

УПРАВЛЕНИЕ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ

УДК 004.75, 519.876.5

Д. С. Донченко, Д. В. Завьялов

ПОДХОДЫ К МАТЕМАТИЧЕСКОЙ ОЦЕНКЕ ВЛИЯНИЯ СУБЪЕКТОВ СОЦИАЛЬНОГО ГРАФА НА ПРИМЕРЕ СОЦИАЛЬНОЙ СЕТИ TWITTER

Волгоградский государственный технический университет

dmitrydonchenko92@gmail.com, sinegordon@gmail.com

В статье проведен анализ подходов к оценке влияния субъектов социального графа. Предложен алгоритм для вычисления влиятельности субъектов социального графа и приведены результаты работы алгоритма на примере социального графа группы пользователей социальной сети Twitter.

Ключевые слова: социальный граф, социальные сети, поиск в ширину, анализ данных.

D. S. Donchenko, D. V. Zavyalov

APPROACHES TO THE MATHEMATICAL ASSESSMENT OF SUBJECTS IN SOCIAL GRAPH ON EXAMPLE OF TWITTER SOCIAL NETWORK

Volgograd State Technical University

The article analyzes the approaches to assessing the impact of the subjects of the social graph. The algorithm for calculating the influence of subjects of the social graph and the results of the algorithm on an example of the social graph group of users of social network Twitter.

Keywords: social graph, social networks, breadth-first search, data analysis.

Популярные социальные сети (Facebook, Twitter, VK и др.) за годы своего существования сумели стать важной частью нашего общества и превратились в мощный политический, экономический и социальный инструмент. В качестве модели социальной сети используется социальный граф. Под социальным графиком понимают граф, вершинами которого являются социальные объекты, в качестве которых могут выступать пользовательские профили, имеющие всевозможные атрибуты (такие, как имя, город проживания, дата рождения и т. д.), сообщества пользователей, медиа-контент (выложенные пользователями социальных сетей фотографии, видеозаписи, текстовые материалы и т. п.), а ребрами – отношения (социальные связи) между пользователями. Актуальной задачей является выявление наиболее влиятельных пользователей социальных сетей.

Однако, для ответа на вопрос о том, как следует определять наиболее влиятельных пользователей социальных сетей, сначала следует определиться с самим понятием влияния. Чем влиятельнее пользователь социальной сети, тем

в большей степени он определяет мнения других пользователей. Кроме того, влиятельность пользователя в значительной степени определяет возможность распространения этим пользователем определенной информации и количество аудитории, которое он способен охватить. Социальное влияние проявляется, когда человек приспосабливает свое поведение, отношение и мнение к определенным явлениям к поведению и отношениям других людей в социальной системе [1]. Влияние людей в обществе друг на друга было темой множества маркетинговых исследований начиная с 1960-х годов. Влияние не обязательно предполагает прямое взаимодействие между двумя людьми и, скорее, определяется на основе информации о других людях [2]. В онлайн-сообществах информация передается между пользователями сообществ в форме цифрового контента. Таким образом, мы можем сделать вывод о типе социального влияния, характерном для социальных сетей: мерой влиятельности пользователя социальной сети является количество пользователей, которые меняют свое поведение в социаль-

ной сети в ответ на изменения в поведении данного пользователя, а также частота, с которой они это делают.

В данной работе мы будем рассматривать в качестве модели социальной сети социальный граф, вершинами которого являются пользователи данной социальной сети, а ребрами – социальные связи между ними, выраженные в виде различных в зависимости от социальной сети способов взаимодействия между пользователями (примером таких взаимодействий для социальной сети Twitter служат репосты сообщений, размещаемых другими пользователями). Ребра социального графа являются ориентированными. Вес ребер характеризует интенсивность социальных связей между пользователями.

В качестве примера рассмотрим социальный граф, состоящий из двух вершин А и В, каждая из которых является представлением некоторых пользователей социальной сети. Факт взаимодействия пользователя А с профилем пользователя В или с предоставленным им цифровым контентом означает изменение в поведении пользователя А, следовательно, можно говорить о том, что пользователь В оказывает влияние на пользователя А. В модели социального графа это будет означать, что вершины графа, представляющие пользователей А и В, связаны ребром В->А, вес которого определяет силу влияния пользователя В на пользователя А.

Как уже было сказано, зачастую влияние не требует прямого взаимодействия между субъектами социального графа. Таким образом, при создании алгоритма для определения влиятельности субъекта социального графа важно учесть не только его прямые связи с другими субъектами, но также и интенсивность социальных взаимодействий субъектов, с которыми он связан, с другими.

Также при определении влиятельности субъектов социального графа следует учитывать актуальность совершенных социальных взаимодействий. В настоящее время большинство исследований рассматривают влияние на основе статических атрибутов, таких, как количество друзей/фолловеров в социальной сети, количество отправленных твитов/полученных ретвитов, упоминания/ответы в графе деятельности. Все данные исследования делают неявные предположения о лежащем в основе динамическом процессе, в котором пользователи социальной сети могут влиять только на своих друзей/фолловеров, с равной степенью влияния

и одинаковой скоростью распространения; в других, влияние предполагается принять за «случайное блуждание» на соответствующей статической сети [4]. В соответствии с теорией диффузии, информационные каскады исходят от общественных лидеров к их последователям. В большинство моделей диффузии, распространители имеют определенные вероятности оказать влияние на своих последователей, которые, в свою очередь, также имеют определенный порог, при котором они поддаются влиянию. Главной целью большинства диффузионных моделей является нахождение общественных лидеров, имеющих наибольшее влияние на своих последователей. Данные методы являются основой для т. н. вирусного маркетинга, целью которого является активация цепной реакции влияния и, таким образом, охват наибольшей аудитории при минимальной цене [5].

Резюмируя данные критерии, можно сказать, что задача вычисления влиятельности субъектов социального графа делится на два этапа. На первом этапе необходимо определить веса ребер социального графах [7]. Для этого необходимо не только произвести вычисления интенсивности социальных связей между соединяемыми ими вершинами, но также и учесть, насколько актуальными по времени и постоянными были данные взаимодействия, и на основе этого вычислить вероятность появления таких взаимодействий в будущем. Таким образом, после первого этапа мы получим социальный граф, вершины которого соединены направленными ребрами, каждое из которых имеет вес, определяющий степень влияния одной вершины на другую. На втором этапе необходимо произвести вычисление весов вершин графа, которые являются мерой влиятельности данных вершин в рамках всей системы. При этом, вес каждой вершины необходимо вычислять не только на основании весов исходящих из нее ребер, но также учитывая веса ребер, исходящих из вершин, связанных с «последователями» исходной вершины. При этом, сумма весов исходящих ребер вершин, связанных с «последователями» исходной вершины, в меньшей степени влияет на влияние, оказываемое исходной вершиной в рамках всей системы. Таким образом, предлагаемый алгоритм также должен учитывать фактор «затухания», который будет определять коэффициент для вычисления сумм весов ребер вершин, не состоящих в прямой связи с исходной.

Данные требования к алгоритму позволяют провести его параллель с известным алгоритмом PageRank. В PageRank также учитывается фактор «затухания» (Damping factor). PageRank предполагает существование «случайного посетителя» (random surfer), который оказался на определенной веб-странице и случайным образом переходит по ссылкам, расположенным на этой странице. Damping factor является вероятностью того, что случайному посетителю в определенный момент станет скучно, и он перейдет на другую веб-страницу. В различных исследованиях были протестированы различные значения Damping factor, но в большинстве случаев на практике значение этого фактора устанавливают около 0.85 [6].

В данном исследовании мы полагаем, что эффект затухания определяется как мера уменьшения влияния субъекта социального графа по отношению к субъектам, который не состоят с ним в прямых связях, но имеют общих знакомых, с которыми активно взаимодействуют. Предположим существование случайного пользователя социальной сети Twitter, который, оказавшись на некоторой странице профиля другого пользователя социальной сети, случайным образом выбирает страницы профилей пользователей, репост записей которых произвел текущий пользователь. Таким образом, эффект затухания – вероятность того, что случайный пользователь выберет перейдет на другой случайно выбранный профиль. По аналогии с алгоритмом PageRank возьмем данный эффект равным 0.85. Приведем составленный с использованием данных критериев алгоритм для вычисления весов вершин социального графа:

$$W_n = d^1 \sum_{i=1}^{i=N_1} W_{e_i} + d^2 \sum_{i=1}^{i=N_2} W_{e_i} + \dots + d^{C_{max}} \sum_{i=1}^{i=N_{C_{max}}} W_{e_i}, \quad (1)$$

где C – количество ребер между узлами (степень удаления узла); C_{max} – максимальная степень удаления узла; N_c – количество ребер, идущих из узла степени C ; W_n – вес узла; W_{e_k} – вес ребра; d – коэффициент затухания, равный 0.85; k – порядковый номер ребра (из общего числа N_c).

Для обхода графа используется алгоритм поиска в ширину. Структурой данных для хранения графа был выбран список смежных вершин, так как данная структура, в отличие от матрицы смежности, значительно снижает затраты по памяти при выполнении алгоритма.

Сложность алгоритма для каждой вершины в графе, таким образом, составляет $O(N+M)$, где N – число вершин графа; M – число ребер.

Таким образом, можно вычислить сложность алгоритма для всех вершин графа:

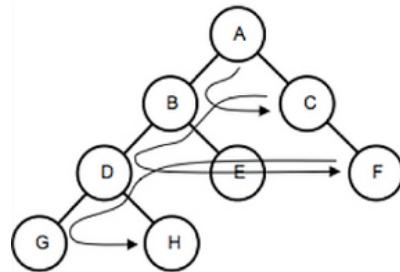
$$N * \sum_{i=0}^{i=N} N + Mi, \quad (2)$$

где N – число вершин графа; i – номер вершины; Mi – количество ребер, исходящих из i -й вершины.

Вследствие этого, при большом количестве вершин графа вычислительная сложность алгоритма будет достаточно высокой. При этом, так как алгоритм поиска в ширину является достаточно быстрым, исполняющимся за линейное время алгоритмом, если граф хранится как список смежных вершин, наилучшим решением для уменьшения времени вычисления алгоритма является его распараллеливание.

В результате этот алгоритм вычисляет вес каждой вершины графа, который является мерой измерения влияния пользователя, представленного данной вершиной. При этом, алгоритм учитывает не только влияние пользователя на пользователей, являющихся его друзьями, но также, в свою очередь, влияние этих пользователей на своих друзей и т. д.

Рассмотрим пример обхода графа в ширину с применением данного алгоритма:



Пример графа с указанием социальных связей

Обход изображенного на рисунке графа для вершины А:

1. А – стартовая вершина графа (узел с информацией о пользователе социальной сети)
2. Завести переменную – вес вершины А, задать ей значение, равное нулю
3. Посетить вершины, непосредственно соединенные с вершиной А.
4. Прибавить к весу вершины А сумму весов всех ребер, исходящих из вершин, которые с ней соединены(в данном примере этими ребрами являются АВ и АС), деленную на порядок удаленности этих вершины от стартовой вер-

шины графа. Вершины связаны с вершиной А непосредственно, поэтому порядок удаленности равен единице.

5. Посетить вершины, которые соединены с вершиной В(вершины D и E)

6. Прибавить к весу вершины А сумму весов всех ребер, исходящих из вершины В, (в данном примере этими ребрами являются BD и BE), деленную на порядок удаленности, равный двум для этих вершин.

7. Посетить вершину, которая соединена с вершиной С (эта вершина – F).

8. Прибавить к весу вершины А сумму весов всех ребер, исходящих из вершины С (ребра CF), деленную на порядок удаленности, равный 2 для этой вершины.

9. Посетить вершины, связанные с вершиной D(этими вершинами являются вершины G и H).

10. Прибавить к весу вершины А сумму весов всех ребер, исходящих из вершины D, (ребра DG и DH), деленную на порядок удаленности, равный 3 для этих вершин.

В качестве примера для тестирования данного подхода была использована социальная сеть Twitter. В качестве метрики для вычисления весов ребер было выбрано количество репостов, осуществленное пользователями дан-

ной социальной сети. Для построения социального графа было использовано API социальной сети Twitter.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Leenders, Roger Th. A. J. Modeling Social Influence Through Network Autocorrelation: Constructing the Weight Matrix // Social Networks, Vol. 24(1), 21–48 (2002).
2. Robins, G. Models for Social Influence Processes / G. Robins, P. Pattison, P. Elliot // Psychometrika, 66(2), 161–90 (2001).
3. Trusov, M. Determining influential users in Internet social networks / M. Trusov, A. Bodapati, R. Bucklin // Journal of Marketing Research Vol. XLVII(August 2010), 643–658.
4. Rogers, E.M. Diffusion of Innovations // Free Press, Vol. 27 (2003), 127–129.
5. Jingxuan, L. Wei, P. Tao, L. and Tong, S. (2013) Social Network User Influence Dynamics Prediction, APWeb 2013, LNCS 7808, pp. 310–322: Springer-Verlag, Berlin Heidelberg 2013.
6. Sergey Brin and Lawrence Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Science Department, Stanford University, Stanford, CA 94305, USA.
7. Укустов, С. С., Кравец, А. Г. Подход к решению задачи идентификации влиятельных разработчиков в социальной сети ГИТХАБ // Известия Волгоградского государственного технического университета. 2012. Т. 15. № 15(102). С. 61–66.
8. Кравец, А. Г., Ле Суан Куен. Протокол для сохранения конфиденциальности профиля в социальных сетях // Известия Волгоградского государственного технического университета. 2014. Т. 22. № 25(152).

УДК 656.13.052:004

M. N. Дятлов¹, О. А. Шабалина¹, Ю. Я. Комаров¹, Р. А. Кудрин²

РАЗРАБОТКА ТЕСТОВЫХ ЗАДАНИЙ ДЛЯ КОМПЬЮТЕРНОЙ ДИАГНОСТИКИ СТЕПЕНИ РАЗВИТИЯ СЕНСОМОТОРНЫХ РЕАКЦИЙ С УЧЕТОМ ОСОБЕННОСТЕЙ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ ВОДИТЕЛЕЙ

¹Волгоградский государственный технический университет

²Волгоградский государственный медицинский университет

makdyatlov@yandex.ru, o.a.shabalina@gmail.com

В статье рассмотрены основные требования, которые предъявляются в настоящее время к будущим водителям пассажирского автотранспорта на этапе профессионального отбора в России. Выделены наиболее значимые профессиональные качества водителей. Разработаны имитационные тестовые задания для аппаратно-программного комплекса диагностики психофизиологических качеств на этапе профессионального отбора водителей.

Ключевые слова: водитель пассажирского автотранспорта, профессионально-важное качество, имитационное тестовое задание, психофизиологическое качество, диагностика сенсомоторных качеств.

M. N. Dyatlov¹, O. A. Shabalina¹, Yu. Ya. Komarov¹, R. A. Kudrin²

DEVELOPMENT OF TEST TASKS FOR COMPUTER DIAGNOSTICS DEGREE OF DEVELOPMENT OF THE SENSOMOTOR REACTIONS TAKING INTO ACCOUNT FEATURES OF DRIVERS PROFESSIONAL PERFORMANCE

¹Volgograd State Technical University

²Volgograd State Medical University

The paper considers basic requirements for professional driver's license applicants in Russia. The most significant driver professional qualities are distinguished. A set of tests for hardware-software complex for psychophysiological qualities diagnostics is developed.

Keywords: passenger vehicle driver, professionally important skills, simulation test, psychophysiological feature, sensorimotor skills diagnostics.