Report on Hyperpolic Community detection

October 18, 2019

1 Introduction

The objective of this repport is to give tools and insight to execute and run the current project. Learning embedding within the poincare model is a difficult task much more than euclidean usual learning. Optimization in hyperbolic space often leads to machine precision issues and tweeking very precisely hyperparameters. Furthermore, coordinate of vectors is meaningfull contrary to euclidean, thus most centered element often means most general concept or in our work a most central nodes (e.g nodes with high degree relatively to others). This last point is especially important when develloping a cost function using negative sampling, indeed sampling negatives nodes randomly will accentuate the initial behaviour while sampling nodes based on the degree will soften it.

In this document we will go through issues encountered, by showing examples of errors that can occures and behaviour depending on model tunning.

2 KMeans and EM algorithm within Poincaré model

2.1 Barycenter hyper-parameters

2.2 EM: Machine precision

What to do : never clamp unormalised wik

2.3 EM: Updating parameters

Updating μ : Because of the gradient descent (or ascent) used for updating μ,this step is subject to failure or taking more or less time to process. Let be the two initalisation possible to start the gradient descent:

- $b_k = \frac{1}{n} \sum_{i=1}^{n} x_i$
- $b_k = \frac{1}{\sum_{j=1}^{n} w_{jk}} \sum_{i=1}^{n} x_i w_{ik}$

Obviously the second initalisation will lead to make less iteration to reach the convergence.

3 Learning embeddings

3.1 Batch, Mini-batch, sum or mean

3.2 Negative sampling

3.3 Compareason of optimization methods

3.4 Going out of the ball border