План

1)основные понятия

2)алгоритм лейдена

3)сравнение алгоритмов

Понятия и обозначения

Пусть граф с узлами и ребрами. Разбиение состоит из сообществ, где каждое сообщество Ci Vсостоит из множества узлов , для любых . Для двух множеств S и R будем использовать следующие обозначения:

Функция качества или сообщества графа . Нам необходимо найти разбиение с максимально возможным качесвтом. Основываясь на разделении , граф Gможет быть агрегирован в новый граф . Граф называется базовым, граф . называется агрегированным. Узлы агрегированного графа являются сообществами в разбиении базового графа , то есть . Ребра агрегированного графа являются многогранниками. Количество ребер между двумя узлами агрегированного графа равняется числу ребер между узлами в двух соответствующих сообществах базового гафа . Следовательно,где мультимножество. Функция качества должна обладать свойством: , где обозначает одноэлементное разбиение агрегированного графа . Это гарантирует, что функция качества дает одинаковые результаты для базового и агрегированного графа.

Обозначим разделение, которое получается когда мы начинаем из разделения и перемещаем узел в сообщество . Обозначим изменение функции качества после перемещения узла в сообщество: . Другими словами: .. Аналогично, обозначим изменение функции качества после перемещения множества узлов в сообщество: . Пустое сообщество обозначим . Следовательно, изменение функции качества после перемещения множества узлов в пустое, т.е. новое, сообщество: .

Теперь рассмотрим сообщество , состоящее из двух частей S1 и S2,таких что и . Предположим, что S1 и S2 не связаны. Другими словами, между узлами из S1 и S2 нет ребер. Нам требуется функция качества, обладающая свойствами: и . Это гарантирует, что разделение всегда может быть улучшено, путем разделения сообщества на его связные компоненты.

Определим мощность множества : , где если s не является множеством. Операция сглаживания множества определяется как: , где если s не является множеством. Такое множество называется сглаженным.

Операция сглаживания для разбиения определяется , как:

Следовательноопределяет операцию в которой каждое сообщество сглажено. Такое сообщество называется сглаженным. Для любого разделения агрегированного графа, эквивалентное разделение базового графа может быть получено путем применения операции уплощения.

Обозначим как подграф, образованный сообществом, т.е. и . Сообщество называется связным, если связный граф. И наоборот, сообщество называется несвязным, если несвязный граф.

Модульность

Одним из способов нахождения сообществ является модульность. Этот метод максимизирует разницу между фактическим и ожидаем количеством ребер в сообществе C. Оценка качества поиска ожидаемого числа ребер выражается функцией: , где сумма степеней вершин в сообществе С, m количество ребер в сообществе, , параметр разрешения. Алгоритмы, рассматриваемые в данной работе используют такую функцию оценки, как CPM(Constant Potts Model). : , где количество узлов в сообществе С. Такая модель имеет преимущество по сравнению с модульностью, поскольку не обладает проблемой передела параметра разрешения Данный параметр разрешения выполняет роль порога: сообщества должны иметь плотность не ниже , а между сообществами плотность должна быть ниже . Более высокой значение приводит к большему количеству сообществ и наоборот.

Алгоритм Лувенского

Данный алгоритм основан на оптимизации функции качества, а именно модульности. Обычно алгоритм начинается с одноэлементного раздела, т.е. каждый узел является сообществом(однако можно запустить алгоритм и с другим разделение. Например, чтобы получить лучший результат, можно выполнить несколько последовательных итераций алгоритма и, используя разделения на сообщества из предыдущих итераций.) и включает в себя 2 фазы:

1. Локальное перемещение узлов, в котором отдельные узлы перемещаются в сообщество, дающее наибольшее увеличение функции качества
2. Агрегация сети. На данном этапе каждое полученное сообщество, построенное на этапе 1 становится узлом

Алгоритм работает до тех пор, пока функция качества не может увеличиться.

Алгоритм Лувенского может находить и “плохо” связанные сообщества. В частности, данный алгорим может находить несвязные сообщества. Т.е. одна часть такого внутренне разъединенного сообщества может достигнуть другой только по пути, выходящему за пределы сообщества. Узел, являющийся мостом между различными компонентами сообщества может быть перемещен в другое сообщество. Удаление такого узла приводит к появлению несвязного сообщества. Можно ожидать, что другие узлы старого сообщества также будут перемещены в другие сообщества. Однако это не всегда так, так как оставшиеся узлы все еще могут быть достаточно сильно связаны между собой несмотря на то, что сообщество стало несвязным. В этом заключается недостаток алгоритма Лувенского: он может находить сколько угодно слабо связанных сообществ. Данный алгоритм дает гарантию раздела на сообщества, для которых гарантируется невозможность объединения в сообщество. Другими словами, сообщества гарантированно хорошо разделены. Более сильные гарантии можно получить, повторяя алгоритм, т.е. используя предыдущие разбиения в качестве отправной точки следующей итерации. Таким образом качество разделения будет увеличиваться до тех пор, пока алгоритм не сможет вносить дальнейшие изменения. На этом этапе гарантируется оптимальное назначение каждого отдельного узла, т.е. алгоритм обеспечивает два варианта гарантии: 1) никакие сообщества не могут быть объединены; 2) никакие узлы не могут быть перемещены. Однако повторение алгоритма усугубляет проблему слабо связанных сообщетсв.Это можно объяснить так: после первой итерации алгоритма было получено некоторое разбиение. Далее на первом шаге следующей итерации алгоритм снова переместит отдельные узлы. Некоторые из узлов могут быть мостами между узлами сообщества и перемещая их появляются слабо связанные сообщества. Более того алгоритм не обладает механизмом фиксации таких сообществ. Таким образом повторение алгоритма с одной стороны улучшает разделение сообществ, а с другой ухудшает его.

Алгоритм Лейдена

В отличие от предыдущего алгоритма, алгоритм Лейдена гарантирует, что сообщества хорошо связаны. Он частично основан на алгоритме Лувенского и является ее усовершенствованной версией. Данный алгоритм состоит из трех фаз: 1) локальное перемещение узлов; 2) уточнение разделения;3) агрегация сети, основанная на основе уточненного разделения с использованием неуточненного разделения для создания потенциального разделения агрегированной сети. Идея фазы уточнения состоит в том, чтоб создать разделение , являющиеся уточнением разделения . Сообщества в могут быть разделены на несколько подсообществ в . Агрегированная сеть создается на основе разделения , однако начальное разделение для агрегированной сети основано на . Создавая совокупную сеть на основе , алгоритм имеет больше возможностей для идентификации высококачественных связей.

Уточненное разделение получается следующим образом: Изначально настроен на одноэллементный раздел, в котором каждый узел находится в своем сообществе. Затем алгоритм локально объединяет узлы в : узлы которые сами по себе находятся в сообществе в , могут быть объединены с другим сообществом. Важно отметить, что слияния выполняются только внутри каждого сообщества разделения . Кроме того, узел объединяется с сообществом в только в том случае, если оба они достаточно хорошо связаны со своим сообществом в . После завершения фазы уточнения сообщества в часто будут разделены на несколько сообществ в ,но не всегда.

На этапе уточнения узлы не обязательно жадно сливаются с сообществом, которое дает наибольший прирост функции качества. Вместо этого узел может быть объединен с любым сообществом, для которого функция качества увеличивается. Сообщество с которым объединяется узел выбирается случайным. Чем больше увеличение функции качества, тем больше вероятность того, что сообщество будет выбрано. Степень случайности выбора сообщества определяется параметром . Случайность выбора сообщества позволяет более широко исследовать пространство разбиения. Слияния узлов, приводящие к уменьшению функцию качества, не учитываются. Это делает этап уточнения более эффективным.

Еще одним важным отличием алгоритмов является реализация локальной движущейся фазы. В отличие от алгоритма Лувенского, алгоритм Лейдена на этом этапе использует процедуру быстрого локального перемещения. Алгоритм Лувенского продолжает посещать все узлы до тех пор, пока не перестанут перемещаться узлы, увеличивающие функцию качества, при этом он посещает и те узлы, которые нельзя перемещать в другое сообщество. В процедуре быстрого локального перемещения посещаются только узлы, окрестности которых изменились. Процедуру быстрого локального перемещения можно резюмировать следующим образом. Начнем с инициализации очереди со всеми узлами в сети. Узлы добавляются в очередь в случайном порядке. Затем мы удаляем первый узел из начала очереди и определяем, можно ли повысить функцию качества, переместив этот узел из его текущего сообщества в другое. Если мы перемещаем узел в другое сообщество, мы добавляем в конец очереди всех соседей узла, которые не принадлежат новому сообществу узла и еще не находятся в очереди. Мы продолжаем удалять узлы из начала очереди, возможно, перемещая эти узлы в другое сообщество. Это продолжается до тех пор, пока очередь не станет пустой.

После разделения данный алгоритм гарантирует, что: 1) все сообщества γ-разделимы; 2) все сообщества γ-связаны. В этих свойствах γ относится к параметру разрешения в оптимизированной функции качества, которая может быть либо модульностью, либо CPM. Свойство γ-разделения также гарантируется алгоритмом Лувенского. В нем говорится, что нет сообществ, которые можно объединить. Свойство γ-связности является несколько более сильным вариантом обычной связности. Как обсуждалось ранее, алгоритм Лувенского не гарантирует связность. Поэтому он также не гарантирует γ-связность.

Итерация алгоритма Лейдена, при котором разбиение не меняется, называется стабильной итерацией. После стабильной итерации алгоритм гарантирует, что: 1) все узлы локально оптимально назначены, т.е. нет отдельных узлов которые можно было бы переместить; 2) все сообщества являются γ-плотными по подразбиениям, т.е. сообщество можно разделить на две части так, что: 1) эит две части хорошо связаны; 2)ни одна из частей не может быть отделена от своего сообщества; 3)каждая часть сама также γ-плотна по подразбиениям. γ-плотностьподразбиения не означает, что отдельные узлы локально оптимально назначены. Это означает только то, что отдельные узлы хорошо связаны со своим сообществом.

В случае алгоритма Лувенского после стабильной итерации все последующие итерации также будут стабильными, значит дальнейшие улучшения невозможны. В отличие от алгоритма Лейдена, в котором после стабильной итерации могут быть дальнейшие улучшения в более поздних итерациях. Фактически, когда мы продолжаем повторять алгоритм Лейдена, он сойдется к разделу, для которого гарантируется, что: 1) все сообщества равномерно γ-плотны; 2) все сообщества оптимальны для подмножества.

Сообщество называется равномерно γ-плотным, если нет подмножеств сообщества, которые можно отделить от него. Равномерная γ-плотность означает, что независимо от того, как сообщество разделено на две части, эти две части всегда будут хорошо связаны друг с другом. Кроме того если все сообщества в разделении равномерно γ-плотны, качество раздела не слишком далеко от оптимального. Сообщество является оптимальным по подмножеству, если все подмножества сообщества локально оптимально назначены, т.е. никакое подмножество не может быть перемещено в другое сообщество. Оптимальность подмножества — самая надежная гарантия предоставляемая алгоритмом Лейдена, из нее вытекают γ-плотность и все остальные свойства.