

УДК 621.396.41

**Е.Б. Грибанова, А.В. Катасонова**

## Модель оценки групп социальной сети для реализации маркетинговых мероприятий

Приводится описание модели оценки групп социальной сети «ВКонтакте» с точки зрения размещения рекламы. Представлен расчет интегрального показателя групп, при этом для определения коэффициентов относительной важности был использован метод анализа иерархий. Рейтинговая оценка может быть рассчитана как для одной группы, так и для набора групп. Также предложена модель классификации групп социальной сети. Представлено решение задачи целочисленного программирования для определения значений индикаторов групп.

**Ключевые слова:** социальная сеть, реклама, рейтинг, регрессионная модель, целочисленное программирование.

**doi:** 10.21293/1818-0442-2017-20-2-68-72

В последние годы анализу социальных сетей уделяется большое внимание в литературе из-за их широкого распространения, большого числа пользователей и огромного количества доступных данных. Пользователи сети обмениваются друг с другом информацией и публикуют в открытом доступе личные сведения, а также вступают в группы, всё это позволяет делать выводы об их интересах и находить потенциальных клиентов. Также среди преимуществ использования фирмами социальной сети отмечают более простую реализацию обратной связи: участники могут делиться мнениями о деятельности организации, оставляя комментарии, и таким образом могут указать пути развития и совершенствования продукции. Кроме того, социальные сети располагают широким набором инструментов для предоставления сведений рекламного характера: видеоролики, аудиозаписи, графические материалы. Все эти факторы делают их перспективной площадкой для проведения маркетинговых мероприятий с целью продвижения своих товаров и услуг.

Современными исследователями [1] отмечается, что использование социальной сети может способствовать значительному увеличению продаж фирмы. В отличие от классических моделей торговли в социальных сетях возрастает роль связей участника с другими пользователями при выборе товара или услуги или получении информации о них. Таким образом, в силу большого объема информации, представленной в сети, а также обусловленности результатов мероприятий сложным и зачастую непредсказуемым поведением людей и их взаимодействием возникает необходимость в инструментах и моделях, позволяющих оценивать и принимать решения экономическими агентами.

Из-за быстрого распространения социальных сетей и их влияния на социальные, экономические и политические процессы, исследованиями в этой области занимаются ученые многих стран, а также существуют научные лаборатории. Среди существующих направлений исследований можно отметить следующие: изучение методов пропаганды, информационных войн, выявление центров распространения информации (например, источника сплетен [2]).

исследование причин и признаков распространения ложной информации участниками сети [3], механизмов формирования репутации и доверия к другим участникам, оценка влияния социальных сетей на поведение людей, их здоровье, экономические и политические процессы [4], классификация элементов сети [5], выявление элементов, относящихся к определенной группе [6], мониторинг социальных процессов с помощью социальных сетей и др.

Среди работ, связанных с маркетингом и деятельностью организаций, можно отметить следующие.

Одно из направлений исследований посвящено вирусному маркетингу. При этом существуют различные определения данного понятия [7]. В одном из них под вирусным маркетингом понимают деятельность компаний, направленную на использование общения клиентов между собой посредством сети для продвижения и сбыта продукции. Также под вирусным маркетингом понимают передачу рекламной информации от пользователя к пользователю посредством сети. При этом значительную роль в распространении информации о товаре (причем как положительную, так и отрицательную) могут сыграть «лидеры общественного мнения». Выявление в сети таких лидеров мнений также является предметом изучения.

В работе [8] приводятся результаты ряда исследований в области интернет-маркетинга. В частности, изучаются следующие вопросы: «Какие товары лучше всего продвигать через социальные сети?», «На кого должна быть направлена реклама?» и т.д. В статье сделан вывод, что социальные сети являются удачным способом распространения информации о книгах, ресторанах, фильмах, т.е. тех продуктах, которые может купить пользователь, как только получит информацию о них даже от одного участника, и о которых люди часто беседуют между собой. Однако для товаров, выбор которых осуществляется только после принятия их большинством из окружения (программного обеспечения, операционных систем), наличие большого числа связей может служить сдерживающим фактором, и использование социальной сети не является удачным инструментом. Также в работе сделан вывод, что для товаров перво-

го вида (книг, фильмов и т.д.) реклама должна быть направлена на людей, обладающих небольшим количеством связей, а для товаров второго вида (операционные системы, новые технологии) лучше, когда информацией делится человек, обладающий значительным количеством связей.

В ряде исследований изучается использование целевой ценовой стратегии в социальных сетях, когда стоимость продукта для потребителя зависит от того, сколько друзей его приобрели, а также рассматриваются вопросы определения оптимальной цены. Например, в работе [9] цена пропорциональна показателю центральности участника сети. В статье [10] рассматривается использование стратегии вирусного маркетинга для максимизации дохода. В ней приводятся два способа поощрения распространения информации: возврат части стоимости покупки в случае, если человек рекомендовал продукт своим друзьям; предложение скидки клиентам с целью побудить их купить продукт после получения рекомендаций. Размер скидки и наличие рекомендации вместе с внешними факторами (качество продукта, состояние рынка) формируют вероятность покупки отдельных участников в предложенной модели. Решается задача определения размера скидки и возврата для заданной сетевой структуры.

Социальные сети могут влиять на экономические процессы, в этом случае важным является определение силы и направления связи между явлениями, а также построение прогнозных моделей. Например, в работе [11] рассматривается прогнозирование поведения цен на фондовой бирже с помощью информации в Twitter.

Одним из популярных видов реализации маркетинговых мероприятий является размещение информации рекламного характера в существующих группах. При этом для того, чтобы данное мероприятие было эффективным и привлекло как можно большее число клиентов, организации должны выполнить анализ характеристик групп. В настоящее время существуют онлайн-сервисы (<http://allsocial.ru/communities/>, <https://sociate.ru/>), которые предоставляют статистические данные групп: число подписчиков, стоимость рекламы, охват и т.д., однако они не предоставляют комплексную оценку групп, и не позволяют анализировать выбор нескольких сообществ, учитывая пересечения множеств участников.

Данная работа посвящена разработке модели оценки групп социальной сети «ВКонтакте», которая в настоящий момент является наиболее многочисленной в России (её аудитория на начало 2017 г. составляет 88–92 млн человек в неделю), для размещения рекламы. Для этого было необходимо разработать интегральный показатель группы, а также модель классификации групп социальной сети.

Для выполнения расчетов были получены данные групп, относящихся к городу Томску, предлагающие возможность платного размещения рекламы, при этом не рассматривались неактивные

группы, в которых запись администратором была сделана более месяца назад, а также группы, где записи на странице могут сделать все желающие. Всего было отобрано 45 групп, удовлетворяющих заданным требованиям.

Программа реализована в среде Visual Studio, язык программирования – C#. Для реализации доступа к данным социальной сети была использована библиотека VkNet.

### Группа как элемент социальной сети

Если социальную сеть представить в виде неориентированного графа, то группа будет представлять собой элемент, связанный ребрами с некоторым количеством соседних вершин (рис. 1), которые называются подписчиками группы или участниками. В сетевом анализе число участников сообщества называется степенью (число соседних вершин). Этот показатель является одним из основных при оценке групп, т.к. он определяет размер аудитории, которой можно донести информацию. Также это количество позволяет судить о том, сколько людей заинтересовано в данном сообществе и его продуктах/услугах. На рис. 1 степень группы 1 равна 4.

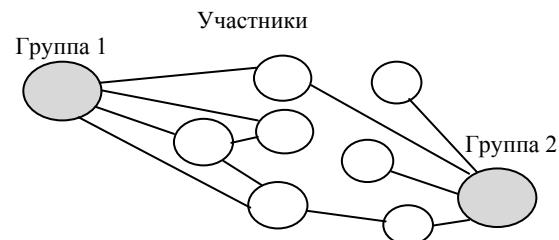


Рис. 1. Группы социальной сети

Объединению участников в группы способствует наличие схожих интересов, связанных с содержанием информации, представленной в группе. Вершины-участники группы также могут быть связаны ребрами между собой. Администратор группы может делиться с участниками информацией, размещая её на странице группы. Подписчики в свою очередь могут оставлять комментарии к записям, отмечать их лайками или размещать их у себя на странице, делая репост.

По характеру предоставляемой информации группы делятся на новостные (например, освещающие события города) и тематические (посвященные автомобилям, рукоделию и т.д.). С точки зрения размещения рекламы более предпочтительной является тематическая группа ввиду возможно большей заинтересованности участников в продукте/услуге, с другой стороны, новостные группы обеспечивают, как правило, более широкий охват аудитории.

В качестве характеристик группы могут быть использованы как стандартные метрики сетевого анализа (степень, коэффициент кластеризации и т.д.), так и специфические оценки (популярность, активность).

### Рейтинговая оценка групп социальной сети

Для формирования рейтинговой оценки необходимо определить индикаторы и целевую функцию

[12]. Были выбраны следующие характеристики групп:

- число подписчиков;
- активность – среднее количество записей на стене группы в день;
- популярность – сумма лайков, репостов и комментариев, деленная на число записей.

Для расчета активности и популярности в программе используются последние 100 записей.

Создание интегральной характеристики подразумевает определение коэффициентов важности показателей. Для их оценки был использован метод анализа иерархий [13], который позволяет придавать количественные выражения качественным понятиям. В качестве группы экспертов были выбраны менеджеры фирм, занимающиеся размещением рекламы в группах. Для заполнения им была предложена таблица, в которой необходимо было указать степень важности показателей по шкале от 1 до 9 (1 – показатели одинаково важны, 3 – первый показатель незначительно важнее, чем второй, 5 – первый показатель значительно важнее, чем второй, 7 – первый показатель явно важнее, чем второй, 9 – первый показатель по своей значимости абсолютно превосходит второй). Полученные значения были сведены в одну таблицу (табл. 1).

Таблица 1  
Матрица попарных сравнений

	Число под- писчиков	Популяр- ность	Активность
Число под- писчиков	1	5	7
Популярность	1/5	1	7
Активность	1/7	1/7	1

Полученные значения коэффициентов относительной важности: 0,701; 0,24; 0,059. Для их расчета были вычислены максимальное собственное значение матрицы попарных сравнений и соответствующий ему собственный вектор. Для вычисления коэффициентов было выполнено деление элементов вектора на общую сумму элементов.

Для формирования рейтинговой оценки был использован метод эталонного значения, а для нормирования переменных выполнено деление всех индикаторов на максимальное значение:

$$x_{ij} = \frac{a_{ij}}{\max a_{ij}},$$

где  $x_{ij}$  –  $i$ -й стандартизованный коэффициент  $j$ -й группы;  $a_{ij}$  –  $i$ -й индикатор  $j$ -й группы.

Все полученные в результате данного преобразования коэффициенты принимают значения от 0 до 1, причем чем выше значение стандартизованных оценок, тем лучше показатель группы.

Интегральный показатель вычисляется как сумма произведений квадратов полученных нормированных значений на коэффициенты относительной важности  $K$ :

$$R_j = x_{1j}^2 K_1 + \dots + x_{nj}^2 K_n.$$

На рис. 2 представлен фрагмент сформированного рейтинга групп города Томска.

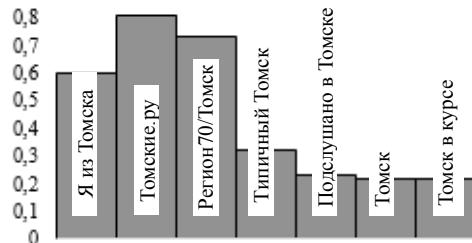


Рис. 2. Рейтинг семи групп города Томска

Полученные значения являются интегральными характеристиками, отражающими деятельность и сетевую структуру группы. Для того чтобы учесть еще один важный показатель – стоимость, была сформирована интегральная оценка двух показателей с равными значениями коэффициентов важности. При этом для показателя «стоимость размещения рекламы» лучшим с точки зрения реализации маркетинговых мероприятий является минимальное значение, поэтому осуществляется деление минимальной величины на значения показателей:

$$x_{ij} = \frac{\min a_{ij}}{a_{ij}}.$$

Для получения сведений о стоимости размещения рекламы в группах города Томска был использован сайт [sociate.ru](#), а также были сделаны запросы администраторам групп.

В случае выбора более одной группы используются следующие данные для расчета интегрального показателя:

- определяется общее число подписчиков групп с учетом их возможного пересечения (например, если подписчик входит в две рассматриваемые группы, то в этом случае он учитывается только один раз). Так, при выборе групп 1 и 2 (см. рис. 1) общее число подписчиков составит 7;

- стоимость размещения рекламы определяется как сумма стоимостей размещения рекламы в каждой группе;

- активность и популярность определяются путем расчета среднего значения данных показателей по выбранным группам.

Данный показатель определяется для выбранных наборов групп с целью оценки их интегрального показателя и сравнения между собой вариантов размещения рекламы.

#### Модель классификации групп социальной сети

Для отобранных групп был выполнен корреляционно-регрессионный анализ с целью оценки влияния разных факторов на величину стоимости рекламы и формирования уравнения зависимости. Была выполнена процедура отбора факторов с расчетом скорректированного индекса детерминации.

Наилучшей с точки зрения значения индекса детерминации оказалась логарифмическая зависимость

$$cost = -776,144 + 110,213 \ln(num\_p), \quad (1)$$

где  $cost$  – стоимость размещения рекламы в группе;  $num\_p$  – число подписчиков сообщества.

Индекс детерминации равен 0,519;  $F$ -статистика равна 46,35; уравнение является статистически значимым на 99%-ном доверительном интервале;  $t$ -статистика для фактора равна 6,8.

При этом было выявлено, что при равном количестве подписчиков стоимость рекламы в группах может значительно отличаться. Причины могут быть различны: более актуальная тематика; нет других групп, посвященных данной тематике; при поиске группа появляется одной из первых (что свидетельствует о хорошем соответствии названия поисковому запросу, популярности группы, отсутствии ботов и т.д.). Эти факторы приводят к увеличению стоимости рекламы. Так, при введении в модель (1) ещё одной переменной, выступающей индикатором, появляется ли группа среди первых при поиске и отсутствуют ли конкуренты, индекс детерминации модели увеличился. Полученная модель имеет вид

$$cost = -724,704 + 174,892f + 92,874 \ln(num\_p), \quad (2)$$

где  $f$  – переменная, которая принимает значение 1 в случае, если группа не имеет конкурентов и появляется среди первых при поиске (относится к первой категории), и 0 – в противном случае (относится ко второй категории).

Индекс детерминации равен 0,705;  $F$ -статистика равна 50,09; уравнение является значимым на 99%-ном доверительном интервале;  $t$ -статистика для факторов: 5,14; 6,99.

Коэффициент при переменной  $f$  показывает, насколько будет больше стоимость рекламы группы первой категории по сравнению с группой второй категории при равном количестве подписчиков. Таким образом, модель (2) можно представить в виде двух моделей:

$$cost = -724,704 + 92,874 \ln(num\_p), \text{ если } f = 0,$$

$$cost = -549,812 + 92,874 \ln(num\_p), \text{ если } f = 1.$$

При реализации маркетинговых мероприятий более актуальной является обратная задача: выявить из существующего набора группы первой (группы-лидеры) и второй категории. Это означает, что необходимо определить значение переменной  $f$  для каждой группы таким образом, чтобы индекс детерминации полученного уравнения был максимальным. Представленная задача является задачей целочисленного программирования, в которой искомые аргументы могут принимать только два значения: 0 или 1:

$$R^2(f) = 1 - \frac{\sum_{i=1}^n (cost_i - \Theta_0 - \Theta_1 f_i - \Theta_2 \ln(num\_p))^2}{\sum_{i=1}^n (cost_i - \overline{cost})^2} \rightarrow \max,$$

где  $f_i \in \{0,1\}$ ;  $\overline{cost}$  – среднее значение стоимости размещения рекламы в группе;  $\Theta$  – вектор параметров, определяемый с помощью метода наименьших квадратов:

$$\Theta = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y},$$

где  $\mathbf{X}$  – матрица объясняющих переменных (первый столбец – единицы, второй столбец – значения  $f_i$ , третий столбец – значения  $\ln(num\_p)$ );  $\mathbf{Y}$  – вектор, состоящий из значений  $cost_i$ .

На рис. 3 представлено полученное в результате решения оптимизационной задачи уравнение регрессии с переменной структурой для выбранных исходных данных.

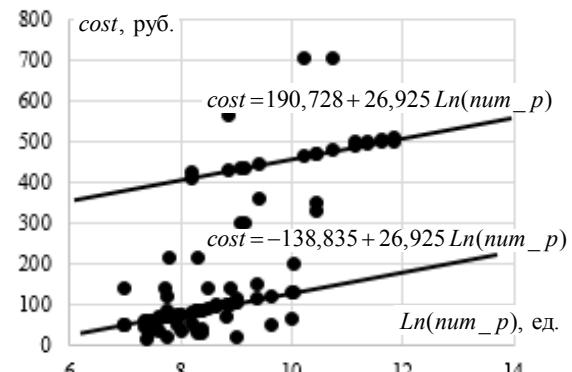


Рис. 3. Уравнение регрессии

В табл. 2 приведены значения индикатора  $f$  для пяти групп.

Таблица 2

#### Результат вычислений

№ группы	Значения показателей		
	Стоимость рекламы	Число подписчиков	$f$
1	500	110509	1
2	500	141133	1
3	499	69775	1
4	200	22749	0
5	50	15343	0

#### Заключение

В статье приводится описание модели оценки групп социальной сети «ВКонтакте» с точки зрения реализации маркетинговых мероприятий. Описано формирование интегрального показателя с помощью метода эталонного значения, для определения весовых коэффициентов был использован метод анализа иерархий. Для выполнения расчетов была использована информация о группах города Томска, удовлетворяющих заданным требованиям. Также в статье рассмотрена задача классификации групп с помощью решения задачи целочисленного программирования, предполагающей максимизацию индекса детерминации модели регрессии. Представлены результаты расчетов программы.

Представленные модели могут быть использованы для принятия решений при реализации маркетинговых мероприятий в социальной сети экономическими агентами.

*Литература*

1. Goyal S. Social networks and the firm / S. Goyal, J. Gagnon // Revista de Administracao. – 2016. – Vol. 51. – P. 240–243.
  2. Gossip: Identifying Central Individuals in Social Network / A. Banerjee, A. Chandrasekhar, E. Duflo, M. Jackson // The national bureau of economic research. – 2014. – № 20422. – P. 1–30.
  3. Self curatio, social partitioning, escaping from prejudice and harassment: the many dimensions of lying online / M. Kleek, D. Murray-Rust, A. Guy et al. // Proceedings of the 7<sup>th</sup> Association for Computing Machinery Web Science conference. – 2015. – P. 1–9 [Электронный ресурс]. – Режим доступа: <http://www.www2015.it/documents/proceedings/companion/p371.pdf>, свободный (дата обращения: 29.06.2017).
  4. Althoff T. Online actions with offline impact: how online social networks influence online and offline user behavior / T. Althoff, P. Jindal, J. Leskovec // Proceedings of the 10<sup>th</sup> Association for Computing Machinery International Conference on Web Search and Data Mining. – 2017. – P. 537–546 [Электронный ресурс]. – Режим доступа: <http://www.wsdm-conference.org/2017/proceedings/>, свободный (дата обращения: 29.06.2017).
  5. Hemalatha I. Social networks analysis and mining using machine learning techniques / I. Hemalatha, G. Saradhi Varma, A. Govardhan // AETS. – 2013. – P. 603–608.
  6. An army of me: Sockpuppets in online discussion communities / S. Kumar, J. Cheng, J. Leskovec, V. Subrahmanian // 26<sup>th</sup> International World Wide Web conference. – April 2017. – P. 1–10 [Электронный ресурс]. – Режим доступа: <http://papers.www2017.com.au.s3-website-ap-southeast-2.amazonaws.com/proceedings/p857.pdf>, свободный (дата обращения: 29.06.2017).
  7. Petrescu M. Viral advertising: definitional review and synthesis / M. Petrescu, P. Korgaonkar // Journal of internet commerce. – 2011. – Vol. 10. – P. 208–226.
  8. Galeotti A. Influencing the influencers: a theory of strategic diffusion / A. Galeotti, S. Goyal // Rand Journal of economics. – 2009. – Vol. 40. – P. 509–532.
  9. Bloch F. Pricing in social network / F. Bloch, N. Querou // Games and Economic behavior. – 2013. – Vol. 80. – P. 263–281.
  10. Pricing strategies for viral marketing on social networks / A. David, M. Rajeev, S. Aneesh, X. Ying // Internet and Network economic. – 2009. – Vol. 5929. – P. 101–112.
  11. Bollen J. Twitter mood as a stock market predictor / J. Bollen, H. Mao // Computer. – 2011. – Vol. 44. – P. 91–94.
  12. Грибанова Е.Б. Информационная система рейтинговой оценки объектов экономики / Е.Б. Грибанова, А.Н. Алимханова, П.Э. Тугар-оол // Доклады Том. гос. унта систем управления и радиоэлектроники. – 2016. – № 2. – С. 51–55.
  13. Саати Т. Принятие решений. Метод анализа иерархий. – М.: Радио и связь, 1993. – 278 с.
- 

**Грибанова Екатерина Борисовна**

Канд. техн. наук, доцент каф. автоматизированных систем управления (АСУ) ТУСУРа  
Тел.: +7 (382-2) 70-15-36  
Эл. почта: katag@yandex.ru

**Катасонова Анастасия Владимировна**

Студентка 4-го курса каф. АСУ ТУСУРа  
Тел.: +7 (382-2) 70-15-36  
Эл. почта: katasonova.a.v@yandex.ru

Gribanova E.B., Katasonova A.V.

**System to evaluate social network groups for the implementation of marketing activities**

The article provides a description of the evaluation model for groups in the social network VKontakte. The calculation for an integral indicator of the group is provided where the importance coefficients are defined with the method of pairwise comparisons. The ranking score can be calculated as for a single group and group set. Also the classification model for social network groups is proposed. A solution for the integer programming problem related to determining the indicator values for the groups.

**Keywords:** social network, advertising, rating, regression model, integer programming.