

CS5232 Project Proposal

Antoine Francois Pascal Creux (A0123427M)

March 18th, 2014

1 Concept Study

Goodreads is a website launched in early 2007, which lets “people find and share books they love... [and] improve the process of reading and learning throughout the world.”[3]. Two months after Flixster[1], a website where users share and rate movies they watch, Goodreads let anyone record and mark the books they have read, and share with everyone their reaction after reading the last best-seller they’ve just devoured.

Besides, goodreads has developed a social network above this book sharing experience. This platform connects people to books and authors by share ratings and recommendations, and thus gather people together. Readers see what their friends have read, but they can also meet new people sharing the same literature taste. In 2015, Goodreads platform gathers 34 million reviews from 30 million members on 900 million books[3].

2 Problem Description

Even if goodreads already has already developed a social network within their website (You can have a friendship relation with another reader), at first we would like to generate, study and compare all hidden networks we can generate based on book read, book reviews and authors followers. Based for instance on a book-user network which is bipartite, we would like to produce a user-user network. Then, we want to highlight any difference between those, or strengthen relations between readers if those networks share similar patterns. In order to carry out this experiment, we will make use of the best suited clustering algorithms such as K-nearest in our study. Finally, based on our generated graph, and if we have enough time to implement it, we can study a user’s influence over its friends. Indeed, we can see when a user has read a book, and then, see if one user of its circle may have recommended him a book.

3 Underlying Assumptions

4 Requirements

Goodreads data is not available online, we have to acquire all data through their APIs. Stanford students have already studied Goodreads platform by building a product recommendation. [2]. However, they have not published the data they gathered and we think that their data acquisition is wrongly implemented based on erroneous assumptions.

5 Project Objectives

6 Proposed Contributions

7 Success Measures

8 Project Plan

9 Deliverables

References

[1] Flixter. <https://www.flixter.com>, 2007.

[2] Book Recommendations on GoodReads.com. <http://cs229.stanford.edu/proj2008/IsaacsonSebastian-GoodReadsRecommendations.pdf>, 2008.

[3] goodreads - About us. <https://www.goodreads.com/about/us>, 2012.