

UNIVERSITY OF
CAMBRIDGE

MATHEMATICS TRIPOS

Part III Essay

**Walking Deeper on
Dynamic Graphs**

December 10, 2019

Written by
JOSHUA SNYDER

Contents

1	Motivation	2
2	Summary	3
3	DeepWalk	4
4	Matters of Convergence	5
5	DeepWalk Sucks	5
6	Dynamic Deep Walking	5
7	Discussion of an application	5

1 Motivation

The motivation for studying nodes in graphs and their representations comes from the desire to understand networks of people or objects and their relations. The motivating question for this essay is

Question (Motivating Question). Given a large network of people how can we quantify the relationships between them?

The natural way to go about this is to let nodes represent persons and let edges between them represent connections from which we can induce some understanding of relationship or trust between two people. This task is difficult; even with a relatively small number of people it is not a task on which humans perform very well and the task rapidly becomes difficult as the number of nodes in the graph increases.

Many of the networks which we wish to study are dynamic; that is they change with time. In a large network it is common that small changes occur during each epoch of time that over time cause larger changes to the network structure. How do we quantify these changes and their impact on the relationships in the network without having to completely re-analyse the network after every epoch of time, losing the information that we already learned. A frequent example upon which we can draw similarity is that of social networks. A large social network has new users (nodes) being added and new relationships (edges) formed in each epoch of time, however in any given short time period the graph representing the social network does not change substantially. It would be both foolish and costly to re-analyse the graph after each epoch however most of the previous literature has focused on static graphs. In the latter half of this essay I will give an account of recent progress into the application of DeepWalk and similar algorithms to dynamic graphs.

2 Summary

There are three main objectives of this essay:

- Firstly the essay will give a mathematical outline of the DeepWalk algorithm and it's application to social representation learning. We will briefly discuss the benefits and drawbacks of the algorithm.
- Secondly we will look at an implementation of DeepWalk to dynamic graphs. The challenge here is to develop an unbiased representation of a graph at time $t+1$ given it's representation at time t , without re-analysing the entire network. This is a very important task since many networks, especially social ones, are constantly changing. However in any given epoch of time the network structure is unlikely to undergo dramatic change and so a computationally effective algorithm will not re-analyse the network at each step.
- Thirdly, we will exhibit an implementation of the outlined dynamic DeepWalk application to a social data set [which data set?] and give suggestions as to good applications of Dynamic DeepWalking. [talk here about the application once I have found one]

3 DeepWalk

This section gives an outline of the social representation learning algorithm DeepWalk, first introduced in the seminal paper DeepWalk: Online Learning of Social Representations by B. Perozzi et. al. [1]. The method proposed in this paper not only demonstrated performance improvements from previous methodologies but also motivated an entirely different approach. At the time of the paper being written, significant advancements were being made in natural language processing (NLP) and the idea of word embeddings was becoming popular through an embedding algorithm known as word2vec [2, 3]. DeepWalk implements this algorithm but replaces the idea of context of a word in a sentence with the context of a node in a random walk.

The original paper on DeepWalk is lacking in a mathematical underpinning and in this section we will define it mathematically. It is suggested that the reader is familiar with the concepts outlined in the paper by Perozzi et. al. prior to reading this (more) mathematical exposition.

Let G denote the

4 Matters of Convergence

5 DeepWalk Sucks

6 Dynamic Deep Walking

7 Discussion of an application

Essay Descriptor

Walking Deeper on Dynamic Graphs: Learning Latent Representations with Random Walks for Image Classification

In the era of big data, graph representation is a natural and powerful tool for representing big data in real-world problems [1],[2],[4]; some examples include data coming from medical records, social networks, recommendation systems and transport systems. A challenging question when using graph representation is how to learn latent representations on multi-label networks for several classification tasks, and a seminal algorithm for this is the DeepWalk technique using random walks [1].

We propose two questions for investigation in this essay. Firstly, we hope that students will develop a rigorous mathematical underpinning for the DeepWalk algorithm, in the spirit of convergence guarantees.

Secondly, we seek to investigate the connection of DeepWalk to dynamic graphs. Many realworld events are dynamic - for example, in a social network new users are constantly added- while most of the body of literature is based on the unrealistic assumption that the graph is static. From the learning point of view, this assumption has a negative impact in the computations, as the graph has to be re-learned each time that an instance changes. We also hope that students will also discuss some open questions that they find interesting.

Relevant Courses

Useful: Background knowledge in Machine Learning and Statistics is helpful, as is probability to the level of Part II Applied Probability. Some content from Part III Mixing Times of Markov Chains, on the long-time behaviour of random walks on graphs, may also be useful.

References

- [1] B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk: Online learning of social representations,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 03 2014.
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 2013.
- [3] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” 2013.