Project Plan

Signe Djernis Olsen (s206759)

Motivation

Many of the problems we are facing today can be modelled as graphs, for example, Network Models are often used to describe inner relations among units, which could be a social network where the vertices are users and edges are connections between two users.[3]
The study of Graph Representation Learning has increased over the last years leading to a truly impressive rise and growth of the field. The field is still evolving, leading to new findings, understandings, and approaches. A quickly rising subset of Graph Representation Learning is the Generative Model. [2]
In this project, I want to explore one Generative model, the Latent Distance Model, which emphasises the geometric position of the vertices. The probability of an edge between two vertices depends on the position of the two vertices in an unobserved latent space. [1]. I will model the Latent Distance Model on a real social network, where the vertices will be users in a ”social space”. The motivation for the project is to see how well the Latent Distance Model performs on the dataset and if new findings or aspects could be made.

Overall goal

The project’s overall goal is to examine the Latent Distance Model and use the findings on the dataset Slashdot social network, which is data from a social news site where users can be either friends or enemies.[4]
The main goal is to implement a Latent Distance Model using Bayesian statistics, involving the estimation of the parameters from a sigmoid function applied to a Bernoulli distribution to predict the presence or absence of an edge in the network, with the addition of a Gaussian prior. The optimisation of parameters is intended to be achieved using gradient descent.
The implementation will be built from scratch to explore different important aspects, such as the dimension of the Latent Space, and the choice of adding a prior and starting positions for the vertices in the Latent Space, of the model.
Furthermore, I seek to evaluate the performance of the model on a small synthetic dataset, and then scale up to a bigger Social Network Dataset, which is the Slashdot social network containing almost 1 million users. I will evaluate based on how well the model captures the friendships between two users on Slashdot. I will analyse and discuss the results, leading to a thorough conclusion and

recommendations of what is important when analysing and programming Latent Distance Models on social networks.

Scientific questions

How does the choice of the dimension of the vertices affect the Latent Distance Model's complexity and performance? How well does a final model capture the network and provide accurate predictions for whether two users are friends on the social network Slashdot? How will the addition of a Gaussian prior affect the model's performance?

Litterateur

[1] Catherine A. Calder Anna L. Smith Dena M. Asta. "The Geometry of Continuous Latent Space Models for Network Data". I: (2017).
[2] William L. Hamilton. "Graph Representation Learning". I: Synthesis Lectures on Artificial Intelligence and Machine Learning 14.3 (), s. 1–159.
[3] Adrian E. Raftery Peter D. Hoff og Mark S. Handcock. "Latent Space Approaches to Social Network Analysis". I: Journal of the American Statistical Association (2002), s. 1090–1098. doi: 10.1198 /016214502388618906.
[4] Ryan A. Rossi og Nesreen K. Ahmed. "The Network Data Repository with Interactive Graph Analytics and Visualization". I: AAAI. 2015. url: http://networkrepository.com. 3

Project Timeline Gantt Chart:

Processes	Week 9 26. feb - 3. mar	Week 10 4. mar - 10. mar	Week 11 11. mar - 17. mar	Week 12 18. mar - 24. mar	Week 13 25. mar - 32. mar	Week 14 1. apr - 7. apr	Week 15 8. apr - 14. apr	Week 16 15. apr - 21. apr	Week 17 22. apr - 28. apr	Week 18 29. apr - 5. may	Week 19 6. may - 12. may	Week 20 13. may - 19. may	Week 21 20. may - 26. may	Week 22 27. may - 2. jun	Week 23 3. jun - 9. jun
Project Planning															
Project Plan															
Literature research															
Coding															
Implement the gradient method															
Implement a prior															
Implement the full model															
Test on real data															
Report writing															
Examine method															
Method section															
Analysis and results															
Proofreading															
Submit															