

Кластеризация сообществ социальной сети «ВКонтакте»

И. Ю. Анохина, Е. В. Рощина

Донецкий национальный технический университет

Ingatula@mail.ru

Анохина И. Ю., Рощина Е. В. Кластеризация сообществ социальной сети «ВКонтакте». Рассматриваются вопросы анализа социальных сетей. Проведен мониторинг и анализ тридцати восьми групп различных тематик, определены статистические показатели для оценки количественных и качественных характеристик участников групп. Проведена оценка уровня политизации групп. Проведен анализ существенности различий в личностных характеристиках участников сообществ различных групп. Методами кластерного анализа проведено разбиение групп на кластеры по уровню политизации.

Ключевые слова: социальные сети, ВКонтакте, кластерный анализ, структурные характеристики, мониторинг социальных сетей, политическая активность, политизация социальных сетей.

Введение

Социальные сети прочно интегрировались в жизнь современного человека.

Теперь это не просто средство общения и связи, а полноценная медиакультура, в рамках которой обсуждаются новости, происходят продажи и покупки, рекламируются товары и формируется общественное мнение, это трибуны для политических партий и средство организации, координации сообщества, это площадки для раскручивания кампании и брендов.

Социальные сети можно классифицировать по типу: лично/деловое общение; видео/аудио/фото; развлечения; покупки/продажа; новости; тематические и многое другое.

Они могут быть открытыми и закрытыми, распространяться по всему миру, одной стране, региону или не зависеть от географии.

Социальную сеть можно рассматривать, как часть общества, на которую пытаются повлиять для достижения определенной цели. Это может быть формирование общественного мнения по определенному вопросу, а может использоваться при проведении маркетинга и бизнес - аналитики.

Для решения таких задач необходимо знать, каким образом максимально эффективно донести информацию до сообществ и их участников, определить максимально эффективные каналы распространения информации по сети, найти пользователей, готовых с большой вероятностью действовать в соответствии с разработанной доктриной.

Помимо формирования общественного в сети можно проводить оценку любого нововведения, будь то сервис, сообщество или

рекламная компания. Такая оценка дает возможность разрабатывать сервисы не наугад, а целенаправленно повышать пользовательскую активность.

По мнению генерального директора «SECL Group» Н. Семенова более 80% компаний по всему миру используют социальные сети для развития бизнеса. Около 78% людей доверяют информации из социальных сетей [1].

Популярность сетей обуславливает необходимость в их изучении, анализе, моделировании.

Постановка задачи

Анализ социальных сетей стали применять во второй половине двадцатого века как дополнение к стандартному набору инструментов социальных исследователей.

Одним из первых направлений было исследование Social Network Analysis, рассматривающее социальные взаимосвязи в виде моделируемых графов и сетей. Модели строили на основе различных данных из печатных источников, дополнительных опросов и анкетирования.

Одной из главных тенденций современного развития сетей является их стремительная политизация. Уход в виртуальное пространство предполагает формирование различных сетевых политических сообществ со своими авторитетами и ценностями, причем среди них могут быть как провластные (патриотически-настроенные), так и оппозиционные (от конструктивных критиков до радикалов и экстремистов) [2].

Таким образом, социальные сети становятся не только неисчерпаемым источником бизнес – возможностей, но и источником риска для власти, включая

«цветные» революции и другие виды переворотов.

И, если сильно политизированной является, как правило, лишь незначительная часть общества, то используя социальные сети как инструмент, возможно привлечение все новых участников именно в политические группы. Как сказал античный философ Перикл: «Если Вы не интересуетесь политикой, то это не значит, что политика не интересуется Вами!»

Благодаря появлению социальных сетей стали публичными персональные данные участников сообществ, их биографии, аудио-, видео-, фотоматериалы, что сделало социальные сети отличным инструментом для получения информации как об отдельном индивидууме, так и группах лиц, объединенных общими интересами. Полученная информация используется для моделирования социальных, экономических политических и других процессов[3].

Как показали исследования, наиболее популярной социальной сетью в России является Вконтакте. Число пишущих авторов на лето 2017г. составило 27.5 млн. человек, основная возрастная группа – авторы в возрасте 25-34лет (37%), авторы женского пола составляют 58% . Наиболее популярной сетью является в

Петербурге (44.9% от общего населения), на втором месте – Мурманская область (30%), Москва – 28% [4].

В 2016г. в России реализован научный проект – «Глобальное исследование политизации социальных сетей», в ходе которого анализировались интернет - сообщества стран Западной Европы, Северной Америки, Латинской Америки, Азии и Африки. Анализировался контент сообществ.

В Западноевропейском кластере наиболее политизированы Германия и Италия. В Германии наиболее многочисленны сообщества, выступающие против политики ЕС, за отставку Меркель и пр.

В США (Англосаксонский кластер), среди оппозиционных групп наиболее популярны группы, разжигающие ненависть между Севером и Югом. В то же время в Канаде популярны проправительственные группы, см. табл.1.

Интересен тот факт, что в Индии резко увеличивается число участников, вовлеченных в политические сообщества. Только за январь 2016г. количество подписчиков официального блога премьер-министра увеличилось практически на миллион.

Таблица 1. Политизация сообществ в социальных сетях

Кластеры Интернета	Политические субкультуры		
	провластная	оппозиционная	радикальная
Западноевропейский	38%	58%	4%
Англосаксонский	54%	33%	13%
Латиноамериканский	81%	18%	1%
Азиатский	59%	35%	6%
Африканский	41%	57%	2%

В социальных сообществах обсуждаются новости и события.

Появление оппозиционных и радикальных политизированных сообществ Интернета отчетливо видно на примере США («южане» и «северяне») и Германии.

Однако в таких странах, как Китай, Индия и Южная Корея интернет- сообщества служат поддержкой для власти.

Таким образом, как пишет известный испанский социолог Мануэль Кастельс, современная легитимность власти уже не может учитывать только силовые методы и заставляет власть искать новые сетевые приемы для работы с электоратом.

Отсюда становится понятно значение современной виртуализации политической сферы и важность изучения этой части

социальных сетей.

Мы исследовали российский кластер сообществ.

Для анализа выбрана социальная сеть ВКонтакте.

Чтобы оценить уровень политизации социальных сетей, возможности пересечения их участников, вероятность перемещения участников из одной группы в другую, мы рассматривали три типа сообществ.

1. Полностью **нейтральные** к политике группы по интересам (рыбалка, кулинарные рецепты, группы книголюбов, путешественников, любителей животных и пр.). Такие группы здесь и далее определены нами как «**группы по интересам**», при кластеризации отмечались как группы с

- показателем $gr = -1$.
2. **Новостные** сообщества и сообщества, созданные для обсуждения политических новостей, сообщества партий, сообщества телеканалов, в целом дающие новости в позитивном плане, здесь и далее «**политические группы**», показатель $gr = 0$.
 3. Третий вид групп аналогичен сообществам второго вида, но с ярко выраженной критической нотой, «**критики**», $gr = +1$.

Поставлена задача, провести статистический анализ участников сообществ с целью проведения кластеризации и выявления возможных тенденций миграции участников одного из видов сообществ к другому.

Исследования

Для решения задачи нами рассматривались 38 сообществ. По каждому сообществу анализировались данные, приведенные в табл. 1.

Исследовали возраст, пол, город проживания и ряд других параметров, указанных в таблице.

Например, в столбце 2 таблицы указано, что 27.4% участников от общего числа в группе имеют возраст до 25 лет.

Данные исследовались в процентном отношении числа участников, соответствующих заданному критерию, к общему числу участников группы.

Номера столбцов в таблице соответствуют данным: 2- название группы, 3 – возраст до 25 лет, 4 – возраст от 25 до 35 лет, 5 – от 35 до 45; 6 – от 45 до 55; 7 – старше 55; 8 и 9 – количество женщин и мужчин в группе соответственно.

В столбцы 10 и 11 заносились процент участников, указавших в качестве религиозных взглядов православие (10) или светский гуманизм (11). На первом этапе обработки статистических данных мы рассматривали другие религии (буддизм, ислам...), но их процент был столь незначителен, что в дальнейшем оставили только два варианта.

Сообщество ВКонтакте предлагает участникам определить главное в жизни, выбирая из нескольких вариантов. В таблице эти варианты указаны в столбцах: 12 - совершенствование мира, 13 - семья и дети, 14- карьера и деньги, 15 – развлечения и отдых, 16 - наука и исследования, 17 – саморазвитие, 18 – красота и искусство, 19 – слава и влияние.

Нами на основании данных столбцов 3- 7 вычислялся средний возраст (столбец 20). Мы также ввели дополнительную характеристику, определив ее как степень открытости участников сообщества, оценив отношение количества

заполненных анкет к общему числу участников сообщества (столбец 21).

Важным показателем реально функционирующего сообщества является процент ботов² (столбец 22) и активность подписчиков, определяемая как количество лайков, комментариев, репостов сообщений. (23). Эти показатели были определены с помощью on-line сервиса <http://smmur.ru>[6].

В табл. 2 приводятся выборочно данные по одной группе из каждого типа сообществ. Группа «Интересная планета» насчитывает более трех миллионов подписчиков, не содержит политической и новостной информации. Группа «Народный журналист» относится к группам критиков (около 6 тысяч подписчиков). «Первый канал» – более миллиона подписчиков.

Группы критиков имеют, как правило, незначительную численность. Самая большая по численности из проанализированных нами была группа «Сводки ополчения Новороссии», на момент анализа имеющая около полумиллиона подписчиков.

В группах по интересам преобладают женщины (42.4%), мужчины – 40.5%.

Политические группы в большей мере привлекают мужчин (50.1%) против 32.5% (женщины).

В группах критиков эта тенденция усиливается. 63.1% в группах – мужчины и 26.2% - женщины, т.е. количество мужчин-участников почти в три раза больше, чем женщин.

Как видим, не все участники сообществ указали свой пол. Немногим более 10% не дали ответ даже на этот вопрос.

С учетом этого нами и была введена характеристика открытости участников сообществ, т.е. их готовность предоставлять информацию о себе.

На рис.1. представлены гистограммы, отображающие распределение участников групп по возрастам (рис.1а) и характеристики непосредственно групп (процентное содержание ботов и активность посетителей в сутки), рис. 1б.

Анализируя данные по распределению возрастных категорий, следует отметить, что в группах по интересам преобладают молодые люди в возрасте до 25 лет, в то время, как в политических группах возрастная шкала смещена к 35 годам и старше.

² Боты - страницы ВКонтакте, наполненные ложной информацией, либо, взломанные страницы реальных людей. Сферы применения : рассылка спама, накрутка подписчиков, лайков, репостов.

Таблица 2. Пример исходных данных для анализа сообществ

№	Группа	Возраст						Пол		Отноше- ние к религии	Главное в жизни								Дополнительные характеристики			
		3	4	5	6	7	8	9	10		11	12	13	14	15	16	17	18	19	20	21	22
1	Интересная планета	27.4	44.4	11.2	3.4	3.4	50.5	33.6	14.5	0.7	1.3	13.5	0.7	1.0	0.2	5.2	0.6	0.2	30	23	13.5	1.1
2	Первый канал	31.4	28.7	12.1	5.6	6.5	49.8	31.7	12.3	0.5	1.0	13.3	0.8	0.8	0.3	3.4	0.6	0.2	32	20	13.4	0.38
3	Народный журналист	7.5	24.7	29.1	17.3	16.3	31.6	58.6	14.5	3.3	5.8	9.9	0.5	0.4	0.9	8.1	0.9	0.5	42	27	6.50	19.7

В группах критиков во всех возрастных категориях старше 35 количество участников превышает аналогичные показатели в остальных видах групп.

Минимальное количество ботов и активность подписчиков в сутки (8%) в группах критиков, что более, чем в два раза превышает активность в остальных видах групп.

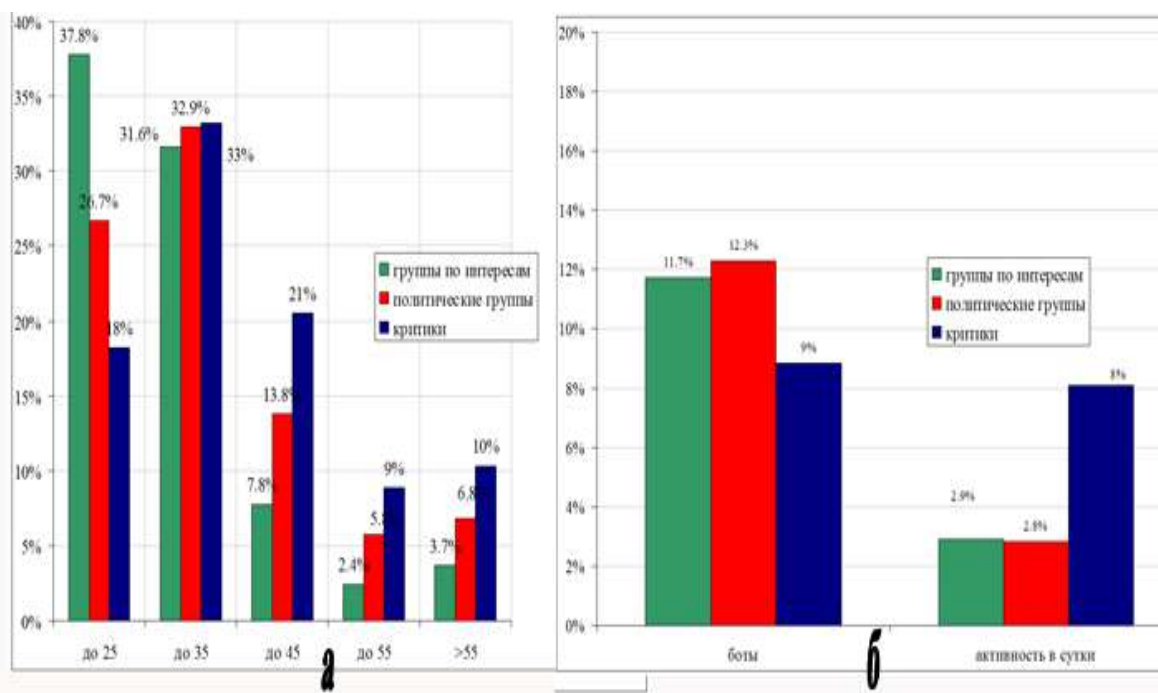


Рисунок 1 – Количественные характеристики групп. а – распределение по возрастному признаку; б – активность участников групп

В целях оценки существенности различий между группами в среде пакета Statistica был проведен t – тест с уровнем значимости $p=0.05$. В табл.3 приведены результаты тестирования.

Для всех групп рассчитывались значения математического ожидания (столбцы 1, 2, 5, 6, 9, 10), определялась величина коэффициента Стьюдента (t-value) и уровень значимости p.

Как видно из таблицы, наличествуют существенные различия по одиннадцати характеристикам между контингентом групп по интересам и участниками групп-критиков (выделены красным).

Состав участников политических групп и групп по интересам отличается в меньшей степени и только по шести признакам, в основном связанным с возрастом и отношением к религии.

Наиболее существенно различаются политические группы и группы –критики (четырнадцать позиций).

Таким образом, первичная обработка данных позволила сделать вывод о наличии существенных различий между контингентом анализируемых групп.

Таблица 3. Т- тест определения существенности различий между выборками данных

	Сравнение групп по интересам и групп-критиков				Сравнение групп по интересам и политических групп				Сравнение групп-критиков и политических групп			
	1	2	3	4	5	6	7	8	9	10	11	12
	gr=-1	gr=1	t-value	p	gr=-1	gr=0	t-value	p	gr=1	gr=0	t-value	p
До 25	0.38	0.18	3.62	0.00	0.38	0.27	2.25	0.03	0.18	0.27	-3.05	0.01
До 35	0.32	0.33	-0.45	0.66	0.32	0.33	-0.36	0.72	0.33	0.33	0.15	0.89
До 45	0.08	0.21	-6.03	0.00	0.08	0.14	-4.75	0.00	0.21	0.14	3.20	0.00
До 55	0.02	0.09	-4.89	0.00	0.02	0.06	-5.73	0.00	0.09	0.06	2.29	0.03
Старше 55	0.04	0.10	-7.57	0.00	0.04	0.07	-4.83	0.00	0.10	0.07	3.52	0.00
Ж	0.42	0.26	2.71	0.01	0.42	0.33	1.57	0.13	0.26	0.33	-1.57	0.13
М	0.41	0.63	-3.45	0.00	0.41	0.50	-1.44	0.16	0.63	0.50	3.24	0.00
Православие	0.12	0.14	-1.96	0.06	0.12	0.16	-3.56	0.00	0.14	0.16	-1.22	0.23
Светский гуманизм	0.01	0.03	-5.74	0.00	0.01	0.01	-2.60	0.02	0.03	0.01	3.57	0.00
Совершенство мира	0.01	0.07	-3.24	0.00	0.01	0.02	-1.59	0.13	0.07	0.02	2.66	0.01
Семья и дети	0.16	0.14	0.48	0.63	0.16	0.14	0.55	0.59	0.14	0.14	-0.11	0.91
Карьера	0.01	0.01	0.11	0.92	0.01	0.01	-0.03	0.98	0.01	0.01	-0.20	0.84
Отдых	0.02	0.01	1.24	0.23	0.02	0.01	1.57	0.13	0.01	0.01	0.29	0.77
Наука	0.00	0.01	-4.22	0.00	0.00	0.00	-0.97	0.34	0.01	0.00	2.97	0.01
Саморазвитие	0.05	0.09	-3.16	0.00	0.05	0.05	-0.10	0.92	0.09	0.05	3.54	0.00
Искусство	0.01	0.01	-0.47	0.64	0.01	0.01	1.51	0.15	0.01	0.01	3.19	0.00
Слава	0.00	0.01	-3.28	0.00	0.00	0.00	-1.41	0.17	0.01	0.00	2.12	0.05
Открытость	0.83	0.90	-1.95	0.06	0.83	0.84	-0.31	0.76	0.90	0.84	3.17	0.00
Боты	0.13	0.09	1.99	0.06	0.13	0.12	0.24	0.81	0.09	0.12	-2.28	0.03
Активность	0.04	0.08	-1.75	0.09	0.04	0.03	0.71	0.49	0.08	0.03	2.34	0.03

Анализируя выборки, состоящие из нескольких групп, мы доказали наличие существенных различий, но в выборках в целом. В то же время нельзя исключать, что отдельные группы, входящие допустим в выборку по интересам, не могут быть по характерным признакам отнесены к той или иной политической группе.

Вряд ли можно утверждать, что люди, вступившие в группу, допустим, путешественников, книголюбов, никогда не интересовались политикой и не будут интересоваться ею впредь.

Поэтому далее нами был применен

кластерный анализ с целью выявления возможных тенденций миграции или добавления участников одной выборки в другую.

При проведении кластерного анализа рассматривалась возможность наличия от трех до шести кластеров, на которые могла быть разбита вся выборка, состоящая из 38 групп. В качестве переменных использовались все переменные, указанные в табл.1. за исключением тех, в которых не были зафиксированы существенные различия между группами.

На первом этапе кластерного анализа определялось оптимальное число кластеров. Для этого использовали метод иерархического

кластерного анализа Joining (Tree clustering). В качестве метрики, определяющей расстояние между кластерами Amalgamation Rule, был

выбран метод ближнего соседа или одиночная связь (single linkage)[7,8]. На рис.2. показана часть обработанных кластеров.

CLUSTER ANALYSIS linkage distance	Single Linkage Euclidean distances					
	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6
.0828254	x_6	x_8				
.0898790	x_33	x_34				
.0916961	x_31	x_37				
.0938040	x_17	x_23				
.0948636	x_17	x_23	x_20			
.0951687	x_31	x_37	x_33	x_34		
.0958407	x_24	x_25				
.1023266	x_18	x_31	x_37	x_33	x_34	
.1089584	x_28	x_36				
.1135806	x_6	x_8	x_28	x_36		
.1292735	x_26	x_38				
.1294454	x_24	x_25	x_26	x_38		
.1325940	x_5	x_30				
.1415378	x_5	x_30	x_24	x_25	x_26	x_38

Рисунок 2 – Матрица расстояний между центрами кластеров

Как видно из рис.2, на первом этапе в один кластер объединяются объекты с номерами 6 (группа «Путешествия и туризм») и 8 («Интересная планета»). Ясно, что люди, склонные к путешествиям, не могут не интересоваться планетой, географией и пр.

В другой кластер попадают группы 31(КПРФ – Коммунистическая партия РФ), 33 («Вежливые люди»), 37 – «Телеканал Звезда».

Аналогичным образом нами были проанализированы остальные составляющие кластеров.

На основании анализа было принято решение разделить группы на пять кластеров. Для определения, какие именно группы входят в кластеры, был применен метод k-средних (k-

means). Суть метода состоит в следующем: заранее определяют количество классов (k), на которые необходимо разбить имеющиеся наблюдения, и первые k – наблюдений становятся центрами этих классов. Для каждого следующего наблюдения рассчитываются расстояния до центров кластеров и анализируемое наблюдение относят к тому кластеру, расстояние до которого было минимальным. После чего для этого кластера (в котором увеличилось количество наблюдений) рассчитывается новый центр тяжести по всем включенным в кластер наблюдениям. Показателем правильности разбиения на кластеры является выводимая таблица расчета дисперсий и уровня значимости p (см. табл.4).

Таблица 4. Таблица анализа дисперсий

Показатели	Возраст					Пол		Характеристики группы		
	<25	<35	<45	<55	>55	Ж	М	открытость	боты	активность
Between SS	0.43	0.05	0.13	0.04	0.03	0.64	0.89	0.072	0.018	0.043
Within SS	0.26	0.16	0.04	0.02	0.01	0.21	0.20	0.169	0.063	0.071
F	13.81	2.57	24.68	20.14	18.45	25.77	36.98	3.531	2.405	5.038
p	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.017	0.049	0.003

В таблице приняты обозначения: дисперсия между кластерами **Between SS**, дисперсия внутри кластеров (**Within SS**), F-критерий для проверки гипотезы о неравенстве дисперсий F, значение уровня p [9].

Как следует из анализа данных табл.3, можно считать достоверным предложенное разбиение на кластеры. Уровень p только для

одной возрастной группы (от 25 до 35) достигает критического значения 0.05 и для характеристики, связанной с наличием ботов в группах, равен 0.049, т.е. приближается к критическому значению.

Аналогично оценили степень достоверности по критерию Фишера. Следовательно, мы можем принять и считать

допустим предложенное деление на кластеры.

На рис. 3 дана графическая интерпретация полученного деления на кластеры. Как видно из рисунка, наиболее существенное деление в кластерах происходит по возрастному, половому признакам и по степени активности участников групп.

Отметим, что участники всех групп в качестве основной составляющей жизни

указывали семью. Совершенствованием мира готовы заниматься в пять раз больше участников групп-критиков, чем участники остальных групп. Эта же группа отличается стремлением к самосовершенствованию и тяге к научным исследованиям. А вот карьеру и деньги в качестве основной цели назвали участники политических групп (gr=0).

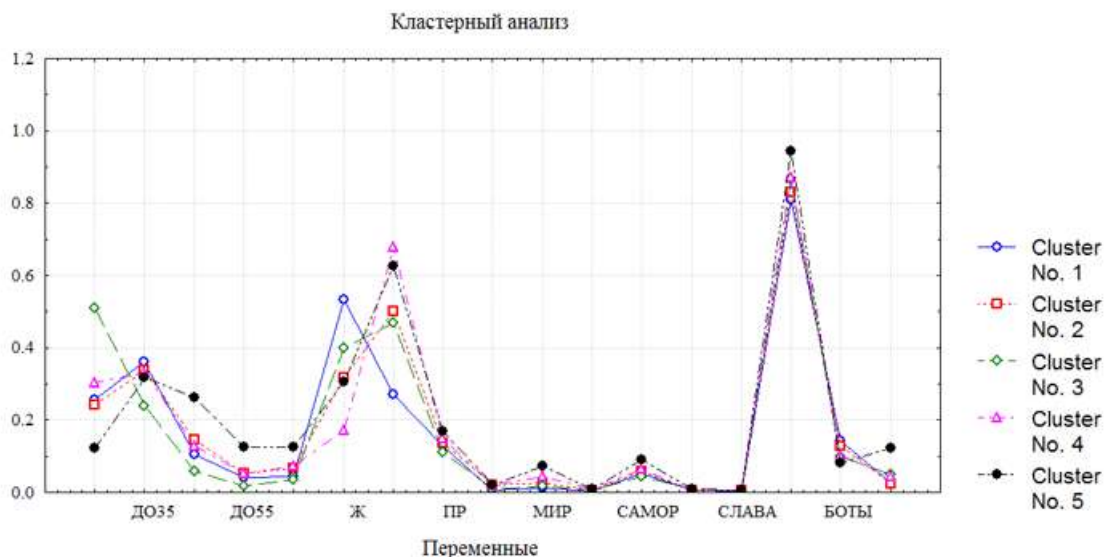


Рисунок 3 – Средние значения переменных для каждого кластера

На основании расчетов и анализа данных следующим образом произведено распределение по кластерам:

1. В первый кластер вошли группы, созданные для лиц, увлекающихся путешествиями, географией, здоровьем, домашними животными и домашним хозяйством. Все эти группы были отмечены у нас как группы по интересам. В этот же кластер вошли группы «Первый канал», «Типичный Донецк» и «Donbass.media Group». Т.е. первый кластер объединил участников без видимого интереса к политике, что показывает и то, что перечисленные три группы, определенные нами ранее как политические, в полной мере таковыми не являются.
2. Во второй кластер вошли группы, которые мы определили как созерцателей. Это группа «Киномания», «Discovery», «Интересная планета», «Бумажный самолетик». Во всех группах, как правило, размещаются красивые фотографии, описания интересных мест. Все перечисленные группы входили в группы по интересам.
3. В третий кластер вошли наиболее остро критикующие группы: «Сводки ополчения Новороссии», «Хроника вставания с колен», «Злюкен Енотен», «Еноты Новороссии», «Очищение», «Народный журналист». Ранее все группы были отнесены к группам критиков.

4. В четвертый кластер вошли группы «Оппозиция», КПРФ, ЛДПР, «Высокие технологии», «Рыбалка», «Преступная Россия», «Другая Россия», «Левый фронт». Отметим, что в этот кластер вошли, как нейтральные группы, так и группы –критики и политические группы. Группы «Другая Россия», «Левый фронт» считаются группами оппозиционных партий.

5. В пятый кластер вошли группы «Единая Россия», «Народное ополчение Павла Губарева», «Донецкая республика», «Комсомольская правда», «МИД России», «Луганск. Новости ЛНР», «Политика», «Права человека», это группы политические, воспринимающие все в позитивном ракурсе.

Можно сделать вывод, что в целом рассмотренный контингент участников делится на пять кластеров. Кластер сообществ абсолютно индифферентных к политике (кластер 2); кластер лиц имеющих свои увлечения и предпочитающих получать краткие политические новости (кластер 1). Кластер критикующих (кластер 3) и два кластера политической направленности, причем в одном из них в группы выкладываются только позитивные новости (кластер 5), в другом возможна незначительная критика (кластер 4).

Общее количество участников во всех проанализированных группах более 33млн. чел. Как видно из табл.5, почти 90% сообщества не

интересуются политическими темами, 1.24% политикой интересуются, но в целом их критикуют происходящие события, 9% устраивает текущее положение дел.

Таблица 5. Процентное соотношение кластеров

Кластер	Группы по интересам	Созерцатели	Критики	Нейтральные группы, критики, позитивные группы	Позитивные группы	Количество участников
№	1	2	3	4	5	
Участники	1 2416 300	16 972 105	411 020	2 148 030	1 199 613	33 147 068
%	37.46%	51.2%	1.24%	6.28%	3.62%	100%

Полученные данные не противоречат данным других источников [10,11]. Как правило, политически активными является лишь незначительная (до 10%) часть общества, однако, как показывают исследования, политическая активность населения растет [12].

Отметим, что если участники второго кластера не имеют склонности к обсуждению политических новостей, то вошедшие в третий кластер сообщества могут стать политизированными или уже являются ими, а значит могут перейти или в разряд позитивных, проправительственных групп или стать оппозицией.

Выводы

Политизация социальных сетей делает важным вопрос их изучения именно с точки зрения поддержки или отрицания власти.

Литература

1. SECL Group. Н. Семенов. Все о социальных сетях. Влияние на человека. Режим доступа: <https://secl.com.ua/article-vse-o-socialnyh-setjah-vlijaniye-na-cheloveka.html>

2. Федорченко С.Н. Глобальное исследование политизации социальных сетей // Обозреватель -Observer, 2016. №8(319). С. 57-67.

3. Тематическая классификация сообществ в социальной сети «ВКонтакте» как нового средства массовой информации. Морозова А.А.: Сборник Сучасная медиясфера: практика трансфармації, тээрэтычнае асэнсаванне, інстытуцыянальныя перспектывы матэрыялы I Міжнароднай навукова-практычнай канферэнцыі. С. В. Дубовік (адказны рэдактар). 2017. С. 160-166.

4. Brand Analytics. Социальные сети в России, лето 2017: цифры и тренды. Режим доступа: <http://blog.br-analytics.ru/sotsialnye-seti-v-rossii-letu-2017-tsfiry-i-trendy/>

В 2010 году в США блогосфера признана самостоятельным направлением реализации внешней политики США, что подчеркивает важность роли Интернета и необходимости активного присутствия в нем политических партий и организаций [13].

Чтобы побеждать в информационной войне необходимо выстраивать многомерные сети на основе информационных, культурных, политических и других компонентов.

Чтобы выстраивать такие сети, необходимы разработки моделей сообществ, изучения личностных характеристик их участников, необходимы моделирование их поведения и прогнозирования тенденций вовлечения нейтральных, неполитизированных членов сообществ в политические группы.

5. Мануэль Кастельс. Информационная эпоха: экономика, общество и культура. - ГУ ВШЭ, 2000.-608с.

6. Анализ сообществ. On-line сервис Smmup.ru. Режим доступа: <http://smmup.ru/activity.php>

7. Боровиков В.П., Боровиков И.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. - М.: Филинь, 1998.- 608с.

8. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. -М.: Горячая линия, 2016. -288с.

9. Statsoft. Электронный учебник по статистике. Режим доступа: <http://statsoft.ru>

10. Вайдлих. Социодинамика: системный подход к математическому моделированию в социальных науках.- М.: Едиториал, 2004. - 480с.

11. Манипуляция обществом или истинный размер пяти процентов. Режим доступа: <https://alex-leshy.livejournal.com/480772.html>

12. Рощина Е.В., Анохина И.Ю. Жизненный цикл групп в социальных сетях/Информатика, управляющие системы, математическое и компьютерное моделирование в рамках III форума «Инновационные перспективы Донбасса» (ИУСМКМ – 2017): VIII Международная научно-техническая

конференция, 25 мая 2017, г. Донецк: / Донец. национал. техн. ун-т; Донецк: ДонНТУ, 2017, с. 73-77.

13. Филимонов Г., Градосельская Г. Поле битвы — соцсети. Особенности нынешней информационной войны. Режим доступа: <http://svpressa.ru/society/article/147378/>

Анохина И. Ю., Рощина Е. В. Кластеризация сообществ социальной сети «ВКонтакте». Рассматриваются вопросы анализа социальных сетей. Проведен мониторинг и анализ тридцати восьми групп различных тематик, определены статистические показатели для оценки количественных и качественных характеристик участников групп. Проведена оценка уровня политизации групп. Проведен анализ существенности различий в личностных характеристиках участников сообществ различных групп. Методами кластерного анализа проведено разбиение групп на кластеры по уровню политизации.

Ключевые слова: социальные сети, ВКонтакте, кластерный анализ, структурные характеристики, мониторинг социальных сетей, политическая активность, политизация социальных сетей.

I. Yu. Anokhina, E. V. Roshchina. Analysis of communities in social networks. The analysis of social networks on the example of the social network "VKontakte" is considered. The example of thirty-eight groups of different subjects identified statistical indicators to assess the quantitative and qualitative characteristics of groups. Groups of three types were considered: apolitical groups, groups of political orientation without a critical component and a group of critics whose participants criticized the majority of the events under consideration. A t-test of communities was carried out, the presence of significant differences between the contingent of groups was revealed. By cluster analysis methods, all groups are divided into clusters, composition and percentage of clusters are determined, their characteristics are analyzed. The percentage of clusters that characterizes the degree of politicization of communities is determined.

Keywords: social networks, VKontakte, cluster analysis, structural characteristics, monitoring of social networks, political activity, politicization of social networks.

Статья поступила в редакцию 15 мая 2018 г.
Рекомендована к публикации профессором Павлышом В. Н.