

АНАЛИЗ СОЦИАЛЬНЫХ СЕТЕЙ

В. С. Ковалевская

*Белорусский государственный университет, пр. Независимости, 4,
220030, г. Минск, Беларусь, varvajonok@gmail.by
Научный руководитель — Л. Л. Голубева, кандидат физико-математических наук,
доцент*

В эпоху цифровых технологий социальные сети превратились в одну из основных платформ для общения, развлечения, поиска информации и маркетинговых коммуникаций. Они предоставляют уникальную возможность для изучения поведения пользователей, анализа их предпочтений и исследования социальных взаимодействий. Анализ социальных сетей выделяется как важная область исследований, обеспечивающая глубокое понимание структуры, динамики и влияния связей между людьми и организациями в онлайн-пространстве. Целью данной работы является изучение и анализ социальных сетей с использованием современных методов и инструментов анализа данных, а также исследование взаимосвязи между различными характеристиками контента и уровнем вовлеченности пользователей.

Ключевые слова: анализ социальных сетей; HAR (HTTP Archive) скрепнинг; анонимизация данных; граф социальной сети; анализ данных.

ВВЕДЕНИЕ

Анализ социальных сетей – это научный подход к изучению социальных связей, применяющий принципы теории сетей для понимания структуры и динамики социальных отношений. Основными элементами этого анализа являются понятия узлов и ребер. Узлы символизируют отдельных объектов, участников социальной сети, а ребра (или связи) отражают различные виды отношений между ними, такие как дружба, родственные или профессиональные связи и другие формы взаимодействия.

На 2024 год топ пять самых популярных социальных сетей в мире являются Facebook, YouTube, Instagram, WhatsApp и TikTok [4]. В данной работе проводился анализ социальной сети Instagram.

СБОР И ПОДГОТОВКА ДАННЫХ

Использование открытых источников данных ставит перед пользователями как юридические, так и этические вопросы [1]. Даже с такими данными могут возникать определенные сложности, а их использование подчиняется определенным принципам и правилам, которые важно изучить и соблюдать. Одним из ключевых юридических аспектов является вопрос о правах на интеллектуальную собственность. Поэтому даже для использования открытых данных, может потребоваться получение согласия от владельцев. Необходимо учитывать ряд ключевых законов о защите данных. Эти

законы варьируются в зависимости от географического региона, но один из наиболее важных является общий регламент по защите данных. В Беларуси использование персональных данных регулируется Законом Республики Беларусь «О защите персональных данных» № 99-З от 07.05.2021 [3].

Одним из ключевых способов в обеспечении защиты личной информации при проведении исследований является анонимизация данных. Учитывая текущие ограничения, связанные с политикой конфиденциальности, существует несколько методов получения информации из Instagram: использование Instagram Graph API, Web Scraping, использование сторонних сервисов и инструментов, ручной сбор данных.

Важно отметить, что перед сбором данных было получено разрешение у пользователей, анализ чьих данных проводился.

В данной работе использовался метод HAR (HTTP Archive) скрэпинга [2] для сбора общей информации некоторых пользователей, а также информации об их подписках. Из всех записей HAR файла выбирались только те, которые соответствуют URL запросам к API Instagram для получения списка подписчиков. Для каждого пользователя были извлечены общие данные и добавлены в список вершин. Связи были добавлены в список ребер, который содержит идентификатор текущего пользователя и идентификатор подписки. Также данные собирались с помощью библиотеки Instaloader. Для использования библиотеки необходима аутентификация.

ГРАФ СОЦИАЛЬНОЙ СЕТИ

Для построения графа социальной сети была использована библиотека NetworkX (рис. 1) [5]. Визуализация служит для анализа структуры социальной сети, выявления ключевых узлов и понимания взаимосвязей между пользователями.

Размер узлов пропорционален степени узла – количеству соединений с другими узлами. Размер узлов позволяет быстро идентифицировать пользователей с наибольшим количеством связей, что может указывать на ключевых участников сети или более популярных пользователей. Цвета узлов графа определяются в зависимости от значения поля `is_private`. Если значение `is_private` равно `True`, то узел окрашивается в красный цвет, иначе – в синий. Это позволяет визуально различать приватные и публичные аккаунты пользователей.

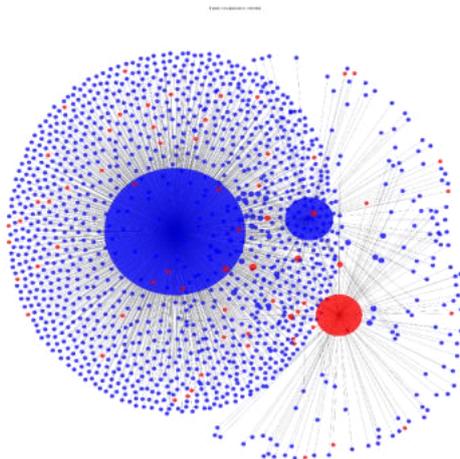


Рис. 1. Граф социальной сети Instagram

АНАЛИЗ ДАННЫХ

Набор данных, полученный с помощью библиотеки Instaloader, был сгруппирован по дню недели и был произведен расчет суммарного количества комментариев и отметок «нравится», а затем построена столбчатые диаграммы. Из временного тренда отметок «нравится» (рис. 2а) видно, что наибольшее их количество приходится на четверг и пятницу. В субботу в собранном наборе данных вообще отсутствуют лайки. Из временного тренда комментариев видно, что больше всего люди комментируют посты в понедельник (рис. 2б).

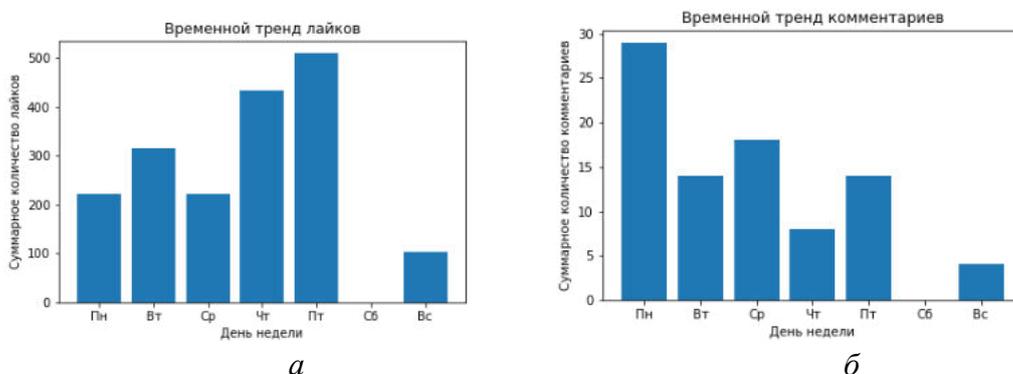


Рис. 2. Временной тренд
а – отметок «нравится»; б – комментариев

Диаграмма рассеивания (рис. 3) представлена в виде набора точек и демонстрирует взаимосвязь между количеством медиаэлементов (количество медиа элементов в посте) и двумя показателями взаимодействия аудитории: количеством лайков и количеством комментариев. Отметки «нравится» обозначены синим цветом, комментарии – красным.

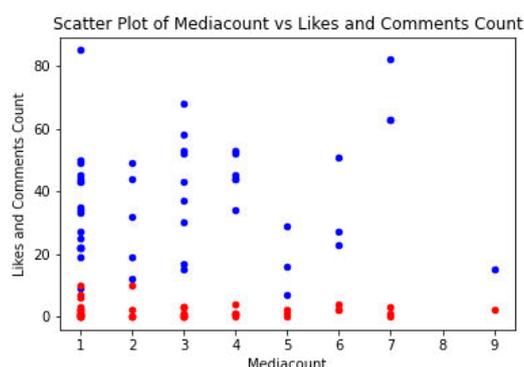


Рис. 3. Диаграмма рассеивания

Наблюдается общая тенденция к увеличению количества лайков по мере роста числа медиаэлементов в посте. Это указывает на потенциальную корреляцию между включением медиаконтента в публикации и повышением их вовлеченности.

Также из графика видно, что рост в количестве лайков более выражен, чем в количестве комментариев. Это может указывать на то, что пользователи более склонны реагировать на контент с помощью лайка, чем оставляя комментарий.

Несмотря на общую тенденцию, график демонстрирует значительный разброс точек, что свидетельствует о вариативности данных. Это может означать, что фактор количества медиаэлементов не является единственным определяющим вовлеченность аудитории.

Библиографические ссылки

1. И. А. Щеглова. Этические и правовые аспекты использования данных из социальных медиа. [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/eticheskie-i-pravovye-aspekty-ispolzovaniya-dannyh-iz-sotsialnyh-media/viewer>
2. Collect, Process, Visualize - Programming Social Graphs (Instagram, Python, Gephi). [Электронный ресурс]. URL: https://www.youtube.com/watch?v=CgHDt_aKyCc
3. Национальный правовой Интернет-портал Республики Беларусь. [Электронный ресурс]. URL: <https://pravo.by/novosti/novosti-pravo-by/2022/november/72150/>
4. Most popular social networks worldwide as of April 2024, ranked by number of monthly active users. [Электронный ресурс]. URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
5. Python 3.12.4 documentation. [Электронный ресурс]. URL: <https://docs.python.org/3/>