БОДРОВ И.Ю., МОНАХОВ Ю.М., АБРАМОВ К.Г.

# АЛГОРИТМЫ ИССЛЕДОВАНИЯ СОЦИАЛЬНЫХ СЕТЕЙ.

Владимирский государственный университет им. А.Г. Н.Г. Столетовых

Сложность исследования состоит в невозможности выбрать подсеть, полностью состоящую из открытых узлов, достаточно большого диаметра, чтобы её можно было считать репрезентативной. Причина проблемы состоит в низкой вероятности нахождения такого подграфа в выборке даже большего размера и является следствием малого отношения количества открытых узлов к закрытым, а также случайности самой выборки.

В рамках научной работы был разработан алгоритм, позволяющий проводить вычисления с частично представительным подмножеством.

Первый этап заключается в выделении частично представительного подмножества обходом переформатированного исходного графа в ширину (Рисунок 1). Поскольку исходный формат данных представляет собой сильно разреженное множество открытых узлов и их соседей, было принято решение выделить связи закрытых узлов, явно не представленные в этом формате, с целью снижения разреженности, являющейся препятствием для точного вычисления коэффициента кластеризации. Переформатирование производилось с одновременной индексацией, поэтому время случайного доступа снижено, относительно исходного формата.

Обход осуществляется с поддержанием частичной консистентности (Рисунок 2). На каждом шаге алгоритма (уровне глубины обхода) считается, что дочерние узлы, еще не присутствующие на карте, являются закрытыми.

Для нахождения клики в произвольном множестве узлов необходимо чтобы минимум два узла являлись открытыми смежными и имели общий смежный узел. Для некоторого открытого узла можно посчитать количество образуемых при его участии клик, как мощность множества паросочетаний смежных ему узлов, включающего только пары также смежных узлов.

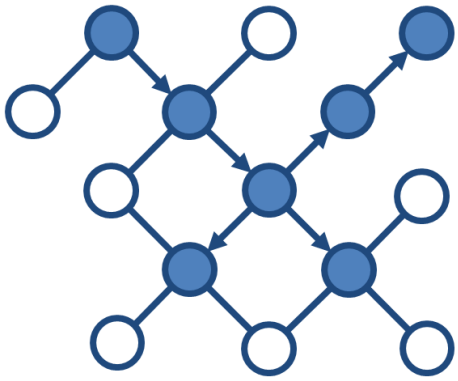

Рисунок 1 Обход в ширину

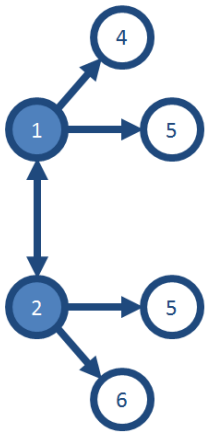


Рисунок 2 Частичная консистентность

Кроме того, для корректного и простого подсчета количества клик необходимо прибавлять к общей сумме 1, если клика образована парой открытых узлов и одним закрытым, и 0.5, если клика образована тремя открытыми узлами. Поскольку подсчет ведется для некоторого открытого узла, то перебор всех пар смежных ему узлов дважды учтет те пары, в которых оба узла открыты.

Описанный алгоритм учитывает только клики, образуемые одним или двумя открытыми соседями некоторого узла.

Предлагаемый алгоритм экстраполяции основан на предположении о стохастичности распределения образуемых клик между открытыми и закрытыми соседями случайно выбранного узла, следовательно, о том, что способность образовывать клики и состояние (открыт/закрыт) узла – независимы.

Эта способность характеризуется вероятностью, с которой рассматриваемый узел может образовать клику с парой смежных узлов, иными словами, кластерным коэффициентом.

Всего узел может образовать P(O)+P(C)+O*C клик, где P(O) - количество всех возможных паросочетаний его открытых соседей, P(C) – закрытых, O и C – количество открытых и закрытых соседей соответственно.

Поскольку приведенный выше алгоритм подсчитывает количество клик образованных парами открытых узлов, а также открытыми и закрытыми

узлами, и не учитывает только клики проходящие через пары закрытых соседей, вероятность образования клики можно вычислить как отношение количества найденных клик к количеству всех возможных паросочетаний, за вычетом количества закрытых пар: K = kliks / ( P(ALL) – P(C) ).

Тогда усредненное количество всех клик, образуемых некоторым узлом, можно вычислить как произведение вероятности образования клики на количество всех возможных паросочетаний узлов, смежных рассматриваемому: ExKliks = K * P(ALL).

В результате практического применения разработанного подхода было получено частично репрезентативное подмножество из 50 тысяч открытых узлов, для которого был рассчитан экстраполированный кластерный коэффициент, составивший ~ 0.166.

Как было сказано ранее, прямо связана с проблемой выделения достаточно большого репрезентативного подмножества, в то же время, высокая средняя степень связности сети, увеличивает объем обрабатываемых данных. Следовательно, необходимо распределить большой объем данных между вычислительными ресурсами.

Для организации наблюдения за состоянием сети, изменяющимся распределено, необходимо дифференцировать этапы моделирования и вводить условное квантование по времени.

Одним из основных подходов к организации распределенных вычислений является разбиение множества задач и исходных данных на непересекающиеся подмножества.

Множество единовременно инфицированных узлов полностью разбивается на непересекающиеся подмножества, однако множества смежных с ними узлов пересекаются с большой вероятностью. Поэтому, по окончанию каждого кванта моделирования, для каждого узла, смежного с теми, которые были затронуты моделированием в этот квант, должен быть агрегирован эффект от всех распределенных моделирующих процессов.

С другой стороны возможна конвейеризация процесса путем организации избыточной предвыборки, вероятно необходимых на следующем шаге данных.

Распределение непересекающихся подмножеств в начале каждого кванта и агрегация результатов моделирования по его окончанию обеспечивается специализированным процессом, который также может принять решение о неэффективности распределения текущего объема вычислений и провести их самостоятельно.

В каждый квант карта хранит только инфицированные узлы, соседи которых распределено заражаются на дочерних процессах. Сами инфицированные узлы иммунизируются в этот квант также распределено. Для всех узлов оставшихся инфицированными по окончании агрегации запускается новый квант, а результаты иммунизации сохраняются отдельно.

## Литература


1. Абрамов К.Г., Монахов Ю.М. Моделирование распространения нежелательной информации в социальных медиа // Труды XXX Всероссийской научно-технической конференции. Проблемы эффективности и безопасности функционирования сложных технических и информационных систем. Часть IV, секция №6. - Серпуховский ВИ РВ. - 2011. - 376 с.; - С. 178-182. - ISBN 978-5-91954-029-8
2. Ladislav Novak, Alan Gibbons, Hybrid graph theory and network analysis, Cambridge university press, 1999, 186 c., ISBN 0521461170
3. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. Российская Академия Наук. Институт проблем управления. М.: МЦНМО, ФИЗМАТЛИТ, 2010. - 228 с. ISBN 978-5-94057-669-3.

## Сведения об авторах

Бодров Иван Юрьевич, Владимирский государственный университет, магистрант кафедры ИЗИ, cantreg@gmail.com

Монахов Юрий Михайлович, Владимирский государственный университет, доцент кафедры ИЗИ, к.т.н., unklefck@gmail.com

Абрамов Константин Германович, Владимирский государственный университет, инженер кафедры ИЗИ, аспирант, abramovkostya@mail.ru