

Использование средств Wolfram Mathematica для оценки эффективности сообщества социальной сети “ВКонтакте”¹

Тякунов А.С., Славский В.В.
Югорский государственный университет
atyakupov@mail.ru, slavsky2004@mail.ru

Аннотация

В статье рассматривается проблема анализа взаимодействия и передачи информации посредством социальных сетей. Предложен вариант оценки эффективности сообщества в социальной сети на основе открытых данных, полученных из API-запросов. Проведена техническая реализация и обоснование полученных результатов.

1. Методы распространения информации в сообществах социальной сети

Развитие социальных сетей, помимо выполнения функций поддержки общения, обмена мнениями и получения информации их членами, в последнее время все чаще становятся объектами и средствами информационного управления, а также существенным инструментом влияния, в том числе, в целях координации личных взглядов, социальных групп и общества в целом.

Таким образом, необходимость качественного анализа социальных взаимодействий посредством социальных сетей является важной задачей современности, в том числе и с математической точки зрения. Для ее решения требуются инструменты, позволяющие обрабатывать запросы к социальным сетям с целью получения понятных результатов анализа. Одним из таких инструментов является среда Wolfram Mathematica.

Математическое программное обеспечение Mathematica, известно как самое мощное в мире вычислительное приложение. Более того, Mathematica служит единой платформой для проведения исследований и вычислений, которые находят свое отражение в живых интерактивных документах издательского качества.

В формате социальной сети распространение информации (в больших масштабах) производится посредством разнообразных тематических сообществ-групп, к которым пользователи сети имеют возможность присоединиться, тем самым получая доступ к публикуемому в сообществах информационному контенту (наполнению).

Важным показателем активности страницы сообщества является количество произведенных пользователями оценок заинтересованности, так называемых “лайков” (англ. like – нравится), а также актов копирования записи сообщества на свою страницу – “репостов” (англ. repost – повторное сообщение).

В рамках социальной сети “ВКонтакте” данные способы оценки имеют наименование “Мне нравится” и “Поделиться” соответственно. Пользователи социальной сети, положительно оценивая представленный на странице контент (записи, фотографии и т.д.), представляют отметку, тем самым выражая свое отношение к опубликованному. При репосте

¹Работа выполнена при поддержке Совета по грантам Президента РФ (грант НШ 2263.2014.1), Правительства РФ (госконтракт 14.В25.31.0029), РФФИ 15-41-00092 р-урал-а, 15-41-00063 р-урал-а, 15-01-06582 А.

отмеченное сообщение появляется в ленте пользователя и в ленте новостей, следящих за пользователем людей.

Анализ оценок “Мне нравится” и, в особенности, репостов позволяет, в первую очередь, оценить заинтересованность пользователей социальной сети в той информации, которая размещается на странице сообщества, а также произвести попытку спрогнозировать рост числа пользователей (и, как следствие, возможной популярности сообщества) с учетом тех пользователей, которые могут увидеть данный контент.

Разумеется, абсолютное количество участников сообщества позволяет говорить о размере охвата пользователей социальной сети. Однако, при сравнении нескольких групп только по количеству членов нельзя сделать однозначного вывода об эффективности того или иного сообщества. Дело в том, что активное использование репостов даже в небольшом сообществе позволяет увеличивать количество просмотров контента в разы и, напротив, большое сообщество с низким количеством репостов не позволит обеспечить аналогичный охват. Важным аспектом функционирования сообщества (в первую очередь, речь идет об информационных или “продающих” группах) является также и выявление того влияния, которое оказывают репосты пользователей социальной сети на распространение контента.

2. Оценка эффективности сообщества

В качестве метода оценки популярности и эффективности сообщества в социальной сети “ВКонтакте” предлагается использовать отношение реального количества фиксированных просмотров контента N к мощности потенциального множества прямых и косвенных элементов сообщества N_0 . Под прямыми элементами сообщества понимаются непосредственно участники (те пользователи, которые являются подписанными на сообщество и видят информацию от него в своей ленте новостей). В качестве косвенных участников следует учитывать “друзей” подписанных на сообщество пользователей, а также подписчиков подписанных пользователей (followers). Подписчики элемента отличаются от “друзей” тем, что элемент еще не подтвердил наличие связи между ним и данным пользователем, но, тем не менее, они также как и “друзья” могут видеть обновления, которые элемент произвел. В нашем случае нам важна возможность подписчиков увидеть репосты записей исследуемого сообщества.

Таким образом, потенциальное множество пользователей социальной сети, которые могут получить доступ к контенту сообщества формируется из суммы числа участников сообщества M , “друзей” этих участников F и подписчиков этих участников U . Однако важным дополнением является то, что множества M , F и U могут пересекаться в различных комбинациях. Так, например, некий пользователь может быть подписчиком сообщества и одновременно одним из друзей другого подписчика этого же сообщества.

Во избежание неверного определения множества N_0 необходимо не производить учет дублированных учетных записей. Следовательно, N_0 (в контексте исследования и для уменьшения количества вычислительных операций) определяется следующим образом:

$$N_0 = M \cup A \cup B, \quad \text{где} \quad A = F \setminus M; \quad B = (U \setminus M \cap U \setminus F)$$

При использовании среды Wolfram Mathematica оптимальным вариантом будет объединить исходные множества M , F и U , после чего использовать функцию удаления дубликатов. Количество элементов полученного множества $|N_0|$ будет показателем того, какое количество уникальных пользователей социальной сети в данный момент потенциально могут иметь представление о контенте социальной сети.

Реальное количество просмотров пользователями, составляющими множество N , предлагается считать совокупность следующих множеств:

M – участники сообщества;

M_F – “друзья” тех участников сообщества, которые сделали репост записи;

M_U – подписчики тех участников сообщества, которые сделали репост записи;

L – пользователи, поставившие оценку “Мне нравится” и при этом не являющиеся членами сообщества;

R – пользователи, сделавшие репост и при этом не являющиеся членами сообщества;

R_F – “друзья” пользователей множества R ;

R_U – подписчики пользователей множества R .

Таким образом, по каждой записи, размещенной в сообществе, осуществляется подсчет всех пользователей социальной сети N , которые получили доступ к размещенному контенту. Следует отметить, что для исключения дублированных записей аналогично тому, как это делалось при нахождении N_0 следует учитывать то, что перечисленные множества могут пересекаться. Множество всех пользователей без удаления дубликатов обозначим N^* , в большинстве случаев мощность N^* значительно превосходит N , что обусловлено наличием некоторого количества повторений пользователей в выборке.

Итоговое значение S (удельное количество реальных уникальных просмотров) есть отношение мощности найденного множества уникальных пользователей, осуществивших реальные просмотры, к мощности множества потенциальных получателей информации. Значение S^* (удельное количество реальных неуникальных просмотров), в свою очередь, показывает отношение количества неуникальных пользователей к множеству потенциальных пользователей.

$$S = \frac{|N|}{|N_0|}, \quad S^* = \frac{|N^*|}{|N_0|}.$$

Полученные коэффициенты позволяют оценить эффективность публикаций, размещаемых на стене сообщества вне зависимости от количества его участников. Таким образом, можно определить, насколько используется потенциал сообщества, либо сравнивать результативность работы нескольких групп с разным количеством участников. В целом, достаточно хорошим результатом для группы будет показатель S больше 0.1 (что можно трактовать как “более 10 процентов потенциально возможных получателей контента совершили просмотры записей сообщества”).

Коэффициент уникальности пользователей D есть отношение S к S^* , и, в свою очередь, указывает на то, насколько широким является потенциальный охват пользователей сети публикациями сообщества.

$$D = \frac{S}{S^*} = \frac{|N|}{|N_0|} \cdot \frac{|N_0|}{|N^*|} = \frac{N}{N^*}.$$

Значение D лежит в промежутке $(0; 1]$ и позволяет оценить, насколько велик вклад в общее количество просмотров отдельными пользователями. К примеру, $D < 0.3$ говорит о существенном однообразии массива пользователей (пользователи сообщества, их друзья и подписчики имеют большое количество связей разного рода между собой, вследствие чего номера пользователей часто повторяются в находимых множествах); $D > 0.6$, напротив, говорит о том, что они связаны в малой степени.

Формат API-запроса социальной сети “ВКонтакте” wall.get, который позволяет получать записи стены сообщества, имеет ограничение в 100 сообщений, что более чем достаточно для анализа. Варьируя глубину запроса (k), можно получать различные значения $|N_k|$, позволяющие оценить эффективность группы за короткий ($k = 10$), средний ($k = 50$), и большой ($k = 100$) промежутки времени. При этом вариант учета непосредственно по количеству последних записей, а не по календарным периодам выбран исходя из потребности оценивания эффективности группы вне зависимости от частоты публикаций. Следует отметить, что при выполнении работы использовались исключительно запросы, не тре-

бующие ключей доступа, а анализ производился на группах, обращение к которым не требует подтверждения (“открытых”).

Полученные в результате значения N_k и соответствующие им N_k^* позволяют рассчитать итоговые значения S_k , S_k^* и D_k .

3. Применяемые технологии Wolfram Mathematica

Средствами Wolfram Mathematica 10.0 производились API-запросы по получению списка членов группы, их "друзей" и подписчиков, сообщений на стене сообщества, а также другой необходимой информации. Выгрузка производилась в формате JSON, который очень удобен для обработки программными средствами, поскольку представляет собой упорядоченный набор данных вида “ключ-значение”. В частности, запрос списка участников сообщества в самом общем виде представляет собой строку вида “http://api.vk.com/method/groups.getMembers?group_id=**ID**&v=5.53&lang=en”, где ID - идентификатор группы.

Результатом выполнения запроса является список идентификаторов участников сообщества. При этом, в тех случаях, когда количество участников группы превосходит максимальное количество, возможное для запроса (1000), необходимо производить выгрузку идентификаторов пользователей со смещением (ключ запроса **offset**). Выгрузка списков друзей и подписчиков участников сообщества производится аналогично с использованием соответствующих API-запросов.

Важными функциями Mathematica при обработке массивов данных (без явного использования циклов для перебора) являются Intersection, Complement и DeleteDuplicates – пересечение множеств, разность и удаление повторяющихся значений соответственно.

4. Практическое применение метода

В качестве опытной группы было выбрано сообщество Государственной библиотеки Югры (vk.com/okrlibrary, идентификатор сообщества – 34233469). Важными критериями при выборе являлись: наличие достаточно большого количества подписчиков, а также то, что на стене данной группы регулярно публикуются записи. Для того, чтобы оценить динамику эффективности сообщества, эксперимент был проведен трижды с интервалами в 20 дней.

При выборе исследуемой группы, в том числе, необходимо руководствоваться вычислительными способностями компьютера, на котором производятся вычисления. Исследование группы с количеством подписчиков более ста тысяч требует значительного количества времени, которое будет потрачено на запросы к социальной сети. С другой стороны, данное исследование, в первую очередь, может быть интересно администраторам небольших групп.

В ходе работы программы в каждом случае был оценен потенциальный охват группы (N_0), исходя из числа участников, друзей участников и подписчиков участников. Рассчитанные показатели приведены в таблице 1.

Таблица 1

Потенциальный охват группы

№ опыта	M	F	U	N_0
1	515	103532	29765	133812
2	537	109756	32765	143058
3	542	118383	43587	162512

Далее, для каждого опыта был осуществлен расчет показателей на глубину запроса в

10, 50 и 100 записей стены. Результат расчетов приведен в таблице 2 и на рисунках 1, 2, 3

Таблица 2

Расчетные данные

№ опыта	k	N^*	N	S^*	S	D
1	10	10619	4662	0.079	0.035	0.44
	50	66344	28704	0.496	0.215	0.43
	100	110477	40240	0.826	0.301	0.36
2	10	6635	3136	0.046	0.022	0.48
	50	42086	13266	0.294	0.093	0.32
	100	108199	40310	0.756	0.282	0.37
3	10	8471	2656	0.052	0.016	0.31
	50	34058	4962	0.210	0.031	0.15
	100	86409	17121	0.532	0.105	0.20

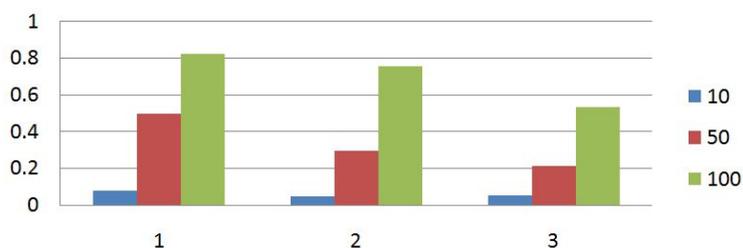


Рисунок 1. Удельное количество реальных просмотров S^*

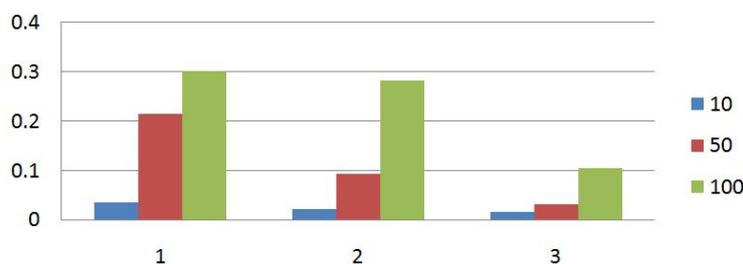


Рисунок 2. Удельное количество уникальных реальных просмотров S

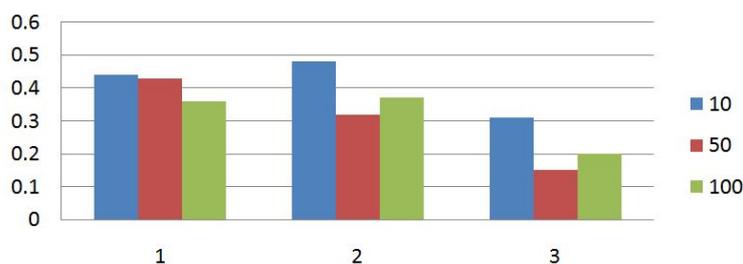


Рисунок 3. Коэффициент уникальности D

На основании полученных данных сделан вывод о том, что в целом эффективность исследованной группы находится на среднем уровне, при этом наблюдается тенденция к

потере пользователями интереса к записям сообщества. Во всех экспериментах коэффициент уникальности достаточно высок, что позволяет говорить об активном взаимодействии различных пользователей социальной сети с контентом, представленным в сообществе. При этом следует заметить, что среди записей, представленных на странице группы в краткосрочном периоде на текущий момент (эксперимент №3), достаточно велико количество тех, которые не вызвали активной реакции пользователей, то есть не были интересны аудитории.

Список литературы

1. Hanneman R.A., Riddle M. Introduction to Social Network Methods. — University of California, Riverside, 2005.
2. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели информационного влияния и информационного управления в социальных сетях. — М. : ФИЗМАТЛИТ, 2010.
3. Abdel-Ghany M.M. Identifying opinion leaders using social network analysis, a study in an egyptian village // Russian Journal of Agricultural and Socio-Economic Sciences. — 2012. — no. 4(4).