

Адаптивный рандомизированный алгоритм выделения сообществ в графах

Тимофей Проданов

Доброе утро, я Тимофей Проданов и я буду рассказывать об адаптивном выделении сообществ в графах.

Слайд 1

Последние двадцать лет развивается изучение сложных сетей, то есть графов с неправильной, сложной структурой. Сложные сети были с успехом применены в различных областях, например в биоинформатике и социологии. Было показано, что такие сети, построенные по реальным системам часто имеют характеристики, усложняющие анализ классической теорией графов или статистический анализ.

Слайд 2
Сложные сети

Сложные сети часто разбиваются на группы тесно связанных узлов графа, сообщества. Способность находить и анализировать подобные группы узлов даёт большие возможности в изучении реальных систем, например, сообщества в интернете представляют распространённые темы.

Винсент Блондель в 2008 году разбил на сообщества два миллиона абонентов бельгийской телефонной компании. На рисунке размер круга отвечает за размер сообщества, а цвет за основной язык — французский или голландский. Заметно, как зелёные и красные сообщества распределились по разным сторонам сети, а в центральных сообществах цвета оказались смешанными.

Слайд 3
Блондель

В 2004 году Марк Ньюман представил целевую функцию модулярность, показывающую качество разбиения графа на сообщества.

Слайд 4
Алгоритмы

Эффективными оказались рандомизированные алгоритмы максимизации модулярности. В 2010 году Овельённе и Гейер-Шульц предложили рандомизированный жадный алгоритм выделения сообществ с параметром k . На каждой итерации рассматривается k случайных сообществ и их соседей, и затем соединяется лучшая пара.

В 2012 году те же учёные победили на конкурсе DIMACS со схемой кластеризации основных групп графа, которая заключается в следующем: Сначала s начальных алгоритмов выделяют сообщества, а те узлы, которые начальные алгоритмы поместили в разные сообщества, затем разбивает по сообществам финальный алгоритм. В итеративной схеме узлы, относительно которых начальные алгоритмы разошлись во мнении, снова распределяют по сообществам начальные алгоритмы.

Однако не существует набора параметров, при которых эти алгоритмы хорошо выделяют сообщества в каждом графе.

Слайд 5
Постановка задачи

Для решения этой проблемы в работе рассматривается создание адаптивных версий алгоритмов, то есть приспособляющихся к входным данным. И целью является создание алгоритма, дающего хорошие результаты на большем количестве графов.

Для создания адаптивных алгоритмов применяется стохастический градиентный спуск к рандомизированному жадному алгоритму и к схеме кластеризации основных групп графа.

Слайд 6
SPSA

Алгоритм *SPSA* предполагает разбиение улучшаемого алгоритма на два n шагов. В течение нечётного шага алгоритм использует один параметр, а во время чётного — другой, после чего подбираются следующие два значения параметра.

Предлагается адаптивный рандомизированный жадный алгоритм, почти идентичный рандомизированному жадному алгоритму с параметром k . Однако действие ARG разбито на шаги длиной в σ итераций, после окончания шага параметр k подстраивается.

Слайд 7
 ARG

Для получения следующих двух параметров k некоторая текущая оценка параметра возмущается в обе стороны. Следующая оценка будет ближе к тому возмущению, которое дало лучший результат. Оценка результата зависит от k и от медианы прироста модулярности за σ шагов.

В отличие от RG , имевшего один параметр, ARG имеет пять параметров, однако от чувствительности и количества итераций в шаге результат зависит слабо, а размер возмущения и начальная оценка имеют значения, дающие хороший результат на всех графах. Существует параметр, отвечающий за значимость времени, при его увеличении время работы уменьшается, однако при уменьшении модулярности.

Слайд 8
Параметры
 ARG

Для сравнения качества рандомизированного жадного алгоритма и его адаптивной версии сопоставлялась медианная модулярность разбиений графов с конкурса DIMACS. Левее пунктирной линии указаны RG с разными параметрами, правее — ARG с разными параметрами. Чем более зелёный цвет — тем разбиение лучше, и чем более красный — тем хуже. Видно, что RG даёт плохие результаты чаще, чем ARG . [RG_{10}]

Слайд 9
Сравнение
 RG и ARG

Представляется адаптивная схема кластеризации основных групп графа, схожая с неадаптивной схемой, но на первом этапе в качестве начальных алгоритмов используется RG с подстраиваемыми параметрами.

Слайд 10
 $ACGGC$

Затем в создании промежуточного разбиения участвуют не все начальные разбиения, но только некоторое количество лучших.

$ACGGC$ имеет несколько параметров, однако можно подобрать набор параметров, достаточно хорошо работающих на всех тестовых графах.

Слайд 11
Параметры
 $ACGGC$

Предложено два механизма снижения времени работы. Первый из них заключается в возможности увеличения параметра значимости времени, уменьшая время и модулярность.

Слайд 12
Снижение
времени работы

Второй механизм устанавливает ограничение на максимальную оценку параметра k , этот вариант более стабилен и как можно увидеть из тепловых карт — время всегда уменьшается при уменьшении максимальной оценки, и часто модулярности становятся выше при уменьшении максимальной оценки. Красный цвет тут обозначает не такие плохие разбиения, как на предыдущем рисунке.

Как и $CGGC$, $ACGGC$ можно итерировать. На таблице изображены сравнение итерационных схем кластеризации основных групп графа. В первых двух столбцах изображена адаптивная схема с разными параметрами, в третьей — неадаптивная и в последней — комбинация этих двух итеративных схем. Как можно увидеть, неадаптивная итерационная схема редко давала лучшее разбиение, а также в отличие от остальных один раз приняла очень плохое значение.

Слайд 13
Сравнение
 $ACGGC_i$ и
 $CGGC_i$

В рамках работы были исследованы современные методы выделения сообществ в сложных сетях, предложен новый адаптивный рандомизированный жадный алгоритм, исследованы его параметры и произведено сравнение результатов с неадаптивным рандомизированным жадным алгоритмом. Адаптивный алгоритм даёт более стабильные результаты и хорошо работает на большем количестве входных графов.

Слайд 14
Результаты

Также представлена адаптивная схема кластеризации основных групп графа, проанализированы её параметры, предложены механизмы снижения времени работы и результаты сопоставлены с результатами неадаптивной схемы. На тестовых графах в большинстве случаев адаптивная схема дала лучшие результаты и ни разу не дала очень плохого разбиения на сообщества.

Были рассмотрены адаптивная итеративная схема кластеризации основных групп графа и комбинирование адаптивной и неадаптивной схемы кластеризации основных групп графа.