

XV Международная научно-практическая конференция студентов аспирантов и молодых учёных
«Молодёжь и современные информационные технологии»

СРАВНЕНИЕ СПОСОБОВ ПОИСКА ВЛИЯТЕЛЬНЫХ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНЫХ СЕТЕЙ В ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ

Замятина В. С., Лунева Е.Е.

Научный руководитель: Лунева Е.Е.

Томский политехнический университет

zamiatina.v@gmail.com

Введение

На данный момент существует несколько эффективных способов решения задачи KPP POS (Key players problem positive) [1, 2], к которой также относится задача выявления пользователей приложений работающих по принципу социальных сетей, влиятельных в заданной предметной области. Однако данные способы в основном применяются на невзвешенных неориентированных графах [1]. Применение таких способов к вышеупомянутой задаче требует адаптации этих способов к работе с ориентированными графами, в которых направление ребра показывает направление влияния одного пользователя на другого, а вес отражает степень этого влияния. Кроме этого, оценка эффективности данных способов на реальных данных осложнена отсутствием фактических данных, т.е. какой пользователь в заданной предметной области является наиболее влиятельным. Именно поэтому тестирование результатов выявления влиятельного пользователя целесообразно производить на лабораторных графах, где заранее известен узел, представляющий собой наиболее влиятельного пользователя.

Цель данной работы – сравнить известные способы [1, 2] применительно к задаче выявления влиятельных пользователей социальной сети, а также определить наиболее эффективный из них.

Теоретический обзор

В качестве тестируемых способов выявления влиятельных пользователей были взяты: метод с вычислением коммуникационной эффективности узлов, метод Боргатти, а также метод с вычислением информационной энтропии [1, 2]. Для этого в генерируемых лабораторных графах введем два типа узлов: узлы–«писатели» и узлы–«читатели». Узлы–«писатели» отображают пользователей социальной сети, которые делятся своим мнением относительно какой-либо темы, узлы–«читатели» представляют собой модель пользователей, читающих, комментирующих и публикующих записи пользователей–«писателей». Для упрощения модели ситуация, когда пользователи–«писатели» и пользователи–«читатели» могут меняться ролями, не учитывается. Поэтому найденные влиятельные пользователи должны будут принадлежать к числу «писателей». Ребра между графами направленные и имеют определенный вес, отображающий степень влияния одного пользователя на другого. Подробное построение модели социального графа, а также использование метода вычисления коммуникационной эффективности узлов описано в работе [1].

Метод с вычислением коммуникационной эффективности узлов

При тестировании данного метода на лабораторных графах обнаружился его существенный недостаток: в качестве наиболее влиятельных узлов могут выступать узлы–«читатели». Сам метод основан на выявлении узлов, при удалении которых нарушается распространение информации в сети, то есть понижается «коммуникационная эффективность». Вполне закономерно, что влиятельными пользователями в данном случае будут являться не только пользователи, которые представляют собой источник информации, но и пользователи, которые способствуют распространению этой информации.

В качестве примера рассмотрим следующий граф (рисунок 1):

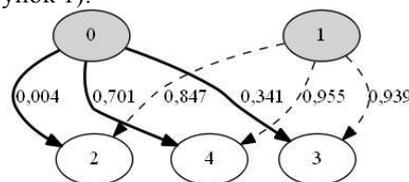


Рис. 1. Пример тестового графа сети

Здесь узлы 0 и 1 являются узлами–«писателями», а узлы 2,3,4 – узлами–«читателями». В данном примере наиболее влиятельным является узел 1. Значения коммуникационной эффективности при удалении из сети каждого из узлов представлены в таблице 1.

Таблица 1. Значения коммуникационной эффективности

Номер узла	Эффективность
0	3,38
1	251,68
2	6,54
3	250,97
4	252,49

На рисунке 2 представлено графическое отображение изменения коммуникационной эффективности при удалении того или иного узла.

Как видно из таблицы 1 и рисунка 2, минимальное значение коммуникационной эффективности наблюдается при удалении узла 0 из сети. В этой точке на графике можно наблюдать «провал» кривой.

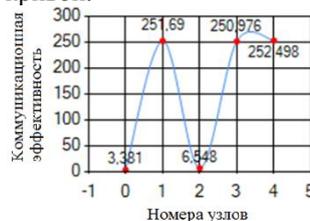


Рис. 2. Значения коммуникационной эффективности

Это означает, что узел 0 представляет собой модель наиболее влиятельного пользователя. Однако «провал» кривой, то есть минимум значения коммуникационной эффективности наблюдается также в узле 2, который является узлом-«читателем».

Метод с вычислением информационной энтропии

Ключевое понятие, используемое в данном подходе – энтропия. Энтропия – это показатель, позволяющий отобразить количество информации, которое может быть передано через канал передачи информации [2].

Значения энтропии каждого из узлов представлены в таблице 2.

Таблица 2. Значения энтропии узлов графа

Номер узла	Энтропия
0	18,59
1	81,40
2-4	0

На рисунке 3 представлено графическое отображение значений энтропии.

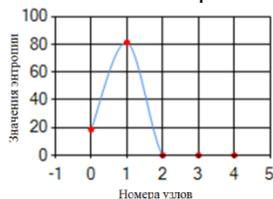


Рис. 3. Значения энтропии узлов графа

Как видно из таблицы 2 и рисунка 3, максимальное значение энтропии у узла 1. Несмотря на то, что узел 1 принадлежит к числу узлов-«писателей», в действительности этот узел не отображает наиболее влиятельного пользователя. Таким образом, метод вычисления информационной энтропии не позволяет найти наиболее точное решение поставленной задачи.

Метод Боргатти

Метод Боргатти основывается на вычислении центральности каждой вершины графа [4]. Результаты вычислений показателей центральности подробно описаны в работе [5]. На лабораторных графах метод Боргатти показал достаточно устойчивые результаты.

Результат выявления наиболее влиятельного пользователя методом Боргатти для графа, представленного на рисунке 1, графически отображен на рисунке 4. Можно увидеть, что данный метод также выбрал узел 0 в качестве наиболее влиятельного (на графике в данной точке наблюдается максимальное значение центральности вершины), второе место после него занимает узел 1.

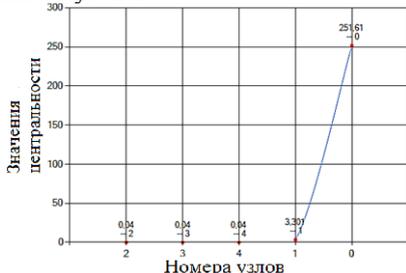


Рис. 4. Значения центральности узлов графа

Однако значительным недостатком данного метода является его ресурсозатратность. Значительное «отставание» по времени от других методов наблюдается уже для графа, состоящего из 80 узлов: 30 узлов-«читателей» и 50 узлов-«писателей». Среднее время поиска наиболее влиятельного пользователя для различных масштабов графов представлено в таблице 3.

Таблица 3. Время поиска наиболее влиятельного пользователя (выражено в мс)

Наименование метода / Количество узлов графа	0	80	100	180
Коммуникационная эффективность	2	1870	4718	32757
Энтропия	3	1876	4741	32793
Боргатти	3	1900	4786	33017

Заключение

Среди рассмотренных методов наиболее точное решение предоставляет метод Боргатти. Однако в отличие от остальных методов, данный метод задействует наибольшее количество вычислительных ресурсов. Этот недостаток является достаточно существенным, так как при росте объема обрабатываемой информации, значительно увеличится и время обработки этой информации. Подобного рода проблема может быть решена с помощью нейронной сети, которая будет обучена находить наиболее влиятельных пользователей методом Боргатти.

Список использованных источников

1. Ortiz-Arroyo D. Discovering Sets of Key Players in Social Networks // Computer Communications and Networks, Т. 1 – 2009, с. 27-47
2. Замятина В.С., Лунева Е.Е. Оценка влиятельности пользователей социальной сети в заданной предметной области// Молодёжь и современные информационные технологии, Т. 1 - 2017, С. 171-173
3. Информационная энтропия [Электронный ресурс]/Wikipedia.URL:https://ru.wikipedia.org/wiki/Информационная_энтропия, свободный. Яз. Рус.
4. Centrality [Электронный ресурс] / Wikipedia. URL: https://en.wikipedia.org/wiki/Centrality, свободный. Англ. Рус.
5. Stephen P Borgatti. On the Robustness of Centrality Measures// Under Conditions of Imperfect Data. – 2006. С 22-34.

Работа выполнена при финансовой поддержке РФФИ (проект №17-07-00034 А).