

Л. Ф. Панченко, Луганський національний університет імені Тараса Шевченка

ПІДГОТОВКА СТУДЕНТІВ УНІВЕРСИТЕТУ ДО АНАЛІЗУ СОЦІАЛЬНИХ МЕРЕЖ

Панченко Л. Ф.

Підготовка студентів університету до аналізу соціальних мереж

Стаття присвячена питанням аналізу соціальних мереж. Аналізується динаміка популярності програмного забезпечення аналізу соціальних мереж, можливості вільного програмного забезпечення NetLogo, Gephi, R для навчання аналізу соціальних мереж студентів різного фаху.

Ключові слова: соціальні мережі, аналіз, підготовка студентів

Панченко Л. Ф.

Подготовка студентов университета к анализу социальных сетей

В статье обсуждаются вопросы анализа социальных сетей. Анализируются динамика популярности программного обеспечения анализа социальных сетей, возможности свободного программного обеспечения NetLogo, Gephi, R для обучения анализу социальных сетей студентов разных специальностей.

Ключевые слова: социальные сети, анализ, подготовка студентов

Аналіз соціальних мереж (social network analysis, SNA) – це міждисциплінарне поле дослідження, яке лежить на перетині таких галузей, як „соціологія”, „математика” („теорія графів”), „психологія”, „інформатика”, „теорія складних систем”, „соціальна комунікація”, „статистика”.

Дослідження динаміки популярності терміну „social network analysis” з 2004 по 2012 рр. за допомогою Google Trend [1], здійснене автором на момент написання статті, показало сталий тренд (див. рис. 1) запитів, які тримаються біля 80-ох умовних одиниць, у той час, як відмітка 100 відповідає найбільшому обсягу запитів.

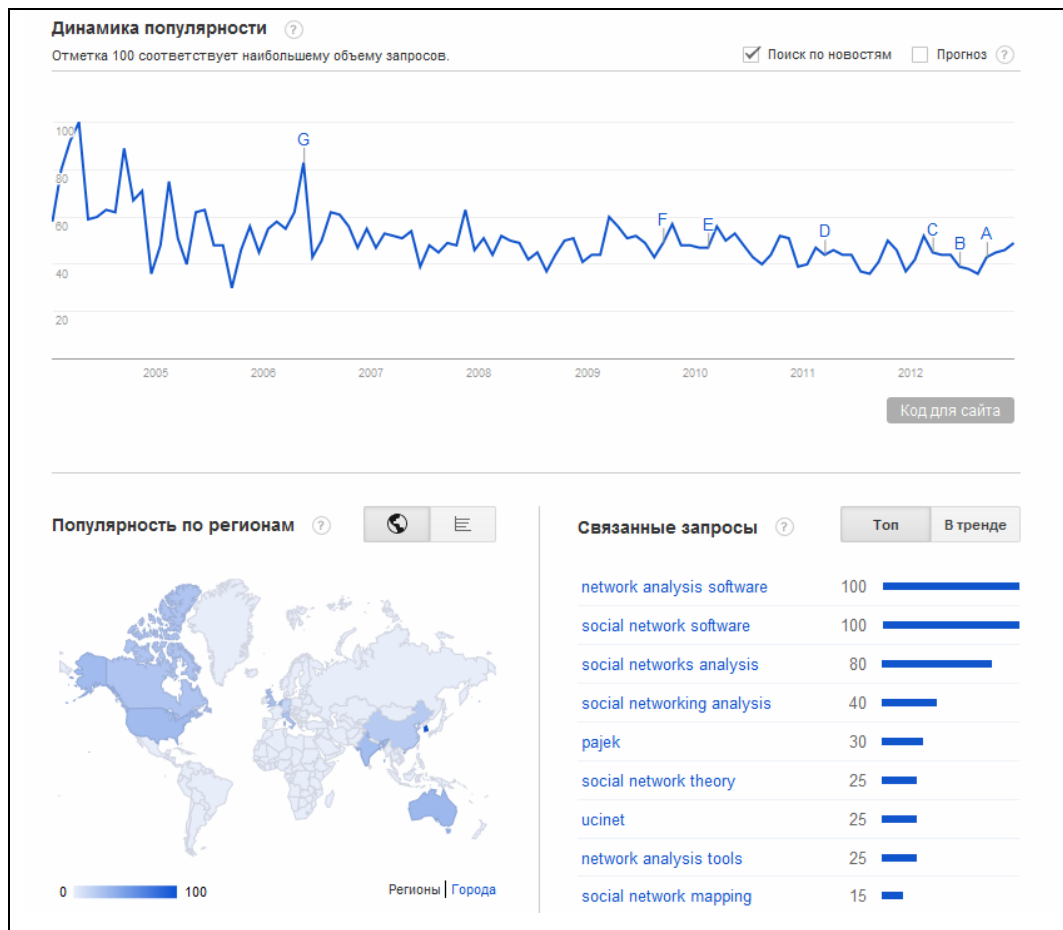


Рис. 1. Динаміка популярності запита „social network analysis” за Google Trend

Аналіз наукових джерел [4 – 8] свідчить, що у сучасному світі попит на вміння аналізувати та інтерпретувати дані соціальних мереж, орієнтовані на різні предметні галузі, такі як соціологія, соціальні комунікації, маркетинг, економіка, психологія, педагогіка, кримінологія тощо надалі зростає. Отже, важливу роль у підготовці конкурентоспроможного фахівця відіграє вивчення аналізу соціальних мереж, яке включає опанування студентами основних понять мережевого аналізу та комп’ютерних програм аналізу соціальних мереж. Ці програмні засоби надають можливість фахівцям з різних галузей, студентам створювати моделі мереж та процесів, що там перебігають, вивчати їхні статистичні та структурні властивості, взаємозв’язок акторів, прогнозувати поведінку мереж, яка визначається зміною структурних властивостей.

Аналіз соціальних мереж традиційно пов'язаний з іменами західних соціологів Дж. Морено [2], Дж. Барнса [3]; методи та застосування аналізу соціальних мереж розглядали С. Вассерман та К. Фауст [4]. Математичні моделі мереж будували та досліджували П. Ердеш, А. Рен'ї, А. Л. Барабаші [5]. Цей напрямок стрімко розвивається у США зусиллями групи „Social Network Analysis” Стенфордського університету [6], проекту Coursera Масачусетського технологічного інституту (MIT); INSNA – професійної асоціації дослідників, зацікавлених в аналізі соціальних мереж та ін. Асоціація INSNA (International Network for Social Network Analysis) [7] видає журнали „Journal of Social Structure”, „Social Networks Journal”, організує конференції і семінари, формує список розсилки SOCNET .

Навчання основам аналізу соціальних мереж стало складовою програми підготовки фахівців в університетах [5;13;21]. У Росії цей напрямок стосовно соціології розвивається в дослідженнях Г. Градосельської [8], А. Чуракова [9] та ін. На підґрунті власних досліджень Г. Градосельська розробила навчальний курс „Мережеві вимірювання в соціології”, який з 2004 р. викладається студентам „Вищої школи економіки”.

На жаль, в Україні засоби аналізу соціальних мереж недостатньо використовуються в підготовці фахівців у сфері вищої освіти. Ми знайшли лише одну програму курсу „Аналіз соціальних мереж” у Мелітопольському державному педагогічному університеті [10]. Практичні заняття з цього курсу зовсім не передбачають застосування програмних засобів аналізу соціальних мереж.

Мета статті – дослідити можливості сучасних засобів аналізу соціальних мереж, динаміку їх популярності з метою включення цих засобів у відповідні курси для навчання студентів різних спеціальностей.

Зазначимо, що процесу залучення комп'ютерних засобів аналізу соціальних мереж в університетські курси повинна передувати теоретична підготовка наступного змісту.

Соціальна мережа (англ. social network) – це структура яка складається з вузлів, які представляють соціальні об'єкти (людей, спільноти, організації, країни та ін.) і зв'язків між ними, що символізують соціальні відношення. У теорії складних мереж виділяють три напрямки дослідження мереж [11]: 1) дослідження атрибутів, які характеризують поведінку мережі; 2) створення або генерація моделей мереж; 3) передбачення поведінки мережі при зміні її структури. Згідно з теорією графів, при цьому досліджуються параметри окремих вузлів, мережі в цілому та мережеві підструктури.

Для окремих вузлів обчислюють наступні параметри: вхідний/вихідний ступінь вузлів; відстань від даного вузла до кожного з інших; середня відстань від даного вузла до інших; ексцентричність (eccentricity) — найбільшу з мінімальних відстаней від даного вузла до інших; посередництво (betwenness), що показує кількість найкоротших шляхів, що проходять скрізь даний вузол; центральність — загальна кількість зв'язків даного вузлу по відношенню до інших.

Для розрахунків параметрів мережі в цілому використовують число вузлів, число ребер, геодезичну відстань між вузлами, середню відстань від одного вузлу до іншого, щільність – відношення числа ребер в мережі до максимально можливої кількості ребер для даного числа вузлів; кількість триад, діаметр мережі (найбільшу геодезичну відстань).

Структурний аналіз мереж включає: *виявлення клік* (підгруп, що пов'язані між собою міцніше, ніж з вузлами інших клік); *виявлення компонент* мережі; знаходження *мостів* (вузлів, при видаленні яких мережа розпадається на незв'язні частини); *груп еквівалентних вузлів* (які мають максимально схожі профілі зв'язку) [11].

Одним з напрямків аналізу складних мереж є їх візуалізація, яка дозволяє отримати важливу інформацію про структуру і властивості мережі без точних розрахунків.

Програмні засоби аналізу соціальних мереж повинні підтримувати розрахунки усіх описаних параметрів вузлів, мережі в цілому, забезпечувати її структурний аналіз та візуалізацію, працювати з різними форматами даних.

На момент написання статті в англійській Вікіпедії [12] ми нарахували 89 посилань на різні програми аналізу соціальних мереж. Для відбору найбільш популярних програм ми звернулись до аналізу програмних засобів, що їх застосовано у курсах з аналізу соціальних мереж у провідних університетах світу [5; 6; 21], власного досвіду, що його отримано в результаті навчання на он-лайн курсі „Аналіз соціальних мереж” Університету Мічиган, в класі професора Лади Адамик [13], протягом вересня-листопада 2012 р., а також інструмента аналітики Google Trend.

Найбільш популярні запити за тематикою „аналіз соціальних мереж” (див. рис. 1) спостерігаються в таких регіонах, як Північна Америка, Австралія, Азія, Західна Європа. Серед переліку пов’язаних запитів зустрічаються назви двох програмних засобів аналізу соціальних мереж *pajek* та *ucinet*.

Відомості про програмні засоби, які використовуються в провідних університетах світу та науковій літературі, зведено в таблицю 1.

Таблиця 1

Використання програмних заходів аналізу соціальних мереж в провідних університетах світу

Програмний засіб	Адреса	Викладач, курс, університет				
		Лада Адамик, Університет Мічиган, курс «Аналіз соціальних мереж»	Ланде Д.В та ін.	McFarland D. Standford university	Hanneman, Robert A. and Mark Riddle	Batagelj V., Mrvar A.:
Geghi	http://gephi.org/	++				
NetLogo	http://ccl.northwestern.edu/netlogo	++				
Igraph	http://igraph.sourceforge.net/	++		++		
Pajek	pajek.imfm.si	+				++
UCINet	www.analytictech.com/ucinet/	+		+	++	++
NodeXL	http://nodexl.codeplex.com/	+				
NetDraw	https://sites.google.com/site/netdrawsoftware/home				+	

NetViz	http://www.ca.com/us/content/Integration/netviz.aspx		+			
Inflow	http://orgnet.com/		+			
Touch Graph	www.touchgraph.com		+			
R	http://www.r-project.org/	+		++		
NetworkX	http://networkx.lanl.gov/	+				
SoNIA	http://www.stanford.edu/group/sonia/				++	

Примітка: ++ – використовується в курсі, + – згадується в курсі або публікації.

Таким чином до складу найбільш популярних інструментів аналізу соціальних мереж увійшли наступні [11-18]: Gephi, NetLogo, Igraph, Pajek, UCINet, NodeXL, NetDraw, NetViz, Inflow, Touch Graph, R, NetworkX, SoNIA.

Ми знову звернулися до Google Trend і отримали таку картину при порівнянні п'яти програмних засобів аналізу соціальних мереж: Pajek, UCINet, NodeXL, Gephi, igraph (рис.2).

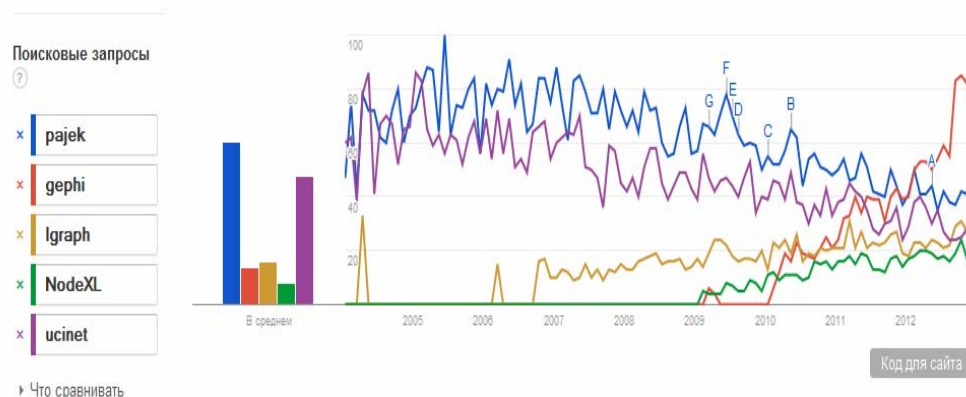


Рис.2. Порівняння динаміки популярності запитів програмних засобів аналізу соціальних мереж за допомогою Google Trend

Аналіз, проілюстрований на рис.2, свідчить, що в середньому найбільш популярними є запити на програмне забезпечення Pajek та UCINet, але динаміка цих запитів така, що, починаючи з 2008 р., кількість запитів поволі знижується. У той же час, приблизно з 2010 р. почала зростати популярність таких засобів, як NodeXL, Gephi, igraph. Особливо це стосується Gephi, попит на який стрімко пішов вгору. У таблиці 2 наведено розподіл популярності gerphi-запитів за регіонами, а в таблиці 3 за змістом пов'язаних запитів, які ми

отримали із застосуванням Google Trend. Можна спостерігати найбільшу популярність у Франції, США, Німеччині та Великобританії. Пов'язані запити стосуються побудови графа за допомогою gephi, пошуків тьюторіалу для роботи з програмою, загрузки програмного засобу, аналізу даних соціальних мереж Facebook та Twitter в Gephi.

Таблиця 2

Країна	Популярність запитів
Франція	100
США	77
Німеччина	65
Великобританія	64

Таблиця 3

Пов'язані запити	Популярність запитів
gephi graph	100
gephi tutorial	60
download gephi	45
gephi facebook	35
gephi twitter	25

Слід зазначити, що серед описаних вище 5-ти засобів аналізу соціальних мереж вільно поширюються всі, крім UCInet, який можна безкоштовно використовувати 90 днів.

Дещо окремо від описаних програмних засобів аналізу соціальних мереж відстоїть NetLogo. NetLogo [20] не є спеціальним інструментом аналізу соціальних мереж, це – мультиагентна система, призначена для моделювання та дослідження явищ у різних галузях. Бібліотека моделей NetLogo (версії 5.0RC3) містить секцію „Networks” з 6 моделями, які ілюструють структуру та процеси, що перебігають у мережах:

- *Diffusion on a Directed Network* (Дифузія у спрямованих мережах)
- *Giant Component* (Гігантський компонент)
- *Preferential Attachment* (Домінантне зв'язування)
- *Small Worlds* (Малі світи)
- *Team Assembly* (Збір команди)
- *Virus on a Network* (Вірус в мережі).

Зупинимося докладніше на останній моделі [21]. Ця модель демонструє процес розповсюдження вірусу в мережі (рис. 2). Хоча модель абстрактна, однією її інтерпретацією може бути така. Кожний вузол представляє комп'ютер, і ми моделюємо процес розповсюдження вірусу в комп'ютерній мережі. Кожен вузол може знаходитися в одному з трьох станів: чутливий, інфікований, не чутливий. В академічній літературі така модель відома як SIR-модель для епідемій.

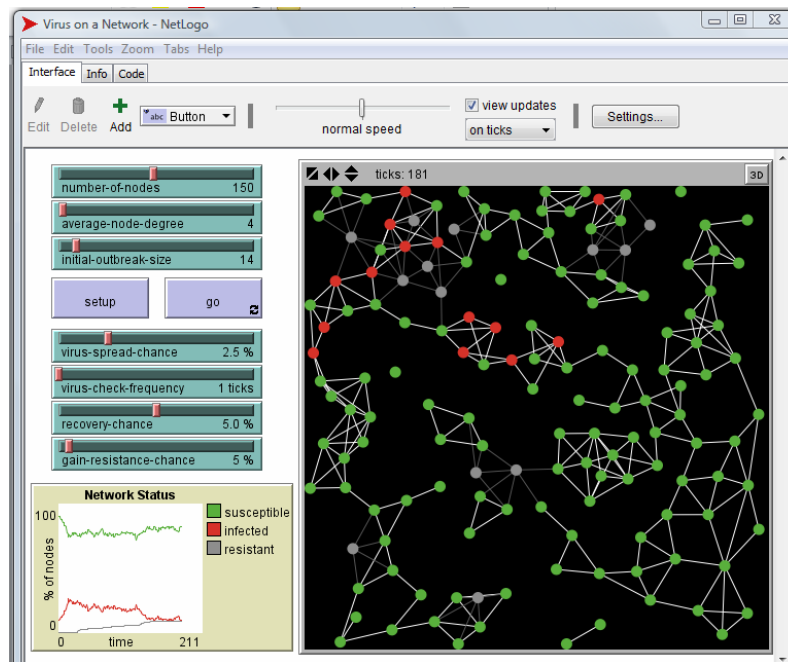


Рис. 2. Модель розповсюдження вірусу в NetLogo

У будь-який момент часу, кожний інфікований вузол (червоного кольору) намагається інфікувати його сусідів. Чутливі сусіди (зеленого кольору) будуть інфіковані з імовірністю, яка задається слайдером VIRUS-SPREAD-CHANCE (шанс розповсюдження вірусу). Це може відповідати імовірності, з якою якась чутлива система може бути інфікована доданим до e-mail файлом.

Не чутливі вузли (сірого кольору) не можуть бути інфіковані. Це може відповідати оновленому програмному забезпеченню антивірусу, яке зробило комп'ютер нечутливим до цього виду вірусу.

Інфіковані вузли не одразу проявляються. Слайдер VIRUS-CHECK-FREQUENCY (частота контролювання вірусу) визначає частоту контролю

вузла на інфікованість. Це відповідає або частоті запуску процедури сканування на віруси, або тому, як людина стежить за поведінкою комп'ютера. Якщо вірус виявлено, імовірність, що його буде нейтралізовано, задається слайдером RECOVERY-CHANCE (шанс відновлення).

Якщо вузол відновлено, є деяка імовірність, що він буде стійкий до інфекції в майбутньому, яка задається слайдером GAIN-RESISTANCE-CHANCE. Коли вузол стає відновленим, зв'язки між ним та сусідами стають темними, оскільки вірус не може розповсюджуватися далі.

Крім дослідження готових моделей, студенти можуть створювати власні моделі мереж і досліджувати їх властивості. Значну кількість таких моделей до кожної теми курсу було складено Л. Адамик в курсі „Аналіз соціальних мереж” [13].

Розглянемо тепер можливості середовища R. R містить декілька пакетів [6; 12], релевантних до аналізу соціальних мереж (див. табл. 4).

Таблиця 4

Пакети R для аналізу соціальних мереж

Назва пакету	Призначення
<i>igraph</i>	Генерація й аналіз мереж
<i>sna</i>	Соціометричний аналіз мереж
<i>network</i>	Управління та показ мережевих об'єктів
<i>tnet</i>	Аналіз зважених мереж, лонгітюдних мереж
<i>ergm</i>	Засоби аналізу та симуляції мереж, основаних на експоненціальному випадковому графі
<i>Bergm</i>	Засоби для байсового аналізу експоненціального випадкового графа
<i>hergm</i>	Реалізація ієрархічних експоненціальних випадкових графів
<i>latentnet</i>	Функції для латентного аналізу і кластерних моделей
<i>degreenet</i>	Статистичне моделювання розподілу ступеню мережі
<i>networksis</i>	Засоби для симуляції мереж дводольних графів із фіксованими маргіналами

Пакет *igraph* – це написана на C бібліотека для аналізу великих мереж. Вона включає швидку реалізацію як класичних завдань теорії графів, так і нещодавно розроблених методів мережевого аналізу. Інтерфейс високого рівня є для R, Python і Ruby.

Група Стенфордського університету – Social Network Analysis” використовує *igraph* і *sna* пакети R для вивчення мережевих феноменів, фокусуючись при цьому на мережах шкіл, університетів, фірм, організацій, історичних подій, економічних транзакцій, он-лайн спільнот, груп тварин, епідеміології тощо.

Розроблене нами методичне забезпечення аналізу соціальних мереж в середовищі NetLogo та середовищі Gephi складається з лекцій-презентацій, циклу лабораторних робіт, завдань для самостійної роботи й може використовуватися викладачами й студентами для дослідження готових моделей мереж, їх обговорення, розширення, доповнення, складання своїх моделей (в NetLogo), аналізу мережевих даних, готових чи власноруч отриманих (Gephi). Методичне забезпечення було апробоване у навчальному процесі магістратури й бакалавріата студентів різних спеціальностей.

Отже, комп’ютерні засоби аналізу соціальних мереж надають можливість розуміння студентами зв’язків у сучасних мережах, моделювати і вивчати поведінку мереж, розвивають „мережеве” мислення майбутніх фахівців, вміння робити прогнози та передбачення. У результаті дослідження було виявлено та проаналізовано динаміку популярності програмних засобів аналізу соціальних мереж Pajek, UCInet, NodeXL, Gephi, *igraph*; розроблено та апробовано навчально-методичне забезпечення для аналізу соціальних мереж в середовищах вільно-поширених програмних засобів NetLogo та Gephi.

Подальший напрямок роботи – розширення створеного навчально-методичного забезпечення за рахунок підключення нових комп’ютерних засобів аналізу соціальних мереж, зокрема *igraph* [14], пошук масивів даних для аналізу в різних галузях.

Література

1. **Google Trends** [Електронний ресурс]. – Режим доступу : <http://www.google.com/trends>

2. **Морено Я. Л.** Социометрия: Экспериментальный метод и наука об обществе / Я. Л. Морено / пер. с англ. А. Боковинова. – М. : Академический Проект, 2001.
3. **Burt R.** 2001. The Social Capital of structural holes // Guillen M. F., Collins R., England P., Meyer M. (eds.). New Directions in Economic Sociology. N.Y. : Russel Sage Foundation. – P. 201 – 246.
4. **Wasserman S.** Social Network Analysis: Methods and Applications / S.Wasserman, K. Faust. – Cambridge University Press, 1994.
5. **A course** on complex network analysis by Albert-László Barabási Network Science [Электронный ресурс]. – Режим доступа : <http://barabasilab.neu.edu/courses/phys5116/>
6. **McFarland D.** Social Network Analysis Labs in R / Daniel,McFarland, Solomon Messing, Michael Nowak, Sean J. Westwood. – Stanford University. – 2010. [Электронный ресурс]. – Режим доступа: <http://sna.stanford.edu/rlabs.php>
7. **INSNA** [Электронный ресурс]. – Режим доступа : <http://www.insna.org/>
8. **Градосельская Г. В.** Сетевые измерения в социологии : учеб. пособие / Г. В. Градосельская / М. : Изд. Дом „Новый учебник”, 2004.
9. **Чураков А. Н.** Вероятностные модели социальных сетей / А.Н.Чураков // Социологические исследования. – 2001. – № 9. – С. 99-114.
10. **Аналіз** соціальних мереж [Электронный ресурс]. – Режим доступа : http://www.lib.mdpu.org.ua/e-book/analiz_soc/official/index.htm
11. **Ландэ Д. В.** Интернетика в сложных сетях: модели и алгоритмы / Д. В. Ландэ, А. А. Снарский, И. В. Безсуднов. – М. : Либроком, 2009. – 264 с.
12. **Social network analysis software** [Электронный ресурс]. – Режим доступа : http://en.wikipedia.org/wiki/Social_network_analysis_software
13. **A course** on social network analysis by Lada Adamic [Электронный ресурс]. – Режим доступа: <https://class.coursera.org/sna-2012-001/class/index>

14. **Borgatti S. P.** Ucinet for Windows: Software for Social Network Analysis / S. P. Borgatti, M. G. Everett, L. C Freeman. – Harvard, MA : Analytic Technologies, 2002.

15. **Bastian M.** Gephi: an open source software for exploring and manipulating networks / M. Bastian, S. Heymann, M. Jacomy // International AAAI Conference on Weblogs and Social Media. – 2009. – [Электронный ресурс]. – Режим доступа : <http://gephi.org/publications/gephi-bastian-feb09.pdf>

16. **Gephi** [Электронный ресурс]. – Режим доступа : <https://gephi.org/>

17. **W. de Nooy.** Exploratory Social Network Analysis with Pajek /W. de Nooy, A. Mrvar, V. Batagelj. – CUP, 2011. – 442 p.

18. **Pajek** [Электронный ресурс]. – Режим доступа : <http://pajek.imfm.si/doku.php>

19. **Stonedahl F.** NetLogo Virus on a Network model / F. Stonedahl, U. Wilensky. – Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. – 2008. [Электронный ресурс]. – Режим доступа : <http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork>.

20. **Wilensky, U.** NetLogo. – Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. –1999. – [Электронный ресурс]. – Режим доступа : <http://ccl.northwestern.edu/netlogo/>.

21. **Hanneman R. A.** Introduction to social network methods / Robert A Hanneman, Mark Riddle. – Riverside, CA: University of California. – 2005. [Электронный ресурс]. – Режим доступа : <http://faculty.ucr.edu/~hanneman/>

Panchenko L. Ph.

Student Training for Social Network Analysis

The article addresses the questions of social network analysis. Possibilities of the free software are analyzed : NetLogo, Gephi, Rusage in teaching of modeling to student of various specialities are offered.

Key words: social network, analysis, student training.

Відомості про автора

Панченко Любов Феліксівна – кандидат педагогічних наук, доцент кафедри теоретичної і прикладної інформатики Луганського національного університету імені Тараса Шевченка. Основні наукові інтереси: інформаційні технології в освіті, інформаційно-освітнє середовище університету, комп'ютерний аналіз даних в педагогіці, психології, соціології.

Стаття надійшла до редакції 13.12.2012 р.

Прийнято до друку 21.12.2012 р.