. A network perspective provides valuable insights into their structure and behavior. In this inquiry networks representing real world systems from different domains are analyzed using concepts of network theory and statistical generative network models—SBM and LSM. This is done to various application scenarios to express the properties of these systems. The findings highlight the unique features and trends seen in each domain.

13. Empirical analysis of synthetic and real networks. International Journal of Information Technology (2019). (Springer)

DOI:https://doi.org/10.1007/s41870-019-00344-4

Abstract With increasing digitization a wide variety of systems from diverse domains such as computer science, transportation, social science have become available in the form of networks. It is argued that to understand complex systems a deep understanding of the networks behind them is needed. A network theoretic perspective provides valuable insights into the structure and trends of systems. Data-sets belonging to different domains have their own unique features and behavioural trends and the current inquiry aims to highlight this. In this inquiry, a comprehensive analysis of synthetic and real-world published benchmark data-sets, evaluation methods, and open source projects is performed. The aim is to provide novice and expert users with tools for algorithmic designs and methodologies. Empirical studies are used to compare the performance of network theoretic tools on common data-sets. Finally, limitations of the network perspective on systems are listed and research directions to facilitate future study are elaborated.

Research Publications in Conferences (Publications-7)

1. Empirical analysis of data clustering algorithms. Procedia Computer Science, 125, 770-779. Elsevier. (2018)

DOI:https://doi.org/10.1016/j.procs.2017.12.099

Scopus Citescore 2020:1.48

Abstract Clustering is performed to get insights into the data whose volume makes it problematic for analysis by humans. Due to this, clustering algorithms have emerged as meta learning tools for performing exploratory data analysis. A Cluster is defined as a set of objects which have a higher degree of similarity to each other compared to objects not in the same set. However there is ambiguity regarding a suitable similarity

metric for clustering. Multiple measures have been proposed related to quantifying similarity such as euclidean distance, density in data space etc. making clustering a multi-objective optimization problem. In this paper, different clustering approaches are studied from the theoretical perspective to understand their relevance in context of massive data-sets and empirically these have been tested on artificial benchmarks to highlight their strengths and weaknesses.

2. A novel heuristic for evolutionary clustering. Procedia Computer Science, 125, 780-789. Elsevier. (2018)

DOI:https://doi.org/10.1016/j.procs.2017.12.100

Scopus Citescore 2020:1.48

Abstract Clustering is considered a challenging problem of data mining due to its unsupervised nature. The literature is inundated with algorithms and concepts related to determining the most suitable clustering structure in data. These techniques have a mathematical model of a cluster and attempt to obtain a result that shall represent this model as closely as possible. However as the problem of clustering is NP hard such strategies have disadvantages such as converging to local optima or suffering from the curse of dimensionality. In such scenario, meta heuristics could be more suitable strategies. Such techniques utilizes biologically inspired techniques such as swarm intelligence, evolution etc. to traverse the search space. Due to their inherent parallel nature, they are most robust towards converging to a local optima. The objective (cost) function used by such meta heuristics is responsible for guiding the agents of the swarm towards the best solution. Hence it should be designed to achieve trade-off between multiple objectives and constraints and at the same time produce relevant clustering. In this paper, a cost function is proposed (PSO-2) to produce compact well separated clusters by using the concept of intra-cluster and inter-cluster distances. Experiments have been performed on artificial benchmark data-sets where performance of the particle swarm optimizer using the proposed cost function is evaluated against other evolutionary and non evolutionary algorithms. The clustering structures produced by the methods have been evaluated using distance based and internal cluster validation metrics to demonstrate that the performance of PSO-2 is comparable to other techniques.

3. A comparative analysis of community detection algorithms on social networks. In Computational Intelligence: Theories, Applications and Future Directions-Volume I (pp. 287-298). Springer, Singapore. (2019)

DOI:https://doi.org/10.1007/978-981-13-1132-1_23

Abstract Social networks display community structures of interest, which have to be uncovered to understand the latent relationships present in them. These networks are composed of loosely connected small clusters, whose structure is more convenient for analysis. Graph clustering algorithms have been developed to identify communities in real or artificial networks using structural characteristics of the network. Determining the efficiency of such techniques with respect to accuracy and computational time is an open question in the absence of ground truth labels for the underlying communities as is seen in most real-world situations. In this study, performance of eight state-of-the-art graph clustering algorithms are demonstrated on small egocentric graphs, obtained from Facebook. The results are used for objective evaluation and a critical discussion is presented.

4. Performance of internal cluster validations measures for evolutionary clustering. In Computing, Communication and Signal Processing (pp. 305-312). Springer. (2019)

   DOI:https://doi.org/10.1007/978-981-13-1513-8_32

   Abstract Clustering is an NP-hard grouping problem and thus there are advantages of using a metaheuristic (swarm intelligence) strategy to find the near global optimal solution to it. To effectively guide the agents of the swarm in the metaheuristic strategy, a suitable cost function is needed for successful outcome. The current inquiry focuses on the use of internal validation criteria as cost functions as they achieve the dual goals of clustering which are compactness and separation. Out of the multiple internal validation criteria included in the literature, two are identified for this purpose, viz. BetaCV and Dunn index. These were used as cost functions of the swarm optimizer metaheuristic (PSO-BCV and PSO-Dunn). To demonstrate the validity of the proposed technique, it was compared with other metaheuristics differential evolution as well as the traditional swarm optimizer based on distance-based criteria (PSO). The analysis of the results obtained on clustering benchmark datasets highlighted the suitability of this approach.

5. Community detection using node attributes: A non-negative matrix factorization approach. In Computational Intelligence: Theories, Applications and Future Directions-Volume I (pp. 275-285). Springer. (2018)

   DOI:https://doi.org/10.1007/978-981-13-1132-1_22

   Abstract Community Detection uses graph topology and ignores node attributes while decomposing the network into coarse-grained descriptions. Popular algorithms that proliferate in the literature highlight important aspects of the network by detecting modules either by using the modularity maximizing approach or through information theoretic approaches. Each of these techniques has a different optimization criteria and objective function, and hence lead to different community structures. In the past few years, a hybrid category of algorithms were proposed that jointly model network topology and node attributes to detect communities in the network. There are significant challenges to this line of study as additional information has to be factored in but at the same time efficiency constraints of time and space have to be respected. In this paper, a variant of the BIGCLAM model is implemented for detecting communities in graphs. It uses the existing conceptual framework of Affiliate Graph Models and modifies it to consider attributes and not community affiliations as the basis for creating a bipartite graph for partitioning. A mathematical model of this novel approach is provided as well as experimental results on data have been presented to justify the use of this technique

6. Analysis of probabilistic models for influence ranking in social networks. In Computing, Communication and Signal Processing (pp. 215-223). Springer. (2018)

   DOI:https://doi.org/10.1007/978-981-13-1513-8_23

   Abstract Influence is a phenomenon occurring in every social network. Network science literature on Influence ranking focuses on investigation and design of computational models for ranking of nodes by their influence and mapping the spread of their influence in the network. In addition to this contemporary literature seeks efficient and scalable influence ranking techniques that could be suitable for application on massive social networks. For this purpose joint and conditional probabilistic models could be a way forward as these models can be trained on data rapidly making them ideal for deployment on massive social networks. However identification of suitable predictors that may have a correlation with influence plays a major role in deciding the successful outcome for these models. The present investigation proceeds with the

intuition that interaction is positively correlated with influence. Furthermore, through extensive experimentation it identifies a joint probabilistic model and trains it on interaction characteristics on nodes of a social network for influence ranking. A qualitative analysis of these models is presented to highlight its suitability.

7. Representation learning for social networks using Homophily based Latent Space Model. In Proceedings of the International Conference on Omni-Layer Intelligent Systems (pp. 38-43). ACM. (2019)

DOI:https://doi.org/10.1145/3312614.3312627

Abstract Representing data in the form of a graph (network) is becoming an increasingly common approach for modelling complex systems. Graph representation models have thus emerged as a unified vocabulary across scientific domains. The advantage of graphs for visualization of systems is that graph theoretic literature can be used for their analysis. Therefore, systems from online social networking, economics, biology, internet, citation and e-commerce are being modelled as graphs (entities as nodes and relationships as edges). Inspite of many advantages in analysis and inferencing, there are several disadvantages associated with network models. To overcome these disadvantages, Network Representation Learning (NRL) has emerged as a popular solution.