Machine Learning for Social Network Analysis

Theory & Applications

Justin Baltazar, Lucy Bodtman, Manan Talwar

Introduction

- Machine Learning (ML)
  - Application of artificial intelligence providing machines the ability to learn without explicitly being programmed.
- The Math behind Machine Learning
  - Extract a set of observable features from user information.
  - Interplay of Statistics (data analysis), Linear Algebra (data representation), Probability (predictions), and Calculus (optimization).
- Graph Based Machine Learning
  - Data Clustering
  - Deep Learning Neural Networks

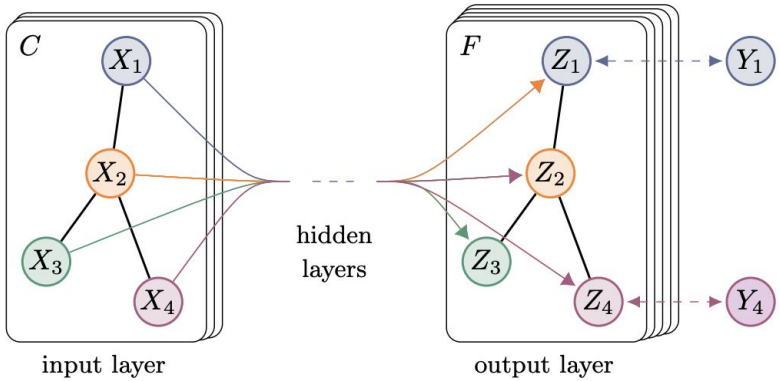


Fig 1: An example of Graph Neural Network

Images courtesy Math Works ©

Introduction

- Social network analysis (SNA)
  - The process of investigating social structures through the use of networks and graph theory.
- Applications of SNA
  - Information spread on social networks
  - Friend and business networks
  - Disease transmission modeling
  - Data aggregation and mining
- Research Questions
  - How is Machine Learning used as a tool in large scale Social Network Analysis?
  - How are ML algorithms designed for large scale and online Social Network Analysis?

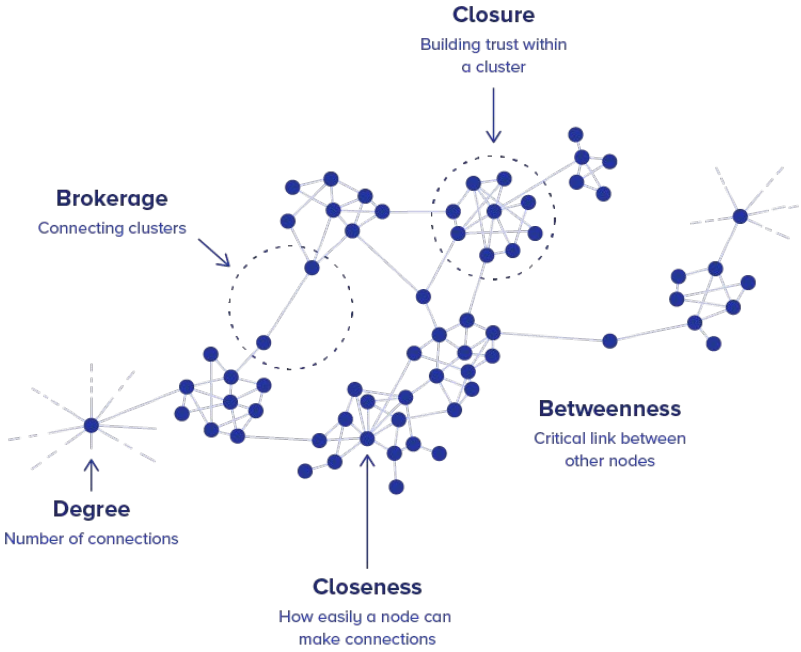
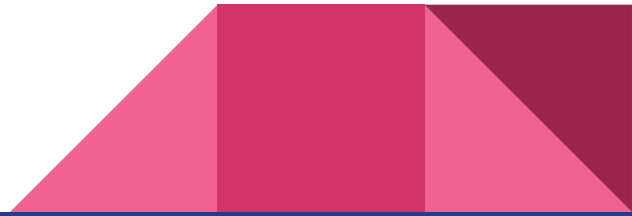


Fig 2: Use of Graph characteristics in SNA

Image courtesy KnowledgeBrief ©

Aims

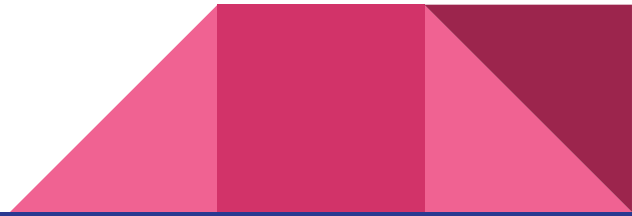
- To provide a basic understanding of Machine Learning for Social Network Analysis.
- To show how Machine Learning is relevant for Social Network Analysis, especially in the domain of online social network data analytics.
- To discuss the scope and applications of Machine Learning in Social Network Analysis.



Background

- Prior work

  Classical approaches such as regression and classification are not sufficient for modeling social network characteristics. **Why?**
  - Graph based inputs
  - Big Data analytics is computationally expensive
  - Dimensionality and Space
- Current state-of-the-art techniques
  - Encoder - Decoder Deep Learning
  - Network Embedding Deep Learning
  - Autoencoder Techniques
  - Graph Convolutional Techniques



Methods

For both our research questions, we conducted an extensive literature review.

- Source of Academic Literature
  - Google Scholar, IEEE and Frontiers in Big Data
- Inclusion Criteria
  - Empirical studies that reported validated statistically significant conclusions.
- Selection Process
  - We first performed keyword search using ‘Social Network Analysis’ and ‘Machine Learning’.
  - We manually reviewed the abstracts of the results for consideration.
  - We selected 19 journal articles that we read and assessed.
  - Based on the inclusion criteria, we included 6 in our review

Results

For our first research question, we identified the following potential **applications of Machine Learning in SNA:**

- Network embedding models can encode high dimensional data into low-dimensional representations without significant loss of information.
- Unsupervised learning models can be used for anomaly detection, data clustering and can detect features and trends that may not seem obvious.
- Deep learning models transform SNA into an optimization problem, which can be solved by gradient descent techniques, Newton-Raphson method and other convex optimization methods.

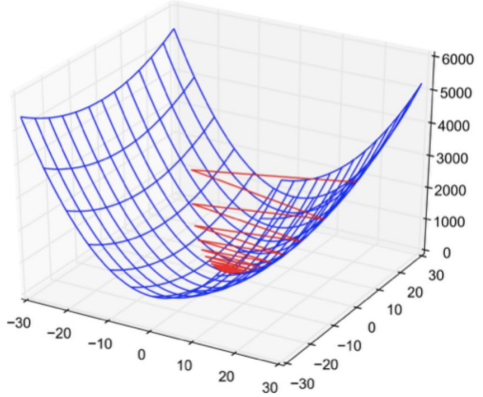


Fig 3: Convex Optimization in Deep Learning

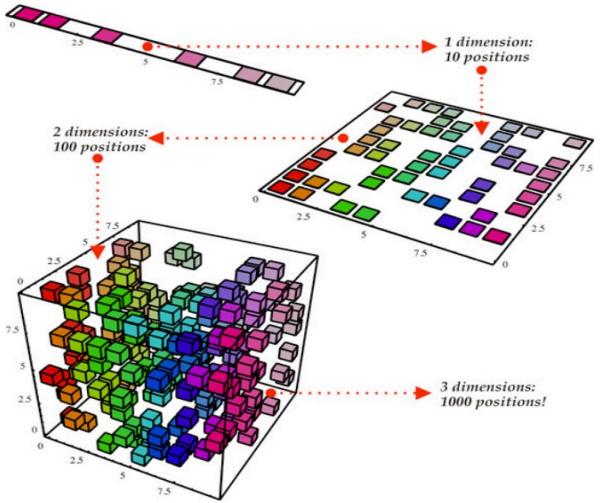


Fig 4: Dimensionality Reduction

Images courtesy Towards Data Science ©

Results

For our second research question, we identified the following as decisive parameters for **algorithmic techniques for ML based SNA**:

- The most commonly used deep learning methods convert network problems into convex optimization problems. An established solution is the the encoder - decoder framework to convert a graph into a Laplacian matrix, given by equation (1):

$$\min_{\Psi} \sum_{\phi \in \Phi_{tar}} \mathcal{L}(\psi_{dec}(\psi_{enc}(\mathcal{V}_{\phi})), \phi | \Psi) \qquad \textit{(Eq. 1)}$$

- A subset of deep learning models use autoencoders that generally have extremely large number of hidden layers to model network graphs. Gradient Descent can then be applied within the inner layers (significantly fewer dimensions) and aggregated at the output layer.

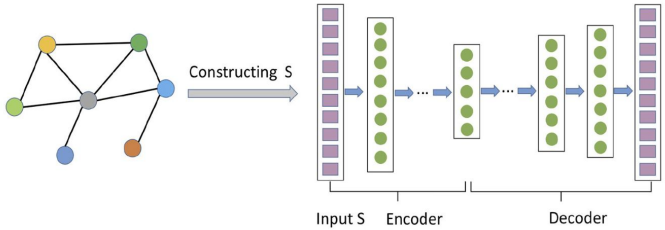


Fig 5: Autoencoder model

Images courtesy Frontiers in Big Data ©

Results

- Recent developments have proposed Graph Convolutional Neural Network based models as the way forward. Data is aggregated as node attributes by techniques such as near neighbour search. Gradient descent is then applied on the regularized multilayer perceptrons.

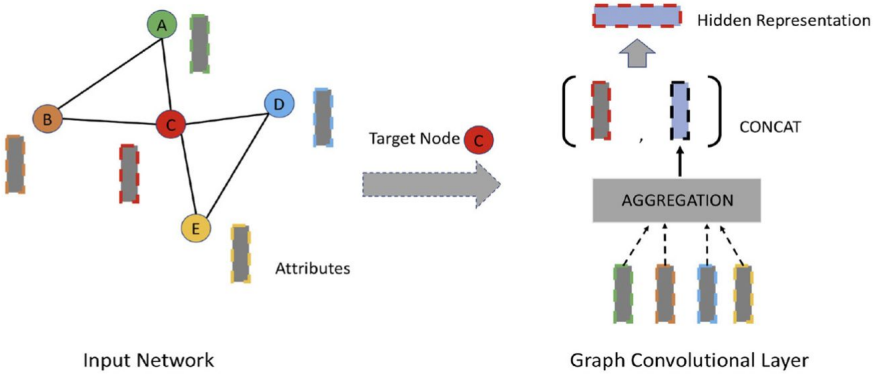


Fig 6: GNN Model

Images courtesy Frontiers in Big Data ©

Conclusion

- SNA is essential to understand the components and behaviour of of social networks.
  - Machine Learning is one approach in this analysis that has recently gained popularity.
- Machine Learning algorithms can be effective and efficient in SNA.
  - Embedding models can reduce data dimensionality by compression.
  - Deep Learning models can extract features and trends that may not seem obvious.
  - Computations can be carried out in parallel and in streams to reduce space complexity.
- Algorithm design for SNA is guided by principles of convex optimization.
  - Algorithms must convert graphs into laplacian matrices and/or other convex objectives.
  - Gradient descent and Newton - Raphson methods are key mathematical ideas behind the neural network architecture for SNA.

Acknowledgements

- Yuxin Ding (Harbin Institute of Technology)
- Shengli Yan (Harbin Institute of Technology)
- YiBin Zhang (Harbin Institute of Technology)
- Runa Ganguli (University of Calcutta)
- Qamar Rayees Khan (University of Calcutta)
- Akib Mohi Ud Din Khanday (Baba Ghulam Shah Badshah University)
- Qiaoyu Tan (Texas A&M University)
- Ninghao Liu (Texas A&M University)
- Xia Hu (Texas A&M University)

References

[1]

A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, “Identifying propaganda from online social networks during COVID-19 using machine learning techniques,” *Int. j. inf. tecnol.*, vol. 13, no. 1, pp. 115–122, Feb. 2021, doi: 10.1007/s41870-020-00550-5.

[2]

Y. Ding, S. Yan, Y. Zhang, W. Dai, and L. Dong, “Predicting the attributes of social network users using a graph-based machine learning method,” *Computer Communications*, vol. 73, pp. 3–11, Jan. 2016, doi: 10.1016/j.comcom.2015.07.007.

[3]

R. Ganguli, A. Mehta, and S. Sen, “A Survey on Machine Learning Methodologies in Social Network Analysis,” in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Jun. 2020, pp. 484–489. doi: 10.1109/ICRITO48877.2020.9197984.

[4]

M. Nurek and R. Michalski, “Combining Machine Learning and Social Network Analysis to Reveal the Organizational Structures,” *Applied Sciences*, vol. 10, no. 5, p. 1699, Jan. 2020, doi: 10.3390/app10051699.

[5]

A. Sapountzi and K. E. Psannis, “Social networking data analysis tools & challenges,” *Future Generation Computer Systems*, vol. 86, pp. 893–913, Sep. 2018, doi: 10.1016/j.future.2016.10.019.

[6]

“Machine learning, explained,” *MIT Sloan*. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained (accessed Oct. 18, 2022).