

УДК 316.354.4

Симуляция эхо-камер в сетевом анализе онлайн-сообществ в социальной сети «ВКонтакте» на примере петербургского политического сегмента сети

Суслов Сергей Игоревич

Аспирант,

кафедра политических институтов и прикладных политических исследований,

Санкт-Петербургский государственный университет,

191060, Российская Федерация, Санкт-Петербург, ул. Смольного, 1/3;

e-mail: s.suslov@spbu.ru

Аннотация

В статье проведен сетевой анализ 111 политических петербургских онлайн-сообществ (либеральных, националистических, социалистических и ЛГБТ-сообществ) в социальной сети «ВКонтакте». Выборка сообществ составлялась методом снежного кома: от найденных по названиям петербургских сообществ к новым сообществам из ссылок и из указания источника перепоста публикаций. Поле собрано с помощью метода `groups.getMembers VK.API`. Вершинами графа являются политические онлайн-сообщества, а ребрами – количество общей аудитории. Из полученных данных была составлена математическая модель взвешенного графа. Экспериментально по метрике модулярности определен порог для симуляции эхо-камер. Проведена кластеризация двумя алгоритмами кластеризации графов: алгоритмом «распространения частотных меток» (Label Propagation Algorithm) и алгоритмом «глухой телефон» (Chinese Whispers Clustering Algorithm). Первый разбил граф на четыре кластера, совпадающие с идеологическими группами, а второй разделил на три, расколов социалистов. Присутствует высокая смежность аудитории между сообществами националистов и большинством сообществ социалистов. Для каждой кластеризации было произведено сравнение метрик центральности кластеров. В результате исследователи обнаружили кластеры либералов и националистов как центры возможной поляризации.

Для цитирования в научных исследованиях

Суслов С. И. Симуляция эхо-камер в сетевом анализе онлайн-сообществ в социальной сети «ВКонтакте» на примере петербургского политического сегмента сети // Теории и проблемы политических исследований. 2016. № 3. С. 136-146.

Ключевые слова

Сетевой анализ, симуляция, взвешенный граф, целевая аудитория, смежная аудитория, центральность, онлайн-сообщества, эхо-камеры, социальные сети, политические сообщества.

Введение

Сетевой анализ исходит из предположения, что связи, которые образуют предметы анализа, определяют их многие характеристики. В основе сетевого анализа лежит математическое моделирование, заключающееся в построении графа отношений между акторами. В качестве объекта исследования выступают политические онлайн-сообщества города Санкт-Петербурга в социальной сети «ВКонтакте». Предметом анализа являются 111 сообществ, различных по своей идеологической направленности: либеральные, националистические, социалистические и ЛГБТ. Выборка сообществ составлялась методом снежного кома: от найденных по названиям петербургских сообществ к новым сообществам из ссылок и из указания источника перепоста публикаций.

Вершинами графа являются политические онлайн-сообщества, а ребрами – количество общей аудитории. Для извлечения множества подписчиков использовался метод API VK groups.getMembers [Список методов VKAPI, www]. Осуществлялось попарное сравнение множеств подписчиков в квадратной матрице размерностью 111 выше главной диагонали матрицы. Таким образом, был получен взвешенный неориентированный граф, где атрибутом веса ребра выступает количество общей аудитории.

Отличительной особенностью сетевого анализа является подсчет центральности (centrality) – выявление наиболее важных узлов или вершин в пределах графа. В сетевых исследованиях центральность применяется для определения влиятельных лиц в организациях, выявления ключевых узлов инфраструктуры, распространителей болезней, информации и другого. Многие идеи и термины в рамках центральности как метода пришли из общественных наук и отражают социологическое происхождение [Newman, 2010, 168]. Примеры подобных терминов – сплоченность, солидарность, членство [Hansen, Shneiderman, Smith, 2010, 3]. Узлы на основании своих связей характеризуются мерами центральности, то есть расположением в сети: степень узла (degrees), промежуточная центральность (betweenness centrality) [Freeman, 1977], центральность по близости (closeness centrality) [Bavelas, 1950], центральность по престижу (eigenvector centrality) [Jackson, 2010, 65]. Кроме того, граф можно охарактеризовать рядом параметров: диаметр графа, средняя степень, средняя длина пути, модулярность, количество связанных компонент, средний коэффициент кластеризации и т. п.

Другой важной частью исследования является симуляция эхо-камер (echo chambers). Касс Санстейн утверждает [Sunstein, 2001], что Интернет способствует созданию «эхо-

камер», в которых пользователи с общими интересами самостоятельно группируются в небольшие коллективы со схожими взглядами. Это приводит к поляризации мнений в обществе из-за фильтрации новостей и информации, которая не соответствует уже существующим взглядам. Центр Беркмана Гарвардского университета в своем сетевом анализе гиперссылок российской блогосферы пришел к выводу, что в российском сегменте Интернета отсутствует четкое разграничение на политические эхо-камеры [Etling et al., 2010]. Тем не менее возможность поляризации в отсутствие сильной многополярной политической структуры может рассматриваться как повод для беспокойства. Алан Цанг (Alan Tsang) и Кейт Ларсон (Kate Larson) полагают, что социальные сети способствуют явлению гомофилии [Newman, 2006], которая способствует распространению одной идеологии среди друзей [McPherson, Smith-Lovin, Cook, 2001].

Изучая нашу модель, мы пришли к выводу, что большинство ребер графа имеют низкий вес и неравномерное распределение. Медиана составляет 12, арифметическое среднее – 47,52, первый квантиль – 4, третий квантиль – 37. Стандартное отклонение равно 135,95. Основываясь на этом, можно сказать, что, убрав слабые по весу связи, мы сможем провести симуляцию возможности поляризации на эхо-камеры и изучить их возможные характеристики.

Для нахождения порога появления эхо-камер нами был проведен эксперимент. Последовательно с шагом 5 удалялись все значения веса ребер графа от 0 до 150 с постоянным

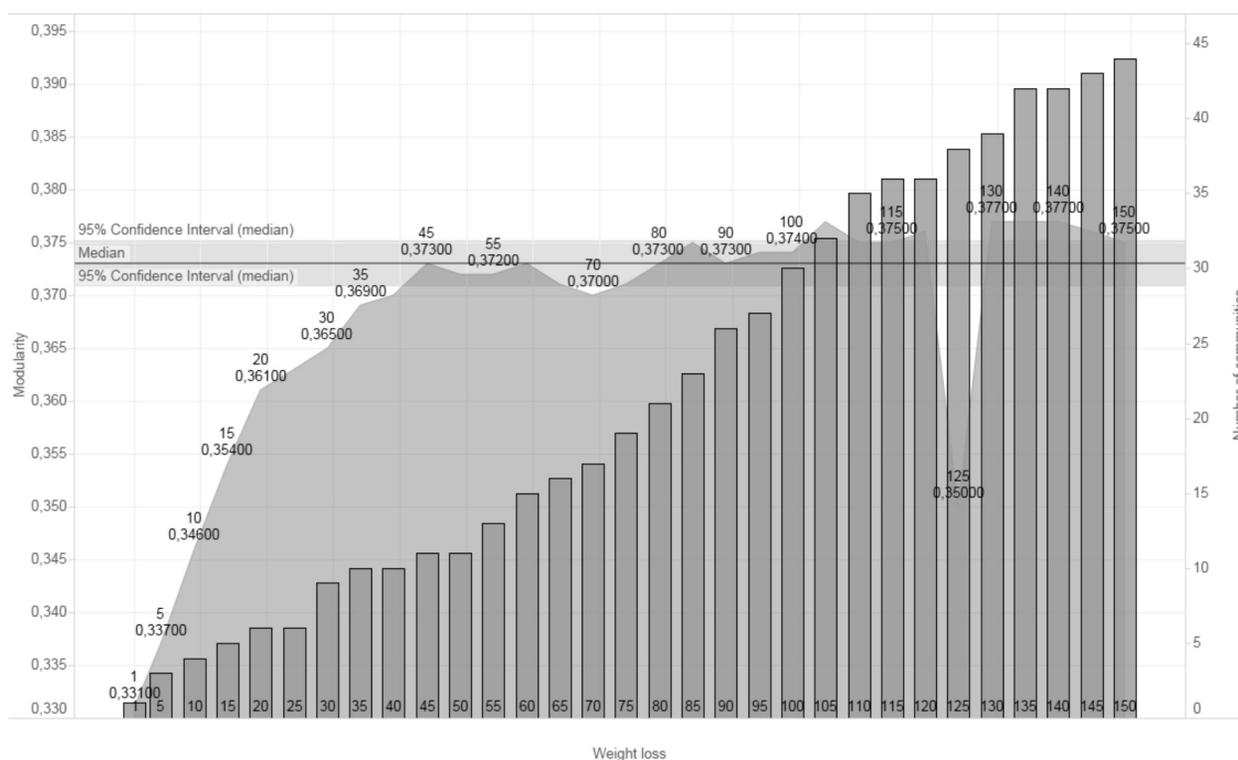


Рисунок 1. График зависимости модулярности от веса ребер, подвергнутых удалению из графа. Аппроксимация точек в уравнение кубической функции (столбец гистограммы определяет количество сообществ-изолятов)

замером модулярности количества сообществ [Tsang, Larson, 2016], на которые распадается граф. Таким образом, был создан график зависимости модулярности от шага, где была проведена аппроксимация значений в кубическую функцию.

На графике заметно, что функция перестает увеличиваться на 11 шаге в значении веса 50. Кроме того, количество сообществ в данном значении остается стабильным. Немаловажно, что данное значение близко к арифметическому среднему. Данное значение было выбрано идеальным значением для порога, после которого в графе останутся только сильно связанные узлы. Восемь сообществ из либеральных («Петербург с Украиной!», «МПГ Санкт-Петербург», «Взаимопомощь оппозиции») и социалистических («Классовый профсоюз «Защита» – СПб и ЛО», «Санкт-Петербургская городская пионерская организация», «Союз воинствующих безбожников – Ленинград», «Движение в защиту детства С-Пб и Лен область», «ОКП и АВРОРА (Ораниенбаумский плацдарм)») оказались изолятами после данной процедуры: все их связи имели вес менее 50 общих подписчиков.

Кластеризация алгоритмом «распространения частотных меток»

Алгоритм «распространения частотных меток» (Label Propagation Algorithm) разделил граф на четыре кластера, которые полностью совпадают с вышеописанным идеологическим разделением сообществ на группы. Таким образом, модели онлайн-дискурса зависят от уже существующих политических характеристик конкретных обществ. Более того, это подтверждает верность выборки с методом снежного кома.

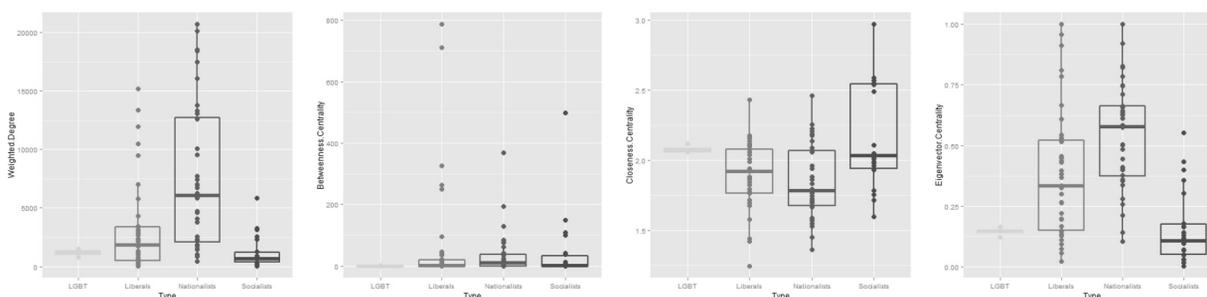


Рисунок 2. Диаграммы размаха метрик центральности кластеров по алгоритму «распространения частотных меток» взвешенного графа политических петербургских онлайн-сообществ «ВКонтакте»

Обратимся к мерам центральности кластеров в графе. Диаграмма размаха (box plot) взвешенной степени узлов графа (weighted degree) показывает, что самый большой размах, а также межквантильный размах мы можем заметить у националистов, а меньший – у либералов. С другой стороны, распределение у националистов ближе к нормальному. Медиана у националистов равняется 6056, тогда как у либералов – 1879, социалистов – 703,5,

ЛГБТ – 1200. Взвешенная степень узла в данном случае свидетельствует о двух вещах: как плотно связаны сообщества в рамках кластера, а также каково количественное измерение смежности аудиторий, интенсивность такой связи в виде наличия общих потребителей информационных благ или участников коммуникации. Группа националистов показывает высокую центральность по взвешенной степени: первый квантиль равняется 2083, когда только медиана в группе либералов достигает значения 1879. Более того, для группы националистов характерно наличие крупных сообществ – распределение наблюдений после медианы имеет более пологую структуру, чем до медианы. Таким образом, мы уже можем увидеть особенность националистической группы – сильно развитую сплоченность. Множество пересекающихся аудиторий формирует широкое метасообщество. Кроме того, попарное сравнение непараметрическим U-тестом Манна-Уитни взвешенной степени узлов кластера националистов показывает отличие этого кластера от остальных, а также отличие либерального кластера от кластера социалистов.

Диаграмма размаха метрики промежуточной центральности (*betweenness centrality*) отличается от предыдущей и подчеркивает сетевые отличия тактик разных идеологических групп. Высокими показателями отличаются социалисты и националисты. Медиана по данной центральности самая высокая у националистов – 9,49, затем идут социалисты – 1,82, либералы – 1 и ЛГБТ – 0,23. Примечательно, что наблюдения от медианы до третьего квантиля у националистов и социалистов имеют большой разброс. Так, третий квантиль у националистов – 69,97, у социалистов – 31,83, у либералов – всего 19,47. В совокупности для социалистических и националистических сообществ оказалась выгодна тактика посредничества для политически заинтересованной аудитории.

Центральность по близости также дает представления об особенностях расположения сообществ относительно центра-периферии политического дискурса. Самыми заметными являются социалисты и ЛГБТ как аутсайдеры в политическом сообществе социальной сети. Половина наблюдений социалистов имеет высокую метрику (что хуже, нежели низкую) от 3 до 2, когда только приблизительно 25% либеральных и националистических сообществ имеют показатели от 2,5 до 2. Из-за этого социалисты имеют такую высокую медиану – 2,03, тогда как у либералов – 1,92 и националистов – 1,78, которая чуть выше первого квантиля либералов (1,76). Медиана ЛГБТ сообществ – 2,06.

Центральность по престижу доказывает еще раз, что социалисты являются периферией данного графа. 75% наблюдений имеют низкие показатели центральности – 0,1, тогда как только 25% наблюдений либералов имеют схожее значение. Медианы кластеров выявляют значительные отличия между кластерами. Самая высокая у националистов – 0,57 (она даже выше третьего квантиля либералов 0,52), затем либералы – 0,33 (которая ниже, чем первый квантиль националистов 0,37), затем ЛГБТ – 0,14 и социалисты – 0,1. U-тест показывает, что практически все выборки являются независимыми, кроме попарного сравнения либералов и ЛГБТ-сообществ, а также ЛГБТ-сообществ и социалистов.

Кластеризация алгоритмом «глухой телефон»

В отличие от предыдущего алгоритма, кластеризация методом «глухой телефон» (Chinese Whispers Clustering Algorithm) разделила граф на три кластера. Алгоритм прежде всего ориентируется на модулярность – 0,372 и число сообществ – три. Группа социалистов была разбита между либералами и националистами. Прежде всего, потому что сообщества группы социалистов имели низкие показатели центральности по взвешенной степени узла, но при этом имели высокую промежуточную центральность практически на уровне либералов и в два раза выше, чем у группы ЛГБТ, а также высокую (что хуже, нежели низкую) центральность по близости: медиана выше, чем у либералов. Три сообщества из группы социалистов: «РСД – Петербург – Российское социалистическое движение», «Студенческое Действие – Петербург» и «Социальный марш в Петербурге» были посчитаны как относящиеся к кластеру с большинством с либералами, а остальные 19 сообществ были посчитаны в составе кластера с 36 националистическими сообществами, образовав национал-социалистический кластер. На наш взгляд, провести демаркационную линию между социалистическими сообществами без глубокого анализа было бы затруднительно. По количеству подписчиков сообщества заметно отличаются: меньшинство из социалистических сообществ, которые перешли в либеральный кластер, малочисленны: сумма их неуникальных подписчиков равняется 1609, в пересчете на одно сообщество – 536,33; большинство социалистических сообществ, создавшие условно называемый национал-социалистический кластер, многочисленнее: количество неуникальных подписчиков – 34356, тогда как как в среднем численность равняется 1808,21. По метрикам центральности больших отличий нет. Средняя взвешенная степень на один узел практически одинаковая: меньшинство – 1142,66, большинство – 1179,84. Центральность по близости на один узел имеет небольшое различие: меньшинство – 1,93, большинство – 2,2. Центральность по престижу на один узел: меньшинство – 0,22, большинство – 0,14. Отличие есть только в промежуточной центральности. В подгруппе большинства хоть и находится множество узлов с нулевой промежуточной центральностью, но значение на один узел получается 48,76, тогда как у меньшинства – 13,55. Более того, кластеризующий коэффициент меньшинства в два раза больше коэффициента большинства: 0,33 и 0,14. Такая разница объясняется лишь разной топологией сети, которая выступает основным критерием для определения кластера. Любопытным выглядит тот факт, что возможная поляризация разделила идеологическое сообщество.

Обратимся теперь к диаграммам размаха мер центральности в полученных кластерах. В основном изменения коснулись нового кластера, собранного из националистической и большинства социалистической групп и в меньшей степени либерального кластера, ЛГБТ остались без изменений. Во взвешенной степени узла произошли следующие изменения: количество наблюдений до медианы в национал-социалистическом кластере заметно возросло, от этого медиана немного «просела» и приблизилась к значению либерального кла-

стера: 21 и 17. Квантили, тем не менее, немного выше значений либерального кластера. К националистам прибавились слабосвязанные с националистическим ядром и малочисленные сообщества.

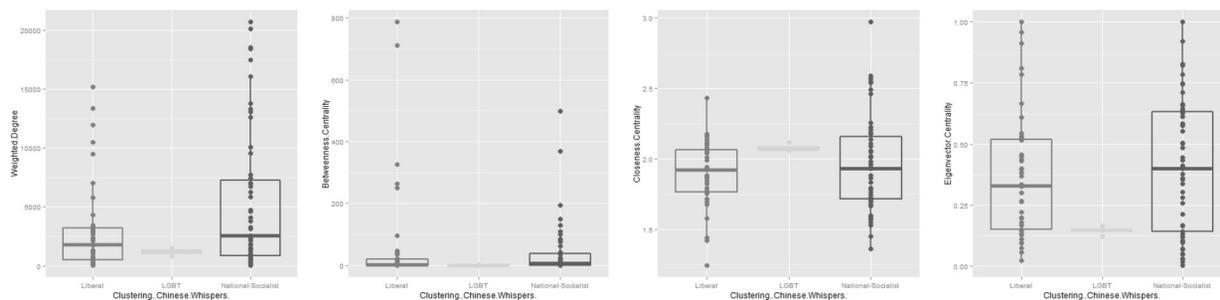


Рисунок 3. Диаграммы размаха метрик центральности кластеров по алгоритму «глухого телефона» взвешенного графа политических петербургских онлайн-сообществ «ВКонтакте»

Единственные высокие показатели промежуточной центральности подгруппы «большевиков» в совокупности дали новому национал-социалистическому кластеру заметное преимущество: медиана у либералов –1,24, у национал-социалистов – 4,87. Кроме того, третий квантиль у либералов – 19,91, у национал-социалистов – 37,56. С другой стороны, либералы имеют по-прежнему выдающиеся наблюдения с максимальными значениями: из-за этого среднее арифметическое у либералов – 61,22, когда у национал-социалистов – всего 40,65.

В диаграмме размаха центральности по близости показатели национал-социалистов изменились в худшую сторону. Медиана либералов и национал-социалистов практически сравнялась: 1,92 и 1,97. Практически одинаковые показатели мы наблюдаем в первом и третьем квантилях: 1,76 и 2,06 – у либералов, 1,71 и 2,15 – у национал-социалистов.

Считая метрику центральности по престижу, можно заметить, что у националистического ядра появилась слабосвязанная периферия. Из-за этого медианы либерального и национал-социалистического кластера практически совпали: 0,33 и 0,39. Разброс между первым и третьим квантилями в национал-социалистическом кластере стал заметно шире, даже больше, чем у либералов.

Приведена таблица основных метрик графа (табл. 1). Для каждого кластера были подсчитаны основные метрики субграфа отдельно от остальных кластеров. Что касается ЛГБТ сообществ, их всего четыре, поэтому показатели выглядят такими высокими. В целом, выдающимися является только националистические и либеральные сообщества. Главная отличительная особенность националистов – количество и сплоченность, поэтому средняя степень графа у них даже выше основного графа, средняя взвешенная степень – выше, чем у основного графа, и в несколько раз выше, чем у либералов. С другой стороны, результат

кластеризации в виде «национал-социалистического» в отдельном рассмотрении распадается на два кластера, что подчеркивает высокую разницу между социалистами и националистами в топологии сети. Диаметр графа равен двум и показывает, что топология сети сообществ такая же, как и топология эго-сетей. Либеральные сообщества показывают короткую среднюю длину пути, но она еще короче в националистических сообществах. Также отличие либералов и националистов заключается в плотности графа: у либералов – 0,42, у националистов – 0,67. В симуляции эхо-камер это говорит только о том, что остались лишь крупные пересечения аудиторий и их много в обоих кластерах. Кроме того, либералы и националисты могут похвастаться высокой взвешенной кластеризацией. Только в этих сообществах она наблюдается выше, чем у главного графа. Объяснить это можно лишь большим количеством сторонников, для которых цифровые технологии Web 2.0 способствуют сотрудничеству через социальные сети, а также являются прекрасной площадкой для распространения идей [Ferguson, 2008; Jackson, Lilleker, 2009].

Таблица 1. Основные метрики главного графа и разных субграфов (кластеры, группы)

	Главный	Либералы	Либералы*	ЛГБТ	Националисты	Национал-социалисты	Социалисты
Средняя степень графа	22,03	17,02	17	3	23,66	18,98	6,63
Средняя взвешенная степень	4157,24	2497,26	2427,13	689,5	6799,66	4869,2	760
Диаметр графа	4	3	3	1	2	4	3
Средняя длина пути	1,93	1,58	1,61	1	1,32	1,91	1,72
Плотность графа	0,21	0,42	0,39	1	0,67	0,35	0,31
Средний коэффициент кластеризации	0,8	0,85	0,85	1	0,86	0,81	0,77
Средний взвешенный коэффициент кластеризации	0,83	0,86	0,86	1	0,93	0,83	0,71
Коэффициент кластеризации	0,77	0,81	0,81	1	0,86	0,76	0,66

Заключение

Симуляция эхо-камеры помогла сделать интересные выводы. В первую очередь, заметно, что симуляция поляризации показывает мощные политические объединения. В наших результатах – это либеральные и националистические сообщества. Метрики их субграфов показывают самые высокие показатели, что позволяет судить о центрах поляризации. Метрики центральности говорят о том, что они в целом схожи, но имеют основное отличие в степени сплоченности.

Кроме того, данная симуляция показала периферийные компоненты: ЛГБТ-сообщества и социалистические сообщества. Если ЛГБТ-сообщества существуют изолиро-

ванно и сплоченно, то социалистические сообщества занимают промежуточное положение в политической сети, а метрики их субграфа показывают самый низкий коэффициент кластеризации, низкую плотность графа и одну из самых высоких средних длин пути. Ожидается, что они могут быть расколоты, – это показывает один из алгоритмов кластеризации.

Появление двух крупных кластеров именно либерального и национал-социалистического (национал-патриотического) толка не является новым для изучения виртуальных сообществ Рунета. В ходе изучения российской блогосферы Центром Беркмана также были выделены схожие по идеологической окраске кластеры.

Библиография

1. Список методов VKAPI. URL: <https://vk.com/dev/groups.getMembers>.
2. Bavelas A. Communication patterns in task-oriented groups // *The Journal of the Acoustical Society of America*. 1950. № 22 (6). P. 725-730.
3. Etling B., Alexanyan K., Kelly J. et al. Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization. URL: https://cyber.law.harvard.edu/publications/2010/Public_Discourse_Russian_Blogosphere
4. Ferguson R. Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing // *Journal of Consumer Marketing*. 2008. № 25 (3). P. 179-182.
5. Freeman L.C. A set of measures of centrality based on betweenness // *Sociometry*. 1977. № 40. P. 35-41.
6. Hansen D., Shneiderman B., Smith M.A. Analyzing social media networks with NodeXL: Insights from a Connected World. Morgan Kaufmann, 2010. 304 p.
7. Jackson M.O. Social and Economic Networks. Princeton University Press, 2010. 520 p.
8. Jackson N., Lilleker D.G. Building an architecture of participation? Political parties and web 2.0 in Britain // *Journal of Information Technology & Politics*. 2009. № 6. P. 232-250.
9. McPherson M., Smith-Lovin L., Cook J.M. Birds of a Feather: Homophily in Social Networks // *Annual Review of Sociology*. 2001. № 27. P. 415-444.
10. Newman M.E.J. Modularity and community structure in networks // *Proceedings of the National Academy of Sciences of the United States of America*. 2006. № 103 (23). P. 8577-8696.
11. Newman M.E.J. Networks: An Introduction. Oxford University Press, 2010. 720 p.
12. Sunstein C. Echo chambers. Princeton University Press, 2001. Digital book.
13. Tsang A., Larson K. The Echo Chamber: Strategic Voting and Homophily in Social Networks // Thangarajah J., Tuyls K., Jonker C., Marsella S. (eds). Fifteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016). Singapore, 2016. P. 368-375.

Echo chamber's simulation in social network analysis of online communities in social network "VKontakte" by the example of the St. Petersburg political network segment

Sergei I. Suslov

Postgraduate,
Department of political institutions and applied political research,
Saint Petersburg State University,
191060, 1/3 Smol'nogo str., St. Petersburg, Russian Federation;
e-mail: s.suslov@spbu.ru

Abstract

The author puts to use social network analysis of 111 St. Petersburg online political communities (liberal, nationalist, socialist and LGBT community) in the social network "VKontakte". The researchers took the chain-referral sampling (or snowball sampling) to derive the list of communities' ids from links and reposted messages made in first community. The data mining was realized by groups.getMembers VK.API method. We pieced the mathematical model of weighted graph out of these data. Communities' IDs are represented as the nodes in graph. The value of intersection sets of both communities' audience (number of common followers) are represented as weight of the edges. Experiment has shown the optimal step of echo chamber's simulation in relation with the modularity metric. Label Propagation Algorithm and Chinese Whispers Clustering Algorithm separate the graph in different ways. The first divide graph into four clusters that interconnected with the ideological groups. The second divide graph into three cluster, splitting the socialist's communities. Moreover, there is a high interconnection between nationalist's communities and a majority of the socialist's communities. For each clustering researchers produced a comparison of metrics of centrality. As a result, communities of liberals and nationalists are regarded to be centers for the possible polarization.

For citation

Suslov S.I. (2016) Simuljatsiya eho-kamer v setevom analize onlain-soobshhestv v social'noi seti "VKontakte" na primere peterburgskogo politicheskogo segmenta seti [Simulation of echo chambers in network analysis of online communities in social network "VKontakte" on the example of the St. Petersburg political network segment]. *Teorii i problemy politicheskikh issledovanii* [Theories and Problems of Political Studies], 3, pp. 136-146.

Keywords

Social network analysis, simulation, weighted graph, target audience, audience adjacency, centrality, online community, echo chamber, social networks, political community.

References

1. Bavelas A. (1950) Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22 (6), pp. 725-730.
2. Etling B., Alexanyan K., Kelly J. et al. (2010) *Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization*. Available at: https://cyber.law.harvard.edu/publications/2010/Public_Discourse_Russian_Blogosphere [Accessed 17/04/2016].
3. Ferguson R. (2008) Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing. *Journal of Consumer Marketing*, 25 (3), pp. 179-182.
4. Freeman L.C. (1977) A set of measures of centrality based on betweenness. *Sociometry*, 40, pp. 35-41.
5. Hansen D.L., Shneiderman B., Smith M.A. (2010) *Analyzing social media networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann.
6. Jackson M.O. (2010) *Social and Economic Networks*. Princeton University Press.
7. Jackson N., Lilleker D.G. (2009) Building an architecture of participation? Political parties and web 2.0 in Britain. *Journal of Information Technology & Politics*, 6, pp. 232-250.
8. McPherson M., Smith-Lovin L., Cook J.M. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, pp. 415-444.
9. Newman M.E.J. (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (23), pp. 8577-8696.
10. Newman M.E.J. (2010) *Networks: An Introduction*. Oxford University Press.
11. *Spisok metodov VKAPI* [The list of methods VKAPI]. Available at: <https://vk.com/dev/groups.getMembers> [Accessed 17/04/2016].
12. Sunstein C. (2001) *Echo chambers*. Princeton University Press.
13. Tsang A., Larson K. (2016) The Echo Chamber: Strategic Voting and Homophily in Social Networks. In: Thangarajah J., Tuyls K., Jonker C., Marsella S. (eds). *Fifteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)*. Singapore, pp. 368-375.