

every year many IT professionals develop a series of forecasts that include new trends in cyber security, predicting new threats and the impact of modern information attacks on the activities of people, organizations and society as a whole.

УДК 004

Код РИНЦ 20.00.00

ИНСТРУМЕНТЫ И МЕТОДЫ АНАЛИЗА СОЦИАЛЬНЫХ СЕТЕЙ

© 2020 Ефимова Татьяна Борисовна

кандидат экономических наук, доцент

Самарский государственный экономический университет

E-mail: TB_Efimova@mail.ru

Ключевые слова: информация, социальные сети, методы, анализ, алгоритм Гирвана-Ньюмана, граф, вершина.

Статья посвящена методам и инструментам анализа социальных сетей, которые представляются в виде графов. Рассмотрены возможности применения различных методов для анализа, сделаны выводы об их достоинствах и недостатках.

В настоящее время роль социальных сетей значительна не только для традиционного общения коллег и единомышленников, но и для образования, и для бизнеса. Большой интерес представляют точки соприкосновения науки и социальных сетей - в частности, социальная сеть "Scitable", созданная в 2009 году в США, объединила ученых генетиков для более эффективного продвижения научного знания¹.

Социальные сети - это эффективный канал взаимодействия с целевой аудиторией, позволяющий получать быструю обратную связь. Через "лайки" и комментарии становится возможным определить реакцию клиентов на изменения политики компании². С точки зрения получения прибыли от продвижения товаров и услуг в социальных сетях важным является анализ деятельности подписчиков.

Существует много интересных научных исследований по вопросу сбора информации в социальных сетях для различных целей, и данная работа посвящена методам анализа графов сетей.

Многие исследователи предлагают описывать социальную сеть как граф. Действительно: люди выступают в качестве сущностей (вершин графа), а между ними имеется как минимум одна связь (ребро), кроме того, каждый участник социальной сети старается распределить своих друзей по группам (коллеги, друзья, близкие друзья и т.п.), то есть имеет место степень связи (метка). Кроме того, можно отметить присутствие локальности - наличие кластеров.

Рассматривая различные примеры сетей, можно выделить (помимо сетей "друзей") телефонную сеть ("ребро" возникает, если происходит телефонный

звонок в течение некоторого периода времени, а весом можно считать количество звонков за период), почтовую сеть (вершины графа - адреса, ребра - факты отправления писем), инфраструктурную сеть и многие другие.

Но можно выделить еще одну группу сетей: в них встречается более одного типа сущностей, и их принято описывать k -дольными графами.

Рассмотрим возможности применения кластеризации к изучаемому вопросу. В этом случае в качестве метрик можно использовать метки. Но если они в графе отсутствуют, то задача существенно усложняется: в ряде работ предлагается считать вершины "близкими" и пометить их "0", если они связаны ребром, и "1..n" - в противном случае, хотя здесь необходимо сделать ряд уточнений.

Существует две группы методов в рамках кластеризации: иерархические и опирающиеся на отнесение точек. При иерархической кластеризации происходит объединение вершин, у которых общее ребро, а затем - случайным образом выбирается ребро, не соединяющее вершины одного кластера, объединяются кластеры, которым принадлежат вершины данного ребра. Но данный метод чреват наличием неточностей и получением ошибочных результатов, так как можно выбрать ребра вершин, не относящихся к одному сообществу. Конечно, можно несколько раз провести кластеризацию и выбрать наиболее удачный вариант, но это непродуктивно. Такая же проблема возникает и при использовании метода отнесения точек.

Известны более специализированные методы, позволяющие обнаружить сообщества. В частности, алгоритм Гирвана-Ньюмана оперирует понятием "промежуточности" ребер: чем больше этот показатель, тем более вероятно, что участники принадлежат разным сообществам. Идея алгоритма состоит в том, что просматривается каждая вершина графа, для нее определяется количество кратчайших путей в другие вершины.

Конечно, данный метод достаточно информативен и пользуется популярностью, но существует ряд других, позволяющих, например, находить сообщества поиском подмножества вершин, которые соединяются большим количеством ребер. В этом случае можно попытаться найти большую клику (множество вершин, каждая пара которых соединяется ребром), но это крайне сложная в решении задача. Поэтому на первый план выходит метод использования двудольных графов для нахождения сообществ в социальной сети.

Чтобы отыскать такие графы, целесообразно разбить граф на левую и правую доли: левая будет содержать "предметы", а правая - "корзины" (и то, и другое - вершины). Элементы "корзины" - те "предметы", с которыми она соединена ребрами. Поэтому поиск можно рассматривать как поиск частых предметных наборов размера t с поддержкой s^3 .

Еще один подход к решению поставленной задачи - спектральный метод, предполагающий минимизировать число ребер, соединяющих разные компоненты. В этом случае нужно определиться с тем, каким образом выполнить "разрез" графа: он должен быть минимальным, при этом размеры разделяемых подмножеств должны оставаться примерно одинаковыми. Имеет место и явление

"нормализации" разреза, который получается, если взять его отношение к объемам получившихся двух подмножеств и сложить их. Данный подход вполне жизнеспособен, так как размер разреза мал, и деление на подмножества получается действительно равным.

Для получения "хорошего" разрезания используется матричная алгебра, а именно: матрица смежности, степенная матрица и матрица Лапласа, которая представляет собой разность между двумя предыдущими. Для нее необходимо определить наименьшие собственные значения и собственные векторы. Доказано, что наименьшее собственное значение равно нулю, а собственный вектор состоит только из единиц. Чтобы получить "хороший" разрез, можно поместить в одно множество вершины, соответствующие положительным элементам собственного вектора второго снизу по величине собственного значения, в другое - те вершины, которые соответствуют отрицательным элементам³.

Данный метод отличается высокой эффективностью, но необходимо помнить, что необязательно помещать в одну группу вершины, соответствующие отрицательным элементам вектора, а в другую - положительным, кроме того, данный метод можно применять при разрезании графа на большее число частей.

Рассматриваемые методики позволяют отыскивать сообщества, но в жизни сообщества пересекаются, и плотность ребер возрастает в пересечении сообществ.

Рассмотрим модель графа принадлежности: идея состоит в том, что если два человека находятся в одном сообществе, то определяется соответствующая вероятность этого события. Поэтому легко определить вероятности для всего графа. Кроме того, находим такое распределение членов по сообществам, при котором максимальное правдоподобие для этого распределения, наибольшее для любого распределения. Далее можно найти вероятности, ассоциированные с каждым сообществом (метод градиентного спуска)³. Несмотря на кажущуюся простоту и адекватность метода, найти наилучшую оценку членства в сообществах невозможно, так как членство - дискретная величина.

Решить эту проблему можно улучшением моделирования сообществ: если участник не является членом сообщества, то ему ставится в соответствии "0", а вероятность существования ребра определять как функцию от силы членства вершин в общих сообществах³.

Еще одна техника анализа графов социальных сетей - Simrank - заключается в предположении, что два объекта похожи, если они связаны с аналогичными объектами⁴.

Рассмотрим наиболее популярные сервисы, упрощающие работу в социальных сетях.

Popsters - это платный сервис для аналитики контента и интересов сообщества, работающий с ВКонтакте, Одноклассники, Instagram, Twitter, YouTube, Google+ и др. Он не собирает персональные данные, обеспечивает маркетинг в социальных сетях, позволяет анализировать конкурентов. Существующие

аналоги - LiveDune, Locowise, FeedSpy, PandaRank, Falcon, SmmBox, SmartMetrics и другие.

AgoraPulse позволяет посчитать вовлечение в социальную сеть.

Hootsuite позволяет анализировать социальную сеть.

Quintly помимо похожих функций позволяет находить лидера мнений.

NetBase используют крупные компании, данный сервис обеспечивает поддержку бизнес-решений.

Oktopost используется для B2B менеджмента, позволяет рассчитать экономическую эффективность маркетинга.

Social Mention - бесплатный сервис, позволяющий найти упоминание о бренде в социальных сетях.

SumAll можно использовать для получения графиков корреляции между показателями страниц бренда на различных платформах⁵.

Можно сделать ряд выводов о перспективах развития социальных сетей. Во-первых, получат распространение сервисы, созданные на их инфраструктуре. Во-вторых, получит дальнейшее развитие AR-контент. В-третьих, именно в соцсетях качество общения существенно повысится благодаря снижению "информационного шума".

¹ Павленко В.В. Социальные сети как фактор развития науки // ИНТЕЛПРОС - интеллектуальная Россия. 2013. № 4 (URL - http://www.intelpros.ru/readroom/credo_new/k4-2013/21500-socialnye-seti-kak-faktor-razvitiya-nauki.html).

² Анализ российского рынка фитнес-услуг / Ефремова М.В., Чкалова О.В., Бошман Т.К. // Экономический анализ: теория и практика. - 2015. - №21 (420). - С.25-37.

³ Лесковец Ю., Раджараман А., Ульман Д. Анализ больших наборов данных. - М.: ДМК Пресс, 2016. - 498 с.

⁴ <https://en.academic.ru/dic.nsf/enwiki/11373150>

⁵ <https://popsters.ru/blog/post/35-instrumentov-dlya-analitiki-socsetey>

TOOLS AND METHODS FOR ANALYZING SOCIAL NETWORKS

© 2020 Efimova Tatyana Borisovna
Candidate of Economics, Associate Professor
Samara State University of Economics
E-mail: TB_Efimova@mail.ru

Keywords: information, social networks, methods, analysis, Girvan-Newman algorithm, graph, vertex.

The article is devoted to methods and tools for analyzing social networks, which are presented in the form of graphs. The possibilities of using various methods for analysis are considered, conclusions about their advantages and disadvantages are made.