

## ЭВРИСТИЧЕСКИЕ ВОЗМОЖНОСТИ КАРТОГРАФИРОВАНИЯ СЕТЕЙ И КАЧЕСТВЕННОГО КОДИРОВАНИЯ В ПОЛИТОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ СОЦИАЛЬНЫХ МЕДИА

*Д.С. Жуков*, доцент Тамбовского государственного университета им. Г.Р. Державина, канд. ист. наук, *ineternatum@mail.ru*

Рецензент: Д.В. Михлик, нач. отд., ФГБОУ ВО «Тамбовский государственный университет», канд. техн. наук, *dim\_tmb@mail.ru*

*В статье рассмотрены некоторые эвристически продуктивные инструменты для изучения сетевых кластеров – наборов связанных сообществ в социальных медиа. Автор представляет возможности картографирования сети и качественного кодирования для мониторинга политизированных виртуальных групп. Для картографирования изучаемых кластеров предложен алгоритм «разматывания клубка» (модификация метода «снежного кома»); для анализа состава кластера – приемы качественного кодирования с закрытым и открытым (пополняемым) наборами кодов. Обсуждаются различные наборы кодов (как номинальных, так и ранговых), позволяющие зафиксировать ключевые качественные свойства сообществ и, соответственно, получить количественно-качественные характеристики изучаемого кластера.*

**Ключевые слова:** сетевой анализ, политизированные сообщества, инструментарий, картографирование сети, качественное кодирование.

## HEURISTIC POSSIBILITIES OF NETWORK MAPPING AND QUALITATIVE CODING IN POLITICAL SCIENCE RESEARCH OF SOCIAL MEDIA

*D.S. Zhukov*, Associate Professor, Tambov State University named after G.R. Derzhavin, Doctor of History, *ineternatum@mail.ru*

*The article considers some heuristically productive tools for studying network clusters – sets of connected communities in social media. The author presents the possibilities of network mapping and qualitative coding for monitoring politicized virtual groups. For mapping the studied clusters, the algorithm of «unwinding the tangle» (a modification of the «snowball» method) is proposed; for analyzing the cluster composition, qualitative coding techniques with closed and open (replenishable) code sets are proposed. Different sets of codes (both nominal and ranked) are discussed, which allow to fix key qualitative properties of communities and, accordingly, to obtain quantitative-qualitative characteristics of the studied cluster.*

**Keywords:** network analysis, politicized communities, toolkit, network mapping, qualitative coding.

### Введение

Методология и инструментарий исследования социальных медиа переживают взрывообразный рост в течение двух последних десятилетий. Это связано не только с появлением и расширением социальных сетей, но и с их интеграцией в социополитические отношения современных обществ. Воздействие сетевых феноменов на реальную политику стало зримым и признанным. Эта статья сфокусирована на исследовательских техниках, которые позволяют идентифицировать и атрибутировать политизированные сообщества в социаль-

ных сетях и их связанные скопления (кластеры). Мы рассмотрим две группы инструментов — картографирование и качественное кодирование, обсудим принципы их использования в политологическом предметном пространстве.

Потребность в пополнении методического арсенала социополитических наук связана с необходимостью обработки массовых источников и, в некоторых случаях, больших данных для реконструкции сетевых кластеров и их отдельных элементов. Качественные и количественные методы изучения локальных феноменов (будь то отдельные сообщества, персоны или даже целые информационные кампании) представляют безусловный интерес как основной поставщик рабочих гипотез, но довольно редко могут дать общее представление о структуре больших сегментов сети (сетевых кластеров) и о процессах, в них протекающих.

Картографирование позволяет не только визуализировать сеть, но и применять к ней разнообразные метрики сетевого анализа, тогда как качественное кодирование дает материал для широкого спектра математико-статистических методов, в частности для корреляционного анализа. И картографирование, и качественное кодирование имеют множество вариаций (являются, по существу, группами методов), выбор между которыми зависит от исходных данных, исследовательских задач и ресурсов, вовлеченных в работу (включая трудовые затраты, специализированное программное обеспечение и пр.). Эти методы мы применили для исследования политизированных сообществ (групп) в VK — под ними мы подразумеваем технически обособленные объединения пользователей, которые (вне зависимости от своей основной тематики) содержат высказывания о текущей политике (и/или политическом устройстве, политизированной истории и политическом целеполагании) не менее чем в 5% постов и/или комментариев. Очевидно, что сетевые сообщества «живут» не изолированно, а в некоторой среде — своеобразной экосистеме, с которой обмениваются смыслами и ресурсами (в том числе влиянием) и пр. Такая экосистема — это кластер, в котором сообщества связаны каналами рефлексивности. Рефлексивность — способность воспринимать и отражать (распространять) сетевые сообщения — должна быть зафиксирована и формализована (измерена) при построении кластера. Эта задача представляет собой отдельную принципиально важную проблему.

### **Литература**

В российской политологии распространяются представления о возрастающем (и в некоторых случаях — определяющем) влиянии сетевых и виртуальных феноменов на политические процессы. Вокруг этой темы формируется весьма обширная литература; вклад в эту область внесли многие авторитетные исследователи. Здесь мы обозначим лишь отдельные работы: это труды И.А. Быкова [1], С.В. Володенкова [2, 3], Л.В. Сморгунова [4], С.Н. Федорченко [5] и ряда других исследователей.

Зарубежные исследования социальных сетей весьма многочисленны. В центре внимания исследователей находятся взаимосвязанные вопросы о конфигурации сетевых кластеров, о каналах распространения информации и о выявлении влиятельных пользователей и групп [6–9]. Интенсивно идет импорт сетевых моделей и идей из физики и теоретической математики [10, 11].

В области исследований социальных сетей, таким образом, накоплено огромное множество предметных и методологических работ — это одно из магистральных направлений в современных социогуманитарных науках.

### **Алгоритм картографирования «Разматывание клубка»**

Эмпирическая база сетевых исследований, как правило, предусматривает использование больших данных (сотни миллионов точек данных). Однако существует доступная аналитическая инфраструктура, которая занимается извлечением и первичной обработкой подобных информационных массивов. Это разнообразные облачные сервисы, которые собирают и предоставляют информацию об интернет-активности групп и пользователей в сетях для

поддержки маркетинговых и рекламных кампаний. Мы можем воспользоваться такими сервисами, чтобы сократить время и усилия, необходимые для сбора исходной эмпирики. Благодаря первичной группировке данных «на стороне» мы можем работать уже с просто обширными данными – до нескольких сотен тысяч точек данных – посредством стандартного программного обеспечения.

Предложенный в наших исследованиях алгоритм «разматывания клубка» формально является модификацией хорошо известного в социологии метода «снежного кома» для составления выборок в условиях труднодоступности объектов исследования. Однако при внешнем сходстве алгоритм «разматывания клубка» не тождественен «снежному кому». Во-первых, «разматывание клубка» (в отличие от «снежного кома») преследует цель не составить выборку, а обнаружить основную структуру сетевого кластера (все наиболее влиятельные узлы, связи между ними, частично периферические узлы). Во-вторых, связи рефлексивности, которые отслеживает алгоритм «разматывания клубка», принципиально важны для сетевого анализа сами по себе (поскольку обеспечивают существенные сетевые взаимодействия), тогда как «снежный ком» использует связи (порой абсолютно несущественные) между респондентами лишь для поиска новых респондентов (и более того, наличие таких связей делает полученную выборку недостаточно качественной с социологической точки зрения).

«Разматывание клубка» представляет собой итерационную – многотуровую – процедуру, которая стартует с некоторых начальных узлов – «точек входа» в сеть. Эти «точки входа» в сеть могут быть заданы произвольно или подсвечены каким-либо образом в качестве значимых сетевых игроков. Однако есть все основания полагать, что процедура картографирования, включающая большое количество туров, практически индифферентна к тому, какие именно «точки входа» выбраны, лишь бы они принадлежали сети.

В ходе первой итерации фиксируются связи «точек входа». Среди связанных узлов выбираются (на основании экспертных оценок и формальных критериев) те узлы, которые рассматриваются как целевые в данном исследовании. В ходе второй итерации фиксируются связи тех узлов, которые получены в ходе предшествующей итерации, и т. д.

Количество итераций ограничено лишь исследовательскими задачами и человеческими ресурсами. Однако, полагаем, отбор из списков связанных узлов тех сообществ, которые являются целевыми, может быть поручен в перспективе искусственному интеллекту. Заметим, что значительное количество итераций, как правило, является избыточным, поскольку процедура быстро выводит исследователя на те сообщества, которые являются значимыми для исследуемой аудитории.

Если задача заключается в том, чтобы выявить политизированный кластер, в котором возникают, развиваются и транслируются политические смыслы, то связи между узлами должны быть каналами рефлексивности. Проявления рефлексивности можно наблюдать непосредственно, например в виде репостов. В этом случае исследователь строит деревья репостов и прослеживает узлы, через которые проходили те или иные сообщения. Несколько лет назад статистика репостов в VK была закрыта. Этот инструмент ныне недоступен. Поэтому для выявления каналов рефлексивности мы предлагаем использовать исследование общей для двух сообществ аудитории.

Пересечение аудиторий двух узлов (сообществ в VK) можно выявить посредством сервиса [vk.barkov.net](http://vk.barkov.net), а именно – скрипта «Группы, где есть целевая аудитория (поиск похожих сообществ)». В наших исследованиях для каждого целевого сообщества мы собирали связанные сообщества, в которых подписчиков не менее 1 тыс. и не более 20 млн. То есть микроскопические группы (менее 1 тыс.) мы не рассматривали. Поисковая выдача была ограничена топ-100 узлов, аудитория которых вносит наибольший вклад в аудиторию целевого сообщества. Долю каждой такой внешней аудитории в аудитории целевого узла можно отождествить с весом входящей связи между узлами, поскольку в этом случае целевое сообщ-

щество следует рассматривать как реципиента аудитории, а связанное — как донора. Одновременно может быть получена противонаправленная связь, вес которой следует определить как долю общих пользователей в аудитории связанного сообщества, которое в данной связи выступает уже не как донор, а как реципиент. Данный вес фигурирует в поисковой выборке при включении настройки «Добавить в выдачу % числа найденных людей из исходной группы к числу участников найденной группы».

Если в число связанных групп попадало сообщество более чем с 300 тыс. пользователей, такая группа в наших исследованиях фиксировалась в базе данных с результатами картографирования, но затем — в течение следующей итерации — мы уже не искали связанных с ней групп. Эта отсечка обусловлена следующими соображениями. Во-первых, скрипт для обработки связи крупных групп требует слишком много времени для выполнения этой операции. Во-вторых, для слишком больших целевых сообществ в топ-100 связанных с ними узлов попадают преимущественно самые популярные группы с развлекательным контентом; тематические политизированные сообщества вследствие этого имеют меньше шансов попасть в выдачу, так как они имеют аудитории, как правило, в диапазоне 10–50 тыс. человек. Поэтому сверхкрупные сообщества видны алгоритму «разматывания клубка», но являются своего рода тупиками, поскольку сами дают минимальное количество входящих связей.

Важно обеспечить полноту картографируемого кластера и непрерывность процедуры картографирования. Для этого нужно выполнить три условия: 1) каждую поисковую выдачу vk.barkov.net (топ-100) нужно рассмотреть и просеять полностью, без изъятия; 2) необходимо довести каждую итерацию до конца (т. е. рассмотреть поисковые выдачи всех без исключения целевых групп, полученных в ходе предшествующей итерации); 3) не следует вносить в базу данных (БД) узел более одного раза (даже если он встречается несколько раз в разных выдачах), чтобы не допустить дублирования. Вместе с тем новые связи, которые обнаруживаются у ранее обнаруженных узлов, следует фиксировать в таблице граней.

### **Качественное кодирование**

Качественное кодирование осуществляется одним или несколькими экспертами. Списки сообществ для экспертного рассмотрения формируются в ходе картографирования. В каждом сообществе эксперт рассматривает определенное количество постов и комментариев к ним за определенный период.

Сущность процедуры качественного кодирования сводится к присвоению сообществам неких ярлыков из разных наборов, что позволяет в дальнейшем получить и проанализировать количественно-качественные характеристики совокупности сообществ — исследуемого кластера. В наших исследованиях применены следующие наборы кодов: «Уровень политизации», «Уровень радикализма», «Уровень нетерпимости», «Уровень русофобии», «Политическая окраска».

Заметим, что все наборы, за исключением последнего, дают возможность эксперту оценить сообщества посредством присвоения им определенных порядковых (ранговых) величин: «максимально радикальные сообщества», «уровень радикализма — средний», «радикальные идеи отсутствуют» и т. п. Все подобные наборы кодов являются непополняемыми.

Набор кодов «Политическая окраска» является пополняемым, поскольку допускает введение экспертом новых ярлыков в том случае, если имеющиеся недостаточно точно маркируют идейно-политическую сущность рассматриваемых сообществ. В данном случае ярлыки представляют собой категориальные (номинальные) дескрипторы: «либералы», «коммунисты», «националисты» и пр.

Результаты картографирования (узлы, связи и их характеристики) накапливаются в реляционной базе данных (схема данных на рис. 1). Посредством запросов к ней можно сформировать массивы данных в формате, пригодном для загрузки в стандартные программы сетевого анализа и визуализации сетей в виде графов (например, Gephi).

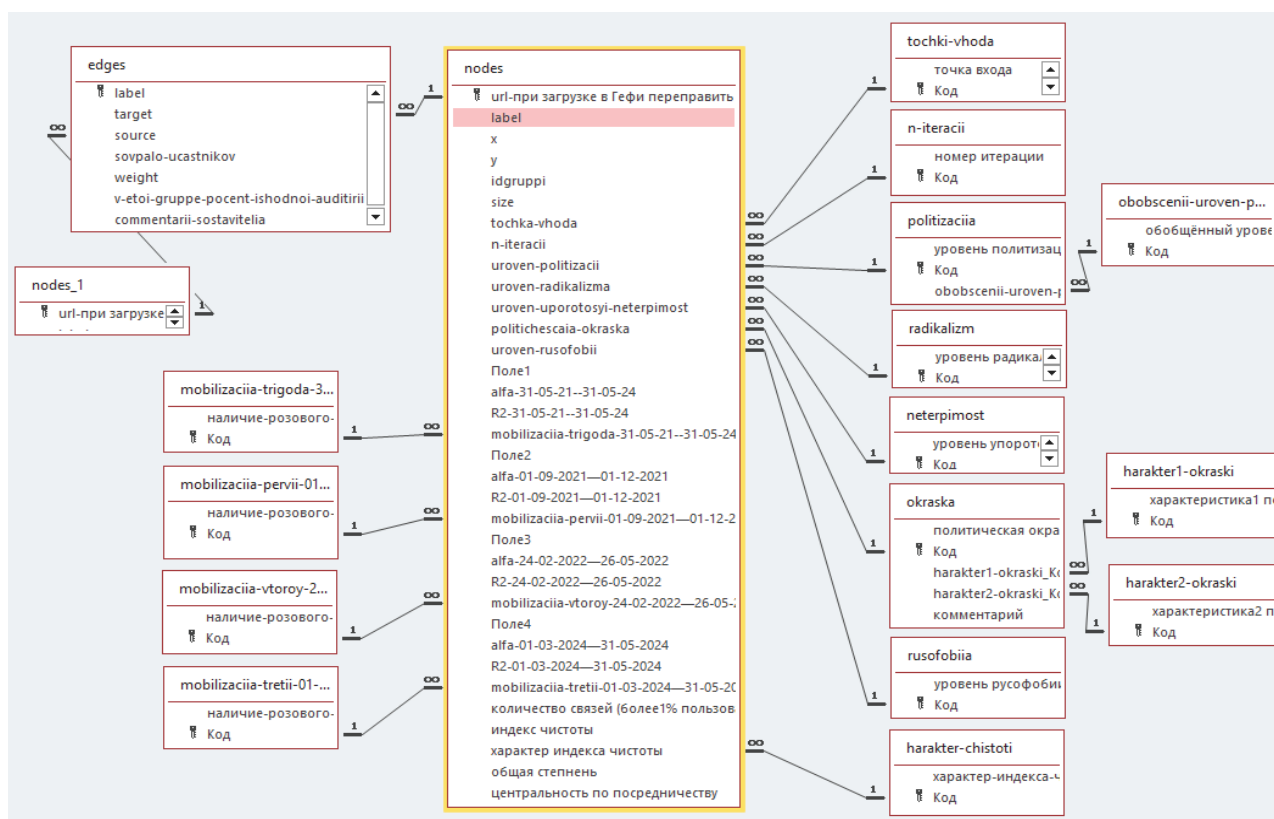


Рис. 1. Схема данных БД, в которой накапливаются результаты картографирования, качественного кодирования, сетевого анализа

### Заключение

Конечный продукт картографирования – граф (геометрическая фигура, состоящая из вершин и связывающих их ребер). В ходе наших исследований мы выявили связанные политизированные сообщества (в сетевом анализе используется термин «узлы»; они же – вершины графа), а также проследили связи между ними (границы, дуги; они же – ребра графа). Кроме того, мы дали количественную оценку политическому составу кластера, приписав узлам некоторые категориальные атрибуты или порядковые величины.

Новизна этой работы (помимо собственно выявления и описания политизированного сегмента в VK) состоит в том, что мы предложили алгоритм картографирования, который условно обозначили как «разматывание клубка». Этот алгоритм не только позволяет быстро выявить наиболее популярные сообщества и связанную с ними периферию, но и, что немаловажно, может быть реализован посредством имеющихся облачных аналитических сервисов. Эти сервисы проводят первичную обработку больших данных, что дает возможность работать в рамках собственно политологического исследования с данными, которые можно назвать просто обширными и которые могут быть проанализированы посредством офисной компьютерной техники и стандартного программного обеспечения (в частности, Statsoft Statistica или MS Excel). Мы также предложили наборы кодов (категорий) для качественного кодирования, которые, полагаем, могли бы помочь быстро и с приемлемыми трудозатратами дать единообразное описание и получить количественные параметры сетевых кластеров, состоящих из сотен узлов.

*Исследование выполнено за счет гранта Российского научного фонда № 24-28-00619.*

### Список литературы

1. Быков И.А. Методологические проблемы исследования политизированных виртуальных сообществ в ВКонтакте // Возможности и угрозы цифрового общества: мат-лы конф. / под ред. А.В. Соколова, А.А. Фролова. Ярославль, 2020. С. 33–36.
2. Володенков С.В. Влияние технологий интернет-коммуникаций на современные общественно-политические процессы: сценарии, вызовы и акторы // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 5 (153). С. 341–364.
3. Володенков С.В. Интернет как технологическое пространство современных политических коммуникаций: перспективы и сценарии развития // Журнал политических исследований. 2017. № 3. С. 79–100.
4. Сморгунув Л.В. Электронные платформы и сетевое научение: как трансформируется публичное пространство // Технологии информационного общества в науке, образовании и культуре: сборник научных статей. СПб.: Санкт-Петербургский нац. исслед. ун-т информационных технологий, механики и оптики. 2014. С. 259–262.
5. Федорченко С.Н. Феномен искусственного интеллекта: гражданин между цифровым аватаром и политическим интерфейсом // Журнал политических исследований. 2020. Т. 4. № 2. С. 34–57.
6. Tufekci Z., Wilson C. Social media and the decision to participate in political protest: Observations from Tahrir Square // Journal of Communication. 2012. Vol. 62. № 2. P. 363–379. DOI: 10.1111/j.1460-2466.2012.01629.x.
7. Onnela J.-P., Reed-Tsochas F. Spontaneous emergence of social influence in online systems // Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences, 2010. Vol. 107. № 43. P. 18375–18380. DOI: 10.1073/pnas.0914572107.
8. Bakshy E., Messing S., Adamic L.A. Exposure to ideologically diverse news and opinion on Facebook // Science. American Association for the Advancement of Science, 2015. Vol. 348. № 6239. P. 1130–1132. DOI: 10.1126/science.aaa1160.
9. Del Vicario M., Bessi A., Zollo F., Petroni F., Scala A., Caldarelli G., Stanley H.E., Quattrociocchi W. The spreading of misinformation online // Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences, 2016. Vol. 113. № 3. P. 554–559. DOI: 10.1073/pnas.1517441113.
10. Mugisha S., Zhou H.-J. Identifying optimal targets of network attack by belief propagation // Phys. Rev. E. American Physical Society, 2016. Vol. 94. № 1. P. 012305. DOI: 10.1103/PhysRevE.94.012305.
11. Lü L., Chen D., Ren X.-L., Zhang Q.-M., Zhang Y.-C., Zhou T. Vital nodes identification in complex networks // Physics Reports. 2016. Vol. 650. P. 1–63. DOI: 10.1016/j.physrep.2016.06.007.

### References

1. Bykov I.A. (2020) *Metodologicheskie problemy issledovaniya politizirovannykh virtual'nykh soobshchestv v VKontakte* [Methodological problems of the study of politicized virtual communities in VKontakte. Opportunities and threats of digital society. Conference Proceedings. Edited by A.V. Sokolov, A.A. Frolov]. Yaroslavl. С. 33–36.
2. Volodenkov S.V. (2019) *Vliyanie tekhnologiy internet-kommunikatsiy na sovremennye obshchestvenno-politicheskie protsessy: stsennarii, vyzovy i aktory* [Influence of Internet communication technologies on modern socio-political processes: scenarios, challenges and actors] *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny* [Monitoring of public opinion: economic and social changes]. No. 5 (153). С. 341–364.
3. Volodenkov S.V. (2017) *Internet kak tekhnologicheskoe prostranstvo sovremennykh politicheskikh kommunikatsiy: perspektivy i stsennarii razvitiya* [Internet as a technological space of modern political communications: prospects and scenarios of development] *Zhurnal politicheskikh issledovaniy* [Journal of Political Studies]. No. 3. С. 79–100.
4. Smorgunov L.V. (2014) *Elektronnyye platformy i setevoye naucheniye: kak transformiruetsya pablichnoye prostranstvo* [Electronic platforms and network learning: how public space is transforming] *Tekhnologii informatsionnogo obshchestva v nauke, obrazovanii i kul'ture: sbornik nauchnykh statey. Sankt-Peterburgskiy nats. issled. un-t informatsionnykh tekhnologiy, mekhaniki i optiki* [Information society technologies in science, education and culture: a collection of scientific articles St. Petersburg National Research University of Information Technologies, Mechanics and Optics]. St. Petersburg. С. 259–262.

5. Fedorchenko S.N. (2020) *Fenomen iskusstvennogo intellekta: grazhdanin mezhdru tsifrovym avatarom i politicheskim interfeysom* [Phenomenon of artificial intelligence: a citizen between a digital avatar and a political interface] *Zhurnal politicheskikh issledovaniy* [Journal of Political Studies]. Т. 4. No. 2. С. 34–57.

6. Tufekci Z., Wilson C. (2012) Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*. Vol. 62. No. 2. P. 363–379. DOI: 10.1111/j.1460-2466.2012.01629.x.

7. Onnela J.-P., Reed-Tsochas F. (2010) Spontaneous emergence of social influence in online systems. *Proceedings of the National Academy of Sciences*. *Proceedings of the National Academy of Sciences*. Vol. 107. No. 43. P. 18375–18380. DOI: 10.1073/pnas.0914572107.

8. Bakshy E., Messing S., Adamic L.A. (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science*. American Association for the Advancement of Science. Vol. 348. No. 6239. P. 1130–1132. DOI: 10.1126/science.aaa1160.

9. Del Vicario M., Bessi A., Zollo F., Petroni F., Scala A., Caldarelli G., Stanley H.E., Quattrociocchi W. (2016) The spreading of misinformation online. *Proceedings of the National Academy of Sciences*. *Proceedings of the National Academy of Sciences*. Vol. 113. No. 3. P. 554–559. DOI: 10.1073/pnas.1517441113.

10. Mugisha S., Zhou H.-J. (2016) Identifying optimal targets of network attack by belief propagation. *Phys. Rev. E*. American Physical Society. Vol. 94. No. 1. P. 012305. DOI: 10.1103/PhysRevE.94.012305.

11. Lü L., Chen D., Ren X.-L., Zhang Q.-M., Zhang Y.-C., Zhou T. (2016) Vital nodes identification in complex networks. *Physics Reports*. Vol. 650. P. 1–63. DOI: 10.1016/j.physrep.2016.06.007.