

МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ЛОЖНОГО СООБЩЕНИЯ В СОЦИУМЕ ПРИ КОМПЛЕКСНОМ ПРОТИВОДЕЙСТВИИ

Петров А.П.

*Институт прикладной математики им. М.В. Келдыша РАН,
Россия, г. Москва, Миусская пл., д.4
petrov.alexander.p@yandex.ru*

Аннотация: Представлена математическая модель распространения ложного сообщения в социуме при противодействии ему путем инокуляции и опровержения. Предполагается, что инокуляция проводится с помощью контактной игры, т.е. инокулированные индивиды сообщают информацию об ее существовании другим индивидам, тем самым, способствуя более широкой инокуляции социума.

Ключевые слова: моделирование слухов, фейк ньюс, информационная инокуляция.

Введение

Противодействие фейковым новостям путем опровержения является малоэффективным. Так, в эмпирическом исследовании [1] показано, что правдивые истории распространяются медленнее и с меньшим охватом, чем ложные. Ввиду этого, в области исследования пропаганды и распространения информации в обществе набирает все больший вес такое направление, как разработка, анализ и оценка эффективности методов информационной инокуляции.

Одна из возрастающе популярных инокуляционных техник – это браузерные игры, такие, как, например, Bad News [2] или Harmony Square [3]. Эти обучающие игры знакомят пользователя с методами информационной манипуляции в шуточной, привлекательной форме. Исследования, представленные в указанных работах, в целом, имеют следующий дизайн. Участникам эксперимента предлагается определить наличие или отсутствие манипуляции в ряд публикаций. Затем одна часть участников играет в инокулирующую игру, другая (контрольная группа) – в Тетрис. После этого им снова предлагается выявить манипуляции: непосредственно после игры, и затем через некоторые промежутки времени. Результаты [2, 3] показывают, что инокуляционные игры действительно повышают способность индивидов выявлять манипуляцию. При этом, в сравнении, например, с медицинскими прививками, информационная инокуляция обеспечивает индивиду менее сильную защиту и эффект ее длится меньшее время.

Еще одним недостатком инокуляционных техник на сегодняшнем этапе их развития является то, что они не являются контактно-заразными. Другими словами, инокулированный индивид не передает свою способность неинокулированному при контакте. В противоположность этому, ложное сообщение может оказаться вирусным (высококонтрактным) и распространяться по социуму в геометрической прогрессии. Неконтактность существенно ограничивает результативность инокуляции [4].

В то же время, нет никаких фундаментальных причин для того, чтобы инокуляция не была контактно-заразной. Нетрудно представить, как пользователь компьютерной игры рассказывает своему неосведомленному другу о положительных эмоциях, полученных во время игры, и этот друг при первой возможности также приступает к игре. Таким образом, имеет место контрактность. Если речь идет об инокулирующей игре, то данный процесс является процессом распространения инокуляции через социальное заражение.

Настоящая работа посвящена моделированию распространения ложного сообщения при противодействии ему путем контактно-заразной инокуляции и опровержения. Отличие от работы [4] состоит в том, что в указанной работе инокуляция предполагалась неконтактно-заразной и предварительной. Другими словами, инокуляция носила разовый характер и предшествовала началу распространения ложной новости. В настоящей работе считается, что категория инокулированных индивидов постоянно пополняется из числа восприимчивых (т.е. неинокулированных и не знакомых с ложным сообщением) индивидов ввиду социального заражения.

Наиболее ранние работы по инокуляционной тематике относятся к 1961 году [5, 6], но активное развитие теории происходит в наши дни (см., напр., [7, 8] и содержащуюся в этих работах библиографию). В более широком контексте, данная область исследований относится к тематике информационных противоборств и общественного мнения. Математическому моделированию и эмпирическому исследованию динамики мнений и информационного влияния посвящена обширная литература – см., например, монографию [9] и статьи [10-18].

1 Модель

Рассматривается модель с дискретным временем; описываемый ею процесс имеет следующий вид. Атакующая сторона распространяет в социуме численности N ложное сообщение, против которого обороняющаяся сторона предпринимает две меры: инокуляция и опровержение.

Ложная информация вбрасывается в момент времени $t = 0$. Инокуляция проводится путем игры, которая начинается раньше, чем происходит вброс ложной информации; таким образом, в момент $t = 0$ некоторое количество индивидов уже инокулировано. Инокуляция продолжается на протяжении всего процесса. Предполагается, что инокулирующая игра является привлекательной для игроков, так что уже сыгравшие индивиды они менее уязвимы для нее, чем восприимчивые. Если восприимчивый или инокулированный член социума принимает ложное сообщение за истинное и интернализирует его, то он становится спредером, т.е. начинает передавать ложное сообщение другим индивидам. Другими словами, спредеры – это лица, которые распространяют ложный слух. Познакомившись со слухом, новые восприимчивые и привитые индивиды становятся спредерами. При этом, будем отдельно учитывать инокулированных спредеров и неинокулированных спредеров (т.е. ставших скептиками из числа восприимчивых или спредеров).

При этом обороняющаяся партия с постоянной интенсивностью b транслирует опровержение ложной информации. Члены социума, интернализовавшие опровержение, становятся скептиками; они распространяют опровержение в виде слуха; таким образом, новые индивиды становятся скептиками. При этом, будем отдельно учитывать инокулированных скептиков и неинокулированных скептиков (т.е. ставших скептиками из числа восприимчивых или спредеров).

Таким образом, рассматривается шесть категорий индивидов. Обозначим численность восприимчивых в момент времени t через $s(t)$, численность инокулированных через $z(t)$, инокулированных распространителей и неинокулированных распространителей – соответственно, $x_z(t), x_s(t)$ инокулированных скептиков и неинокулированных скептиков – соответственно, $r_z(t), r_s(t)$. В начальный момент времени численность инокулированных составляет $z(0)$, восприимчивых : $s(0) = N - z(0)$, для распространителей и скептиков имеем

$$x_s(0) = x_z(0) = r_s(0) = r_z(0) = 0$$

Примем, что динамика численности инокулированных происходит в соответствии со следующими положениями. Восприимчивый может стать инокулированным, если встретится с индивидом, который играл в игру ранее (инокулированный, инокулированный спредер либо инокулированный скептик), узнает от него об игре и сыграет. Будем считать, что если встреча происходит в момент t , то индивид, бывший восприимчивым в момент t , является инокулированным в момент $t+1$.

Тогда численность восприимчивых, ставших инокулированными между этими моментами, имеет вид

$$c_3 s(t) \frac{z(t) + x_z(t) + r_z(t)}{N}$$

Следуя работе [4] примем, что вероятность принятия ложного сообщения за правду характеризуется воспринимаемой достоверностью, которая неодинакова для восприимчивых и инокулированных категориями. Другими словами, инокуляция не защищает индивида от ложной информации полностью, она лишь делает его менее восприимчивым.

Обозначим воспринимаемую достоверность для восприимчивых и привитых индивидуумов через ρ_s и ρ_z соответственно. Пренебрегая убывающим эффектом инокуляции положим, что $\rho_z = \rho_s(1 - E)$, где $E < 1$ — эффективность прививки. Достоверность ρ_d опровергающего сообщения предполагается равной для всех категорий.

Через α обозначим долю индивидов (как среди восприимчивых, так и среди инокулированных), познакомившихся с первоначальной публикацией фейкового новостного сообщения, произошедшей в момент $t = 0$.

Соответственно, имеем

$$s(1) = s(0) - c_3 s(0) \frac{z(0) + x_z(0) + r_z(0)}{N} - \alpha \rho_s s(0) = s(0) - c_3 s(0) \frac{z(0)}{N} - \alpha \rho_s s(0),$$

$$\begin{aligned}
z(1) &= z(0) + c_3 s(0) \frac{z(0) + x_z(0) + r_z(0)}{N} - \alpha \rho_s (1-E) z(0) = z(0) + c_3 s(0) \frac{z(0)}{N} - \alpha \rho_s (1-E) z(0), \\
x_s(1) &= x_s(0) + \Delta \{s(0) \rightarrow x(1)\} = x_s(0) + \alpha \rho_s s(0), \\
x_z(1) &= x_z(0) + \alpha \rho_s (1-E) z(0), \\
r_s(1) &= r_z(1) = 0.
\end{aligned}$$

В момент $t = 1$ начинает распространяться опровергающее сообщение. В предположении, что количество индивидов, получивших в одну и ту же единицу времени более одного сообщения (например, ложное и инокулирующее или ложное и опровергающее), динамика при $t \geq 1$ описывается следующими уравнениями

$$\begin{aligned}
s(t+1) - s(t) &= -c_3 s(t) \frac{z(t) + x_z(t) + r_z(t)}{N} - \rho_s c_1 \frac{x_s(t) + x_z(t)}{N} s(t) - \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) s(t), \\
z(t+1) - z(t) &= c_3 s(t) \frac{z(t) + x_z(t) + r_z(t)}{N} - \rho_s c_1 (1-E) \frac{x_s(t) + x_z(t)}{N} z(t) - \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) z(t), \\
x_s(t+1) - x_s(t) &= \rho_s c_1 \frac{x_s(t) + x_z(t)}{N} s(t) - \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) x_s(t), \\
x_z(t+1) - x_z(t) &= \rho_s c_1 (1-E) \frac{x_s(t) + x_z(t)}{N} z(t) - \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) x_z(t), \\
r_s(t+1) - r_s(t) &= \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) (s(t) + x_s(t)), \\
r_z(t+1) - r_z(t) &= \rho_d \left(b + c_2 \frac{r_s(t) + r_z(t)}{N} \right) (z(t) + x_z(t)).
\end{aligned}$$

Здесь константы c_1, c_2, c_3 характеризуют интенсивность пересказывания информации при межличностной коммуникации: соответственно, ложного сообщения, опровергающего сообщения, информации о существовании игры. При этом будем полагать, что $c_2 \ll c_1$, т.е. правдивая информация передается намного менее активно, чем ложная [19].

2 Вычислительные эксперименты

Приведем результаты нескольких численных экспериментов

Эксперимент 1. Примем следующие значения параметров:

$$\begin{aligned}
N &= 10^7 \text{ (10 млн. чел); } b = 0,02; c_1 = 5; c_2 = 0,001; c_3 = 1; \\
\rho_s &= 0,3; \rho_d = 0,7; E = 0,4; \nu = 0,2; \alpha = 0,1.
\end{aligned}$$

Графики компонент решения представлены на Рис.1. Эти графики показывают, что при данных значениях параметров инокуляция не позволяет существенной роли в противодействии ложной информации. Численность категории инокулированных достигает высоких значений, однако численность индивидов, зараженных ложной информацией достигает высоких значений (на пике: почти 90% от общей численности). При этом, как показывает правый рисунок, существенно большая часть спредеров состоит из инокулированных индивидов.

Эксперимент 2. Примем те же значения параметров, что в Эксперименте 1, за исключением того, что эффективность инокуляции положим равной единице:

$$\begin{aligned}
N &= 10^7 \text{ (10 млн. чел); } b = 0,02; c_1 = 5; c_2 = 0,001; c_3 = 1; \\
\rho_s &= 0,3; \rho_d = 0,7; E = 1; \nu = 0,2; \alpha = 0,1.
\end{aligned}$$

Графики компонент решения представлены на Рис. 2. При данных значениях параметров

инокуляция не позволяет существенной роли в противодействии ложной информации. Максимальная численность индивидов, зараженных ложной информацией, оказывается весьма низкой (около 10%).

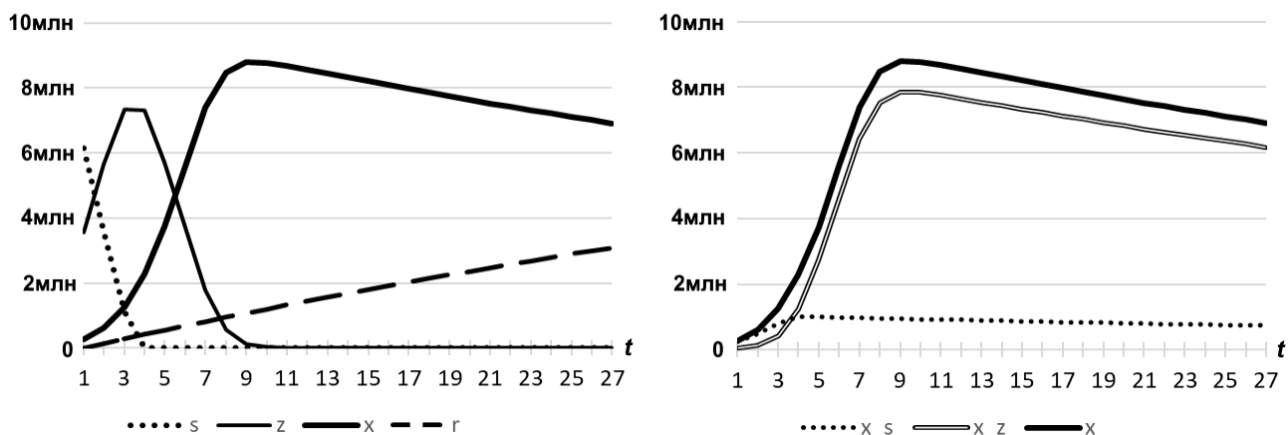


Рис. 1. Эксперимент 1: результаты расчета

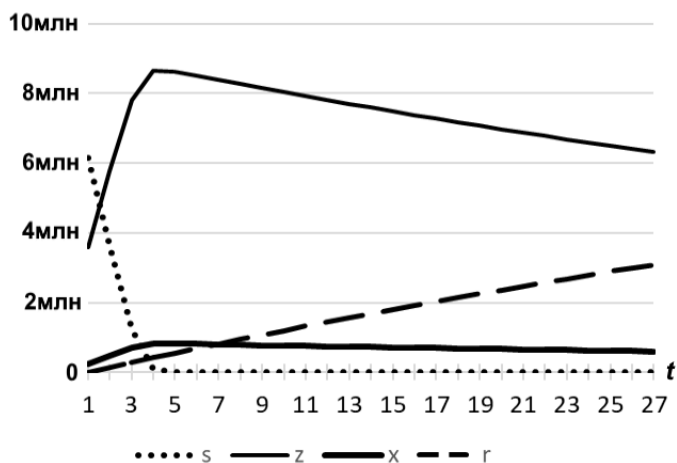


Рис. 2. Эксперимент 2: результаты расчета

Заключение

Вывод из представленных двух экспериментов является тривиальным лишь на первый взгляд. Если инокуляция имеет сравнительно не очень высокую эффективность, то она неспособна успешно противостоять распространению ложного сообщения (в одном из экспериментов, не представленных в данной работе, расчет проводился при значении эффективности $E = 0.8$; максимум численности спредеров ложной информации в этом случае достигает 70% от общей численности индивидов). Если же эффективность равна единице (инокулированный индивид не может стать спредером), то численность спредеров остается сравнительно низкой.

Этот вывод выглядит менее тривиальным, если принять во внимание, что в модели работы [4] даже единичная эффективность инокуляции не позволяет успешно противостоять распространению ложного сообщения. Причина этого в модели работы [4] состоит в том, что она предполагает неконтагиозность инокуляции. Другими словами, в Эксперименте 3 работы [4] инокуляция предполагалась абсолютно эффективной, но численность инокулированных оставалась невысокой ввиду неконтагиозности. В то же время, в Эксперименте 1 настоящей работы численность инокулированных становится высокой, но эффективность инокуляции – низкая. В обоих этих случаях ложная информация охватывает неприемлемо большую часть населения.

Сдержать распространение ложной информации оказывается возможным лишь если имеет место и очень высокая эффективность инокуляции, и контагиозность инокулирующей игры, как это было продемонстрировано в Эксперименте 2 настоящей работы.

Работа выполнена при поддержке РФФИ (проект 20-01-00229).

Литература

1. Vosoughi S., Roy D., Aral S. The spread of true and false news online // *Science*. Vol. 359. 2018. – P. 1146–1151.
2. Roozenbeek J., van der Linden S. The fake news game: Actively inoculating against the risk of misinformation // *Journal of risk research*, vol. 22(5), pp. 570–580, 2019.
3. Roozenbeek J., van der Linden S. Breaking Harmony Square: A game that “inoculates” against political misinformation. *The Harvard Kennedy School Misinformation Review*, 1(8). 2020.
4. Petrov A. Is Inoculation Effective against Fake News? A Mathematical Model // 14th International Conference "Management of large-scale system development" (MLSD), IEEE. 2021, pp. 1-4, doi: 10.1109/MLSD52249.2021.9600188.
5. McGuire W.J., Papageorgis D. The relative efficacy of various types of prior belief-defense in producing immunity against persuasion // *Journal of Abnormal and Social Psychology*. Vol. 62. 1961. – P.327–337.
6. Papageorgis D., McGuire W.J. The generality of immunity to persuasion produced by pre-exposure to weakened counterarguments. *Journal of Abnormal and Social Psychology*. Vol. 62. 1961. P. 475–481.
7. Compton J. Inoculation theory. *The Sage handbook of persuasion: Developments in theory and practice*, Vol. 2. 2013. P. 220-237.
8. Compton J., van der Linden S., Cook J., Basol M. Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories // *Social and Personality Psychology Compass*, e12602. DOI: 10.1111/spc3.12602.
9. Chkhartishvili A.G., Gubanov D.A., Novikov D.A. *Social Networks: Models of information influence, control and confrontation*. Cham, Switzerland: Springer International Publishing, 2019. – 158 p. DOI: 10.1007/978-3-030-05429-8
10. Popova S., Skitalinskaya G. Keyphrase Extraction Using Extended List of Stop Words with Automated Updating of Stop Words List. // *Advances in Intelligent Systems and Computing II*. (CSIT 2017.) Springer, Cham, vol 689, 2018, p. 374-385.
11. Danilova V., Popova S., Alexandrov M. Multilingual Protest Event Data Collection with GATE, In: *Proc. of 21-th Intern. Conf. on Natural Languages and Data Base (NLDB-2016)*, Springer, vol.9612, 2016, p.115-126.
12. Chkhartishvili A.G., Gubanov D.A. On the Concept of an Informational Community in a Social Network. // *Journal of Physics: Conference Series 2021*. Vol. 1864, No. 1, p. 012052. IOP Publishing
13. Kozitsin I.V. Formal models of opinion formation and their application to real data: evidence from online social networks // *The Journal of Mathematical Sociology*. 2022 Apr 3;46(2):120-47.
14. Kozitsin I.V. A general framework to link theory and empirics in opinion formation models // *Scientific reports*, vol. 12, p. 5543, 2022.
15. Boldyreva A., Bayburin B., Alexandrov M., Koshulko O., Mogilev P. Search Queries to Internet as a Tool for Mid-term Forecast of Cryptocurrencies (XRP, Waves, ETH) on GMDH-based models // 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 2021, pp. 375-379, doi: 10.1109/CSIT52700.2021.9648740.
16. Boldyreva A. Authority Changes Constitution and Regions Answer: What Search Queries Show // 2021 29th Conference of Open Innovations Association (FRUCT), 2021, pp. 65-70, doi: 10.23919/FRUCT52173.2021.9435564.
17. Akhtyamova L., Alexandrov M., Cardiff J., Koshulko O. Opinion Mining on Small and Noisy Samples of Health-related Texts. // *Advances in Intelligent Systems and Computing III* (Proc. of CSIT-2018), Springer, AISC, 2019, vol. 871, p. 1-12.
18. Akhtyamova L., Ignatov A., Cardiff J. A Large-scale CNN ensemble for medication safety analysis // *Applications of Natural Language to Inform. Systems* (Proc. NLDB-2027), Springer, Cham. P. 247-253.
19. Maertens R., Roozenbeek J., Basol M., van der Linden S. Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, Vol. 27(1), 2021. P. 1–16.