

DOI: 10.31857/S0869049922010075

## Трибуна молодого ученого TRIBUNE OF YOUNG SCIENTIST

Оригинальная статья / Original Article

### Модель влияния цензуры на общественное мнение в социальных медиа

© К.А. ТОЛОКНЕВ

**Толокнев Кирилл Александрович**, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), kirtoloknev@gmail.com

В настоящее время все актуальнее становятся исследования влияния цензуры на формирование общественного мнения. Значительное количество статей посвящено цензуре в социальных медиа и ее воздействию на политические коммуникации, однако большинство из них не учитывают важную черту пользователей социальных медиа – гомофилию, то есть тенденцию вступать во взаимодействие со схожими людьми. Принимая во внимание сложность исследования данного вопроса при помощи эмпирических методов, в данной статье рассматривается вычислительная модель, которая симулирует формирование общественного мнения в социальных медиа при различных уровнях гомофилии. Предложенная модель показывает, что эффективность цензуры в значительной мере зависит от уровня гомофилии. При низком ее показателе даже крайне слабая цензура эффективна: общественное мнение стабильно сдвигается в сторону не подвергаемых цензуре мнений. Однако при росте гомофилии ситуация меняется. Недостаточное цензурирование либо неэффективно, либо ведет к противоположному эффекту: общественное мнение сдвигается в сторону цензурируемых мнений. В условиях средней и высокой гомофилии только обширная цензура позволяет достичь желательных для цензоров результатов.

**Ключевые слова:** цензура, социальные медиа, агентно-ориентированная модель, вычислительное моделирование, общественное мнение, динамика мнений, политические коммуникации

**Цитирование:** Толокнев К.А. (2022) Модель влияния цензуры на общественное мнение в социальных медиа // Общественные науки и современность. № 1. С. 127–144. DOI: 10.31857/S0869049922010075

# Model of Censorship Influence on Public Opinion Formation in Social Media

© K. TOLOKNEV

**Kirill A. Toloknev**, National Research University “Higher School of Economics” (Moscow, Russia), kirtoloknev@gmail.com

**Abstract.** Importance of research on how censorship influences the formation of public opinion in social media has grown significantly in recent years. However, existing research neglects potential effects of homophily, which is a tendency to create links with like-minded individuals. Considering difficulties conducting such research with empirical methods, the author proposes a computational model. It simulates public opinion formation under different levels of homophily. Modelling shows two main outcomes. If the homophily level is low, even limited censorship effectively pushes public opinion away from censored opinions. Increasing level of homophily makes limited censorship ineffective and, under certain conditions, pushes public opinion towards censored opinions. Only extensive censorship remains effective under this condition.

**Keywords:** censorship, social media, agent-based model, computational modelling, public opinion, opinion dynamics, political communication

**Citation:** Toloknev K. (2022) Model of Censorship Influence on Public Opinion Formation in Social Media. *Obshchestvennye nauki i sovremennost'*, no.1, pp. 127–144. DOI: 10.31857/S0869049922010075 (In Russ.)

Вскоре после того, как сторонники Д. Трампа захватили здание Капитолия в январе 2021 г., крупнейшие социальные медиа заблокировали аккаунты на тот момент действующего президента США и ряда его сторонников. Наблюдатели, в зависимости от их политической позиции, рассматривали произошедшее либо как акт политической цензуры, либо как закономерный результат беспорядков или даже попытки государственного переворота. Блокирование (бан) обычных пользователей или уголовное преследование за публичное выражение нежелательных мнений в социальных медиа уже давно не считается чем-то необычным. Ряд авторитарных государств также стараются усилить контроль над высказываниями в социальных медиа. В любом случае, в последнее время значительно актуализировался вопрос о том, как подобного рода «вмешательство» в политическую коммуникацию влияет на климат мнений в социальных медиа. Эффективно ли оно – и если эффективно, то при каких условиях?

Цензура – ограничение доступа к определенной информации, которую цензурирующие органы считают «нежелательной». Следует отметить, что данное определение включает в себя цензуру не только со стороны государства, но и самих социальных медиа<sup>1</sup>. Значительный объем литературы посвящен цензурированию социальных медиа авторитарными режимами. Исследователям удалось выделить его три основных механизма: «страх», «трение» и «затопление» [Roberts 2020]. Несмотря на рост интереса к исследованиям цензуры (как онлайн, так и офлайн), вопрос об ее эффективности остается открытым. Некоторые работы демонстрируют, что в определенных условиях она не только эффективно подавляет распространение нежелательных мнений, но и сдвигает общественное мнение в пользу «желатель-

<sup>1</sup> К примеру, до недавнего времени в Facebook посты в поддержку К. Риттенхауса компания удаляла, а поиск по имени блокировала [Clayton 2021].

ных». Другие показывают, что цензура неэффективна, и при ее применении цензурируемый контент становится интересен большему количеству пользователей. Однако данные исследования упускают важнейший аспект коммуникации в социальных медиа – гомофилию.

Гомофилия – тенденция к созданию социальных связей с людьми, которые обладают схожим социальным положением, возрастом, уровнем образования и, что самое важное, политическими предпочтениями [McPherson, Smith-Lovin, Cook 2001; Huber, Malhotra 2017]. Гомофилия значительно влияет на онлайн-коммуникации, способствуя формированию сообществ единомышленников – своего рода информационных «пузырей» [Cinelli, De Francisci Morales, Galeazzi, Quattrociocchi, Starnini 2021]. Однако ее воздействие на эффективность цензуры пока мало изучено как при помощи эмпирических методов, так и при использовании вычислительного моделирования. Отсутствие эмпирических работ в целом не вызывает удивления, так как перед исследователями встают как минимум две сложные задачи: измерение уровня гомофилии [Colleoni, Rozza, Arvidsson 2014] и определение эффективности цензуры. В то же время недостаток работ с использованием вычислительного моделирования представляется крупной «лакуной», учитывая то, насколько данный метод распространен в моделировании при формировании общественного мнения (динамики мнений). Если принимать во внимание, что вычислительные модели, в которых поведение системы симулируется в контролируемой модельной среде, позволяют создавать потенциально проверяемые на эмпирическом материале гипотезы, недостаточное количество вычислительных моделей вредит и эмпирическим исследованиям в данной сфере.

В настоящий момент существует две работы, посвященные моделированию цензуры в социальных медиа, обе написаны в 2021 г. [Wang 2021; Lane, McCaffree, Shults 2021]. Ни одна из них не учитывала потенциальное влияние гомофилии. Обе статьи скорее оценивают цензуру как неэффективный инструмент. С. Ванг в своей работе продемонстрировал, что в зависимости от количества не подверженных социальному влиянию агентов цензура может быть как эффективной, так и неэффективной. Дж. Лейн и его коллеги показали, что и централизованное, и децентрализованное цензурирование радикальных мнений ведет к большей радикализации, чего цензоры явно не желают. Эмпирические исследования также демонстрируют, что цензура может быть как эффективна, так и неэффективна.

Цель данного исследования – рассмотреть, как цензура в условиях различных уровней гомофилии влияет на формирование общественного мнения в социальных медиа. Предстоит установить, при каких уровнях гомофилии цензура эффективна и приводит к сдвигу общественного мнения в сторону не подвергаемых цензуре мнений, а при каких уровнях гомофилии она неэффективна – либо никак не влияет на формирование общественного мнения, либо ведет к обратному эффекту, сдвигая общественное мнение в сторону цензурируемых мнений.

### **Цензура в социальных медиа: эмпирические свидетельства**

В своей статье от 2020 г. М. Робертс [Roberts 2020] выделила три основных механизма цензурирования онлайн-коммуникаций: «страх», «трение» и «затопление».

«Страх» – наиболее простой механизм действия цензуры, который предполагает наказание распространителя «нежелательной информации». Таким образом повышается цена доступа к «нежелательной информации» и ее распространения. Стандартный процесс модерации в социальных медиа основан на данном механизме. Такая цензура может быть эффективной и изменить поведение подвергнутых наказанию. Однако она оказывает обратный эффект на социальное окружение ее жертв, повышая их интерес к «нежелатель-

ной информации» [Pan, Siegel 2020]. Как отдельную формы цензуры «страхом» можно выделить так называемый «деплатформинг»<sup>2</sup> – относительно новую форму цензуры, которую проводят сами социальные медиа [Rogers 2020]. Дискуссии о его эффективности пока только начинаются, но можно предположить, что его эффект будет схож с результатом «обычной» цензуры при помощи страха: успешное «подавление» коммуникации репрессированного актора при временном всплеске активности его сторонников.

«Трение» – более продвинутый метод цензуры. В данном случае она представляет собой усложнение доступа к «нежелательной информации» простой блокировкой сайтов и селективным удалением информации (к примеру постов, призывающих к коллективным действиям [King, Pan, Roberts 2013]), а также с помощью манипуляции алгоритмами поисковых систем [Epstein, Robertson 2015] – потенциально наиболее мощной формы цензуры через «трение». Р. Эпштейн и Р. Робертсон продемонстрировали в серии экспериментов, что данный метод способен изменить электоральные предпочтения до 20% неопределившихся избирателей. Цензура через «трение», особенно с помощью манипуляции поисковыми выдачами, менее заметна (если о подобной манипуляции не объявляли публично [Калюков, Чебакова 2021]) и может проводиться как контролирующими социальные медиа государствами, так и самими социальными медиа. При цензурировании с помощью «трения» цена доступа к «нежелательной информации» и ее распространения повышается не за счет запугивания пользователей, а через увеличение необходимых для доступа к «нежелательной информации» времени и навыков (пользователям приходится использовать VPN-сервисы или открывать вторую страницу поисковой выдачи).

«Затопление» значительно отличается от прочих механизмов тем, что его нельзя назвать цензурой в строгом смысле. Он предполагает массированную целенаправленную генерацию контента – дезинформации, проправительственных материалов или развлекательного контента, которые «затапливают» социальные медиа и отвлекают внимание пользователя от поиска «нежелательной информации». Как особую форму «затопления» можно выделить так называемый «цифровой джерримендеринг»: гипотетическую форму манипуляции контентом [Zittrain 2013], когда социальное медиа нацеливает на определенный сегмент аудитории сообщения, которые поощряют или подавляют политическое участие (например, участие в выборах). Потенциальная эффективность такой формы манипулирования подтверждена проведенным в 2012 г. экспериментом [Bond, Fariss, Jones, Kramer, Marlow, Settle, Fowler 2012].

Для всех трех механизмов онлайн-цензуры характерна черта, которая отличает онлайн-цензуру от традиционной. В ее рамках практически не используют предварительное цензурирование: в основном ограничивают уже опубликованный контент. Конечно, есть пример использования списков «запрещенных слов» в китайском сегменте интернета – однако пользователи обходят такие фильтры, используя похожие в звучании или произношении слова<sup>3</sup>. Цензурирование написанных постов требует времени, что позволило Г. Кингу и его коллегам исследовать различия между цензурируемыми и нецензурируемыми сообщениями [King, Pan, Roberts 2014; King, Pan, Roberts 2013].

Цензура может оказывать двойное воздействие на общественное мнение. С одной стороны, иногда она вызывает обратный эффект и повышает интерес общества к цензуриру-

---

<sup>2</sup> Удаление нежелательных пользователей с интернет-платформ. В отличие от обычной модерации, деплатформинг предусматривает перманентную блокировку либо крупных интернет-знаменитостей, либо целых групп пользователей.

<sup>3</sup> Примечательно, что к использованию слов-эвфемизмов перешли так же и группы «антипрививочников» в Facebook [Collins, Zadrozny 2021].

емой теме. Дж. Пан и А. Сигел показали, что после репрессий против твиттер-активистов в Саудовской Аравии количество критических постов со стороны их последователей временно увеличилось [Pan, Siegel 2020]. З. Наби в своем исследовании продемонстрировал, что попытки цензурировать Youtube в Турции и Пакистане привели к росту интереса к подвергаемому цензуре контенту [Nabi 2014]. С. Вонг и Дж. Лианг показали, что цензурированный контент могут воспринимать как заслуживающий большего доверия, чем не подвергнутый цензуре [Wong, Liang 2021]. Следует отметить, что данный эффект проявлялся только тогда, когда цензуру замечали пользователи. На данный момент исследователи не смогли выявить его при использовании более незаметных форм цензуры, к примеру в ситуациях, когда через манипуляции поисковой выдачей создается «трение» или сеть «затапливают» отвлекающим контентом. Весьма вероятно, что именно «заметность» цензуры делает ее неэффективной. Конечно, нельзя утверждать, что очень заметная китайская система онлайн-цензуры не работает, однако вероятно, что на потенциальную эффективность «заметной» цензуры влияет такой фактор, как низкий спрос на не подвергнутую цензуре информацию [Chen, Yang 2019].

С другой стороны, цензура может успешно подавить распространение нежелательной информации и сдвинуть общественное мнение в предпочтительную для цензора сторону. Можно выделить два механизма подобного воздействия, которые используют и в онлайн-, и в офлайн-цензуре:

1. Ограничение доступа к «нежелательной информации», которое способствует возникновению «спирали молчания» [Noelle-Neumann 1974]. Даже если цензурирование ведется не при помощи механизма «страха», который прямо увеличивает риски для сторонников «нежелательных мнений», оно способствует искусственному занижению количества активных сторонников «нежелательной» точки зрения. В результате наблюдаемое общественное мнение может не соответствовать истинному распределению позиций в обществе. «Спираль молчания» может возникать в социальных медиа и без влияния цензуры [Kim 2016], но цензура способствует усилению эффекта [Zhu, Fu 2020].

2. «Сдвиг» общественного мнения в «противоположную» от цензурируемой к более приемлемой для цензора точке зрения, который происходит из-за ограничения доступа к определенному спектру мнений. Цензура не только ограничивает распространение «нежелательной» информации, но и облегчает распространение «желательной». Данный эффект при цензурировании или государственном контроле традиционных медиа выявили М. Адена и ее соавторы в своем исследовании о радиопропаганде Веймарской республики и нацистской Германии [Adena, Enikolopov, Petrova, Santarosa, Zhuravskaya 2015]. Когда по радио агитировали в поддержку правительства и «про-демократических» партий<sup>4</sup> (1930–1932 гг.), а НСДАП не допускали к эфиру, в районах с радиовещанием поддержка НСДАП снижалась, а в 1933 г., когда уже нацисты не допускали своих соперников к вещанию – увеличивалась. Наличие данного эффекта в онлайн-цензуре показали уже вышеупомянутые Эпштейн и Робертсон [Epstein, Robertson 2015].

Таким образом, можно выдвинуть первую гипотезу:

H1: использование цензуры сдвигает общественное мнение в сторону нецензурируемого мнения.

Стоит подвести промежуточный итог. Использовать цензуру в социальных медиа может как государство, так и негосударственные акторы: сами социальные медиа или – через апелляцию к ним – группы активистов. Цензуру в социальных медиа проводят с целью

<sup>4</sup> Всех крупных партий, кроме НСДАП и КПГ.

ограничить доступ к «нежелательной информации», причем необязательно навсегда – зачастую функционал учетной записи, подвергнутой цензуре, ограничивают на время.

Теперь следует обратиться к литературе, посвященной эффективности репрессий, так как цензура – не более чем их форма. В работе 2011 г. Д. Зигель продемонстрировал, что эффективность репрессий зависит от структуры социальной сети, которая связывает людей. Точечные репрессии более эффективны в условиях, когда количество влиятельных акторов в сети ограничено, но не в том случае, если все они обладают высоким уровнем мотивации. Случайные репрессии в таком случае не оказывают желаемого действия [Siegel 2011]. Поскольку цензура (в особенности цензура страхом) сама по себе лишь форма репрессий, следует выдвинуть предположение о том, что эффективность цензуры так же будет зависеть от характеристик связывающей людей сети. Наиболее важны следующие особенности социальных медиа.

Во-первых, влияние в них распределяется неравномерно. В социальных медиа выделяется «цифровая элита», которая обладает диспропорционально большим количеством «друзей» или «подписчиков» по сравнению с рядовыми участниками. Некоторые исследования показывают, что «цифровая элита» диспропорционально сильнее влияет на распространение информации в социальных медиа [González-Bailón, Wang 2016], в то время как другие работы демонстрируют, что влияние «цифровой элиты» и «цифровой периферии» (обычных пользователей) на распространение информации зависит от характера материалов. Обычные мемы и политические сообщения – в особенности призывы к протестному поведению – распространяются по-разному [Guilbeault, Becker, Centola 2018; Romero, Meeder, Kleinberg 2011]. Если в распространении первых высоко влияние «цифровой элиты» [Romero, Meeder, Kleinberg 2011], то для вторых, наоборот, более важную роль играет периферия [Steinert-Threlkeld 2017].

Во-вторых, для социальных медиа характерна гомофилия: схожие люди с большей вероятностью формируют связи друг с другом [McPherson, Smith-Lovin, Cook 2001]. Сходство касается как социально-экономического положения и этнического происхождения, так и политических предпочтений [Huber, Malhotra 2017]. Похожие люди не только скорее окажутся связаны друг с другом, но и с большей вероятностью получают информацию, которая подтверждает их точки зрения [Del Vicario, Bessi, Zollo, Petroni, Caldarelli, Stanley, Quattrociocchi 2016]. В результате формируются относительно гомогенные группы [Cinelli, De Francisci Morales, Galeazzi, Quattrociocchi, Starnini 2021], что может способствовать росту политической поляризации в обществе [Du, Gregory 2016]. Тем не менее, пока в научном сообществе не сформировался консенсус по вопросу о том, насколько гомофилия в социальных медиа влияет на рост политической поляризации. Некоторые исследования показывают, что коммуникации в социальных медиа могут и снижать ее [Barberá 2014]. Весьма вероятно, что противоположные результаты исследований вызваны различием степени гомофилии в зависимости от страны или предмета обсуждения [Barberá, Jost, Nagler, Tucker, Bonneau 2015].

Таким образом, можно выдвинуть вторую гипотезу:

H2: эффективность цензуры зависит от уровня гомофилии в сети, которая связывает подвергаемых цензуре агентов.

Проверить подобные исследовательские гипотезы – в особенности вторую – на наборе эмпирических данных невозможно. В первую очередь потому, что уровнями цензуры<sup>5</sup> и гомофилии нельзя манипулировать. Проведение экспериментов также представляется ма-

---

<sup>5</sup> Конечно, манипуляция возможна на уровне определенных акторов, но они не заинтересованы в проведении научных исследований.

лореальным. Поэтому следует обратиться к агентно-ориентированному моделированию – активно развивающемуся методу, который позволяет симулировать поведение интересующих исследователя систем [Ахременко, Петров, Жеглов 2021]. Агентно-ориентированное моделирование обладает важным преимуществом для решения задач данной статьи в связи с тем, что оно позволяет симулировать возникновение эффектов на макроуровне (в данном случае – изменений «общественного мнения») как результат взаимодействий на микроуровне (в данном случае – обмена мнениями между агентами и цензурирования отдельных агентов). В следующей части статьи будет дано описание предлагаемой агентно-ориентированной модели.

## Модель

Прежде чем приступить к описанию модели, необходимо сказать несколько слов о применяемом методе.

Агентно-ориентированное моделирование – один из методов вычислительного моделирования. Он предполагает создание модели, которая состоит из конечного количества агентов – автономных объектов, которые взаимодействуют с ограниченным количеством других агентов по заранее определенным правилам. Агенты могут представлять собой как отдельных индивидов [Lane, McCaffree, Shults 2021], так и целые государства [Cederman 1994]. Они обладают определенными характеристиками (параметрами), часть из которых не меняется, а часть трансформируется при взаимодействии. Характеристики обоих типов могут как константными (их задает исследователь до начала симуляции) и извлекаемыми из определенного распределения. Внутри одной модели действуют агенты разных типов – к примеру, в модели возникновения спирали молчания в социальных медиа, разработанной Б. Россом и его коллегами, использовались и агенты-люди, и агенты-боты<sup>6</sup> [Ross, Pilz, Cabrera, Brachten, Neubaum, Stieglitz 2019]. Правила взаимодействия могут отличаться у агентов разных типов. Так, в модели Росса агенты-люди демонстрировали или скрывали свою позицию в зависимости от того, находились они в большинстве или меньшинстве, а агенты-боты раскрывали ее в любом случае. Следует отметить еще одну черту агентно-ориентированных моделей. Агенты взаимодействуют друг с другом локально: либо с соседями в некотором симулируемом пространстве, либо с соседями по социальной сети.

Как верно отметил Д. Зигель, наиболее известный вид моделирования поведения агентов – теоретико-игровое моделирование, в котором агенты взаимодействуют по правилу максимизации собственного выигрыша [Siegel 2018]. Для таких моделей возможен поиск аналитического решения. Однако для агентно-ориентированных моделей, в которых агенты контактируют по другим правилам, поиск аналитического решения невозможен. Анализировать их возможно с помощью «вычислительных экспериментов».

«Вычислительный эксперимент» – множество симуляций с изменяемыми значениями параметров модели. Каждая отдельная симуляция состоит из этапа инициализации, когда устанавливаются значения параметров агентов, и определенного исследователем количества «шагов» – повторяющихся взаимодействий по установленным правилам. Параметры, значение которых устанавливает исследователь до начала симуляций, считаются «входными», «откликом» модели становится один или несколько изменяемых параметров (в предложенной ниже модели – изменения на пространстве мнений). Так как некоторые значения параметров извлекаются из случайных распределений, для каждой

---

<sup>6</sup> Автоматизированные аккаунты социальных медиа.

комбинации параметров проводится множество симуляций, после чего анализируются агрегированные данные.

Модель состоит из 2,6 тыс. агентов, которые объединены в безмасштабную сеть. Архитектура сети похожа на реальные социальные медиа. Относительно небольшое количество агентов связано с диспропорционально большим количеством других агентов, что симулирует существование «цифровой элиты». Параметр  $H$ , который принимает значение от 0 до 1, определяет, насколько сильно различаются мнения агентов, связанных друг с другом в сети. Чем он выше, тем ниже уровень гомофилии.

Эмпирика показывает, что социальное влияние происходит в том случае, если мнения людей уже достаточно схожи [Nickerson 1998; Sarup, Suchner, Gaylord 1991]. До начала симуляции задается параметр  $\alpha$  – «интервал доверия». Он определяет, как сильно могут различаться мнения агентов для того, чтобы при взаимодействии они повлияли друг на друга. Чем выше значение  $\alpha$ , тем более агенты открыты для влияния со стороны своих соседей по социальной сети.

У каждого агента присутствуют следующие индивидуальные параметры: «мнение»  $O$  (принимающее значения между 0 и 1, извлекаемое из равномерного распределения), «внутренняя мотивация»  $m$ , которая определяет готовность агента участвовать в коммуникации вне зависимости от социального контекста (участие в коммуникации агентов-соседей). Вслед за Д. Зигелем предположим, что для каждого агента значение  $m$  извлекается из нормального распределения, ограниченного максимальным значением 1 и минимальным 0 с заранее заданным средним и стандартным отклонением<sup>7</sup>[Siegel 2011].

Мнение  $O$  отражает отношение агента к какой-либо политической проблеме – допустим, при значении  $O = 0$  агент не поддерживает определенную политическую позицию, при значении  $O = 1$  – полностью поддерживает. Общественным мнением в модели представляется распределение всех мнений всех агентов.

При инициализации активируются две группы агентов: все агенты с исходными мнениями в диапазонах от 0 до 0,1 и 0,9 до 1. Таким образом моделируется эффект воздействия алгоритмов социального медиа, сортирующих контент так, чтобы пользователи получили информацию, которая с наибольшей вероятностью вызовет у них отклик [Pariser 2011]. Для подобных агентов предполагается более высокое значение внутренней мотивации  $m$ , устанавливаемое как 0,99<sup>8</sup> – наиболее активные участники социальных медиа отличаются крайне высокой уверенностью в своей позиции. Все последующие шаги симуляции модели проходят в следующем порядке.

Активизация агентов происходит в условиях, когда среднее значение суммы «внутренней мотивации агента» и доли активных соседей с мнением в пределах «интервала доверия» агента ниже извлекаемого из равномерного распределения с минимальным значением 0 и максимальным значением 1 случайного числа. Чем выше «общая мотивация» агента, тем больше вероятность того, что он вступит в коммуникацию.

Если количество активных агентов с близкой позицией обозначить как  $n_s$ , а общее количество агентов-соседей как  $n$ , формально данный процесс можно записать как:

$$\frac{(m + \frac{n_s}{n})}{2} \leq U[0,1]. \quad (1)$$

<sup>7</sup> При использовании равномерного распределения значимых изменений поведения модель не выявила.

<sup>8</sup> Без использования данного параметра результаты симуляций не отличаются от демонстрируемых ниже.

Далее происходит обмен мнениями между активными агентами. Здесь используется модель динамики мнений «ограниченного доверия»<sup>9</sup>. Также предполагается, что влияние оказывается только на тех агентов, чье мнение на момент взаимодействия не изменилось относительно исходного мнения более чем на значение  $\alpha$  со стороны его соседей по сети, чьи мнения попадают в диапазон  $\alpha$ .

Взаимодействия с соседями (обмен мнениями) происходит так:

$$\Delta o_i = \begin{cases} \text{Для агента } i \text{ если } [O_{it0} - O_t] < \alpha \text{ и } |O_{it} - O_{jt}| < \alpha \\ (1 - m_i) \cdot \mu \cdot \omega \cdot (O_{jt} - O_{it}) \\ \text{иначе} \\ 0 \end{cases},$$

где  $\mu$  – сила воздействия,  $\omega$  – доля активных соседей, способных оказывать ассимилятивное влияние относительно количества всех активных соседей,  $O_{jt}$  и  $O_{it}$  – мнения взаимодействующих агентов,  $O_{it0}$  – мнение агента  $i$  на начало симуляции.

Соответственно, мнение агента  $i$  после обмена мнениями:

$$O_{t+1} = O_t + \Delta o_i. \quad (3)$$

Для симуляций без моделирования цензуры на данном этапе каждый конкретный шаг модели заканчивается, повторяясь, пока не возникнет стабильное финальное распределение мнений среди агентов.

Тем не менее, просто симуляция динамики мнений автора интересует мало: подобных работ существует великое множество. Основной объект интереса в данной работе – эффективность цензуры, то есть ее способность «сдвинуть» общественное мнение в сторону нецензурируемого мнения, поэтому следует перейти к описанию механизма моделирования цензуры в симуляции.

Цензура понимается как временное отключение активного агента от коммуникации. Применяется самый широкий подход, который позволяет включить в модель как «офлайн» цензуру со стороны государственных акторов (заклучение обладателя неудобного мнения под стражу или его физическое устранение), так и более «мягкие» формы онлайн-цензуры (временное ограничение доступа к аккаунту в социальном медиа или его функциональности, манипуляции алгоритмами социального медиа и т. д.). Цензура, как и репрессии, может быть нацелена как на случайных обладателей «нежелательных» мнений, так на наиболее влиятельных из них (обладающих наибольшим количеством соседей в сети). Как и реальная цензура в социальных медиа, она не должна происходить мгновенно.

Следовательно, симулируемая цензура должна отражать три основных аспекта реальной онлайн-цензуры. Ей должны подвергаться активные агенты, чьи мнения находятся в определенном промежутке; из промежутка могут выбираться как случайные агенты, так и агенты с наибольшим количеством соседей (наиболее влиятельные агенты); цензура не должна происходить на каждом шаге модели. Чтобы учесть последнюю характеристику, цензурирование происходит на первом и далее на каждом десятом шаге модели. Подвергнутые цензуре агенты временно (на срок до десяти шагов модели) отключаются от коммуникации.

Так же следует отметить, что в модель не встроено отдельный механизм обратной реакции на цензуру: сдвиг «общественного мнения» в сторону цензурируемого мнения. Подобное может произойти в процессе симуляций лишь вследствие определенного уровня гомофилии в симулируемой сети.

<sup>9</sup> Большое семейство моделей «динамики мнений», которые учитывают, что мнения меняются только при взаимодействии с достаточно схожими соседями.

## Результаты вычислительных экспериментов

Моделирование выполнено в программе Netlogo версии 6.2.0. [Wilensky 1999]. Следующие параметры при симуляции варьировались: уровень гомофилии  $H$  (0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1), «интервал доверия»  $\alpha$  (0,1; 0,15; 0,2; 0,25; 0,3), «внутренняя мотивация»  $m$  (N (0,2; 0,15), N (0,5; 0,15), N (0,8; 0,15))<sup>10</sup>. Максимальное количество цензурируемых агентов устанавливалось на уровне 1%, 10% и 20% от количества агентов с «мнениями» в промежутке от 0,9 до 1. Такая цензура направлена на наиболее радикальных агентов, большая часть общества может не замечать ее или поддерживать, как, например, борьбу с «риторикой ненависти»<sup>11</sup>. Отдельно рассматривались ситуации, когда выбор агентов для цензурирования происходит случайно и когда цензуре подвергаются агенты с наибольшим количеством соседей.

Чтобы избежать случайных отклонений, которые возникают при создании сети и распределении параметра «мнение»  $O$  по агентам с различным количеством «соседей», для каждого сочетания параметров  $H$ ,  $\alpha$  и  $m$  проводилось по 50 симуляций длительностью в 500 шагов, потом анализировались агрегированные результаты по множеству симуляций.

Распределение мнений, которое формировалось в симуляциях без цензурирования и крайне низким уровнем гомофилии (от  $H = 1$  до  $H = 0,7$ ), соответствовало традиционным для моделей «динамики мнений» результатам: при низких значениях  $\alpha$  (0,1) изменений на пространстве мнений не происходило. При увеличении значения  $\alpha$  количество групп уменьшилось до трех: две из них можно условно назвать «радикальными» – они формировались в промежутке от 0,05 до 0,15 и 0,85 до 0,95, третья группа, «центристская», формировалась в промежутке от 0,4 до 0,6. При увеличении значения  $\alpha$  количество агентов в первых двух группах снижалось, в третьей – увеличивалось. При  $\alpha = 0,3$  количество агентов в «центристской» группе было столь высоко, что можно говорить о формировании «консенсуса». Медианное «мнение» в системе стабильно оставалось равным 0,5.

При увеличении уровня гомофилии ( $H = 0,4$ ;  $H = 0,3$ ;  $H = 0,2$ ) и низких значениях  $\alpha$  ( $\alpha = 0,1$ ;  $\alpha = 0,15$ ) формировалось 4-5 кластеров со схожими мнениями, однако при дальнейшем увеличении  $\alpha$  картина менялась. Формировалось всего две группы. В центре, где при отсутствии гомофилии появлялась самая крупная группа, возникала «пустота»: симулируемое общество поляризовалось. Что наиболее интересно, для такого эффекта не потребовалось очень высокого уровня гомофилии: при значении  $H = 0,6$  условный агент  $i$  с мнением на уровне 0,4 мог быть связан с любым другим агентом в системе.

Когда в модель добавляется цензура, результаты становятся еще неожиданнее. Основной вывод данной работы – при средней или высокой гомофилии цензура, нацеленная на наиболее радикальные мнения, не только не эффективна (не способна сдвинуть общественное мнение ближе к нецензурируемой области), но и – при значениях  $\alpha$  ( $\alpha = 0,2$ ;  $\alpha = 0,25$ ) – общественное мнение сдвигается в сторону цензурируемого мнения. Впрочем, в условиях отсутствия гомофилии цензура показала свою эффективность.

При высоком значении «интервала доверия»  $\alpha$  ( $\alpha = 0,3$ ) случайное цензурирование даже 1% от количества агентов в группе с мнениями в промежутке от 0,9 до 1 приводило к стабильному сдвигу медианного мнения «влево» (0,44 против 0,5 в условиях отсутствия цензуры). Увеличение количества цензурируемых агентов позволяло добиться такого результата

<sup>10</sup> Ограничено максимальным значением 1 и минимальным 0. Изменения параметра не показали значительных изменений в поведении модели.

<sup>11</sup> Допустим, с призывами к насильственным действиям по отношению к этническим меньшинствам.

Таблица 1.

Формирование групп в пространстве мнений без влияния цензуры

Table 1.

Formation of groups in the opinion space (no influence of censorship)

	H = 0,2	H = 0,3	H = 0,4	H = 0,5	H = 0,6	H = 0,7	H = 0,8	H = 0,9	H = 1
$\alpha = 0,1$	Фрагментация	Фрагментация	Три группы	Нет изменений					
$\alpha = 0,15$	Три группы	Три группы	Три группы	Три группы	Три группы	Три группы	Три группы	Три группы	Три группы
$\alpha = 0,2$	Поляризация	Поляризация	Поляризация	Поляризация	Три группы				
$\alpha = 0,25$	Поляризация	Поляризация	Поляризация	Поляризация	Поляризация	Поляризация	Три группы	Три группы	Три группы
$\alpha = 0,3$	Поляризация	Поляризация	Поляризация	Поляризация	Поляризация	Три группы	Консенсус	Консенсус	Консенсус

Таблица 2.

Влияние цензуры (цензурюются 10% от количества агентов в группе с мнением O > 0,9)

Table 2.

Influence of censorship (10% of agents with opinion O > 0.9 are censored)

	H = 0,2	H = 0,3	H = 0,4	H = 0,5	H = 0,6	H = 0,7	H = 0,8	H = 0,9	H = 1
$\alpha = 0,1$	Неэффективна	Неэффективна	Неэффективна	Нет изменений					
$\alpha = 0,15$	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна
$\alpha = 0,2$	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна
$\alpha = 0,25$	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна	Неэффективна
$\alpha = 0,3$	Неэффективна	Эффективна	Эффективна	Неэффективна	Неэффективна	Эффективна	Эффективна	Эффективна	Эффективна

и при более низких значениях «интервала доверия»  $\alpha$ . Впрочем, при крайне низких значениях  $\alpha$  ( $\alpha = 0,1$ ) эффекта у цензуры не было.

По мере увеличения гомофилии картина менялась. Теперь, как и в условиях отсутствия цензуры, при высоких значениях  $\alpha$  ( $\alpha > 0,15$ ) формируются две группы, при низких – три и более. Даже при условии, что цензуре подвергается 10% или 20% агентов с мнениями в промежутке от 0,9 до 1, которые обладают наибольшим количеством соседей, цензура не способна помешать образованию группы с мнениями, близкими к цензурируемому. При средних значениях параметра  $\alpha$  (0,2 и 0,25) медианное мнение стабильно сдвигается «вправо» – в сторону группы, подвергаемой цензуре. Лишь при значении  $\alpha = 0,3$  и высокой гомофилии цензура оказывается эффективна.

Эффективно цензурировать сеть с высоким уровнем гомофилии при большей части возможных значений  $\alpha$  возможно только тогда, когда значительно увеличивается промежуток на пространстве мнений, в котором агенты подвергаются цензуре (до состояния от 0,6 до 1) и увеличивается время отключения от коммуникации (до как минимум 30 шагов модели). Данный результат объясняет, почему некоторые авторитарные страны вкладывают столь большие ресурсы в цензурирование социальных медиа и репрессии против, на первый взгляд, безвредных обладателей «нежелательных» мнений.

Таким образом, первая гипотеза подтверждается только частично: использование цензуры действительно может сдвигать общественное мнение в сторону цензурируемых мнений, но лишь при определенных условиях. Либо в сети должен быть очень низкий уровень гомофилии, либо цензурироваться должна значительная часть сети. Вторая гипотеза подтверждается полностью: эффект цензуры в значительной степени зависит от уровня гомофилии в сети. Цензурирование радикальных мнений эффективно только в случае низкой гомофилии. Что примечательно, цензура «цифровых элит» – агентов с наибольшими количеством соседей в цензурируемой группе – оказалась неэффективна в условиях высокой гомофилии.

## Выводы

Проведенные симуляции показали достаточно интересный результат. Модель продемонстрировала, что эффективность цензуры в значительной мере зависит от степени гомофилии в социальном медиа. При низкой гомофилии даже крайне слабое вмешательство эффективно: общественное мнение стабильно сдвигается в сторону не подвергаемых цензуре мнений. Однако при росте уровня гомофилии ситуация меняется: цензурирование только радикальных мнений оказывается недостаточным и ведет к противоположному эффекту – общественное мнение сдвигается в сторону цензурируемых мнений. В данном случае только обширная цензура, которая захватывает значительную часть пространства мнений, способна достичь желаемых результатов.

Конечно, следует отметить, что вычислительные модели показывают исключительно упрощенные представления реальности. В используемую модель не включен механизм, который делает цензурируемые «мнения» более интересными для пользователей социальных медиа<sup>12</sup>. Впрочем, как показали симуляции, и без него цензура может быть неэффективной.

Почему в условиях гомофилии цензура становится менее действенной? Модель подсказывает два потенциальных механизма. Во-первых, цензура не приводит к изменению мнений подвергнутых ей агентов: вернувшись в коммуникацию, они продолжают транслировать его на соседей. Во-вторых, в условиях гомофилии близость мнений влияет на

<sup>12</sup> Допустим, как продемонстрировали Вонг и Лианг, подвергнутые цензуре сообщения воспринимались как более достоверные [Wong, Liang 2021].

вероятность образования связи между двумя агентами. Блокирование наиболее радикальных мнений не мешает носителям чуть менее радикальных мнений высказывать их и влиять на свое окружение.

Из данных выводов следует, что, во-первых, цензурирование исключительно радикальных мнений неэффективно в социальных медиа в условиях гомофилии. Во-вторых демонстрируется, что эффективная цензура требует значительных ресурсов (как, например, во внутрикитайских социальных медиа). С одной стороны, можно предположить, что попытки использовать цензуру для борьбы с нежелательными мнениями обречены на провал: сторонники теорий заговоров, бывшего президента США Д. Трампа и «антипрививочных» движений продолжают действовать в социальных медиа и рекрутировать новых сторонников. С другой стороны, неэффективность «ограниченной» цензуры в отношении только самых радикальных мнений со временем станет заметной и для цензоров. В результате они будут вынуждены расширять масштаб своих действий, что рано или поздно сделает цензуру заметной и для пользователей. Между тем о неэффективности такой цензуры позволяет говорить значительный объем накопленных знаний.

## СПИСОК ЛИТЕРАТУРЫ

Ахременко А.С., Петров А.П., Жеглов С.А. (2021) Как информационно-коммуникационные технологии меняют тренды в моделировании политических процессов: к агентному подходу // Политическая наука. № 1. С. 12–45. DOI: <https://doi.org/10.31249/poln/2021.01.01>.

Калюков Е., Чебакова Д. (2021) «Яндекс» объяснил удаление сайта «Умного голосования» из поиска // РБК. 9 сентября. (<https://www.rbc.ru/politics/07/09/2021/613709739a79476fd52e153a>).

Козицин И.В., Щербинина А.А., Бабкина Т.С., Гильмутдинов Б.Н., Гойко В.Л., Сайфулин Э.Р. (2020) Эволюция социальных связей пользователей вконтакте и информационные пузыри // УПРАВЛЕНИЕ РАЗВИТИЕМ КРУПНОМАСШТАБНЫХ СИСТЕМ MLSД'2020. С. 1865–1873.

Суслов С.И. (2016) Кластеры петербургских политических онлайн-сообществ в «ВКонтакте» // Вестник Санкт-Петербургского университета. Социология. №. 4. С. 69–87. DOI: <https://doi.org/10.21638/11701/spbu12.2016.405>.

Шерстобитов А.С., Бряннов К.А. (2013) Технологии политической мобилизации в социальной сети «ВКонтакте»: сетевой анализ протестного и провластного сегментов // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. №. 10-1. С. 196–202.

Шишкина А.Р., Исаев Л.М. (2015) Интернет-цензура в арабских странах // Азия и Африка сегодня. №. 2. С. 40–45.

Adena M., Enikolopov R., Petrova M., Santarosa V., Zhuravskaya, E. (2015) Radio and the Rise of the Nazis in Prewar Germany // The Quarterly Journal of Economics. Vol. 130. № 4. Pp. 1885–1939. DOI: <https://doi.org/10.1093/qje/qjv030>.

Barberá P. (2014) How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US // Job Market Paper, New York University. Vol. 46. ([http://pablobarbera.com/static/barbera\\_polarization\\_APSA.pdf](http://pablobarbera.com/static/barbera_polarization_APSA.pdf)).

Barberá P., Jost T.J., Nagler J., Tucker J.A., Bonneau R. (2015) Tweeting from Left to Right: Is Online Political Communication More than an Echo Chamber? // Psychological Science. Vol. 26. № 10. Pp. 1531–1542. DOI: <https://doi.org/10.1177/0956797615594620>.

Barberá P., Wang N., Bonneau R., Jost T.J., Nagler J., Tucker J., González-Bailón S. (2015) The Critical Periphery in the Growth of Social Protests // PloS one. Vol. 10. № 11. <https://doi.org/10.1371/journal.pone.0143611>.

Bond R.M., Fariss C.J., Jones J.J., Kramer A.D., Marlow C., Settle J.E., Fowler J.H. (2012) A 61-Million-Person Experiment in Social Influence and Political Mobilization // *Nature*. Vol. 489. № 7415. Pp. 295–298. DOI: <https://doi.org/10.1038/nature11421>.

Cederman L. E. (1994) Emergent Polarity: Analyzing State-Formation and Power Politics // *International Studies Quarterly*. Vol. 38. № 4. Pp. 501–533. DOI: <https://doi.org/10.2307/2600863>.

Chen Y., Yang D. Y. (2019) The Impact of Media Censorship: 1984 or Brave New World? // *American Economic Review*. Vol. 109. № 6. Pp. 2294–2332. DOI: [10.1257/aer.20171765](https://doi.org/10.1257/aer.20171765).

Cinelli M., Morales G.D.F., Galeazzi A., Quattrociochi W., Starnini M. (2021) The Echo Chamber Effect on Social Media // *Proceedings of the National Academy of Sciences*. Vol. 118. № 9. DOI: <https://doi.org/10.1073/pnas.2023301118>.

Clayton J. (2021) Facebook Reverses Kyle Rittenhouse Policy // *BBC*. 1 December 2021. (<https://www.bbc.com/news/technology-59486397>).

Colleoni E., Rozza A., Arvidsson A. (2014) Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data // *Journal of Communication*. Vol. 64. № 2. Pp. 317–332. DOI: <https://doi.org/10.1111/jcom.12084>.

Collins B., Zadrozny B. (2021) Anti-Vaccine Groups Changing into ‘Dance Parties’ on Facebook to Avoid Detection // *NBC*. July 22. (<https://www.nbcnews.com/tech/tech-news/anti-vaccination-groups-turn-coded-language-facebook-pledges-crackdown-rcna1480>).

Del Vicario M., Bessi A., Zollo F., Petroni F., Scala A., Caldarelli G., Stanley H.E., Quattrociochi, W. (2016) The Spreading of Misinformation Online // *Proceedings of the National Academy of Sciences*. Vol. 113. № 3. Pp. 554–559. DOI: <https://doi.org/10.1073/pnas.1517441113>.

Du S., Gregory S. (2016) The Echo Chamber Effect in Twitter: Does Community Polarization Increase? // In: *International Workshop on Complex Networks and Their applications*. H. Cherifi., S. Gaito., W. Quattrociochi., A. Sala (eds.). Cham: Springer. Pp. 373–378. [https://doi.org/10.1007/978-3-319-50901-3\\_30](https://doi.org/10.1007/978-3-319-50901-3_30).

Epstein R., Robertson R. E. (2015) The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections // *Proceedings of the National Academy of Sciences*. Vol. 112. № 33. Pp. 4512–4521. DOI: <https://doi.org/10.1073/pnas.1419828112>.

González-Bailón S., Wang N. (2016) Networked Discontent: The Anatomy of Protest Campaigns in Social Media // *Social Networks*. Vol. 44. Pp. 95–104. DOI: <https://doi.org/10.1016/j.socnet.2015.07.003>.

Guilbeault D., Becker J., Centola D. (2018) Complex Contagions: A Decade in Review // In: S. Lehmann, Y.Y. Ahn (eds). *Complex Spreading Phenomena in Social Systems*. Cham: Springer. Pp. 3–25. DOI: [https://doi.org/10.1007/978-3-319-77332-2\\_1](https://doi.org/10.1007/978-3-319-77332-2_1).

Kim M. (2016) Facebook’s Spiral of Silence and Participation: The Role of Political Expression on Facebook and Partisan Strength in Political Participation // *Cyberpsychology, Behavior, and Social Networking*. Vol. 19. № 12. Pp. 696–702. DOI: <https://doi.org/10.1177/20563051211044239>.

King G., Pan J., Roberts M. E. (2013) How Censorship in China Allows Government Criticism but Silences Collective Expression // *American Political Science Review*. Vol. 107. № 2. Pp. 326–343. DOI: <https://doi.org/10.1017/S0003055413000014>.

King G., Pan J., Roberts M. E. (2014) Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation // *Science*. Vol. 345. № 6199. Pp. 891–903. DOI: [10.1126/science.1251722](https://doi.org/10.1126/science.1251722).

Lane J. E., McCaffree K., Shults F. L. R. (2021) Is Radicalization Reinforced by Social Media Censorship? // *arXiv*. (<https://arxiv.org/abs/2103.12842>),

McPherson M., Smith-Lovin L., Cook J. M. (2001) Birds of a Feather: Homophily in Social Networks // *Annual Review of Sociology*. Vol. 27. № 1. Pp. 415–444. DOI: <https://doi.org/10.1146/annurev.soc.27.1.415>.

Nabi Z. (2014) Resistance-Censorship is Futile // *First Monday*. Vol. 19. № 11. (<https://firstmonday.org/article/view/5525/4155>).

- Nickerson R. S. (1998) Confirmation Bias: A Ubiquitous Phenomenon in Many Guises // *Review of General Psychology*. Vol. 2. № 2. Pp. 175–220. DOI: <https://doi.org/10.1037/1089-2680.2.2.175>.
- Pan J., Siegel A. A. (2020) How Saudi Crackdowns Fail to Silence Online Dissent // *American Political Science Review*. Vol. 114. № 1. Pp. 109–125. DOI: <https://doi.org/10.1017/S0003055419000650>.
- Pariser E. (2011) *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin UK.
- Roberts M. E. (2020) Resilience to Online Censorship // *Annual Review of Political Science*. Vol. 23. Pp. 401–419. DOI: <https://doi.org/10.1146/annurev-polisci-050718-032837>.
- Rogers R. (2020) Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media // *European Journal of Communication*. Vol. 35. № 3. Pp. 213–229. DOI: <https://doi.org/10.1177/0267323120922066>.
- Romero D.M., Meeder B., Kleinberg J. (2011) Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter // *Proceedings of the 20th International Conference on World Wide Web*. Pp. 695–704. DOI: <https://doi.org/10.1145/1963405.1963503>.
- Ross B., Pilz L., Cabrera B., Brachten F., Neubaum G., Stieglitz, S. (2019) Are Social Bots a Real Threat? An Agent-Based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks // *European Journal of Information Systems*. Vol. 28. № 4. Pp. 394–412. DOI: <https://doi.org/10.1080/0960085X.2018.1560920>.
- Sarup G., Suchner R. W., Gaylord G. (1991) Contrast Effects and Attitude Change: A Test of the Two-Stage Hypothesis of Social Judgment Theory // *Social Psychology Quarterly*. Vol. 54. № 4. Pp. 364–372. DOI: <https://doi.org/10.2307/2786848>.
- Siegel D. A. (2011) When Does Repression Work? Collective Action in Social Networks // *The Journal of Politics*. Vol. 73. № 4. Pp. 993–1010. DOI: <https://doi.org/10.1017/S0022381611000727>.
- Siegel D. A. (2018) Analyzing Computational Models // *American Journal of Political Science*. Vol. 62. № 3. Pp. 745–759. DOI: <https://doi.org/10.1111/ajps.12364>.
- Steinert-Threlkeld Z. C. (2017) Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring // *American Political Science Review*. Vol. 111. № 2. Pp. 379–403. DOI: <https://doi.org/10.1017/S0003055416000769>.
- Wang C. (2021) Opinion Dynamics with Bilateral Propaganda and Unilateral Information Blockade // *Physica A: Statistical Mechanics and its Applications*. Vol. 566. DOI: [10.1016/j.physa.2020.125646](https://doi.org/10.1016/j.physa.2020.125646).
- Wilensky U. (1999) *NetLogo* // Netlogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (<https://ccl.northwestern.edu/netlogo/>)
- Wong, S. H. W., Liang, J. (2021) Dubious Until Officially Censored: Effects of Online Censorship Exposure on Viewers' Attitudes in Authoritarian Regimes // *Journal of Information Technology & Politics*. Vol. 18. № 3. Pp. 310–323. DOI: <https://doi.org/10.1080/19331681.2021.1879343>.
- Zhu Y., Fu K. (2020) Speaking Up or Staying Silent? Examining the Influences of Censorship and Behavioral Contagion on Opinion (Non-) Expression in China // *New Media & Society*. Vol. 23. № 12. Pp. 3634–3655. DOI: <https://doi.org/10.1177/1461444820959016>.
- Zittrain J. (2013) Engineering an Election // *Harvard Law Review Forum*. Vol. 127. Pp. 335–341. ([http://cdn.harvardlawreview.org/wp-content/uploads/2014/06/vol127\\_Symposium\\_Zittrain.pdf](http://cdn.harvardlawreview.org/wp-content/uploads/2014/06/vol127_Symposium_Zittrain.pdf)).

## REFERENCES

- Adena M., Enikolopov R., Petrova M., Santarosa V., Zhuravskaya, E. (2015) Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics*. vol. 130, no. 4, pp. 1885–1939. DOI: <https://doi.org/10.1093/qje/qjv030>.
- Akhremenko A.S., Petrov A.P., Zheglov S.A. (2021) Kak informatsionno-kommunikatsionnye tekhnologii menyayut trendy v modelirovanii politicheskikh protsessov: k agentnomu podkhodu [How Information and Communication Technologies Change Trends in Modeling Political Processes: to Agent-Based Approach]

mation and Communication Technologies Change Trends in Modelling Political Processes: Towards an Agent-Based Approach]. *Politicheskaya nauka*. no. 1, pp. 12–45.

Barberá P. (2014) How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University*. vol. 46. ([http://pablobarbera.com/static/barbera\\_polarization\\_APSA.pdf](http://pablobarbera.com/static/barbera_polarization_APSA.pdf)).

Barberá P., Jost T.J., Nagler J., Tucker J.A., Bonneau R. (2015) Tweeting from Left to Right: Is Online Political Communication More than an Echo Chamber? *Psychological Science*. vol. 26, no. 10, pp. 1531–1542. DOI: <https://doi.org/10.1177/0956797615594620>.

Barberá P., Wang N., Bonneau R., Jost T.J., Nagler J., Tucker J., González-Bailón S. (2015) The Critical Periphery in the Growth of Social Protests. *PLoS one*. vol. 10, no. 11. DOI: <https://doi.org/10.1371/journal.pone.0143611>.

Bond R.M., Fariss C.J., Jones J.J., Kramer A.D., Marlow C., Settle J.E., Fowler J.H. (2012) A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature*. vol. 489, no. 7415, pp. 295–298. DOI: <https://doi.org/10.1038/nature11421>.

Cederman L. E. (1994) Emergent Polarity: Analyzing State-Formation and Power Politics. *International Studies Quarterly*. vol. 38, no. 4, pp. 501–533. DOI: <https://doi.org/10.2307/2600863>.

Chen Y., Yang D. Y. (2019) The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review*. vol. 109, no. 6, pp. 2294–2332. DOI: [10.1257/aer.20171765](https://doi.org/10.1257/aer.20171765).

Cinelli M., Morales G.D.F., Galeazzi A., Quattrociocchi W., Starinini M. (2021) The Echo Chamber Effect on Social Media. *Proceedings of the National Academy of Sciences*. vol. 118, no. 9. DOI: <https://doi.org/10.1073/pnas.2023301118>.

Clayton J. (2021) Facebook Reverses Kyle Rittenhouse Policy. *BBC*. 1st December 2021. (<https://www.bbc.com/news/technology-59486397>).

Colleoni E., Rozza A., Arvidsson A. (2014) Echo Chamber or Public Sphere? Predicting Political Orientation and Measuring Political Homophily in Twitter Using Big Data. *Journal of Communication*. vol. 64, no. 2, pp. 317–332. DOI: <https://doi.org/10.1111/jcom.12084>.

Collins B., Zadrozny B. (2021) Anti-Vaccine Groups Changing into ‘Dance Parties’ on Facebook to Avoid Detection. *NBC*. July 22. (<https://www.nbcnews.com/tech/tech-news/anti-vaccination-groups-turn-coded-language-facebook-pledges-crackdown-rcna1480>).

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*. vol. 113, no. 3. pp. 554–559. DOI: <https://doi.org/10.1073/pnas.1517441113>.

Du S., Gregory S. (2016) The Echo Chamber Effect in Twitter: Does Community Polarization Increase? In: H. Cherifi., S. Gaito., W. Quattrociocchi., A. Sala (eds.). *International Workshop on Complex Networks and Their Applications*. Cham: Springer. pp. 373–378. DOI: [https://doi.org/10.1007/978-3-319-50901-3\\_30](https://doi.org/10.1007/978-3-319-50901-3_30).

Epstein R., Robertson R. E. (2015) The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections. *Proceedings of the National Academy of Sciences*. vol. 112, no. 33, pp. 4512–4521. DOI: <https://doi.org/10.1073/pnas.1419828112>.

González-Bailón S., Wang N. (2016) Networked Discontent: The Anatomy of Protest Campaigns in Social Media. *Social Networks*. vol. 44, pp. 95–104. DOI: <https://doi.org/10.1016/j.socnet.2015.07.003>.

Guilbeault D., Becker J., Centola D. (2018) Complex Contagions: A Decade in Review. In: S. Lehmann, Y.Y. Ahn (eds). *Complex Spreading Phenomena in Social Systems*. Cham: Springer. pp. 3–25. DOI: [https://doi.org/10.1007/978-3-319-77332-2\\_1](https://doi.org/10.1007/978-3-319-77332-2_1).

Kalyukov E., Chebakova D. (2021) «Yandeks» ob'yasnily udalenie sajta «Umnogo golosovaniya» iz poiska [“Yandex” Explained Removal of “Smart Vote” Site from Search]. *RBK*, September 9th. (<https://www.rbc.ru/politics/07/09/2021/613709739a79476fd52e153a>).

Kim M. (2016) Facebook's Spiral of Silence and Participation: The Role of Political Expression on Facebook and Partisan Strength in Political Participation. *Cyberpsychology, Behavior, and Social Networking*. vol. 19, no. 12, pp. 696–702. DOI: <https://doi.org/10.1177/20563051211044239>.

King G., Pan J., Roberts M. E. (2013) How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review*. vol. 107, no. 2, pp. 326–343. DOI: <https://doi.org/10.1017/S0003055413000014>.

King G., Pan J., Roberts M. E. (2014) Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation. *Science*. vol. 345, no. 6199, pp. 891–903. DOI: 10.1126/science.1251722.

Kozitsin I.V., Scherbinina A.A., Babkina T.S., Gil'mutdinov B.N., Gojko V.L., Sajfulin Eh.R. (2020) Ehvolyutsiya sotsial'nykh svyazej pol'zovatelej v kontakte i informatsionnye puzryi [Evolution of social ties of Vkontakte users and information bubbles]. *UPRAVLENIE RAZVITIEM KRUPNOMASSHTABNYKH SISTEM MLS D'2020*. pp. 1865–1873.

Lane J. E., McCaffree K., Shults F. L. R. (2021) Is Radicalization Reinforced by Social Media Censorship? *arXiv*. (<https://arxiv.org/abs/2103.12842>),

McPherson M., Smith-Lovin L., Cook J. M. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*. vol. 27, no. 1, pp. 415–444. DOI: <https://doi.org/10.1146/annurev.soc.27.1.415>.

Nabi Z. (2014) Resistance-Censorship is Futile. *First Monday*. vol. 19, no. 11. (<https://firstmonday.org/article/view/5525/4155>).

Nickerson R. S. (1998) Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*. vol. 2, no. 2, pp. 175–220. DOI: <https://doi.org/10.1037/1089-2680.2.2.175>.

Pan J., Siegel A. A. (2020) How Saudi Crackdowns Fail to Silence Online Dissent. *American Political Science Review*. vol. 114, no. 1, pp. 109–125. DOI: <https://doi.org/10.1017/S0003055419000650>.

Pariser E. (2011) *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin UK.

Roberts M. E. (2020) Resilience to Online Censorship. *Annual Review of Political Science*. vol. 23, pp. 401–419. DOI: <https://doi.org/10.1146/annurev-polisci-050718-032837>.

Rogers R. (2020) Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media. *European Journal of Communication*. vol. 35, no. 3, pp. 213–229. DOI: <https://doi.org/10.1177/0267323120922066>.

Romero D.M., Meeder B., Kleinberg J. (2011) Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. *Proceedings of the 20th International Conference on World Wide Web*. pp. 695–704. DOI: <https://doi.org/10.1145/1963405.1963503>.

Ross B., Pilz L., Cabrera B., Brachten F., Neubaum G., Stieglitz, S. (2019) Are Social Bots a Real Threat? An Agent-Based Model of the Spiral of Silence to Analyse the Impact of Manipulative Actors in Social Networks. *European Journal of Information Systems*. vol. 28, no. 4, pp. 394–412. DOI: <https://doi.org/10.1080/0960085X.2018.1560920>.

Sarup G., Suchner R. W., Gaylord G. (1991) Contrast Effects and Attitude Change: A Test of the Two-Stage Hypothesis of Social Judgment Theory. *Social Psychology Quarterly*. vol. 54, no. 4, pp. 364–372. DOI: <https://doi.org/10.2307/2786848>.

Sherstobitov A.S., Bryanov K.A. (2013) Tekhnologii politicheskoy mobilizatsii v sotsial'noj seti» VKontakte»: setevoy analiz protestnogo i provlastnogo segmentov [Political Mobilization Technologies in Social Network «VKontakte»: Network Analysis of Protest and Pro-Regime Segments]. *Istoricheskie, filosofskie, politicheskie i yuridicheskie nauki, kul'turologiya i iskusstvovedenie. Voprosy teorii i praktiki*. no. 10-1, pp. 196–202.

Shishkina A.R., Isaev L.M. (2015) Internet-tsenzura v arabskikh stranakh [Internet Censorship in Arab Countries]. *Aziya i Afrika segodnya*. no. 2, pp. 40–45.

Siegel D. A. (2011) When Does Repression Work? Collective Action in Social Networks. *The Journal of Politics*. vol. 73, no. 4, pp. 993–1010. DOI: <https://doi.org/10.1017/S0022381611000727>.

Siegel D. A. (2018) Analyzing Computational Models. *American Journal of Political Science*. vol. 62, no. 3, pp. 745–759. DOI: <https://doi.org/10.1111/ajps.12364>.

Steinert-Threlkeld Z. C. (2017) Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring. *American Political Science Review*. vol. 111, no. 2, pp. 379–403. DOI: <https://doi.org/10.1017/S0003055416000769>.

Suslov S.I. (2016) Klastery peterburgskikh politicheskikh onlajn-soobschestv v «VKontakte» [Saint-Petersburg's Political Online Communities Clusters in “Vkontakte”]. *Vestnik Sankt-Peterburgskogo universiteta. Sotsiologiya*. no. 4, pp. 69–87. DOI: <https://doi.org/10.21638/11701/spbu12.2016.405>.

Wang C. (2021) Opinion Dynamics with Bilateral Propaganda and Unilateral Information Blockade. *Physica A: Statistical Mechanics and its Applications*. vol. 566. DOI: 10.1016/j.physa.2020.125646.

Wilensky U., (1999) NetLogo. Netlogo, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (<https://ccl.northwestern.edu/netlogo/>).

Wong, S. H. W., Liang, J. (2021) Dubious Until Officially Censored: Effects of Online Censorship Exposure on Viewers' Attitudes in Authoritarian Regimes. *Journal of Information Technology & Politics*. vol. 18, no. 3, pp. 310–323. DOI: <https://doi.org/10.1080/19331681.2021.1879343>.

Zhu Y., Fu K. (2020) Speaking Up or Staying Silent? Examining the Influences of Censorship and Behavioral Contagion on Opinion (Non-) Expression in China. *New Media & Society*. vol. 23, no. 12, pp. 3634–3655. DOI: <https://doi.org/10.1177/1461444820959016>.

Zittrain J. (2013) Engineering an Election. *Harvard Law Review Forum*. vol. 127, pp. 335–341. ([http://cdn.harvardlawreview.org/wp-content/uploads/2014/06/vol127\\_Symposium\\_Zittrain.pdf](http://cdn.harvardlawreview.org/wp-content/uploads/2014/06/vol127_Symposium_Zittrain.pdf)).

## Информация об авторе

**Толокнев Кирилл Александрович**, аспирант аспирантской школы по политическим наукам, Национальный исследовательский университет «Высшая школа экономики». Адрес: Мясницкая улица, 20, 101000, Москва. E-mail: [kirtoloknev@gmail.com](mailto:kirtoloknev@gmail.com)

## About the author

**Kirill A. Toloknev**, Postgraduate Student, Doctoral School of Political Science, National Research University “Higher School of Economics”. Address: 101000, Moscow, Myasnitckaya Ulitsa, 20. E-mail: [kirtoloknev@gmail.com](mailto:kirtoloknev@gmail.com)

Статья поступила в редакцию / Received: 28.09.2021

Статья поступила после рецензирования и доработки / Revised: 15.12.2021

Статья принята к публикации / Accepted: 22.12.2021