

Аннотация к курсовой работе
Выделение пересекающихся сообществ во взвешенных графах

Студента группы м15МОС
Славнова Константина Анатольевича.
Москва, 2016.

В данной работе будет рассмотрена задача выделения структуры сообществ — набора таких групп вершин графа, которые плотно связаны между собой, но не с остальными вершинами. На текущий момент известно множество подходов и методов для выделения непересекающихся сообществ как в обычных графах, так и в взвешенном случае. Гораздо меньше внимания уделено проблеме пересекающейся структуре. В данной работе будет рассматриваться проблема в еще более конкретной и наименее изученной постановке: на взвешенных графах с пересекающимися группами вершин. В работе будет предложен новый метод решения проблемы. Алгоритм основан на модели BigClam, которая разработана для случая невзвешенных графов. Можно сказать, что новый метод является обобщением модели BigClam на взвешенный случай.

Работа начинается с постановки задачи и подробного описания метода BigClam. Задача выделения пересекающихся сообществ сводится и рассматривается как проблема неотрицательного матричного разложения с произвольным функционалом качества. Особое внимание уделено методу инициализации. Данный вопрос был подробно изучен, после чего предлагается несколько способов улучшить оригинальный подход. Новые способы инициализации протестированы на реальных и модельных данных и показали прирост как по начальному, так и по конечному значению функционала, а так же по времени сходимости метода.

Далее в работе речь пойдет о подходах обобщения метода на случай взвешенного графа. Рассматривается самое простое, наивное обобщение BigClam, описываются его недостатки. Предлагается перейти к более сложной модели, которая является центральным результатом данной работы. Заканчивается работа экспериментами на модельных данных и сравнением с другими методами решения задачи.

В заключении сформулированы основные выводы, указаны направления дальнейшего исследования. Полученные результаты говорят о том, что метод работает на уровне современных алгоритмов решающих подобную задачу, но не лучше их.

In this work Overlapping Community Detection problem will be considered. A community is typically thought of as a group of nodes with more connections amongst its members than between its members and the remainder of the network. Now it's known a big variety of methods for disjoint community detection problem in weighted and simple case. Much less attention is given to overlapping community detection problem. In this work, the problem will be considered in an even more specific and less studied case: on weighted graphs with overlapping vertex groups. A new method of solving the problem will be presented. The algorithm is based on BigClam model, which was originally developed for unweighted graphs. It can be said that the new method is a generalization of BigClam model in weighted case.

The work begins with a problem statement, and a detailed description of BigClam method. The task is reduced to a non-negative matrix factorization problem with arbitrary quality function. Initialization method has been considered in detail and suggested several ways to improve the original approach. New initialization ways tested on real and simulated data and showed an increase in both the initial and final quality function value, as well as on the method convergence time.

Generalization approaches of the method to the case of weighted graph will be discussed next. The most simple, naive BigClam generalization is considered, and its shortcomings are described. It is proposed to move to more complex model, which is the central result of this work. The experiments on simulated data and comparison with other methods are presented at the end of work.

Finally the basic conclusions and areas for further research are given. The experiments showed that the proposed method works at the level of modern algorithms for this problem, but not better.