**«Факторы, определяющие распространение рекламных сообщений**

**на платформе YouTube»**

Тихомирова К.А., НИУ ВШЭ

1 курс магистратуры «Журналистика данных»

В работе рассматривались факторы распространения сообщений в русскоязычном сегменте YouTube. По итогам работы была создана модель, предсказывающая распространение информации о роликах на YouTube среди лидеров мнений.

Исследование затрагивает нерекламные сообщения (ролики на YouTube) и те из рекламных, которые представляют собой полноценное видео блогера, посвященное бренду или снятое при его спонсорстве. Только в данном случае было возможно оценивать влияние блогера на результат распространения ролика. Итого, в нашу выборку попало 8 сообщений: реклама Veet с Марьяной Ро, серия роликов для Клинского, новогодняя реклама Coca-Cola с блогерами, ролики проекта HypeCamp, клипы “Блокеры” и “Мега звезда”, дисс Лиззки, versus-баттл между Оксимироном и Славой КПСС.

Современные модели распространения информации в сети (MScaleDP, Ву, Чэнь, Сиань и Го (2016); ARP, Ли, Лин и Йе (2017)) используют несколько уровней анализа и включают данные о предыдущих взаимодействиях между пользователями. Факторы, используемые в анализе, можно разделить на три группы: характеристики сообщения, его автора и черты узла, который принимает или отвергает сообщение. Всего в нашем исследовании рассматривалось 23 фактора.

Предыдущее взаимодействие мы учитывали при построении графа блогеров. Связи между узлами в сети показывали, что одному каналу когда-то понравился ролик другого. Всего в сети оказалось более 17 тысяч узлов (каналов) с более чем 70 тысячами связей.

Сетевой анализ показал, что граф блогеров на YouTube по структуре склонен к «маленькому миру». Это означает, что в сообществе нет ярко выраженных центров, а практически все члены могут достичь друг друга. Выделив из сети узлы с входящей мощностью от 20, мы получили 2 структурные группы: тесно связанные между собой комьюнити блогеров и периферию. И рассматривая графы блогеров, которые распространяли то или иное сообщение, эта структурная особенность сохранялась и включала в себя близкие или одинаковые каналы.

Мы проверяли две гипотезы:

1. факторы, связанные со структурой сети, влияют на степень активации и, следовательно, на распространение информации;
2. факторы сообщения и автора первичны (влияют на активацию), факторы структуры – вторичны (влияют на степень активации).

При сетевом анализе мы выявили только влияние рекламного статуса сообщения на его распространение. Если сообщение рекламное, то каналов-последователей с периферии у него больше, чем каналов из связанного коммьюнити.

Для детальной оценки факторов мы использовали модели регрессии и классификации. В качестве зависимой переменной мы использовали степень активации узла сообщением, где “0” - узел не активирован, а число, близкое к 100, - первый последователь после автора. Хороший результат показала линейная регрессия и SVM с радиальным ядром.

Первая гипотеза не опровергается. Сравнивали две линейные регрессии: одна применялась на всех узлах, а другая – только на активированных. Для всех узлов значимыми переменными стали те, которые связаны с путем от узла до автора и показателем PageRank автора. При анализе только активированных узлов в значимые переменные добавляются факторы сообщения. Последний момент опровергает вторую гипотезу.

Основной результат исследования: по данным об авторе сообщения и его положении в сети при оптимальных показателях сообщения и принимающего узла, можно объяснить до 30% успеха распространения сообщения. Таким образом, на основе модели можно создать веб-приложение, которое будет определять подходящих блогеров для старта маркетинговой кампании и предсказывать движение сообщения в начальный период времени.