

УДК-323.2

DOI: 10.17072/2218-1067-2021-4-16-29

**БОТЫ, ВЕДУЩИЕ НАРОД?
МОДЕЛЬ ВЛИЯНИЯ БОТОВ НА ПОЛИТИЧЕСКУЮ МОБИЛИЗАЦИЮ И
ДЕМОБИЛИЗАЦИЮ В СОЦИАЛЬНЫХ МЕДИА**

К. А. Толокнев

Толокнев Кирилл Александрович, аспирант,

Национальный исследовательский университет «Высшая школа экономики», Россия, Москва

E-mail: kirtoloknev@gmail.com (ORCID: 0000-0003-3712-9978. ResearcherID: AAD-3672-2020).

Аннотация

Исследования влияния ботов (автоматизированных аккаунтов) в социальных медиа привлекают все большее внимание исследователей. Автоматизированные аккаунты применяются как государственными, так и негосударственными акторами для манипулирования коммуникациями в социальных медиа. Тем не менее, пока эта проблематика, несмотря на очевидные достижения, остается относительно мало изученной. Весьма заметен недостаток результатов в сфере изучения эффективности ботов для мобилизации (поощрения) и демобилизации (подавления) политических коммуникаций в социальных медиа. Отчасти это связано с попытками рассматривать эту проблему исключительно на основе эмпирических исследований, крайне трудоемких и страдающих из-за большого числа потенциальных объясняющих факторов, без использования методов вычислительного моделирования, способных симулировать исследуемую систему в контролируемой модельной среде. В данной статье предлагается новая абстрактная агентно-ориентированная модель, отличающаяся от более ранних разработок учетом стратегического характера применения ботов – как инструментов и мобилизации сторонников, и подавления мобилизации противников. Было продемонстрировано, что использование ботов, нацеленных на демобилизацию противников, более эффективно и стабильно, чем использование ботов для мобилизации сторонников: при попытке мобилизации сторонников применение излишне «радикальных» ботов способно привести к подавлению мобилизации группы. Также показано, что в условиях высокой гомофилии боты оказываются значительно менее эффективны, чем в условиях низкой гомофилии. Полученные результаты позволяют оценить эффективность ботов и могут служить для создания проверяемых эмпирическим образом гипотез.

Ключевые слова: боты; социальные медиа; политические коммуникации; политическая мобилизация; политическая демобилизация; агентно-ориентированная модель; вычислительное моделирование.

Социальные связи имеют очень большое значение в жизни людей. Через социальные связи распространяется информация, оказывается нормативное давление, распространяются модели поведения (Gould, 1991). Социальное окружение значительно влияет и на политическое участие. Человек, находящийся в меньшинстве, с меньшей вероятностью будет открыто высказывать свою политическую позицию (Noelle-Neumann, 1974). Если его политическая позиция совпадает с позицией большинства, возникает эффект «безопасности в массах» (Kuran, 1991): в таком случае люди, ранее скрывавшие свои политические позиции, с большей вероятностью их откроют. Структура социальных связей также оказывает влияние на политическое участие: уровень политического участия зависит от типа социальной сети, связывающей людей (Siegel, 2009).

В 2010 г. Ларри Даймонд назвал Интернет и социальные медиа «технологиями освобождения» – технологиями, потенциально способными расширить возможности для общественной дискуссии и политического участия в недемократических странах (Diamond, 2010). Ранние исследования, посвященные влиянию социальных медиа на протестную мобилизацию, показывали едва ли не решающую роль коммуникации в социальных медиа для протестной мобилизации: события арабской весны часто упоминались как «твиттерные» или «фейсбучные революции» или «восстания» (Farrell,

2012). Период первоначального оптимизма сменился более сдержанными оценками: согласившись, что влияние действительно есть, исследователи перестали рассматривать социальные медиа как основную «причину» политической мобилизации (Stein, 2017). Большого внимания удостоился потенциал использования социальных медиа самими авторитарными правительствами как для подавления политической коммуникации и мобилизации среди оппозиционеров, так и для мобилизации лоялистов. В качестве примеров тут следует назвать действия «50-центовой армии», распространяющей проправительственный контент в китайских социальных медиа (King et al., 2017), или «кремлеботов», активных в российском сегменте Твиттера (Sanovich et al., 2018).

Впрочем, не следует считать, что попытки манипулировать коммуникацией в социальных медиа являются «эксклюзивной» тактикой авторитарных режимов: в российском сегменте Твиттера так же действует значительное число «оппозиционных» ботов (Sanovich et al., 2018), ботов активно применяли сторонники Евромайдана, «альтернативные правые» и сторонники террористических организаций (Benigni et al., 2019).

Боты – распространенные обитатели социальных медиа. По оценкам исследователей, они могут составлять до 9–15% от общего числа аккаунтов в Твиттере (Assenmacher et al., 2020) и генерировать более половины контента по определенным темам (Stukal et al., 2017). В Фейсбуке боты распространены не столь сильно, по оценкам самой компании разного рода фейковые аккаунты составляют около 5% от общего числа аккаунтов. В прочих социальных медиа также присутствуют боты. Но, несмотря на повышенное внимание к ботам со стороны СМИ, зачастую представляющих ботов в качестве крайне эффективного инструмента влияния на общественное мнение, исследователям пока не удается однозначно ответить на вопрос, эффективно ли использование ботов для политической мобилизации и демобилизации. Именно этот вопрос находится в центре внимания данной статьи.

Боты: эмпирические исследования и математические модели

Анализ влияния ботов на политические коммуникации – относительно новая сфера политических исследований. В настоящий момент пока не сложилось единого общепризнанного определения понятия «бот». Боты могут рассматриваться как в узкой интерпретации, то есть исключительно как автоматизированные аккаунты, так и в расширенной, когда в число «ботов» могут включаться и «тролли»-люди, и «гибриды» – полуавтоматизированные аккаунты. Объем накопленных знаний пока также еще относительно невелик. Тем не менее, определенных, и весьма важных, результатов исследователям ботов достигнуть удалось. Эмпирические исследования ботов проходят по двум основным направлениям.

В рамках первого направления специалисты в области компьютерных наук разрабатывают методы выявления ботов с использованием различных методов машинного обучения. Феррера и коллеги выделили три основных подхода, используемых для обнаружения ботов (Ferrara et al., 2016).

Первый подход к их выявлению предполагает анализ структуры социальной сети на основе предположения, что боты и люди различаются по сетевым характеристикам положения в сети.

Второй подход более прямолинеен – он предполагает классификацию аккаунтов с использованием труда большего числа людей, как экспертов, так и так называемых «турков», то есть работников, выполняющих однотиповые, но не автоматизируемые задачи, (Wang et al., 2012). Ручная классификация продолжает использоваться при исследованиях ботов на этапе создания обучающих выборок при использовании методов машинного обучения с учителем, но из-за трудоемкости процесса не получила распространения как полноценный подход к выявлению большого числа ботов.

Третий подход предполагает использование методов машинного обучения для обнаружения ботов на основе определенного набора их характеристик, таких как количество твиттов (постов), использования геопозиции, соотношения числа друзей и фолловеров (для Твиттера) и прочих. Этот подход позволяет не только отделять людей от ботов, но и классифицировать самих ботов, различая их по типам поведения. К числу таких работ относятся работы Стукала и его коллег (Stukal et al., 2019; Sanovich et al., 2018) или Ферреры (Ferrara, 2017).

В рамках второго направления исследуются результаты действия ботов, их влияние на коммуникации в социальных медиа и стратегии применения ботов.

Первое задокументированное массовое применение ботов в политических целях произошло в 2010 г. на промежуточных выборах в США (Metaxas & Mustafaraj, 2012; Ratkiewicz et al., 2011). Именно тогда впервые были выявлены автоматизированные аккаунты, поддерживающие отдельных политиков. Большой толчок исследованиям влияния ботов на политические коммуникации дало

вмешательство нескольких тысяч предположительно российских аккаунтов в президентские выборы 2016 г. в США и позднейшие попытки применить ботов для воздействия на электоральные процессы в европейских странах, в чем также обвиняли Россию. Несмотря на некоторые сомнения в нейтральности этих исследований, в результате появился набор достаточно ценных эмпирических наблюдений, посвященных поведению ботов в социальных медиа.

Во-первых, было показано, что боты используют разные стратегии поведения. Некоторые «русские боты» пытались представить себя аккаунтами локальных новостных агентств, собирающих и распространяющих новости местного значения. Так называемые «правые боты» поддерживали Трампа, распространяя «правопопулистские твиты», критиковали Демократическую партию и «умеренных республиканцев». В свою очередь, «левые боты» поддерживали Сандерса, обсуждали вопросы «гендерной, сексуальной, религиозной и расовой идентичности» и критиковали Клинтон (Linville & Warren, 2020).

Во-вторых, «правые» и «левые» боты действовали скоординировано: к примеру, 12 сентября 2016 г. «правые боты» активно комментировали состояние здоровья Хиллари Клинтон, «распространяли теорию заговора, что она использует двойников, чтобы скрыть свое плохое состояние здоровья», в то время как «левые тролли» были не особенно активны (Linville & Warren, 2020).

Помимо применения ботов непосредственно для воздействия на коммуникацию также популярно использование ботов для «накрутки подписчиков» – завышения числа последователей в социальном медиа. Как показала работа Сильвы и Прокша, особенно активно для этой цели ботов использовали правые евроскептики (Silva & Proksch, 2021).

Исследования применения ботов в других странах тоже подтвердили существование «политической специализации» у ботов. Во время референдума о независимости Каталонии боты стремились повлиять и на сторонников независимости, и на противников независимости, однако контент, продвигавшийся ботами, различался. Среди сторонников независимости боты распространяли сообщения, способствующие обострению конфликта, «связывавшие обретение «независимости» и «свободы» с необходимостью сражаться с «диктатурой» (испанским правительством)», разжигали ненависть по отношению к полицейским (Stella et al., 2018). Боты среди противников независимости такие сообщения не продвигали.

Впрочем, не следует считать, что боты в социальных медиа практически неотличимы от людей или что их операторы всегда используют продвинутые стратегии манипуляции. Исследование влияния ботов на выборы в Германии в 2017 г. не выявило наличие у ботов ни скоординированной стратегии, ни собственно значительного влияния на ход политических коммуникаций (Brachten et al., 2017), а большая часть доступных для покупки на «рынке» ботов представляют собой примитивные программы, не способные достоверно выдавать себя за людей (Assenmacher et al., 2020).

Несколько особняком стоят исследования, посвященные вычислительному (агентно-ориентированному) моделированию влияния ботов на политические коммуникации. Подобные модели можно разделить на два класса: «факсимильные» модели, стремящиеся как можно точнее симулировать особенности социальных медиа, такие как разные типы связей между людьми (допустим, различия между двухсторонней «дружбой» и односторонним фолловингом) и «абстрактные» модели, где специфические особенности социальных медиа передаются в очень общем виде, а упор сделан на простоте симуляции.

Среди «факсимильных» моделей наибольший интерес представляет работа Бескоу и Карли, рассматривавшая так называемые «дезинформационные маневры» ботов (Beskow & Carley, 2019). Выделяется четыре типа «дезинформационных маневров», выполняемых с использованием ботов: позитивную и негативную манипуляцию «нарративами» – действия, направленные на поощрение или подавление какой-либо дискуссии в социальных медиа, и позитивную и негативную манипуляцию «социальной сетью» – действия, направленные на изменение структуры социальной сети с целью либо поддержать коммуникацию в социальном медиа («позитивная манипуляция»), либо подавить ее («негативная манипуляция») (Carley, 2020). Модель Карли и Беркоу (Beskow & Carley, 2019) использовалась для анализа двух методов «позитивной манипуляции социальной сетью»: использование ботов для поддержки лидеров мнений и использование ботов для создания «мостов» между группами пользователей. Их вычислительные эксперименты показали, что 12% ботов в составе социального медиа достаточно для того, чтобы их воздействие значительно повлияло на коммуникацию.

Среди «абстрактных» моделей популярность получили модели, рассматривающие возникновение в социальных медиа «спирали молчания» – подавления публичного выражения своей позиции меньшинством, вызванного страхом социальной изоляции. Среди данных моделей примечательна

работа Росса (Ross et al., 2019), демонстрирующая, что для возникновения глобальной «спирали молчания» достаточно 2–4% ботов в социальной сети. Впрочем, другая модель «спирали молчания», разработанная Соном (Sohn, 2019), показала, что глобальная «спираль молчания» возникает в социальных медиа только в редких случаях.

Как мы видим, ни эмпирические исследования, ни исследованная, выполненная методами вычислительного моделирования пока не дают четкого ответа на вопрос, насколько эффективно использование ботов¹, и если эффективно, то каков результат от их применения: позволяют ли боты результативно мобилизовать сторонников, подавлять мобилизацию противников или ни то, ни другое. Использование методов вычислительного моделирования обладает значительным преимуществом при исследовании данной сферы, так как в условиях новизны дисциплины и высокой сложности проведения исследований вычислительные модели позволяют генерировать гипотезы, которые в дальнейшем могут быть проверены уже на эмпирическом материале.

Далее будет предложена агентно-ориентированная модель, позволяющая рассмотреть данные вопросы.

Вычислительная модель: свои и чужие боты

Агентно-ориентированное моделирование – один из методов вычислительного моделирования. Основа метода – создание модели, состоящей из некоторого конечного числа автономных агентов, взаимодействующих по определенным и заранее установленным исследователем правилам, находящимся в ограниченном числе состояний (Siegel, 2018). Как отметил Зигель, наиболее известный вид моделирования поведения агентов — это теоретико-игровой подход, где правило поведения агента – максимизация собственного выигрыша (Siegel, 2018). При использовании иных правил поведения агентов получение аналитического решения модели становится невозможным, и в таких случаях поведение модели анализируется при помощи вычислительных экспериментов (Ахременко и др., 2019).

Рассматриваемая модель состоит из 2000 агентов-людей, объединенных в безмасштабную сеть. Безмасштабная сеть – тип графа, в котором распределение степеней вершин (числа соседей) подчинено степенному закону: доля вершин со степенью k приблизительно равна $k^{-\gamma}$, где $2 < \gamma < 3$. Архитектура безмасштабной сети обладает сходством с реальными социальными медиа: относительно небольшое количество агентов оказывается связано с диспропорционально большим числом соседей, симулируя существование «цифровой элиты» (González-Bailón & Wang, 2016), что делает использование безмасштабной сети «стандартом» для «абстрактных» моделей социальных медиа (к примеру, Ross et al., 2019). Однако есть еще один фактор, который необходимо принять во внимание.

Люди предпочитают общество достаточно схожих с ними людей: с одним уровнем образования, социальным происхождением, возрастом и религиозными воззрениями (McPherson et al., 2001). Сходство политических убеждений также является значимым фактором для создания социальных связей: как показали Мослех и его коллеги, пользователи Твиттера с большей вероятностью создавали связи с ботами, чья партийная позиция совпадала с их собственной (Mosleh et al., 2020). Хубер и Малхорта показали, что сходство политических позиций у пользователей сайта для знакомств увеличивало вероятность того, что пользователи начнут коммуникацию, на 9.5% (Huber & Malhotra, 2017). Это явление называется гомофилия. Нельзя сказать, что гомофилия влияет только на само создание социальных связей, благодаря действию этого механизма распространяются общие модели поведения (DellaPosta et al., 2014), происходит рекрутирование новых участников протестных движений (Valenzuela et al., 2014). Для того, чтобы учесть влияние гомофилии, в модель вводится параметр H , принимающий значение от 0 до 1 и определяющий насколько сильно могут различаться позиции агентов, связанных друг с другом в сети. Чем выше значение H , тем ниже уровень гомофилии, и тем более отличающиеся друг от друга агенты могут создавать связи.

Представим себе, что есть несколько человек, объединенных социальной сетью. В каких условиях они могут высказать свою позицию? Во-первых, если они достаточно высоко мотивированы. Во-вторых, в условиях, когда достаточное число их соседей уже высказывают свою позицию, и эта позиция близка позиции молчащего: как под воздействием социального давления² (Klandermans, 1984), так и за счет эффекта «безопасности в массах» (Kuran, 1991). Подобными «соседями» в соци-

¹ Если, конечно, не считать, что «русские боты» значительно повлияли на избрание Трампа.

² «Социальный мотив» к участию по Кландермансу.

альных медиа могут быть как люди, так и боты. В таком случае боты могут способствовать мобилизации, увеличив число политически активных людей. Однако и боты, и люди могут так же и транслировать достаточно далекую позицию, завышая число соседей, поддерживающих противоположную точку зрения. В этом случае на индивида будет оказываться негативное воздействие, побуждающее его скрыть свою позицию.

Чтобы учесть этот фактор – возможность двоякого влияния социального окружения, до начала симуляции устанавливается параметр α , определяющий максимальное различие между позицией агента и усредненной позицией его активных соседей, при котором неактивный агент может присоединиться к активной группе и открыть свою позицию.

Индивидуальны для каждого агента следующие параметры: позиция p (принимающая значения между 0 и 1, извлекаемая из равномерного распределения) и «внутренняя мотивация» m (принимающая значения между 0 и 1, извлекаемая из равномерного распределения)³. Позиция p показывает отношение агента к какой-либо политической проблеме, выступая в роли показателя положения агента на условном «идеологическом спектре», где, например, позиция 0 – полная поддержка определенной политической инициативы, а 1 – ее полное неприятие. Внутренняя мотивация m определяет готовность агента к демонстрации своей позиции и его способность противодействовать негативному воздействию окружения (Noelle-Neumann, 1974).

В момент начала симуляции активируются агенты, для которых выполняется неравенство:

$$(m * |p - 0.5|) > U[0,1].$$

Для сохранения преемственности с более ранними работами (Sohn & Geidner., 2016; Ross et al., 2020; Sohn, 2019) принимается, что публично индивид транслирует не свою точную позицию, а общее отношение «за» или «против». Активированные таким образом агенты со значением позиции $p > 0.5$ образуют группу 1, со значением позиции $p < 0.5$ – группу 2. В дальнейшем, для удобства, входящие в состав группы 2 агенты будут называться «оппозиционеры», входящие в состав группы 1 – «лоялисты».

На последующих шагах симуляции возможность активироваться (открыть позицию и прикнуться к одной из двух групп) получают агенты, для которых среднее значение суммы «внутренней мотивации агента» и доли активных соседей с мнением, находящимся в пределах интервала α , ниже извлекаемого из равномерного распределения с минимальным значением 0 и максимальным значением 1 случайного числа.

Если число активных агентов с близкой позицией (входящих в интервал α) обозначить как n_s , а общее число агентов-соседей как n , то формально это можно записать так:

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

В случае, если условие выше выполняется, агент сравнивает свою позицию p с усредненными позициями p своих активных соседей: он присоединяется к той группе, различия с которой минимальны, но только в том случае, если это различие не превышает заранее установленное значение α . Разделение групп по позиции 0.5, используемое в момент инициализации симуляции, здесь не применяется. Поэтому агент, чья позиция $p = 0.4$ (потенциальный «оппозиционер»), будучи окруженным активными агентами из группы «лоялистов», может публично присоединиться к группе «лоялистов», сфальсифицировав свои предпочтения.

Находясь в неблагоприятном социальном окружении, люди склонны скрывать свои политические позиции (Noelle-Neumann, 1974). Следовательно, активный агент так же должен обладать способностью скрыть свою позицию, если его группа оказывается в меньшинстве. Соотношение числа сторонников и противников агента в его социальном окружении или локальный климат мнений c , определяется для каждого активного агента по следующей формуле:

$$c = \begin{cases} \frac{n_g - n_d}{n_g + n_d} & \text{если } n_d \text{ и } n_g > 0, \\ \text{иначе } 0. \end{cases}$$

где n_g – количество агентов в той же группе, что и данный агент, n_d – число агентов в другой группе. В случае, если для агента i выполняется $m_i - c_i < m_i$, агент скрывает свою позицию.

³ Рассматривалось так же m , извлекаемое из нормального распределения, различий в поведении модели установлено не было.

Помимо агентов, симулирующих поведение людей, в сети присутствуют боты. Боты отличаются от людей-агентов по трем основным характеристикам. Во-первых, они всегда демонстрируют свою позицию и принадлежат либо к числу «лоялистов», либо к числу «оппозиционеров». Во-вторых, боты всегда имеют меньше связей, чем обычные агенты (люди). В-третьих, боты нацеливаются на агентов, чья позиция находится в определенном промежутке значений позиции p . Первые две характеристики являются стандартом, используемым в «абстрактных» агентно-ориентированных моделях социальных медиа. Третья характеристика позволяет учесть стратегический характер применения ботов: боты используются направленно, для воздействия на определенные сегменты аудитории (Stella et al., 2018; Linvill & Warren, 2020).

Сразу отметим, что в данном случае «бот» симулирует не только собственно ботов в узком смысле – автоматизированные аккаунты, но и любой аккаунт в сети, постоянно демонстрирующий свою политическую позицию. Таковыми могут быть как «тролли» (проплаченные политизированные аккаунты), так и интернет-активисты (не проплаченные политизированные аккаунты).

Результаты вычислительных экспериментов

Моделирование выполнено в программе Netlogo версии 6.2.0 (Wilensky & Rand, 2015). Следующие параметры в процессе проведения симуляции варьировались: уровень гомофилии H (0.2, 0.4, 0.6, 0.8, 1), значение α (0.1, 0.15, 0.2, 0.25, 0.3, 0.35⁴). Во всех симуляциях использовалось 2000 обычных агентов и, если рассматривались боты, 40 ботов или 200 ботов (2% и 10% относительно исходного числа людей соответственно). Чтобы избежать случайных отклонений, возникающих при создании сети и распределении параметра позиции p по агентам с различным числом «соседей», проводилось множество симуляций (100 для каждой комбинации параметров) и анализировались агрегированные результаты по множеству симуляций.

Для начала рассмотрим поведение модели в условиях отсутствия ботов. В зависимости от значения гомофилии H выделяются две основных траектории поведения модели. При слабой гомофилии возникают две приблизительно равные по численности группы агентов; при сильной гомофилии возникает «спираль молчания», и одна из групп оказывается значительно меньше другой – большинство подавляет меньшинство. Так же различия в размере возникающих групп возникают при увеличении значения параметра α . Кратко различия между группами представлены в таблице ниже.

Таблица 1

Значение α	Значение гомофилии H					
	$H = 0.2$	$H = 0.4$	$H = 0.6$	$H = 0.8$	$H = 1$	
0.1	100,5	28,3	35,9	80,7	112,1	Различие в среднем количестве агентов в группах
0.15	110,7	45,3	50,2	124,3	187,3	
0.2	92,9	71,6	78,2	220,0	317,9	
0.25		79,0	109,38	276,7	402,9	
0.3		83,3	123,7	323,2	400,3	
0.35		99,0	139,8	388,7	510,8	

Следует отметить, что при усреднении всех прогонов обе группы являются приблизительно одинаковыми по численности: без применения ботов не удастся стабильно создавать группу «оппозиционеров» большего размера, чем группу «лоялистов» или наоборот. Однако при применении ботов картина меняется.

Рассмотрим первую потенциальную стратегию применения ботов – использование ботов для поддержки мобилизации одной из групп. В этом случае актер, использующий ботов, будет стараться «завысить» число активных агентов в социальном круге потенциальных сторонников, стремясь таким образом «мобилизовать» сторонников.

Рассмотрим ситуацию, когда параметр позиции p у ботов равен 0, а боты нацелены на потенциальных «оппозиционеров». Если боты нацелены на группу агентов с позициями между 0 и 0.2 и

⁴ При дальнейшем увеличении параметра значимых изменений в поведении выявлено не было.

10% ботов, то при значениях гомофилии выше $H > 0.8$, число активных «оппозиционеров» стабильно выше числа активных «лоялистов»: при $H = 0.8$ в среднем в 1.8 раз, при $H = 1$ – в 2.5 раза. При 2% ботов эффект скромнее: при $H = 0.8$ «оппозиционеров» в среднем в 1.3 раза больше «лоялистов», при $H = 1$ – в 1.5 раза. Однако при усилении гомофилии эффект пропадает: при значениях $H < 0.8$ значительная разница в количестве агентов между двумя группами отсутствует.

При нацеливании ботов на агентов с позициями в промежутке от 0.15 до 0.35 при 10% ботов в сети стабильного эффекта удается добиться и при $H = 0.6$: активных «оппозиционеров» в среднем в 1.2 раза больше, чем активных «лоялистов», при $H = 0.8$ – в 2.3 раза, при $H = 1$ – в 2.8 раз. Однако при значениях гомофилии $H < 0.6$ эффект все так же отсутствует. При 2% ботов и $H = 0.6$ эффекта снова нет: при $H = 0.8$ – «оппозиционеров» в среднем в 1.2 раза больше, чем «лоялистов», при $H = 1$ – в 1.3 раза.

Однако, если нацелить ботов на агентов с позициями в промежутке от 0.3 до 0.5, ситуация меняется. При 2% ботов этот новый эффект проявляется при очень сильной гомофилии ($H = 0.2$) и очень слабой гомофилии ($H = 1$): число «лоялистов» в среднем оказывается в 1.15 и 1.17 раз больше числа «оппозиционеров». При 10% ботов число «оппозиционеров» и всех значениях H стабильно оказывалось меньше числа «лоялистов». Для примера, при очень сильной гомофилии ($H = 0.2$) в среднем «лоялистов» было в 1.6 раз больше, чем «оппозиционеров», а при очень слабой гомофилии ($H = 1$) – в 1.8 раза, то есть боты, транслируя слишком далекую позицию, ограничили размер группы, мобилизацию которой должны были стимулировать.

Весьма примечательно, что подобный эффект сохранялся в условиях нацеливания ботов с позицией 0 на агентов с позициями от 0 до 0.2 и 0.3 до 0.5. В этом случае только в условиях слабой гомофилии ($H > 0.6$) число «оппозиционеров» превышало число «лоялистов». Это наталкивает на два вывода.

Для удобства сведем полученные результаты в таблицу:

Таблица 2

Позиция ботов	Значение гомофилии H				
	$H = 0.2$	$H = 0.4$	$H = 0.6$	$H = 0.8$	$H = 1$
$p = 0$, 10% ботов					
Боты нацелены на группу от 0 до 0.2	Нет эффекта			Позитивный эффект ⁵	
Боты нацелены на группу от 0.15 до 0.35	Нет эффекта		Позитивный эффект		
Боты нацелены на группу от 0.3 до 0.5	Негативный эффект ⁶				
$p = 0$, 2% ботов					
Боты нацелены на группу от 0 до 0.2	Нет эффекта			Позитивный эффект	
Боты нацелены на группу от 0.15 до 0.35	Нет эффекта			Позитивный эффект	
Боты нацелены на группу от 0.3 до 0.5	Негативный эффект	Нет эффекта			Негативный эффект

Во-первых, боты, используемые для мобилизации сторонников, должны быть «точно нацелены». Ошибка может привести к нежелательным последствиям.

Во-вторых, боты могут использоваться для проведения информационной манипуляции «под чужим флагом». Можно легко представить, как некое правительство или иная организация использует ботов, замаскированных под излишне радикальных сторонников своих противников.

Теперь рассмотрим другую ситуацию, когда параметр позиции p у ботов равен 0.4. Если боты нацелены на группу агентов с позициями между 0 и 0.2, то, в зависимости от значения α , возникает

⁵ Использование ботов увеличило уровень мобилизации в целевой группе.

⁶ Использование ботов снизило уровень мобилизации в целевой группе.

два типа поведения. При низких значениях α (меньше 0.25) «лоялистов» стабильно больше, чем «оппозиционеров» при всех значениях параметра гомофилии H . При высоких значениях α (выше 0.25) и слабой гомофилии ($H > 0.6$) «оппозиционеров» стабильно больше, чем «лоялистов». Это происходило и при низком, и при высоком количестве ботов (2 и 10% соответственно).

При нацеливании ботов на агентов с позициями в промежутке от 0.15 до 0.35 при 10% ботов в сети, как и в предыдущем случае, стабильного эффекта удастся добиться и при $H = 0.6$: активных «оппозиционеров» в среднем в 1.3 раза больше, чем активных «лоялистов», при $H = 0.8$ – в 2.5 раза, при $H = 1$ – в 3 раза. Однако при значениях гомофилии $H < 0.6$ эффект все так же отсутствует. При 2% ботов и $H = 0.6$ эффекта снова нет: при $H = 0.8$ – «оппозиционеров» в среднем в 1.4 раза больше, чем «лоялистов», при $H = 1$ – в 1.6 раз.

При нацеливании ботов на агентов с позициями в промежутке от 0.3 до 0.5 число «оппозиционеров» стабильно превышает число «лоялистов». При 2% ботов при очень сильной гомофилии ($H = 0.2$) «оппозиционеров» в среднем было в 1.3 раза больше «лоялистов», при очень слабой ($H = 1$) – в 2.6 раза. При 10% ботов при очень сильной гомофилии ($H = 0.2$) «оппозиционеров» в среднем было в 1.3 раза больше «лоялистов», при очень слабой ($H = 1$) – в 9.5 раз.

Для удобства сведем полученные результаты в таблицу:

Таблица 3

Позиция	Значение гомофилии H				
	Позиция ботов $p = 0.4$	$H = 0.2$	$H = 0.4$	$H = 0.6$	$H = 0.8$
Позиция ботов $p = 0.4$, 2% ботов					
Боты нацелены на группу от 0 до 0.2	Эффект зависит от значения α : при значениях $\alpha < 0.25$ негативный эффект, при значениях $\alpha > 0.25$ позитивный эффект				
Боты нацелены на группу от 0.15 до 0.35	Нет эффекта			Позитивный эффект	
Боты нацелены на группу от 0.3 до 0.5	Позитивный эффект				
Позиция ботов $p = 0.4$, 10% ботов					
Боты нацелены на группу от 0 до 0.2	Эффект зависит от значения α : при значениях $\alpha < 0.25$ негативный эффект, при значениях $\alpha > 0.25$ позитивный эффект				
Боты нацелены на группу от 0.15 до 0.35	Нет эффекта			Позитивный эффект	
Боты нацелены на группу от 0.3 до 0.5	Позитивный эффект				

Таким образом, можно сделать несколько предварительных выводов. Во-первых, наибольший эффект боты имеют в условиях слабой гомофилии, а при сильной гомофилии их эффект незначителен. Во-вторых, наибольший эффект имеют боты, нацеленные не на наиболее радикальных членов сети и не транслирующие наиболее радикальные мнения. Более того, если боты, транслирующие наиболее радикальное мнение (в данном случае позицию $p = 0$), нацелить на достаточно далеких от них людей, то эффект будет противоположным. Боты подавляют процесс открытия позиций, то есть фактически подавляют потенциальную политическую мобилизацию. В-третьих, что достаточно очевидно, большее число ботов дает больший эффект.

Наиболее примечательны первые два наблюдения. Барбера показал, что уровень политической гомофилии в американском сегменте Твиттера выше, чем испанском или немецком (Barberá, 2014) – при этом именно американский сегмент считается жертвой наиболее известной атаки ботов на данный момент. Впрочем, следует отметить, что степень гомофилии различается как между различными социальными медиа, различными сообществами внутри социальных медиа (сообщества, активно включенные в обсуждение политики, более гомофильны, чем аполитичные сообщества), так и в различные периоды времени: в период роста общественного интереса к политике степень гомофилии снижалась (Bond & Sweitzer, 2018). Можно предположить, что, для целей мобилизации, наиболее эффективно будет использование ботов в менее гомофильных, обычно аполитичных сообществах, в период роста общественного интереса к политике, а боты должны транслировать доста-

точно умеренные сообщения. Для мобилизации узких и радикальных групп единомышленников боты будут менее эффективны.

Теперь рассмотрим вторую потенциальную стратегию использования ботов: для подавления мобилизации (демобилизации) одной из групп. В таком случае актор, применяющий ботов, будет стараться «завысить» число активных противников группы-цели в социальном круге потенциальных сторонников группы-цели.

Снова рассмотрим ботов, нацеленных на потенциальных «оппозиционеров», но на этот раз – транслирующих позиции потенциальных «лоялистов», завышая их количество в социальном окружении потенциальных «оппозиционеров».

Можно сразу отметить, что при такой стратегии результат применения ботов в целом более эффективен и стабилен. Как и при использовании ботов для мобилизации, эффект растет по мере снижения гомофилии, однако он намного более ярко выражен. Для примера, при нацеливании ботов на агентов с позициями в промежутке от 0.3 до 0.5 (позиция «ботов» равна 0.6) в условиях слабой гомофилии ($H = 1$) число активных «оппозиционеров» в среднем в 26.5 раз меньше числа активных «лоялистов». Фактически, возникает «глобальная спираль молчания».

Как и при применении ботов по первой стратегии, наибольшие различия в размере групп достигаются при трансляции ботами позиции, относительно близкой для агентов-целей. При нацеливании ботов на потенциальных «оппозиционеров» с позициями в промежутке от 0.15 до 0.35 и позиции ботов, равной 0.6, то есть при чуть больших различиях между позицией ботов и позициями членов целевой группы, чем в примере выше и прочих равных параметрах, активных «оппозиционеров» в среднем оказывается только в 3.4 раза меньше числа «лоялистов». Так же выявлено и важное различие в результатах применения ботов по двум потенциальным стратегиям: в отличие от попыток мобилизации при помощи ботов, при демобилизации не возникает «обратный эффект», когда боты, транслирующие «слишком далекую» позицию, вместо того, чтобы мобилизовать агентов-людей, подавляют их активность.

Что примечательно, результаты отчасти перекликаются с результатами, полученными в двух различных исследованиях Бескоу и Карли (Beskow & Carley, 2019) и Россом (Ross B. et al., 2020), для эффективного подавления одной из активных групп в работе последнего потребовалось значительно меньше ботов, чем для эффективной «поддержки» ботами в работе первых.

Заключение

Подводя итоги, кратко повторим, что продемонстрировали вычислительные эксперименты. Было показано, что эффект ботов зависит от уровня гомофилии в сети: при сильной гомофилии боты оказываются значительно менее эффективны, чем при слабой. Боты, используемые для мобилизации сторонников, должны быть «точно нацелены» и транслировать достаточно близкие позиции. Ошибка может привести к «неприятным» для манипулятора результатам в виде демобилизации его потенциальных сторонников. Из этого следует, что боты потенциально могут использоваться для проведения информационных манипуляций «под чужим флагом»: легко представить, как некое правительство или другая организация используют замаскированных под излишне радикальных сторонников какого-либо движения ботов, чтобы снизить число его потенциальных сторонников. В-третьих, использование ботов, нацеленных на демобилизацию противников, имеет больший эффект и более стабильно. К тому же оно не чувствительно к «ошибкам нацеливания». Четвертое достойное внимания наблюдение – зависимость эффективности ботов от уровня гомофилии в социальной сети. На основе этих наблюдений могут быть сформулированы проверяемые эмпирически гипотезы.

Полученные результаты соответствуют и некоторым эмпирическим наблюдениям. Китайская «50-центовая армия» в основном размещает контент, «восхваляющий Китай, революционную историю Коммунистической партии и символы режимы» (King et al., 2017). Эта стратегия соответствует рассмотренной выше стратегии «демобилизации» оппонентов при помощи искусственного завышения числа сторонников правящей партии, показавшей при проведении вычислительных экспериментов большую эффективность, чем стратегия мобилизации сторонников. Конечно, члены «50-центовой армии» – не боты, однако, они, как и боты в представленной выше модели, не подвержены негативному воздействию социального окружения. Исследования влияния отечественных «троллей» на дискуссии в «Живом Журнале» показали, что последние более эффективно подавляли антиправительственные обсуждения, чем поддерживали проправительственную повестку (Sobolev, 2019). Важно отметить, что эти эффекты выявлялись только для «политических» обсуждений: при обсуждениях

экономической ситуации «тролли» были неэффективны. Впрочем, с демобилизацией онлайн-дискуссии справлялись и обычные боты: ботам успешно удалось подавить хэштег #YaMeCanse в Мексике в 2014 г. (Suárez-Serrato et al., 2016), похожая тактика применялась для подавления оппозиционных хэштегов в России и Китае (Verkamp & Gupta, 2013). Попытки использовать ботов и «троллей» для поддержки мобилизации сторонников какой-либо политической позиции так же происходят достаточно регулярно (Schäfer et al., 2017; Keller et al., 2020). Впрочем, однозначного вывода об эффективности или неэффективности этой стратегии пока сделать не удастся. Как уже было сказано выше, эффективность действий «троллей» для поддержки проправительственных дискуссий в «Живом Журнале» была низка (Sobolev, 2019), однако эксперимент Цербака и его коллег показал, что потенциально такое влияние может быть относительно велико (Zerback et al., 2021).

Негативное воздействие «радикальных» ботов на мобилизацию так же получает определенное подтверждение в существующей литературе. Литература, посвященная демобилизации социальных движений, в качестве одного из факторов демобилизации выявляет возникновение внутри движения радикальных фракций (Della Porta & Diani, 2006). Как демонстрируют проведенные вычислительные эксперименты, «завысив» число радикалов, можно стабильно демобилизовать потенциально возникающее движение еще на этапе мобилизации. Используются ли боты правительствами разных стран для достижения этой цели? Здесь эмпирические свидетельства отсутствуют, однако следует вспомнить, что и правительства в авторитарных государствах, и правительства в демократиях используют, как выразился Гэри Маркс, «забытых» участников социальных движений: информаторов и провокаторов (Marx, 1974). Предположение, что провокаторы действуют и в социальных медиа не следует считать нереалистичным.

Среда, в которой действуют боты, не нейтральна: социальные медиа с различной степенью успеха пытаются бороться с ботами, в особенности с теми, что распространяют политическую дезинформацию. Кроме того, социальные медиа используют алгоритмы фильтрации контента, подбирающие для каждого пользователя посты, соответствующие его интересам. Пусть эти два механизма и не рассматривались в данной статье отдельно, но на основе полученных результатов можно выдвинуть несколько предположений о том, каким будет это взаимодействие.

У борьбы с ботами могут проявиться и побочные эффекты. Маркировка пользователей и «фактчекинг» и маркировка сообщений как «недостовверных» может вызвать гнев у тех, кто считает позиции «маркированных» близкими. А гнев, как продемонстрировали более ранние исследования, способствует распространению дезинформации (Weeks, 2015). Есть еще два фактора, снижающих эффективность маркировок сообщений и пользователей. Первый связан с особенностями поведения самих пользователей, более склонных распространять те сообщения фактчекеров, которые соответствуют их собственным политическим убеждениям (Shin & Thorson, 2017). Второй негативный фактор – относительно низкая эффективность борьбы с политической дезинформацией по сравнению с прочими видами дезинформации. При этом простая маркировка сообщений, апелляции к «недостовверности» источника или удаление сообщений малоэффективно и для неполитической дезинформации (Walter & Murphy, 2018). Таким образом, можно предположить, что борьба с дезинформацией не сможет полностью обезвредить ботов.

Что касается фильтрации контента, то, по крайней мере для Твиттера, было показано, что алгоритмы фильтрации контента политически нейтральны, хотя и предлагаемая пользователю информация в значительной мере зависит от собственных политических предпочтений пользователя (Chen et al., 2021). Такая ситуация соответствует симулируемым высоким уровням гомофилии, при которых воздействие ботов на мобилизацию отсутствует, а на демобилизацию минимально. Конечно, социальное медиа, с целью борьбы с распространением определенной информации может манипулировать системой фильтрации, «искусственно завысив представленность мнений оппонентов» в выдачах пользователей. Однако, как показали симуляции выше, как раз при уменьшении гомофилии социальное медиа становится более уязвимым для манипуляций при помощи ботов. Впрочем, как уже было отмечено выше, это только предположения, и исследование влияния взаимодействия ботов, методов борьбы с ними и фильтрующих алгоритмов социальных медиа требует отдельных исследований.

Конечно, не следует, опираясь на полученные выше результаты, сразу провозглашать, что при помощи ботов можно контролировать поведение пользователей в социальных медиа, подстегивая мобилизацию одних и подавляя ее среди других. Однако полученные результаты явно демонстрируют, что подобная возможность есть, и следует продолжить исследовать эту проблему уже при помощи эмпирических методов.

Список литературы / References

- Ахременко, А. С., Петров, А.П., Пхоун, Ч. and Жеглов, С. А. (2021) ‘Как информационно-коммуникационные технологии меняют тренды в моделировании политических процессов: к агентному подходу’, *Политическая наука*, (1), сс. 12–45. DOI: <http://www.doi.org/10.31249/poln/2021.01.01>. [Akhremenko, A. S., Petrov, A. P., Phoun Ch., Zheglov, S.A. (2021) ‘How information and communication technologies change trends in modelling political processes: towards an agent-based approach’ [‘Kak informacionno-kommunikacionnye tehnologii menjajut trendy v modelirovanii politicheskikh processov: k agentnomu podhodu’], *Political science* (RU), (1), pp. 12–45. DOI: <http://www.doi.org/10.31249/poln/2021.01.01> (In Russ.)].
- Assenmacher, D., Clever, L., Frischlich, L., Quandt, T., Trautmann, H. and Grimme, C. (2020) ‘Demystifying social bots: On the intelligence of automated social media actors’, *Social Media+ Society*, 6(3), pp. 1–16. DOI: <https://doi.org/10.1177/2056305120939264>.
- Barberá, P. (2014) How social media reduces mass political polarization. Evidence from Germany, Spain, and the US [Online]. Available at: URL: http://pablobarbera.com/staic/barbera_polarization_APSA.pdf (5th November 2021).
- Benigni, M. C., Joseph, K. and Carley, K. M. (2019) ‘Bot-ivism: assessing information manipulation in social media using network analytics’, in: *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, pp. 19–42. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-94105-9_2.
- Beskow, D. M. and Carley, K. M. (2019) ‘Agent based simulation of bot disinformation maneuvers in Twitter’, in: *2019 Winter simulation conference (WSC)*, pp. 750–761. IEEE. DOI:10.1109/WSC40007.2019.9004942.
- Bond, R. M. and Sweitzer, M. D. (2018) ‘Political homophily in a large-scale online communication network’, *Communication Research*, pp. 1–23. DOI: <https://doi.org/10.1177/0093650218813655>.
- Brachten, F., Stieglitz, S., Hofeditz, L., Kloppenborg, K. and Reimann, A. (2017) ‘Strategies and Influence of Social Bots in a 2017 German state election-A case study on Twitter’ [online], [Preprint]. [Viewed date 05 September 2021]. URL: <https://arXiv:1710.07562>.
- Carley, K. M. (2020) ‘Social cybersecurity: an emerging science’, *Computational and mathematical organization theory*, 26(4), pp. 365–381. DOI: <https://doi.org/10.1007/s10588-020-09322-9>.
- Chen, W., Pacheco, D., Yang, K. C. and Menczer, F. (2021) ‘Neutral bots probe political bias on social media’, *Nature communications*, 12(1), pp. 1–10. DOI: <https://doi.org/10.1038/s41467-021-25738-6>.
- Della Porta, D. and Diani, M. (2006) *Social movements: An introduction*. John Wiley & Sons.
- DellaPosta, D., Shi, Y. and Macy, M. (2015) ‘Why do liberals drink lattes?’, *American Journal of Sociology*, 120(5), pp. 1473–1511. DOI: <https://doi.org/10.1086/681254>.
- Diamond, L. (2010) ‘Liberation technology’, *Journal of democracy*, 21(3), pp. 69–83. DOI:10.1353/jod.0.0190.
- Farrell, H. (2012) ‘The consequences of the internet for politics’, *Annual review of political science*, 15, pp. 35–52. DOI: <https://doi.org/10.1146/annurev-polisci-030810-110815>.
- Ferrara, E. (2017) ‘Disinformation and social bot operations in the run up to the 2017 French presidential election’ [online], [Preprint]. [Viewed date 05 September 2021]. URL: <https://arXiv:1707.00086>.
- Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A. (2016) ‘The rise of social bots’, *Communications of the ACM*, 59(7), pp. 96–104. DOI: <https://doi.org/10.1145/2818718>.
- González-Bailón, S. and Wang, N. (2016) ‘Networked discontent: The anatomy of protest campaigns in social media’, *Social networks*, 44, pp. 95–104. DOI: <https://doi.org/10.1016/j.socnet.2015.07.003>.
- Gould, R. V., (1991) ‘Multiple networks and mobilization in the Paris Commune, 1871’, *American Sociological Review*, 56(6), pp.

- 716–729. DOI: <https://doi.org/10.2307/2096251>.
- Huber, G. A. and Malhotra, N. (2017) ‘Political homophily in social relationships: Evidence from online dating behavior’, *The Journal of Politics*, 79(1), pp. 269–283. DOI: <https://doi.org/10.1086/687533>.
- Keller, F. B., Schoch, D., Stier, S. and Yang, J. (2020) ‘Political astroturfing on Twitter: How to coordinate a disinformation campaign’, *Political Communication*, 37(2), pp. 256–280. DOI: <https://doi.org/10.1080/10584609.2019.1661888>.
- King, G., Pan, J. and Roberts, M. E. (2017) ‘How the Chinese government fabricates social media posts for strategic distraction, not engaged argument’, *American political science review*, 111(3), pp. 484–501. DOI: <https://doi.org/10.1017/S0003055417000144>.
- Klandermans, B. (1984) ‘Mobilization and participation: Social-psychological expansions of resource mobilization theory’, *American sociological review*, pp. 583–600. DOI: <https://doi.org/10.2307/2095417>.
- Kuran, T. (1991) ‘Now out of never: The element of surprise in the East European revolution of 1989’, *World politics*, 44(1), pp. 7–48. DOI: <https://doi.org/10.2307/2010422>.
- Linvill, D. L. and Warren, P. L. (2020) ‘Troll factories: Manufacturing specialized disinformation on Twitter’, *Political Communication*, 37(4), pp. 447–467. DOI: <https://doi.org/10.1080/10584609.2020.1718257>.
- Marx, G. T., (1974) ‘Thoughts on a neglected category of social movement participant: The agent provocateur and the informant’, *American Journal of Sociology*, 80(2), pp. 402–442. DOI: <https://doi.org/10.1086/225807>.
- McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001) ‘Birds of a feather: Homophily in social networks’, *Annual review of sociology*, 27(1), pp. 415–444. DOI: <https://doi.org/10.1146/annurev.soc.27.1.415>.
- Metaxas, P. T. and Mustafaraj, E. (2012) ‘Social media and the elections’, *Science*, 338(6106), pp. 472–473. DOI: [10.1126/science.1230456](https://doi.org/10.1126/science.1230456).
- Mosleh, M., Martel, C., Eckles, D. and Rand, D. G. (2021) ‘Shared partisanship dramatically increases social tie formation in a Twitter field experiment’, *Proceedings of the National Academy of Sciences*, 118(7). DOI: <https://doi.org/10.1073/pnas.2022761118>.
- Noelle-Neumann, E. (1974) ‘The spiral of silence a theory of public opinion’, *Journal of communication*, 24(2), pp. 43–51. DOI: <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>.
- Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A. and Menczer, F. (2011) ‘Truthy: mapping the spread of astroturf in microblog streams’, in: *Proceedings of the 20th international conference companion on World wide web*, pp. 249–252. DOI: <https://doi.org/10.1145/1963192.1963301>.
- Ross, B., Pilz, L., Cabrera, B., Brachten, F., Neubaum, G. and 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*, 28(4), pp. 394–412. DOI: <https://doi.org/10.1080/0960085X.2018.1560920>.
- Sanovich, S., Stukal, D. and Tucker, J. A. (2018) ‘Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia’, *Comparative Politics*, 50(3), pp. 435–482. DOI: [10.5129/001041518822704890](https://doi.org/10.5129/001041518822704890).
- Schäfer, F., Evert, S. and Heinrich, P. (2017) ‘Japan’s 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe’s Hidden Nationalist Agenda’, *Big data*, 5(4), pp. 294–309. DOI: [10.1089/big.2017.0049](https://doi.org/10.1089/big.2017.0049).
- Shin, J. and Thorson, K. (2017) ‘Partisan selective sharing: The biased diffusion of fact-checking messages on social media’, *Journal of Communication*, 67(2), pp. 233–255. DOI: <https://doi.org/10.1111/jcom.12284>.
- Siegel, D. A. (2009) ‘Social networks and collective action’, *American journal of political science*, 53(1), pp. 122–138. DOI: <https://doi.org/10.1111/j.1540-5907.2008.00361.x>.
- Siegel, D. A. (2018) ‘Analyzing computational models’, *American Journal of Political Science*, 62(3), pp. 745–759. DOI: <https://doi.org/10.1111/ajps.12364>.
- Silva, B. C. and Proksch, S. O. (2021) ‘Fake It ‘Til You Make It: A Natural Experiment to Identify European Politicians’ Benefit from Twitter Bots’, *American Political*

- Science Review*, 115(1), pp. 316–322. DOI: <https://doi.org/10.1017/S0003055420000817>.
- Sobolev, A. (2019) How social media reduces mass political polarization. Evidence from Germany, Spain, and the US [Online]. Available at: URL: <https://www.wpsanet.org/papers/docs/2019W-Feb-Anton-Sobolev-Trolls-VA.pdf> (5th November 2021).
- Sohn, D. and Geidner, N. (2016) ‘Collective dynamics of the spiral of silence: The role of ego-network size’, *International Journal of Public Opinion Research*, 28(1), pp. 25–45. DOI: <https://doi.org/10.1093/ijpor/edv005>.
- Sohn, D. (2019) ‘Spiral of silence in the social media era: A simulation approach to the interplay between social networks and mass media’, *Communication Research*, pp. 1–28. DOI: <https://doi.org/10.1177/0093650219856510>.
- Stein, E. A. (2017) ‘Are ICTs Democratizing Dictatorships? New Media and Mass Mobilization’, *Social Science Quarterly*, 98(3), pp. 914–941. DOI: 10.1111/ssqu.12439.
- Stella, M., Ferrara, E. and De Domenico, M. (2018) ‘Bots increase exposure to negative and inflammatory content in online social systems’, *Proceedings of the National Academy of Sciences*, 115(49), pp. 12435–12440. DOI: 10.1073/pnas.1803470115.
- Stukal, D., Sanovich, S., Bonneau, R. and Tucker, J. A. (2017) ‘Detecting bots on Russian political Twitter’, *Big data*, 5(4), pp. 310–324. DOI: 10.1089/big.2017.0038.
- Stukal, D., Sanovich, S., Tucker, J. A. and Bonneau, R. (2019) ‘For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia’, *SAGE Open*, 9(2), pp. 1–16. DOI: <https://doi.org/10.1177/2158244019827715>.
- Suárez-Serrato, P., Roberts, M. E., Davis, C. and Menczer, F. (2016) ‘November. On the influence of social bots in online protests’, in: *International Conference on Social Informatics*, pp. 269–278. Springer, Cham. DOI: http://dx.doi.org/10.1007/978-3-319-47874-6_19.
- Valenzuela, S., Arriagada, A. and Scherman, A. (2014) ‘Facebook, Twitter, and youth engagement: A quasi-experimental study of social media use and protest behavior using propensity score matching’, *International Journal of Communication*, 8, pp. 2046–2070.
- Verkamp, J. P. and Gupta, M. (2013) ‘Five incidents, one theme: Twitter spam as a weapon to drown voices of protest’, in: *3rd {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 13)*.
- Walter, N. and Murphy, S. T. (2018) ‘How to unring the bell: A meta-analytic approach to correction of misinformation’, *Communication Monographs*, 85(3), pp. 423–441. DOI: <https://doi.org/10.1080/03637751.2018.1467564>.
- Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H. and Zhao, B. Y. (2012) ‘Social turing tests: Crowdsourcing sybil detection’ [online], [Preprint]. [Viewed date 05 September 2021]. URL: <https://arxiv.org/abs/1205.3856>.
- Weeks, B. E. (2015) ‘Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation’, *Journal of communication*, 65(4), pp. 699–719. DOI: <https://doi.org/10.1111/jcom.12164>.
- Wilensky, U. (1999) NetLogo [online]. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. [Viewed date 05 September 2021]. URL: <http://ccl.northwestern.edu/netlogo/>.
- Zerback, T., Töpfl, F. and Knöpfle, M. (2021) ‘The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them’, *New Media & Society*, 23(5), pp. 1080–1098. DOI: <https://doi.org/10.1177/1461444820908530>.

Статья поступила в редакцию: 22.09.2021

Статья поступила в редакцию повторно, после доработки: 06.11.2021

Статья принята к печати: 15.11.2021

**BOTS LEADING THE PEOPLE?
MODEL OF SOCIAL BOT'S IMPACT ON POLITICAL MOBILIZATION
AND DEMOBILIZATION IN SOCIAL MEDIA**

К. А. Toloknev

K. A. Toloknev, graduate student,
National Research University “Higher School of Economics”, Russia, Moscow
E-mail: kirtoloknev@gmail.com (ORCID: 0000-0003-3712-9978. ResearcherID: AAD-3672-2020).

Abstract

Bots (automated accounts) in social media have gained more attention from social scientists in recent years. The former are employed by both governments and civil society groups in order to manipulate online political discussion in social media. Nevertheless, there is no conclusive evidence on their effectiveness as a tool to bolster (mobilize) or suppress (demobilize) political discussion in social media. This paper presents a novel agent-based model, capable of simulating bot deployment as both mobilization and demobilization tool. Results of the simulations reveal three major effects of bot deployment. First, bots are more effective at demobilizing opponents than mobilizing supporters. Second, transmitting radical opinions via bots may backfire, demobilizing a certain group rather than mobilizing it. Third, the effectiveness of social bots is dependent on homophily: more homophilous networks are less susceptible to bots' influence. Results of modeling may both advance our estimations of bots' effectiveness and serve as a tool to generate potential hypotheses for future empirical research.

Keywords: bots; social media; political communication; agent-based model; protest; political mobilization; political demobilization; computational modelling.