

А.М. ПЕЛЕЩИШИН, Г.О. БАНДРОВСЬКИЙ
Національний університет «Львівська політехніка»

ІНФОРМАЦІЙНИЙ ВПЛИВ У СОЦІАЛЬНОМУ СЕРЕДОВИЩІ ІНТЕРНЕТУ: АНАЛІЗ АКТИВНОСТІ КОРИСТУВАЧІВ ТА ЇХ РЕАКЦІЙ НА ПУБЛІКАЦІЇ

В статті представлено моделювання інформаційного впливу в соціальному середовищі Інтернету, а саме зацікавленості актуальних тем для користувачів на основі аналізу активності користувачів та їх реакції на публікації. Проведено аналіз інформаційного впливу в соціальному середовищі Інтернету. Досліджено активність користувачів як респондентів та визначено вагомні фактори аналізу реакції на публікації, на основі яких побудовано математичну модель для визначення закономірностей та передбачення впливу на думку в соціальному середовищі мережі Інтернет.

Ключові слова: інформаційний вплив, перколяція, теорія ймовірності, статистичний аналіз.

A.M. PELESHCHYSHYN, H.O. BANDROVSKYI
Lviv Polytechnic National University

INFORMATION INFLUENCE IN THE INTERNET SOCIAL ENVIRONMENT: ANALYSIS OF USERS' ACTIVITY AND THEIR RESPONSE TO PUBLICATIONS

The article presents the modelling of information influence in the social environment of the Internet, namely, the interest of relevant topics for users as readers based on the analysis of users' activity and their reaction to publications. The analysis of information influence in the social environment of the Internet is carried out. Users' interaction is considered as a non-linear dynamic system. The activity of users as respondents is investigated and important factors of the analysis of reactions to publications are defined. On this basis, a mathematical model is built to determine the patterns and predict the impact of opinion in the social environment of the Internet. The phenomenon of percolation applied to disseminate information among social network users is taken as the basis of the model. The percolation threshold is taken as an average value for a certain number of the most popular publications for the selected period of time depending on density of publications as a whole. The chosen mathematical problem is considered on the example of the analysis of popularity of topics of publications within the limits of publications of one author on the personal page in a social network. The given mathematical model makes it possible to apparently analyze the reaction of users as readers and to investigate the informational influence of the content published by the author. For further use of this mathematical model it is important to complicate and calibrate it in accordance with a more complex structure of the social network and analysis of the text of respondents' comments on published content. In practical terms, the results can be used in interdisciplinary research, including mathematical modelling, to predict the information impact in the social environment of the Internet.

Keywords: information influence, percolation, probability theory, statistical analysis.

Вступ

В сучасному світі соціальне середовище Інтернету та отримана звідти інформація відіграє невід'ємну роль в житті людей. В XXI столітті окрім таких джерел інформації як радіо, телебачення, газети, книги тощо до ключових джерел інформації додається всевітня система сполучених комп'ютерних мереж. Відповідно, актуалізується дослідження впливу інформації на сприйняття людиною, скерування її на прийняття конкретних рішень та вплив на формування її точки зору стосовно будь-якої проблеми.

Проблема аналізу інформаційного впливу є багатоаспектною. Важливою складовою є вивчення реакції користувачів на інформацію з метою визначення наявності чи відсутності її впливу.

Для дослідження поширення інформації в мережі та аналізу її впливу, на сьогодні є актуальними дослідження в таких розділах математики та інформатики, як математична статистика, теорія ймовірності, теорія динамічних систем та штучний інтелект.

Важливим аспектом аналізу інформаційного впливу в соціальному середовищі є розуміння джерел впливу як математичних формалізованих одиниць. За вихідний момент приймемо таку одиницю соціальної мережі, як публікацію за одинарне джерело інформаційного впливу на користувачів.

Аналіз останніх досліджень

З метою структурування, групування та аналізу публікацій у формальному вигляді спираємось на розв'язання таких задач, як задача кластеризації або задача класифікації, які в даному випадку передбачають також обробку природної мови (НЛП). Максимізація розповсюдження впливу в соціальних мережах досліджена в багатьох роботах [1, 4, 5], зокрема можна відмітити такі роботи в сфері аналізу контенту тексту, як контенту, який є фактором поширення інформації, «Добування даних в соціальних мережах для онлайн маркетингу» [3], «Новий підхід до добування даних з тексту на основі НММ-SVM для класифікації веб-новин» [8], «Класифікація тексту у Twitter для покращення фільтрації інформації» [9]. Математичні моделі поширення інформаційного впливу представлені в роботі «Соціальні мережі: моделі впливу, контролю та конфронтації інформації» А. Чхартішвілі, Д. Губанова та Д. Новікова [7], а також у дослідженнях українських учених [5, 6].

Мета роботи

Мета роботи полягає в побудові гнучкої моделі для аналізу інформаційного впливу на користувачів на основі моделі перколяції, з допомогою якої можливо оцінити досягнення порогу поширення інформаційного впливу серед користувачів, враховуючи їх зворотну реакцію.

Постановка математичної задачі

Розглянемо взаємодію користувачів в соціальному середовищі Інтернету як динамічну систему. Зміни настроїв користувачів в контексті соціальних структур можна описати за допомогою переходів між станами вузлів системи. Множину усіх станів системи позначимо через X . Стан системи, який досліджуємо в момент часу t , можна позначити як x_i ($x_i \in X$). Введемо інтервал часу τ_0 , за який можлива зміна стану x_i . В даному випадку $t = h\tau_0$, де h – номер кроку переходу між станами, $h = 0, 1, 2, 3, \dots, N$. Порядковий стан x_i на кроці h , після переходу на кроці $h+1$ може за рахунок випадкових факторів збільшуватись на деяку величину ε , або зменшуватись на величину ξ , і ставати рівними $x_i + \varepsilon$, або $x_i - \xi$. В нашому випадку ε та ξ будуть сталими величинами для будь-якого кроку h і можуть залежати від інформації.

Введемо поняття ймовірності знаходження системи в певному стані. Нехай, після деякого числа кроків h про систему можна сказати, що:

- $P(x - \varepsilon, h)$ – ймовірність того, що вона перебуває в стані $(x - \varepsilon)$;
- $P(x, h)$ – ймовірність того, що вона перебуває в стані x ;
- $P(x + \xi, h)$ – ймовірність того, що вона перебуває в стані $(x + \xi)$.

Після кожного кроку, стан x_i (далі індекс i можна опускати) може змінюватися на величину ε або ξ . Ймовірність того, що на наступному кроці $(h + 1)$ система буде в стані x позначимо як $P(x, h + 1)$ та запишемо наступним чином:

$$P(x, h + 1) = P(x - \varepsilon, h) + P(x + \xi, h) - P(x, h) \tag{1}$$

Схему можливих переходів між станами даної системи показано на рис. 1.

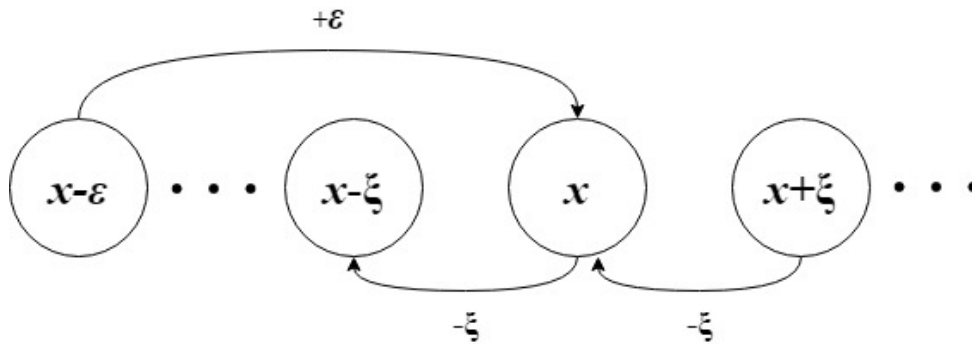


Рис. 1. Схема можливих переходів між станами системи на $h+1$ кроці.

Ймовірність переходу в стан x на кроці h , $P(x, h + 1)$ визначається сумою ймовірностей переходів в цей стан із станів $(x - \varepsilon) - P(x - \varepsilon, h)$ та $(x + \xi) - P(x + \xi, h)$, в яких була система на кроці h без ймовірності переходу ($P(x, h)$) системи із стану x в будь-який стан на $h+1$ кроці.

Треба враховувати на кожному кроці попередні стани системи. Для цього визначимо ймовірності $P(x - \varepsilon, h)$, $P(x + \xi, h)$ і $P(x, h)$ через стан на $h-1$ кроці.

$$P(x - \varepsilon, h) = P(x - 2\varepsilon, h - 1) + P(x - \varepsilon + \xi, h - 1) - P(x - \varepsilon, h - 1) \tag{2}$$

$$P(x + \xi, h) = P(x + \xi - \varepsilon, h - 1) + P(x + 2\xi, h - 1) - P(x + \xi, h - 1) \tag{3}$$

$$P(x, h) = P(x - \varepsilon, h - 1) + P(x + \xi, h - 1) - P(x, h - 1) \tag{4}$$

Підставимо (2), (3) і (4) в рівняння (1) отримаємо:

$$P(x, h + 1) = \{P(x - 2\varepsilon, h - 1) + P(x - \varepsilon + \xi, h - 1) - P(x - \varepsilon, h - 1)\} + \{P(x + \xi - \varepsilon, h - 1) + P(x + 2\xi, h - 1) - P(x + \xi, h - 1)\} - P(x - \varepsilon, h - 1) - P(x + \xi, h - 1) + P(x, h - 1) \tag{5}$$

або

$$P(x, h + 2) = \{P(x - 2\varepsilon, h) + P(x - \varepsilon + \xi, h) - P(x - \varepsilon, h)\} + \{P(x + \xi - \varepsilon, h) + P(x + 2\xi, h) - P(x + \xi, h)\} - P(x - \varepsilon, h) - P(x + \xi, h) + P(x, h) \tag{6}$$

Оскільки $t = h\tau_0$, де t – конкретна одиниця яка відображає момент часу, h – номер кроку, τ_0 – довжина одного кроку, перейдемо від h до t , тобто розвинемо в ряд Тейлора дану функцію від двох змінних. Обмежимося другими частковими похідними:

$$\frac{dP(x, t)}{dt} = a \frac{d^2 P(x, t)}{dx^2} - b \frac{dP(x, t)}{dx} - c \frac{d^2 P(x, t)}{dt^2} \tag{7}$$

де

$$a = \frac{\varepsilon^2 - \varepsilon\xi + \xi^2}{\tau_0} \tag{8}$$

$$b = \frac{\varepsilon - \xi}{\tau_0} \tag{9}$$

$$c = \tau_0 \tag{10}$$

Припустимо, що функція $P(x,t)$ неперервна, тоді можна перейти від ймовірності $P(x,t)$ (рівняння (7)) до густини ймовірності. Відповідно, запишемо густину ймовірності стану x в момент часу t — $\rho(x,t)$:

$$\frac{d\rho(x,t)}{dt} = a \frac{d^2\rho(x,t)}{dx^2} - b \frac{d\rho(x,t)}{dx} - c \frac{d^2\rho(x,t)}{dt^2} \tag{11}$$

Підставивши у рівняння 7 рівняння 11, отримаємо:

$$\frac{d\rho(x,t)}{dt} = a \frac{d^2\rho(x,t)}{dx^2} - b \frac{d\rho(x,t)}{dx} - c \frac{d^2\rho(x,t)}{dt^2} \tag{12}$$

Величина опозиціонерів має бути на відрізку від 0 до величини порогу перколяції для даної системи (позначимо її через l).

Перша гранична умова береться: стан $x=0$ визначає відсутність негативно настроєних людей.

$$\rho(x,t)_{x=0} = 0 \tag{13}$$

Оскільки в момент часу $t=0$ стан системи може бути рівний x_0 , то початкову умову розглянемо:

$$\rho(x,t=0) = \delta(x-x_0) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases} \tag{14}$$

Друга гранична умова береться: $x=L$ (крайній випадок $L=1$):

$$\rho(x,t)_{x=L} = 0 \tag{15}$$

Так як перша гранична умова містить дельта функцію, то розв'язок для $\rho(x,t)$ розіб'ємо на дві області при $x > x_0$ і при $x \leq x_0$. Використовуючи методи операційного числення для густини ймовірності $\rho_1(x,t)$ і $\rho_2(x,t)$ знаходження стану системи в одному із значень на відрізку від 0 до L можна отримати наступну систему рівнянь:

При $x \geq x_0$:

$$\rho_1(x,t) = -\frac{2}{L} e^{-\frac{t}{2\tau_0}} e^{k(x-x_0)} \sum_{n=1}^{\infty} \frac{\sin\left(\frac{\pi n x_0}{L}\right) \sin\left(\frac{\pi n (L-x)}{L}\right)}{\cos(\pi n)} \operatorname{ch}\left(\frac{t}{\tau_0} \sqrt{\frac{k\varepsilon\xi}{2(\varepsilon-\xi)} - \frac{\pi^2 n^2 (\varepsilon-\xi)}{2kL^2}}\right) \tag{16}$$

При $x < x_0$:

$$\rho_2(x,t) = -\frac{2}{L} e^{-\frac{t}{2\tau_0}} e^{k(x-x_0)} \sum_{n=1}^{\infty} \frac{\sin\left(\frac{\pi n (L-x_0)}{L}\right) \sin\left(\frac{\pi n x}{L}\right)}{\cos(\pi n)} \operatorname{ch}\left(\frac{t}{\tau_0} \sqrt{\frac{k\varepsilon\xi}{2(\varepsilon-\xi)} - \frac{\pi^2 n^2 (\varepsilon-\xi)}{2kL^2}}\right) \tag{17}$$

де

$$k = \frac{(\varepsilon - \xi)}{2(\varepsilon^2 - \varepsilon\xi + \xi^2)} \tag{18}$$

Якщо обчислити інтеграл $P(l,t)$ то можна стверджувати, що функція буде давати ймовірність того, що стан системи до моменту часу t буде на відрізку від 0 до l , тобто поріг перколяції l не буде досягнутим або протилежне.

Програмна реалізація

Розглянемо дану математичну задачу на прикладі аналізу популярності тем публікацій в рамках публікацій одного автора на своїй особистій сторінці в соціальній мережі. Вважатимемо, що автор є популярним і в нього є сформована аудиторія. Для попередньої оцінки приймемо поріг перколяції як середнє значення з певної кількості (залежно від щільності публікацій в цілому) найпопулярніших публікацій за вибраний період часу.

Позначимо множину публікацій $Posts$, як множину усіх публікацій за певний період часу:

$$Posts = \{Post_1, \dots, Post_i\} \tag{19}$$

Для оцінки потенційного інформаційного впливу публікації візьмемо як приклад стандартний формат публікації, який широко використовується на сьогоднішній день, де публікація містить в собі такі сутності:

- Текст публікації (*Content*) – сам контент, реакцію на який вважатимемо інформаційним впливом.

- Реакції на публікацію (позначимо як R) – залежно від специфіки мережі може бути реалізовано в різних формах, найпопулярніша форма на сьогодні – реакція «сподобалось» (Like), проте в деяких соціальних мережах є більш комплексна реалізація з різними реакціями, які, окрім позитивного ставлення, можуть бути і негативним, наприклад «обурення». В даній роботі розглянемо аналогічну структуру до соціальної мережі Twitter, де є лише реакція «сподобалось».

- Коментар (C від слова Comment) – коментар користувача до даної публікації.

- Поширення (S від слова Share) – поширення користувачем (читачем) даної публікації.

При розгляданні усіх публікацій основними ідентифікаторами інформаційного впливу вважатимемо обрані три характеристики: Реакції, Коментарі та Поширення. Запишемо одинарну сутність публікації з рівняння (19) як кортеж:

$$Post_i = Content_i, \{R_i\}_{i=1}, \{C_i\}_{i=1}, \{S_i\}_{i=1} \quad (20)$$

Важливо відмітити, що всі три ідентифікатори інформаційного впливу мають різну вагу, адже вони мають різний потенціал поширення серед користувачів. Умовно це можна пояснити тим, що коли користувач коментує чи поширює публікацію на своїй сторінці це може викликати більше інтеракцій з публікацією серед інших користувачів.

У відповідність кожному з ідентифікаторів інформаційного впливу публікації поставимо коефіцієнт, який буде репрезентувати вагомість даного ідентифікатора. Запишемо вагу інформаційний потенціал кожної публікації I_i враховуючи певний коефіцієнт k :

$$I_i = \left(\sum_{i=1}^{N_i^R} R_i k_R + \sum_{i=1}^{N_i^C} C_i k_C + \sum_{i=1}^{N_i^S} S_i k_S \right) \quad (21)$$

Для програмної реалізації було обрано такі бібліотеки мови програмування Python: matplotlib – для візуалізації обчислень, request – для створення запитів та збору інформації з конкретних джерел. На основі рівняння (17) та (21) можна обчислити потенціал інформаційного впливу публікацій в момент ітерації та на основі обчислень проаналізувати наявність переходу стану (наявність досягнення перколяції або ж недосагнення).

Для прикладу розглянемо дві публікації автора, дані будуть розглядатись як кортеж рівняння (20) і оскільки маємо лише дві публікації, то індекси відповідно будуть 1 та 2. Розглянемо стан на часовому періоді з появи публікації та кроком в один день. Статистичні дані зібрані наведемо таблиці 1.

Таблиця 1

| | R_1 | C_1 | S_1 | R_2 | C_2 | S_2 |
|-------|-------|-------|-------|-------|-------|-------|
| t_0 | 88 | 7 | 3 | 205 | 10 | 10 |
| t_1 | 90 | 7 | 3 | 270 | 21 | 13 |
| t_2 | 90 | 8 | 3 | 363 | 27 | 18 |

На основі даних наведених в таблиці 1, візуалізуємо обчислення та переходи стану (досягнення перколяції). На рисунку 2 та рисунку 3 зображені графіки обчислень рівняння (17) для обох публікацій:

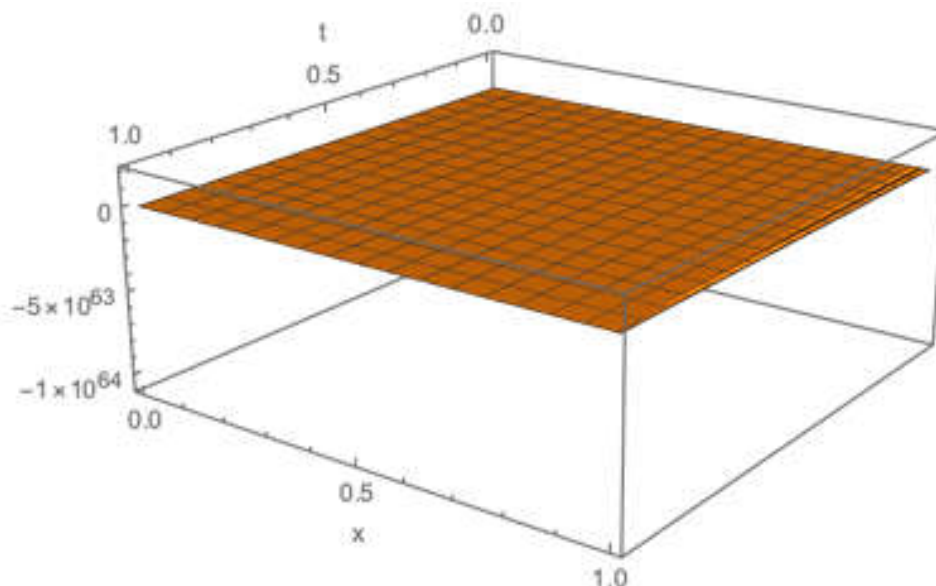


Рис. 2. Обчислені дані проходження перколяції для першої публікації

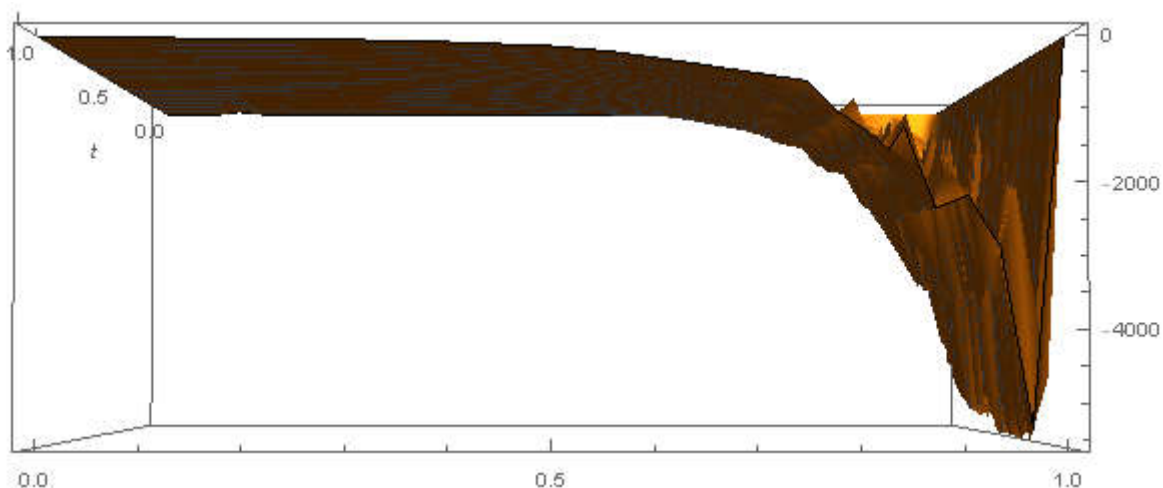


Рис. 3. Обчислені дані проходження перколяції для другої публікації

На основі отриманих результатів можемо зробити висновок, що для першої публікації поріг перколяції не був досягнутий, у випадку другої публікації на основі активності користувачів поріг перколяції був досягнутий. Це дає змогу потенційного з'ясування шляхом аналізу контенту зробити висновки стосовно зацікавленості користувачів.

Аналізуючи одержані результати, приходимо до таких **висновків**:

- Використовуючи дану математичну модель, можливо поверхнево проаналізувати реакцію користувачів як читачів та дослідити інформаційний вплив контенту опублікованим автором.
- Для подальшого використання даної математичної моделі важливим є ускладнення та калібрування відповідно до більш комплексної структури соціальної мережі та аналізу тексту коментарів респондентів на опублікований контент.
- Можливе використання штучних нейронних мереж, а саме навчання з підкріпленням для більш точного результату, адже кожна підзадача має певну специфіку і зумовлена складністю поставленої задачі (прогнозування, чи корекції напрямку та теми публікацій автора для збільшення зацікавленості читачів).

References

1. D. Kempe, J. Kleinberg and E. Tardos, "Maximizing the spread of influence through a social network", Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington: ACM Press, 2003, pp. 137–146.
2. W. Pedrycz, S.-M. Chen, Social Networks: A Framework of Computational Intelligence. Springer, 2013.
3. P. Domingos, "Mining social networks for viral marketing", University of Washington, 2018. Available at: <https://homes.cs.washington.edu/~pedrod/papers/iis04.pdf>
4. W. Chen, Ch. Wang and Ya. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks", Microsoft Research Technical Report, 10. Available at: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/msr-tr-2010-2_v2.pdf
5. R.V. Huminskiy, A.M. Peleshchyshyn, Z. Holub, "Suggestions for informational influence on a virtual community", International Journal of Computer Science and Business Informatics, 2015, p. 47–65.
6. Peleshchyshyn, V. Vus, O. Markovets, and S. Albota, "Identifying specific roles of users of social networks and their influence methods", Proceedings of 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 2018, vol. 2, pp. 39–42.
7. A.G. Chkhartishvili, D.A. Gubanov, and D.A. Novikov, Social Networks: Models of Information Influence, Control and Confrontation. Springer, 2018.
8. G. Krishnalal, S. Rengarajan, S. Babuand, K.G.Srinivasagan, "A new text mining approach based on HMM-SVM for web news classification", International Journal of Computer Applications (0975 - 8887), 2010, Volume 1, No. 19, pp. 103–109.
9. Sriram, D. Fuhry and E. Demir, "Short text classification in Twitter to improve information filtering", SIGIR'10 Proceedings of the 33rd International Conference on Research and Development in Informational Retrieval, Geneva, 2010, pp. 841–842.
10. E.A. Stanley, "Social networks and mathematical modeling", CONNECTIONS, Vol. 27(1), 2006, pp. 43–49. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.6966&rep=rep1&type=pdf>

Рецензія/Peer review : 27.5.2020 р.

Надрукована/Printed : 16.6.2020 р.

Стаття рецензована редакційною колегією