О ВЛИЯНИИ БОТОВ И МОДЕРАЦИИ КОНТЕНТА НА ФОРМИРОВАНИЕ МНЕНИЙ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНОЙ СЕТИ

Губанов Д.А., Чхартишвили А.Г.
Институт проблем управления им. В.А. Трапезникова РАН,
Москва, Россия
dmitry.a.g@gmail.com, sandro_ch@mail.ru

Аннотация. Рассмотрена имитационная модель формирования информационного каскада к посту в социальной сети. У пользователя есть два параметра, характеризующих его мнение и действие. На формирование мнений пользователя влияют ранее написанные комментарии. При этом часть комментариев может быть удалена в ходе модерации.

Ключевые слова: онлайновые социальные сети, информационный каскад, формирование мнений пользователей, информационное управление, боты, модерация контента.

Введение

В последние десятилетия большой интерес теоретиков и практиков привлекают социальные сети. Выявление и прогнозирование динамики предпочтений пользователей онлайновых социальных сетей имеет огромную важность при моделировании информационного управления и информационного противоборства [1]. Эти предпочтения могут иметь социально-политическую, экономическую, психологическую или какую-либо другую природу.

Одним из методов исследования динамики мнений в социальных сетях является имитационное моделирование. Оно основано на задании на микроуровне правила изменения мнения пользователей (агентов) в зависимости от наблюдаемых ими действий других агентов (см., напр., [2]). Отметим, что альтернативным подходом является моделирование динамики на макроуровне, например, при помощи системы дифференциальных уравнений [3]. При моделировании динамики мнений в социальных сетях традиционно считается, что мнение и действие агента (индивида, являющегося узлом сети) отождествляются – см., например, [1]. Применительно к онлайновым социальным сетям это означает, что агент (в данном случае – пользователь сети) без искажения транслирует свое внутреннее состояние, и другие агенты имеют возможность это состояние наблюдать. В последнее десятилетие ситуация меняется в сторону разработки более сложных и реалистичных моделей [4-6].

В данной работе развивается ранее предложенная авторами модель [7] именно такого класса, где мнения (или предпочтения) агентов не наблюдаемы, а наблюдаемые действия не полностью отражают их мнения. При этом на формирование мнений пользователя влияют ранее написанные (в том числе ботами) комментарии, которые видит пользователь. При этом часть комментариев может быть удалена в ходе модерации.

1. Формирование информационного каскада

В данном разделе мы опишем, следуя [7], модель формирования последовательности комментариев к сообщению (посту) в социальной сети. В этой модели считается, что имеется фиксированное множество пользователей социальной сети, являющихся подписчиками информационного источника. В информационном источнике публикуется пост, который рано или поздно увидят все пользователи-подписчики. У каждого пользователя есть мнение, которое он корректирует, прочитав часть ранее оставленных комментариев. После этого пользователь выбирает действие (в соответствии со своим сформированным мнением), ставит лайк соответствующим комментариям (из числа просмотренных), затем с некоторой вероятностью сам пишет комментарий.

Приведем теперь формальное описание модели. В начальный момент имеется множество $N = \{1, \ldots, n\}$ пользователей, которые подписаны в онлайновой сети на данный информационный источник. Считаем, что у каждого пользователя $i \in N$ в начальный момент времени имеются следующие параметры: мнение $x_i \in [0; 1]$, вероятность написать комментарий p_i, а также n_i – максимальное количество комментариев, которые пользователь просмотрит перед выбором своего действия. Также заданы неотрицательные числа $b_{ij}, i, j \in N$, характеризующие степень доверия пользователя i пользователю j.

Формирование последовательности комментариев после появления в информационном источнике сообщения (поста) осуществляется посредством выполнения n однотипных шагов.

На каждом шаге i с равной вероятностью выбирается любой из еще не видевших сообщение

пользователей, не ограничивая общности будем считать его i-м. Он просматривает сообщение и либо первые n_i комментариев, либо, если количество всех имеющихся к данному шагу комментариев меньше n_i, все комментарии. Обозначим множество авторов просмотренных i-м пользователем комментариев через N_i. Каждый комментарий j-го пользователя является отражением его действия $y_i \in \{0,1\}$ – выбора позиции «за» (действие $y_j = 1$) или «против» (действие $y_j = 0$).

Будем считать, что под влиянием просмотренных комментариев i-й пользователь корректирует свое мнение x_i следующим образом:

$$x_i' = \frac{b_{ii}x_i + \sum_{j \in N_i} b_{ij}y_j}{b_{ii} + \sum_{j \in N_i} b_{ij}}. \quad (1)$$

После этого i-й пользователь выбирает свое действие $y_i \in \{0,1\}$ в соответствии с параметром x_i': действие 1 («за») с вероятностью x_i' и действие 0 («против») с вероятностью $(1 - x_i')$. Далее i-й пользователь ставит лайк тем из просмотренных комментариев, которые соответствуют выбранному им действию (т.е. ставит лайк комментарию j-го пользователя при условии $y_i = y_j$). Наконец, в завершение шага i пользователь с вероятностью p_i сам пишет комментарий «за» или «против» в соответствии с выбранным действием (соответственно, с вероятностью $(1 - p_i)$ i-й пользователь не оставляет комментарий под сообщением). Введем параметр $z_i \in \{0,1\}$, который равен 1, если i-й пользователь оставил комментарий, и равен 0 в противоположном случае.

В результате n шагов алгоритма формируется последовательность комментариев. Обозначим через N_z множество оставивших комментарий пользователей. Будем считать, что наиболее важной характеристикой последовательности является доля комментариев «за», т.е. $\delta = \sum_{i \in N_z} y_i / N_z$.

Пользователи социальной сети не являются стратегическими игроками, однако на них нацелено воздействие стратегических игроков. Будем рассматривать стратегических игроков трех типов.

Первый тип – боты. Будем считать, что боты составляют множества $M^0 = \{n+1, \dots, n+m^0\}$ и $M^1 = \{n+m^0+1, \dots, n+m^0+m^1\}$, как бы дополняющие множество пользователей N. Бот отличаются от обычного пользователя тем, что всегда (с вероятностью 1) пишет комментарий, и его действие предопределено заранее: для j-го бота, $j \in M^0$, это действие $y_j = 0$ (и проставление лайков комментариям, соответствующим этому действию); для k-го бота, $k \in M^1$, это действие $y_k = 1$ (и, аналогично, проставление лайков комментариям, соответствующим этому действию). Таким образом, боты из множества M^1 (далее для краткости будем называть их 1-ботами) стремятся увеличить долю комментариев «за», а боты из множества M^0 (их будем называть 0-ботами) – уменьшить.

Второй тип стратегического игрока – администратор страницы информационного источника в сети (далее – администратор). Будем считать, что администратор источника может удалять нежелательные для него (по какой-либо причине) комментарии и лайки пользователей и ботов.

Наконец, третий тип стратегического игрока – сама онлайновая социальная сеть (далее – онлайновая сеть). Будем считать, что онлайновая сеть может управлять параметрами алгоритма ранжирования комментариев к постам.

Опишем теперь стратегии игроков, которые мы будем рассматривать.

Для ботов будем рассматривать два варианта:

(б1) боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи);

(б2) боты являются первыми комментаторами.

Администратор может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев (нежелательными являются либо комментарии «за», либо комментарии «против»). Будем считать, что администратор стирает каждый нежелательный комментарий, как только тот появляется с фиксированной вероятностью $q \in [0; 1]$.

Онлайновая сеть может применять один из трех вариантов показа комментариев:

(с1) в обратном хронологическом порядке – сначала новые, потом старые;

(с2) в порядке убывания количества лайков (при одинаковом количестве лайков – в обратном хронологическом порядке как в п. (с1));

(с3) сначала комментарии «за», затем комментарии «против» (внутри обеих множеств – в обратном хронологическом порядке как в п. (с1)).

2. Анализ имитационной модели

Для введенной выше модели будем оценивать характеристики информационных каскадов при помощи имитационного моделирования, позволяющего рассчитать усредненную долю комментариев

«за» в зависимости от варианта показа комментариев. Будем считать, что в сети n = 100 участников (например, подписчиков данной информационного ресурса), и она представляет собой полный граф, в котором каждый участник одинаково доверяет всем агентам (в том числе самому себе). Мнения агентов в начальный момент времени: равномерно распределены на отрезке [0; 1].

2.1. Результаты одиночного эксперимента

Выполним анализ для одного запуска модели (вероятность написания комментариев одинакова для всех агентов $p = 1$).

Глубина просмотра. На рис. 1 показаны траектории доли комментариев «за» и «против» для глубины просмотра 5 и ∞. Если глубина просмотра максимальная, то алгоритм показа не влияет на результат. В противном случае траектории расходятся, и алгоритм «сначала за» приводит к максимальному δ.

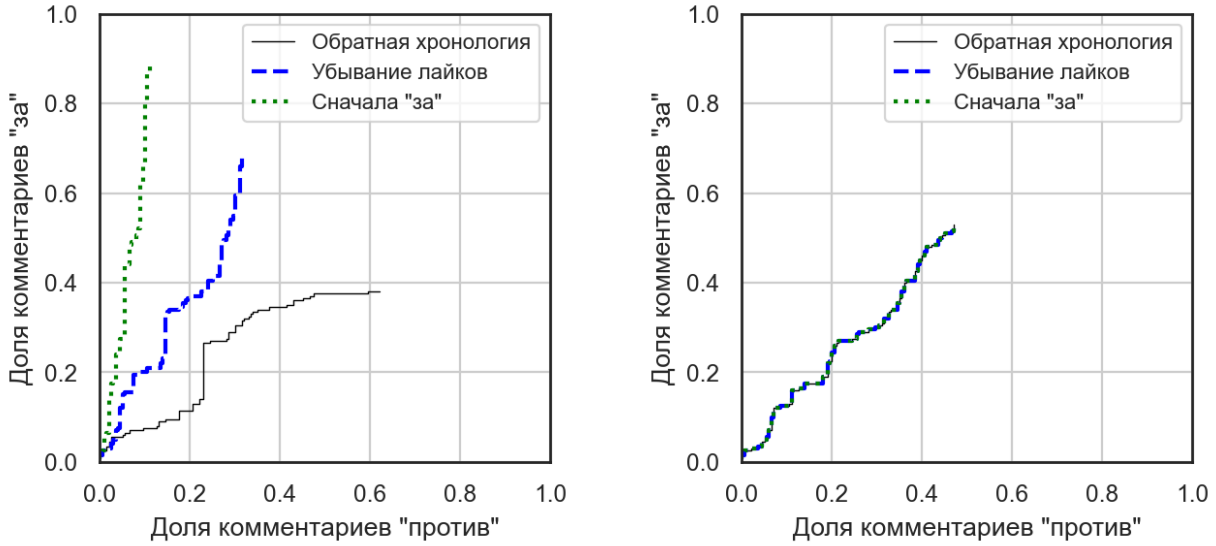

Рис. 1. Траектории доли комментариев «за» и «против»: а) глубина просмотра $n_i = 5$, б) глубина просмотра $n_i = \infty$

Далее глубина просмотра пользователей предполагается равной 5.

Боты «как обычные пользователи». На рис. 2 показаны траектории доли комментариев «за» и «против»: а) для 30% доли 0-ботов (1-боты отсутствуют), б) для 30% доли 1-ботов (0-боты отсутствуют). В первом случае происходит поворот траекторий по часовой стрелке, во втором – против.

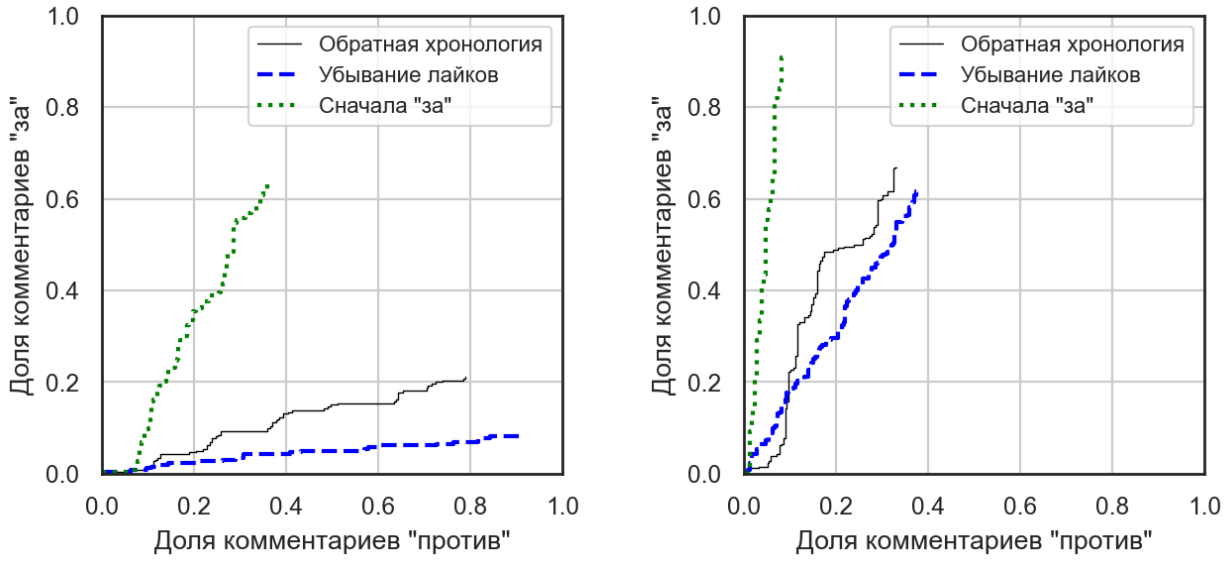

Рис. 2. Траектории доли комментариев «за» и «против»: а) доля 0-ботов 0,3 (1-боты отсутствуют), б) доля 1-ботов 0,3 (0-боты отсутствуют)

Боты являются первыми комментаторами. На рис. 3 показаны траектории доли комментариев «за» и «против»: а) для 30% доли 0-ботов (1-боты отсутствуют), б) для 30% доли 1-ботов (0-боты отсутствуют). Видно, что сначала доминирует позиция ботов, затем начинается расхождение траекторий. В случае (б) траектории для алгоритма «убывание лайков» и «сначала за» совпадают: 1-боты при помощи лайков к своим комментариям полностью определяют «повестку» для последующих пользователей, что аналогично воздействию алгоритма «сначала за».

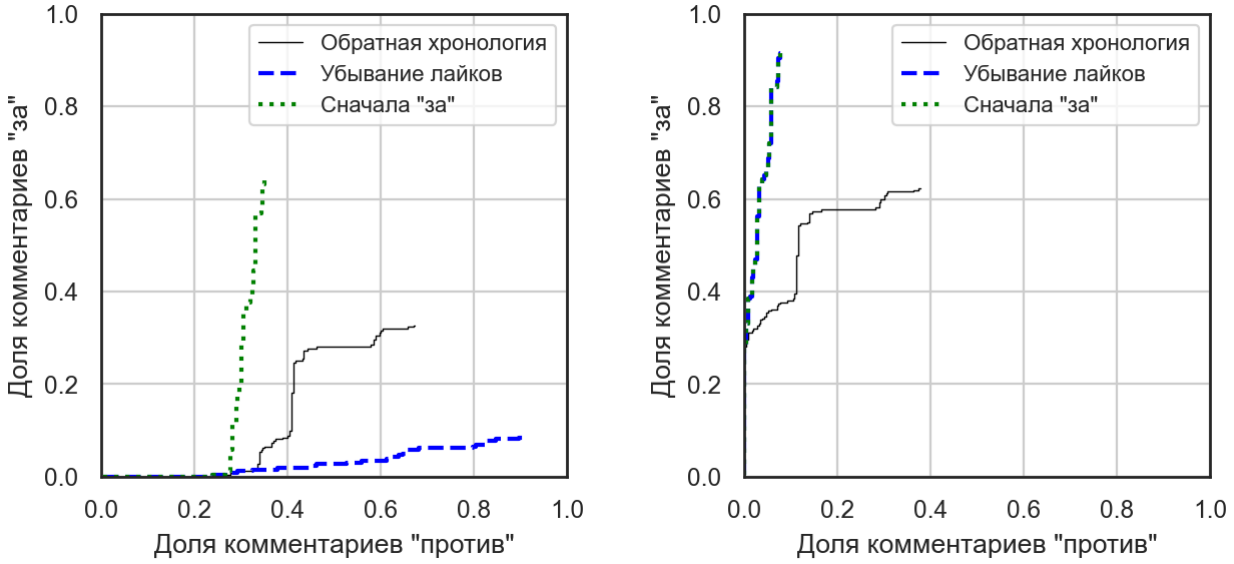

Рис. 3. Траектории доли комментариев «за» и «против»: а) доля 0-ботов равна 0,3, б) доля 1-ботов равна 0,3

Позиция администратора. На рис. 4 показаны траектории доли комментариев «за» и «против»: а) для администратора с позицией «против», б) для администратора с позицией «за». Параметр $q = 0{,}5$. Администратор за счет удаления нежелательных для него комментариев добивается увеличения доли комментариев со «своей» позицией.

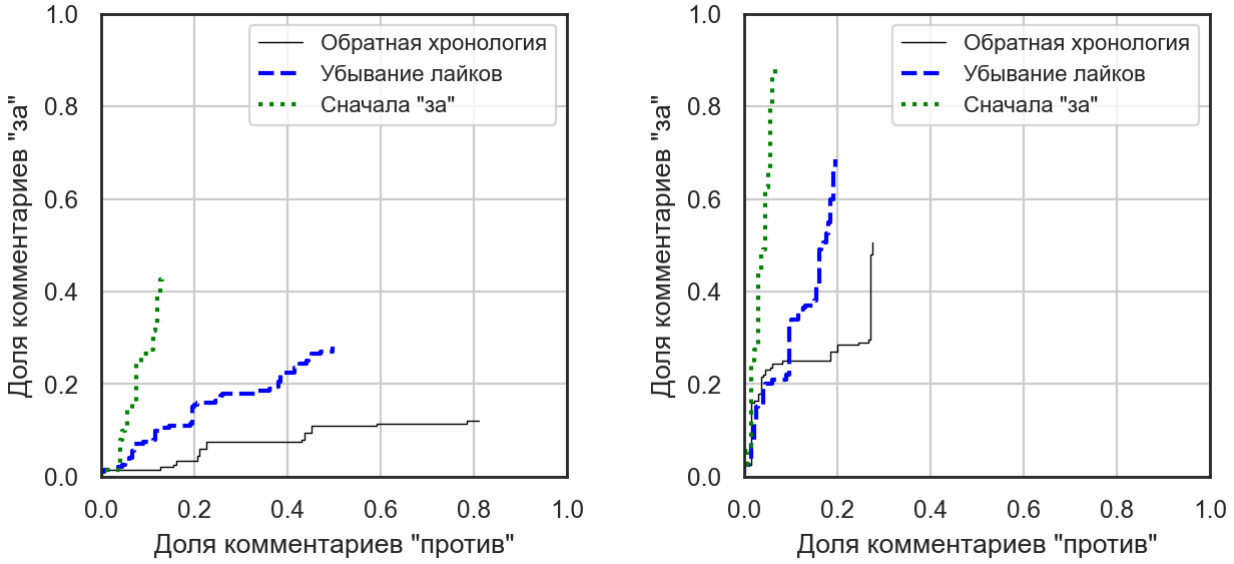

Рис. 4. Траектории доли комментариев «за» и «против»: а) позиция модератора «против», б) позиция модератора «за»

Таким образом на примере одного запуска модели можно заключить, что значения параметров модели влияют на итоговый результат. Возникает вопрос: насколько велико и статистически значимо это влияние? Поскольку модель является стохастической и нелинейной, то в следующем разделе проводится анализ влияния параметров (как по отдельности, так и с учетом эффектов взаимодействия) на дисперсию δ.

2.2. Анализ влияния параметров модели и стратегий игроков

- Рассмотрим следующие параметры имитационного моделирования:
- доля $m_0/n \in [0; 1]$,
- доля $m_1/n \in [0; 1]$
- «глубина» просмотра агентов $n_i \in \{1, \dots, n + m_0 + m_1\}$,
- вероятность написать комментарий $p_i \in [0; 1]$,
- позиция модератора $y \in \{0,1\}$ («против» или «за»).

Стратегия q будет принимать значение из отрезка [0; 1].

Поскольку модель является нелинейной, оценим глобальную чувствительность модели к значениям параметров по методу Соболя (долю объясняемой дисперсии).

Результаты для алгоритма «*в обратном хронологическом порядке*» приведены на рис. 5.

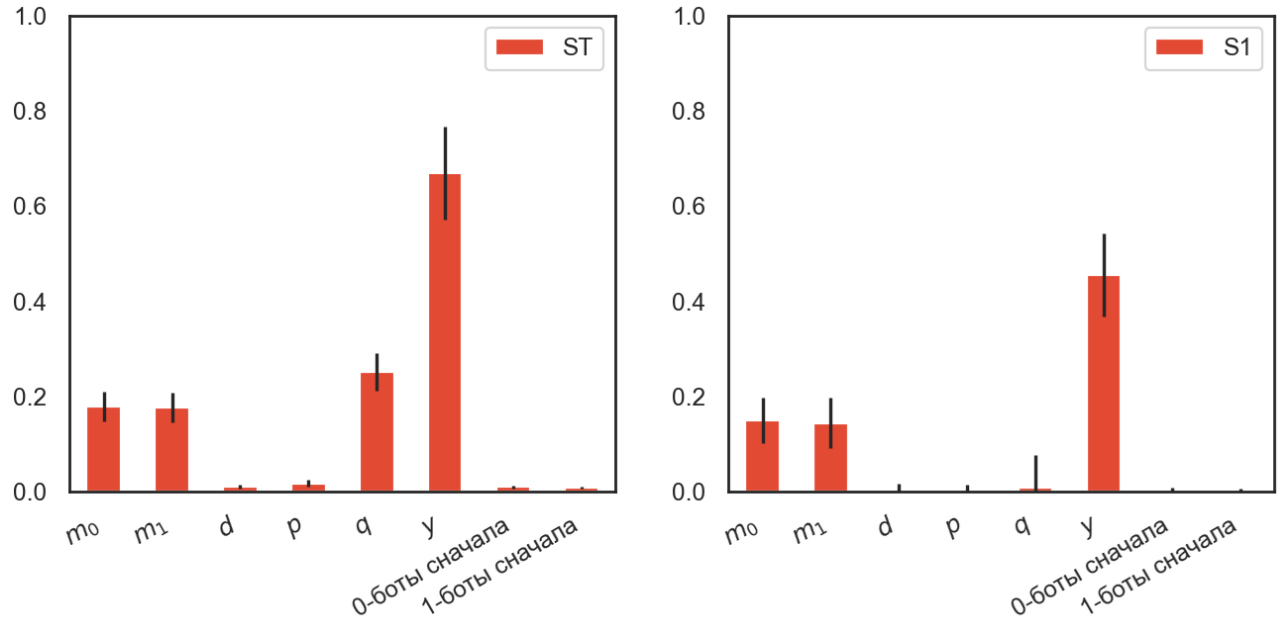

Рис. 5. Диаграмма показателей влияния параметров модели на результаты (S^T – влияние с учетом эффектов взаимодействия с другими переменными, S^1 – индивидуальное влияние переменных)

Позиция администратора (параметр y) играет решающую роль в сети, это видно как по индексу первого порядка 45%, так и по общему индексу 69%. Вероятность удаления нежелательного комментария (параметр q) оказывает незначительное индивидуальное влияние, но во взаимодействии с позицией администратора вклад составляет 22% (S^2), во взаимодействии с другими переменными – 26% (S^T). Количество ботов (m_0 и m_1) значительно влияет на результирующую переменную (δ): индексы первого порядка 14% и 13%, общие индексы 17% и 18%. Глубина просмотра, вероятность написания комментариев, порядок ботов оказывают ограниченное воздействие, что видно из их низких значений как индексов первого порядка, так и общих индексов. Индексы второго порядка показывают, что эффекты взаимодействия параметров незначительны (за исключением q и y).

Другие алгоритмы («*в порядке убывания количества лайков*», «*сначала комментарии за*») показывают качественно схожие результаты. В то же время вклад параметра y: (с2) $S^T = 61\%$ ($S^1 = 37\%$), (с3) $S^T = 65\%$ ($S^1 = 39\%$). Вклад m_0 и m_1: (с2) 15% и 15% (18% и 19%), (с3) 13% и 13% (15% и 18%). Вклад q во взаимодействии с y: (с2) 23%, (с3) 23%.

3. Заключение

В работе рассмотрена модель формирования информационных каскадов, в которой мнения (относительно некоторого вопроса) агентов не наблюдаемы, а наблюдаемые действия не полностью отражают их мнения. Совершаемые агентами действия (написание комментариев) влияют на мнения действующих впоследствии агентов, тем самым формируя информационный каскад мнений и действий. Как показали вычислительные эксперименты, существенное влияние на такой каскад оказывают параметры модели (число ботов, алгоритм ранжирования) и стратегии игроков (администратора и ботов). Анализ чувствительности модели показывает, что роль администратора (y и q) является ключевой в сети. Количество ботов (m_0 и m_1) также значительно влияет на долю комментариев «за» (δ). Выбор алгоритма показа комментариев не столь существенен. Однако одиночный запуск модели продемонстрировал, что алгоритм показа комментариев сильно определяет долю комментариев «за» (если глубина просмотра не максимальна). Можно предположить, что от выбора конкретной последовательности агентов зависит степень выраженности каскадных эффектов и – соответственно – степень воздействия тех или иных значений параметров, включая алгоритм показа. Представляют интерес дальнейшие исследования таких мультипликативных эффектов.

Литература

1. *Губанов Д.А., Новиков Д.А., Чхартишвили А.Г.* Социальные сети: модели информационного влияния, управления и противоборства. 3-е изд., перераб. и дополн. М.: МЦНМО, 2018. – 224 с.
2. *Perra N., Rocha L. E. C.* Modelling opinion dynamics in the age of algorithmic personalisation // Scientific reports. – 2019. – Vol. 9. – №. 1. – P. 1-11.
3. *Petrov A.P., Lebedev S.A.* Online Political Flashmob: the Case of 632305222316434 // Computational mathematics and information technologies. – 2019. – No 1. – P. 17–28.
4. *Новиков Д.А*. Модели динамики психических и поведенческих компонент деятельности в коллективном принятии решений // Управление большими системами. – 2020. – Вып. 85. – С. 206–237.
5. *Губанов Д.А., Новиков Д.А*. Модели совместной динамики мнений и действий в онлайновых социальных сетях. Ч. 2. Линейные модели // Проблемы управления. – 2023. - №3. – С. 40–64.
6. *Banisch S., Olbrich E*. An argument communication model of polarization and ideological alignment //arXiv preprint arXiv: 1809.06134. – 2018.
7. *Губанов Д.А., Чхартишвили А.Г.* О влиянии алгоритмов онлайновой социальной сети на формирование мнений пользователей / Труды 16-й Международной конференции «Управление развитием крупномасштабных систем» (MLSD’2023, Москва). – М.: ИПУ РАН, 2023. – С. 121-125.