

## АНАЛІЗ РОЗПОДІЛУ ЕЛЕКТОРАЛЬНИХ ПОЛІВ З ВИКОРИСТАННЯМ МЕРЕЖЕВИХ СТРУКТУР

Є.О. ТЕРПІЛЬ, О.С. МАКАРЕНКО

**Анотація.** Досліджено підходи на основі математичного моделювання до аналізу розподілу електоральних полів. Розглянуто моделі на основі клітинних автоматів та складних мереж. Клітинні автомати пояснюють утворення стійких електоральних регіонів, однак мало придатні для реального прогнозування. З огляду на можливість урахування нерівномірного поширення суспільної думки перевагу надано складним мережам. Розглянуто зв'язок між спільнотами у мережі та утворенням стійких електоральних регіонів. Запропоновано використати модулярність для знаходження меж спільнот і завдяки яким пояснити існуючі електоральні поля та передбачити місця утворення меж нових електоральних регіонів. Для верифікації обраного підходу проаналізовано розподіл електорального поля півдня України на підставі даних парламентських виборів 2014 р.

**Ключові слова:** клітинні автомати, складні мережі, електоральні поля, модулярність, моделювання суспільної думки.

### ВСТУП

Завдання моделювання суспільної думки, зокрема електоральних уподобань, на тепер є одними з найактуальніших. Через стрімкий розвиток демократії у всьому світу процес виборів став популярним об'єктом дослідження, його стало легко формалізувати. Зазвичай основною метою дослідників є прогнозування результатів майбутніх виборів, або рейтингу партії (політика). Однак цю роботу сфокусовано на географічному розподілі вподобань виборців.

Електоральні мапи з багатьох країн світу характеризуються поділом електорального поля на регіони з домінуванням певної політичної сили. До того ж такі регіони зазвичай дуже стійкі і не змінюються протягом багатьох років. Для прикладу на рис. 1 зображено поділ електорального поля України у період останніх п'ятнадцяти років.



Рис. 1. Поділ електорального поля України на виборах у 2004 – 2012 рр.

Попередні дослідження [1] пов'язували виникнення чітких меж з особливостями ландшафту, географічними характеристиками та історичними особливостями територій. У цій роботі буде застосовано підходи на основі клітинних автоматів та складних мереж до моделювання розподілу електоральних полів, порівняно їх та досліджено місця утворення меж. Показано, що клітинні автомати хоч і дають змогу зрозуміти як утворюються стійкі електоральні регіони, однак мало придатні для реального прогнозування. Обійти цю проблему дозволить запровадження описаної мережевої структури. Для верифікації обраного підходу окремо проаналізовано розподіл електорального поля півдня України на підставі даних парламентських виборів 2014 р.

### ПІДХІД НА ОСНОВІ КЛІТИНИХ АВТОМАТІВ

Одним з підходів до моделювання суспільної думки є клітинні автомати. Вони набули поширення внаслідок інтенсивного розвитку інформаційних технологій, який дозволив швидко і легко проводити числові експерименти з моделями. Крім того, алгоритми таких моделей більш прозорі, ніж, наприклад, диференціальні рівняння, які потребують окремої реалізації числових методів наближення їх розв'язків.

Розглянемо процес формування суспільної думки як множину елементарних взаємодій між індивідами [2]. Припустимо, що всі індивіди керуються одним і тим самим набором правил. Ці правила задають для кожного моменту часу, стан індивіда, який буде в наступний момент часу залежно від поточного стану інших індивідів. Станом індивіда вважатимемо прихильність до однієї з альтернатив (партії чи кандидата) у певний момент часу. Відповідно множина станів є множиною альтернатив. Відкинувши демографічні та міграційні особливості, на невеликих проміжках часу розглянемо множину індивідів як таку, що не змінюється з часом і зафіксована на певній ґратці. Це означає, що для кожного індивіда вказано множину його сусідів (окіл), яка також є фіксованою. Звісно, що це доволі загальні припущення, однак вони дають змогу побудувати найпростішу модель клітинного автомата.

В основу моделі впливу оточення покладено гіпотезу, яка передбачає, що виборчі пріоритети індивідуума визначаються уподобаннями його найближчого оточення. Нехай індивід приймає рішення щодо голосування в момент  $t+1$  як одну з  $n$  альтернатив. За правило переходу візьмемо правило більшості, тобто індивід обирає таку альтернативу, як і більшість у його околі. Як оточення виборця беремо окіл Мура. Тоді функція переходу такого клітинного автомата описується рівнянням

$$y^{i,j}(t+1) = \operatorname{argmax}_{v=1..n} C(v, O(i, j), t), \quad (1)$$

де  $y^{i,j}(t)$  — альтернатива, яку обрав індивід в клітині з індексами  $i, j$  в момент часу  $t$ ;  $O(i, j)$  — окіл клітини з індексами  $i, j$ ;  $C(v, O(i, j), t)$  — кількість клітин зі значенням  $v$  у околі  $O(i, j)$ .

Крім того, можна замість функції максимуму у рівняння (1) розрахувати вагу кожної з альтернатив, які будуть нормованими значеннями

$C(v, O(i, j), t)$  і на їх підставі випадковим чином обирати одну з альтернатив. Типові конфігурації такого автомата зображено на рис. 2.

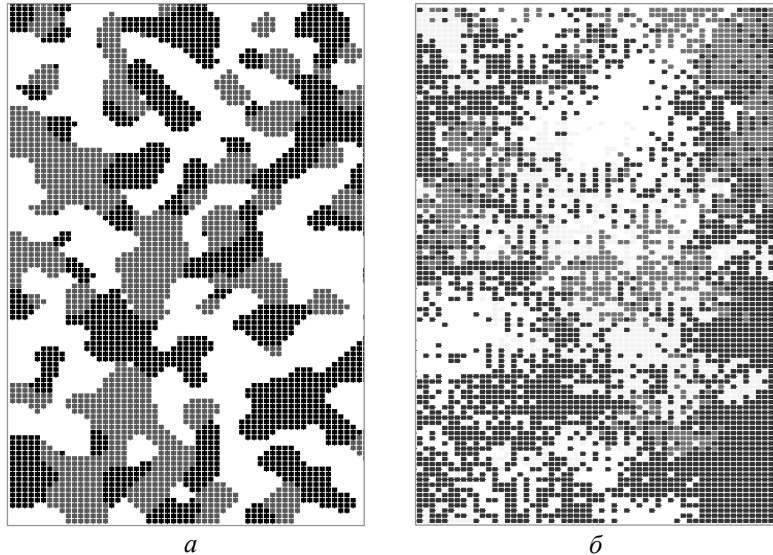


Рис. 2. Еволюція моделі впливу оточення звичайним автоматом (а) та ймовірнісним автоматом (б)

Детермінована модель характеризується швидкою збіжністю (20–30 тактів). У ній утворюються стійкі групи виборців з чіткими та доволі гладкими межами. На відміну від неї стохастична модель не збігається і утворює доволі складні еволюції (рис. 3), які також характеризуються утворенням груп виборців, однак не настільки чітких, як у звичайній моделі.

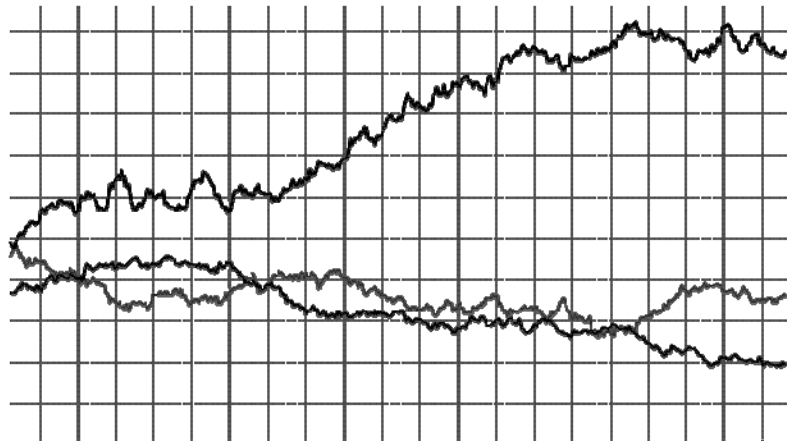


Рис. 3. Динаміка зміни кількості прихильників партій в імовірнісній моделі

Крім наведених вище моделей впливу оточення, існують їх численні розширення, які враховують нерівномірний початковий розподіл, вплив зовнішніх факторів та залежність альтернатив [3]. Однак процес поширення думки в таких моделях відбувається рівномірно і задається тільки початковим розподілом, що не зовсім відповідає електоральним реаліям. Частково цю проблему вирішує модель з «далекими зв'язками».

Грановеттер [4] запропонував поділити зв'язки між індивідами на слабкі та сильні, взявши за означення, що чим більше спільних друзів мають два індивіди, тим сильніший зв'язок між ними. Відповідно до цього він висунув гіпотезу: соціальні зв'язки різних типів мають різні ролі в динаміці соціальної структури та обміну інформацією.

У праці [5] уже розглядалась модель клітинного автомата з «далекими (слабкими) зв'язками». Тобто до околу Мура для кожного індивіда додавалася наперед задана кількість випадково обраних клітин, які відповідали «далеким зв'язкам». Проведені дослідження моделі показали, що на відміну від базової детермінованої моделі швидкість збіжності (набуття до стаціонарних станів) виявилася дуже великою (10–30 тактів), при цьому кінцеві стани свідчать про значну кількісну перевагу електорату, що має на початку незначну перевагу. Однак така модель не дає змоги пояснити утворення стійких електоральних регіонів помірно популярних партій.

## ПІДХІД НА ОСНОВІ СКЛАДНИХ МЕРЕЖ

Альтернативою клітинним автоматам є підхід на основі складних мереж. Замість однорідної ґратки множина індивідів подається у вигляді графу. Причому розглядається не випадковий граф, а з характерними для соціальних мереж характеристиками.

Попередні дослідження [6] виявили, що кількість соціальних зв'язків у індивідів розподілено нерівномірно, зазвичай показниково. Такі мережі називають безмасштабними (scale-free). Відповідно і структура оточення, яке може впливати, змінюється відповідно до положення індивіда в мережі. Це означає, що в мережі є велика кількість вузлів (індивідів) з малою кількістю зв'язків та невелика група індивідів з великою кількістю зв'язків.

Варто зазначити, що зв'язки, окрім поділу на сильні та слабкі, можуть бути несиметричними. Те, що деякі індивіди можуть впливати на вибір конкретного індивіда, не означає, що і він так само впливає на їх вибір. Популярні індивіди з великою кількістю вхідних зв'язків зазвичай мають малу кількість вихідних.

Крім прямого зв'язку, індивіди можуть бути зв'язані між собою опосередковано, тобто через посередника чи групу посередників. У теорії мереж визначають поняття шляху як найкоротшої відстані між вузлами мережі та середнього шляху як середню по всіх парах вузлів найкоротшу відстань між ними. Соціальні мережі зазвичай мають доволі невеликий середній шлях. Такий ефект називають феноменом «малих світів» чи «п'яти рукостикань». Ця властивість проявляється саме через наявність у мережі «далеких зв'язків».

Складні соціальні мережі характеризуються наявністю структури спільнот (communities). Спільнота — група вузлів, які мають високу щільність їх ребер, попри те, що щільність ребер між окремими групами низька. Традиційним методом для виявлення структури спільнот є кластерний аналіз. Існують десятки методів, які базуються на різних вимірах відстаней між вузлами, зважених шляхових індексах між вузлами тощо.

У роботі більш ретельно розглянуто метод на основі оптимізації модularity, запропонований Ньюменом [7]. (Цей метод обрано тому, що він

реалізований у програмному пакеті Gephi і дозволяє швидко виділити спільноти в досить великих мережах.). Значення модулярності дорівнює різниці між часткою ребер всередині спільноти і очікуваної частки зв'язків, якби ребра були розміщені випадково. Мережі з високою модулярністю мають щільні з'єднання між вузлами в межах спільнот і рідкі з'єднання між вузлами з різних спільнот.

Припустимо, що задано мережу з  $n$  вершинами, які поділено на дві групи. Нехай  $s_i = 1$ , якщо вершина  $i$  належить групі 1, і  $s_i = -1$ , якщо вона належить групі 2. Нехай ребра між вершинами  $i$  та  $j$  задано матрицею суміжності  $A_{ij}$ , яка складається з 0 та 1 (розглядається випадок коли ребра без ваг). Утім очікувана кількість ребер між вершинами  $i$  та  $j$ , якщо вершини з'єднати випадковим чином, буде становити  $\frac{k_i k_j}{2m}$ , де  $k_i$  і  $k_j$  — кількість зв'язків відповідних вершин (ступені вершин), а  $m = \frac{1}{2} \sum_i k_i$  — загальна кількість ребер у мережі. Тоді модулярність визначається за таким виразом:

$$Q = \frac{1}{4m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] s_i s_j, \quad (2)$$

який можна переписати у матричній формі

$$Q = \frac{1}{4m} s^T B s,$$

де  $B$  — симетрична матриця:

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2m}.$$

Цю матрицю також називають «матрицею модулярності». Оскільки сума значень по рядках та стовпцях такої матриці дорівнює 0, вона завжди має власний вектор  $(1, 1, 1, \dots)$  і власне число 0.

Перепишемо вираз (2), подавши  $s$  як лінійну комбінацію нормалізованих власних векторів  $u_i$  матриці  $B$  таких, що  $s = \sum_{i=1}^n a_i u_i$ ,  $a_i = u_i^T s$ , тоді

$$Q = \frac{1}{4m} \sum_{i=1}^n a_i u_i^T B \sum_{i=1}^n a_i u_i = \frac{1}{4m} \sum_{i=1}^n (u_i^T s)^2 \beta_i,$$

де  $\beta_i$  — власні числа  $B$ , які відповідають власним векторам  $u_i$  і впорядковані за спаданням ( $\beta_1 \geq \beta_2 \geq \dots \geq \beta_n$ ).

Задача полягає у знаходженні такого вектора розбиття мережі  $s$ , який би максимізував значення  $Q$ . Тобто треба максимізувати  $u_i^T s$ . Для цього можна встановити  $+1$  для всіх елементів  $s$ , якщо відповідний елемент  $u_i$  є додатним і  $-1$ , якщо відповідний елемент  $u_i$  є від'ємним.

Для розбиття мережі більше ніж на дві спільноти [8] модулярність визначається як

$$Q = \frac{1}{2m} \sum_{ij} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j),$$

де  $c_i, c_j$  — спільноти вершин  $i$  та  $j$  відповідно;  $\delta(c_i, c_j)$  дорівнює 1, коли  $c_i = c_j$ , в іншому випадку — 0. Оптимізація складається з двох фаз: на першій вона утворює невеликі спільноти, щоб досягти локальних максимумів модулярності, а на другій будується нова мережа, вузли якої є спільноти, виявлені на першому етапі. Ці дві фази повторюються доти, доки не досягнеться максимум модулярності. Результатом роботи алгоритму є ієрархія спільнот. Детальне математичне обґрунтування запропонованого методу описано у праці [9]. Крім того, в останні роки з'явилося багато оптимізацій і вдосконалень виявлення спільнот на основі модулярності [10, 11].

Існує доволі велика кількість робіт, що пов'язує поляризацію поглядів у суспільстві з утворенням у ньому окремих спільнот [12, 13] та вплив модулярності на поширення інформації [14]. Висунемо гіпотезу, що розбиття електорального поля також пов'язане з розбиттям суспільства на спільноти. Для перевірки припущення розглянемо реальні дані електорального поля півдня України.

#### ОЦІНЮВАННЯ МЕЖІ ЕЛЕКТОРАЛЬНОГО ПОЛЯ В ПІВДЕННІЙ УКРАЇНІ

В електоральному сенсі південь України протягом багатьох попередніх виборів входив до великого південно-східного електорального регіону, однак події 2013–2014 рр. зумовили зсуви межі поля і на парламентських виборах 2014 р. в усіх виборчих округах Херсонської області перемогу здобули прозахідні партії, а натомість у сусідній Запорізькій області — проросійські партії (рис. 4). Результати усереднено за виборчими округами.



Рис. 4. Результати парламентських виборів 2014 р. на основі опрацювання 50% протоколів: ■ — прозахідні партії; ■ — проросійські партії

Цікавим є той факт, що електоральна межа майже збіглася з адміністративною, хоч області і не розділені горами чи великою річкою.

Для перевірки гіпотези було зібрано базу користувачів соціальної мережі «ВКонтакте», які мешкають (за даними профілю) на межі вище від згаданих областей. Цю соціальну мережу обрано з огляду на те, що вона найпопулярніша в Україні і має зручне API. Для отримання вибірки з мережі випадковим чином обрано користувачів, які мешкають у містах та селах, що потрапляють у прямокутник, зображений на рис. 5.

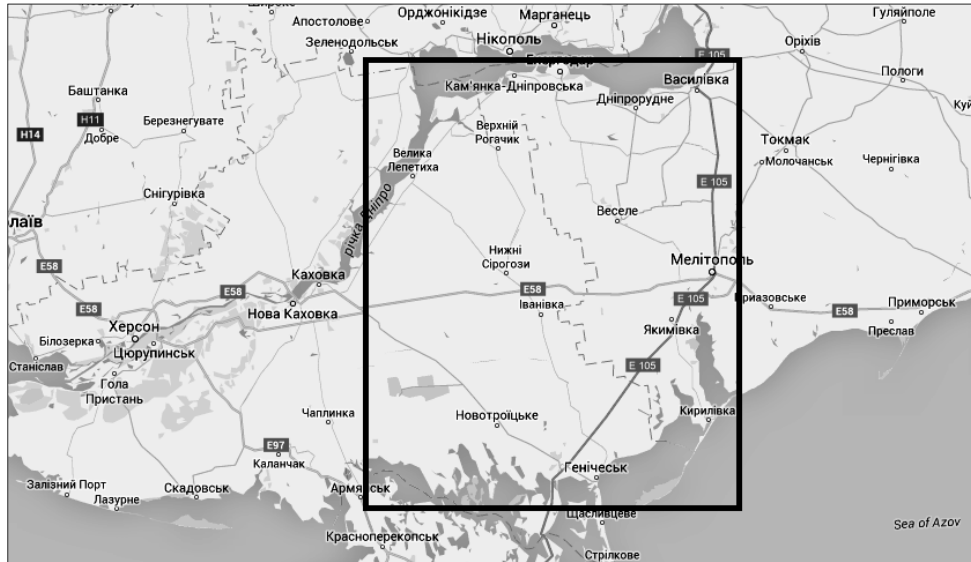


Рис. 5. Ділянка, з якої бралися місця для побудови мережі

Від кожного з таких користувачів розпочиналось випадкове блукання по мережі, запитуючи його друзів. Якщо будь-які два користувачі були друзями, вони поєднані зв'язком. Слід зазначити, що в такій соціальній мережі дружба напрямлена, тобто може не існувати зворотного зв'язку. У результаті отримано граф з вершинами і напрямленими ребрами (рис. 6):

За допомогою описаного вище методу на основі модулярності знайдено спільноти у мережі, які майже збігалися з межами міст та сіл. Однак з урахуванням початкової задачі, мета якої виокремлення лише двох основних спільнот, які могли б бути схожі на розбиття електорального поля. Результати такого розбиття порівняно з поділом за областями показано на рис. 7.

Із рисунка видно, що поділ за областями, хоча і проходить по межах спільнот, незначно відрізняється від розбиття за модулярністю. Зауважимо, що найбільшою спільнотою є місто Мелітополь, яке залучає спільноти сусідніх міст. Поділ за модулярністю об'єднує агломерацію Мелітополя з частиною південних міст Херсонської області.

Якщо порівняти карту території (рис. 5) та отриману мережу (рис. 6), то чітко видно залежність між ступенем зв'язності спільнот від наявності транспортної інфраструктури. Там, де є велика дорога, спільноти мають тісніший зв'язок і перекривають одна одну, а там, де немає, — модулярність більша. Так, наприклад, північ Херсонської та Запорізької областей мають низьку зв'язність, тому і голосують по-різному. У південній частині наявність великої дороги (траси E105) від Мелітополя на південь транслює погляди від великого міста на доволі великі приморські міста Херсонської об-

ласті (наприклад, Геніченськ). Електоральна мапа (див. рис. 4) усереднює результати по виборчих округах. За більш детального розгляду виявляється, що на дільницях Геніченська перевагу отримали проросійські партії. Це доводить ефективність використання модулярності як міри розподілу електорального поля, а отримані на її основі спільноти пояснюють утворення меж електоральних полів (рис. 7).

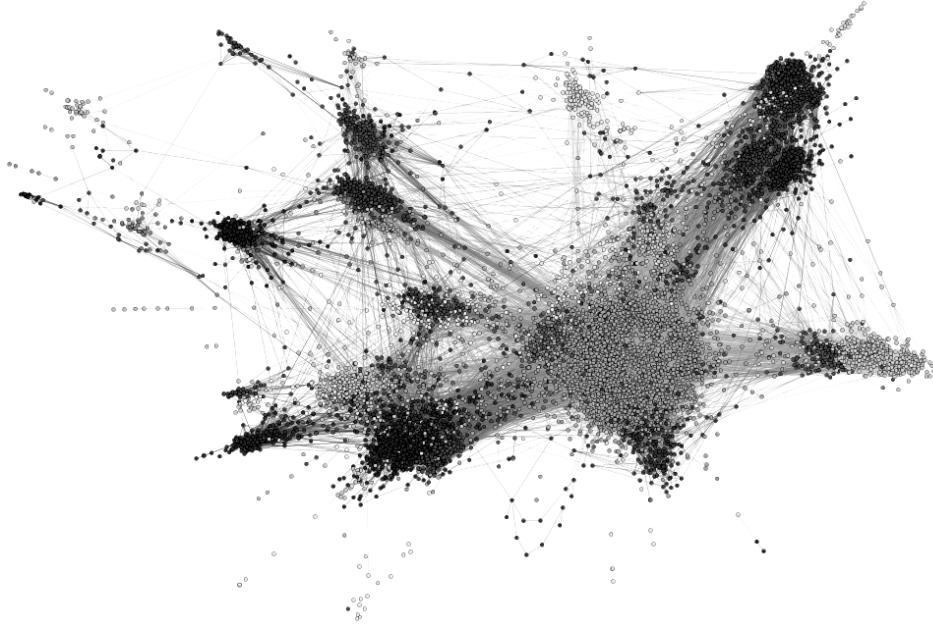


Рис. 6. Мережа користувачів соціальної мережі «Вконтакте» в районі Мелітополя, укладена за допомогою алгоритму ForceAtlas2 (насиченість вершини різниться відповідно до міста (села))



Рис. 7. Розбиття мережі на основі модулярності (а) та за областю (б)

## ВИСНОВКИ

Розглянуто задачу аналізу розподілу електоральних полів. Запропоновано підходи на основі клітинних автоматів та складних мереж. З огляду на можливість врахування нерівномірності поширення суспільної думки та легкість отримання початкових даних з популярних соціальних мереж перевагу надано моделі складних мереж. Для тестування запропонованих підходів по-



будовано граф користувачів соціальної мережі «ВКонтакте» на межі Херсонської та Запорізької областей. В ході експерименту підтверджено зв'язок між спільнотами в моделі складних мереж та межами розподілу електоральних полів. Запропонований метод на основі модулярності можна використовувати для пояснення наявних меж електоральних регіонів та передбачення місць утворення нових електоральних меж.

## ЛІТЕРАТУРА

1. *Ралко О.Н.* Возникновение и развитие электоральной географии в США и Западной Европе / О.Н. Ралко // Ученые записки Таврического национального университета им. В.И. Вернадского. Серия «География». — 2012. — **25** (64). — № 2. — С. 147–152.
2. *Brown T.A.* Nonlinear Politics / T.A. Brown // Chaos Theory in the Social Sciences. — 1996. — P. 119–137.
3. *Terpil I.O.* Simulation of Public Opinion with Ideas of Cellular Automata / I.O. Terpil A.S. Makarenko // Lecture Notes in Computer Science. — **8751**. — P. 518–525.
4. *Granovetter M.S.* The Strength of Weak Ties / M.S. Granovetter // American Journal of Sociology. — 1973. — **78**, Issue 6. — P. 1360–1380.
5. *Горбулін В.П.* Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: моногр. / В.П. Горбулін, О.Г. Додонов, Д.В. Ланде. — К.: Інтертехнологія, 2009. — С. 153–162.
6. *Barabási A.-L.* Statistical mechanics of complex networks / A.-L. Barabási, R. Albert // Reviews of Modern Physics **74**. — 2012. — **74**. — P. 47–97.
7. *Newman M.E.J.* Modularity and community structure in networks / M.E.J. Newman // Proceeding of the National Academy of Sciences of the United States of America. — 2006. — **103**. — P. 8577–8582.
8. *Blondel V. D.* Fast unfolding of communities in large networks / V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre // Journal of Statistical Mechanics: Theory and Experiment. — 2008. — **10**. — P. 1000.
9. *Lambiotte R.* Laplacian Dynamics and Multiscale Modular Structure in Networks / R. Lambiotte, J.-C. Delvenne, M. Barahona // IEEE Transaction on Network Science and Engineering. — 2015. — **1**, Issue 2. — P. 76–90.
10. *Chen M.* Community Detection via Maximization of Modularity and Its Variants / M. Chen, K. Kuzmin, B.K. Szymanski // IEEE Transactions on Computational Social Systems. — 2014. — **1**(1). — P. 46–65.
11. *Azizifard N.* Social Network Clustering / N. Azizifard // I.J. Information Technology and Computer Science. — 2014. — **01**. — P. 76–81.
12. *Adamic L.A.* The political blogosphere and the 2004 u.s. election: divided they blog / L.A. Adamic, N. Glance // Proceedings of the 3rd international workshop on Link discovery. — New York: LinkKDD. — 2005. — P. 36–43.
13. *Guerra P.H.C.* A Measure of Polarization on Social Media Networks Based on Community Boundaries / P.H.C. Guerra, W.M. Jr, C. Cardie, R. Kleinberg // Режим доступу: <https://www.cs.cornell.edu/home/cardie/papers/ICWSM13-Polarization.pdf>
14. *Nematzadeh A., Ferrara E., Flammini A., Ahn YY.* Optimal network modularity for information diffusion / A. Nematzadeh, E. Ferrara, A. Flammini, YY. Ahn // Phys. Rev. Lett. — 2014. — **113**. — P. 8.

*Надійшла 18.11.2015*