

Д. В. Шингалов

Центральноукраїнський національний технічний університет, Кропивницький, Україна

ДОСЛІДЖЕННЯ ПРОГРАМНИХ ЗАСОБІВ ДЛЯ АНАЛІЗУ ТА ВІЗУАЛІЗАЦІЇ СОЦІАЛЬНИХ ГРАФОВИХ СТРУКТУР

Предметом вивчення у статті є процес аналізу та візуалізації соціальних графових структур. **Метою** є дослідження програмних засобів для аналізу та візуалізації соціальних графових структур. **Завдання:** дослідити сучасні платформи для побудови соціальних графів та їх статистичного аналізу, порівняти їх переваги та недоліки, дослідити придатність платформ для застосування до великих наборів даних та доцільність їх використання для аналізу соціальних мереж. Отримані такі **результати:** Досліджено найбільш популярні на ринку платформи для візуалізації графових структур. Здійснено порівняльний аналіз цих платформ з точки зору наявності вільної ліцензії на використання, їх мультиплатформенності та багатоформатності, а також можливості застосування до великих соціальних мереж, та оцінки впливовості агентів, наявності відкритого програмного коду. **Висновки.** Основними перевагами більшості програмних засобів є підтримка багатьох форматів, багато можливостей для математичного та статистичного аналізу та умовно вільна або вільна ліцензія. Серед недоліків слід відмітити, що не усі платформи призначені для роботи в умовах великих динамічних об'ємів даних отриманих з соціальних мереж, також, на жаль, в усіх програмних засобах зовсім відсутня можливість оцінки впливовості агентів соціальних відносин. Для вирішення даної проблеми необхідне створення програмного забезпечення, яке би аналізувало інформаційні потоки та визначало ступінь впливу поширюваної інформації, а також рівень взаємного впливу між агентами соціальних спільнот.

Ключові слова: аналіз соціальних мереж, візуалізація графів, соціальний граф, соціальні мережі, програмні засоби.

Вступ

Аналіз соціальних мереж бере свій початок не тільки в соціальній науці, а й у широкій області мережевого аналізу та теорії графів. Мережевий аналіз пов'язаний з поданням і рішенням проблем мережевої структури та соціальних відносин всередині груп, що дозволяє розкривати неформальні зв'язки між членами групи. Це допомагає аналітично виявляти агентів взаємодії, інформаційні потоки, напрямки поширення знань та ідей, механізми прийняття рішень.

Метою даної роботи є дослідження існуючих програмних засобів для виявлення найбільш придатних з них для аналізу та візуалізації соціальних графових структур. Поставлена мета реалізується шляхом вирішення таких задач: дослідження параметрів аналізу соціальних графів; дослідження пристосованості програмних засобів до роботи з різними операційними системами; дослідження можливості застосування програмного забезпечення для аналізу великих соціальних мереж.

Основна частина

Теорія графів [1] надає набір абстрактних понять і методів для аналізу графів. У поєднанні з іншими аналітичними інструментами і методами, розробленими для візуалізації та аналізу соціальних (і інших) мереж, теорія графів утворює автоматизовані методи аналізу соціальних мереж. В аналізі задіяні не тільки індивідууми і їх атрибути, а також макроскопічні соціальні структури і відносини між окремими особами, групами або соціальними установами.

Вхідними даними для програмного аналізу соціальних мереж є соціальний граф, його вузли та ребра. Серед головних обчислювальних параметрів соціальних мереж, та соціальних спільнот зокрема, на яких зосереджено аналіз [2, 3] є щільність, середня відстань від однієї вершини до інших, діаметр соці-

альної мережі, ступінь центральності, модульність, індекс впливовості агентів.

Щільність – являє собою концентрацію зв'язків у соціальній мережі, і розраховується шляхом ділення числа спостережуваних відносин на загальну кількість можливих зв'язків у даній мережі. Щільність мережі можна розглядати як показник згуртованості. У більш щільній мережі інформація поширюється швидше, ніж у мережі з меншою кількістю зв'язків.

Діаметр соціальної мережі – найдовший шлях між вершинами в графі.

Ступінь центральності – це кількість посилок, які ведуть до вузла або з вузла, у неорієнтованому графі ці посилення ідентичні. Часто використовується як міра ступеня зв'язності вузла і впливає на популярність. Дослідження цієї властивості є корисним при оцінці того, які вузли є центральними щодо поширення інформації і як вони впливають на інші вузли в безпосередніх околицях.

Модульність – це оцінка ступеня розбиття графу на підграфи на основі порівняння між даними реального графу та моделі графу, побудованого випадковим чином з тим же самим очікуваним ступенем розподілу, що і в оригінальному графі.

За функціональними можливостями програми для роботи з соціальними графами можна поділити на 3 класи [4]: програми для аналізу; програми для візуалізації; комплексні (використовуються для аналізу і мають модуль візуалізації).

Мережеві аналітичні інструменти використовуються для подання вузлів (агентів) і ребер (відносин) в мережі, а також для аналізу мережі передачі даних. Як і в інших програмних засобах, дані можуть бути збережені в зовнішніх файлах. Інструменти для аналізу соціальних мереж дозволяють дослідникам аналізувати великі соціальні мережі, передачу вірусної інформації і т.д. Ці інструменти реалі-

зують математичні функції, які можуть бути застосовані до мережевої моделі.

Візуалізація соціальних мереж дозволяє наглядно представити мережеві дані і передавати результати аналізу. Багато програм для мережевого аналізу мають модулі для візуалізації. Дослідження даних здійснюється через відображення вузлів і зв'язків у різних компоновках, і кольорах, розмірі і додаткових установках вузлів. Типове представлення мережевих даних являє собою скомпонований соціальний граф (вузли та зв'язки). Це не дуже легко читатися і не дозволяє наочно інтерпретувати дані.

Програмне забезпечення для аналізу соціальних мереж використовується для ідентифікації, представлення, аналізу, візуалізації чи моделювання вузлів та відносин з різних типів вхідних даних (реляційних і не реляційних), включаючи математичні моделі соціальних мереж. Вихідні дані можна зберігати у зовнішніх файлах. Існують різні формати файлів вводу та виводу. Інструменти аналізу мережі дозволяють дослідникам аналізувати представлення мереж різного розміру – від малих (наприклад, сім'ї, команди проекту) до дуже великих (наприклад, Інтернет, передача хвороб). Більшість з програм аналізу соціальних графів надають різні інструменти математичних та статистичних процедур, які можна застосувати до моделей мереж.

Візуальні уявлення в соціальних мережах важливі для розуміння мережевих даних та передачі результатів аналізу. Візуалізація часто використовується як додатковий або автономний метод аналізу даних. Інструменти візуалізації даних аналізу мережі використовуються для зміни макета, кольорів, розміру та інших властивостей мережі.

Програми аналізу соціальних мереж поділяються за сферою застосування. Для науково-дослідних цілей зазвичай використовують, такі інструменти як UCINET [5], Pajek [6], ORA [7], пакет Statnet [8,9]. Для соціальних мереж, орієнтованих на бізнес, використовують iPoint [10], NetMiner [11], InFlow [12], NodeXL [13]. Для великих мереж з мільйонами вузлів найбільш придатні Sonamine або ORA [7]. Пакети з відкритим вихідним кодом з графічним інтерфейсом для Linux, Windows і Mac - це VisuaLyzер [14] для соціальних мереж або SocNetV [15].

UCINET – це програмний пакет для аналізу даних соціальної мережі. Він оснащений інструментом візуалізації NetDraw. Хоча абсолютний максимальний розмір мережі при аналізі становить близько 2 мільйонів вузлів, на практиці більшість процедур UCINET занадто повільні для запуску мереж, що перевищує приблизно 5000 вузлів. Однак це залежить від конкретного аналізу та роздільної здатності мережі. Наприклад, ступінь центральності може бути визначений в мережах десятків тисяч вузлів, і більшість теоретичних графіків працюють швидше, коли у вас дуже мало зв'язків, незалежно від того, скільки вузлів у вас є.

Pajek – програмний пакет для аналізу, візуалізації та макетування соціальних мереж.

Обчислює багато понять центральності: ступінь, проміжність, близькість. Має можливість підключення

додаткових компонент для аналізу. В пакеті передбачено функцію автоматичного відбір зразків графів та перегляд проміжних результатів аналізу. Передбачено модулі для аналізу мережі Facebook. Корисна інформаційна підтримка споживачів: навчальні посібники, лекції, книга.

ORA – це інструмент оцінки та аналізу динамічних мета-мереж. Він містить сотні функцій для обчислення динамічних мережевих показників, метрики стежок, процедури групування вузлів, визначення локальних шаблонів, порівняння та протиставлення мереж, груп та осіб з динамічної перспективи мета-мережі. Він може визначити основних гравців, групи та вразливості, модельні зміни в мережі протягом певного часу.

Statnet – це набір програмних пакетів для мережевого аналізу, які впроваджують останні досягнення в статистичному моделюванні мереж. Аналітична база побудована на експоненціальних сімейних моделях випадкових графів. Він забезпечує всеохоплюючу основу моделювання мереж на основі моделювання, включаючи інструменти для оцінки моделі, моделювання мережі та візуалізації мережі. Ця широка функціональність забезпечується алгоритмом центрального марковського ланцюжка Монте-Карло.

Statnet має інше призначення, ніж чудові пакети UCINET або Pajek; акцент робиться на статистичному моделюванні мережевих даних. Можливості статистичного моделювання Statnet включають моделі ERGM, латентний простір та приховані кластерні моделі.

iPoint – являє собою набір інструментів аналізу мереж з акцентом на ефективність, портативність та простоту використання. Він є відкритим вихідним кодом і безкоштовним, може бути запрограмований в R, Python і C / C ++.

NetMiner – програмний інструмент для дослідницького аналізу та візуалізації мережевих даних. Він дозволяє візуально та в інтерактивному режимі вивчати дані мережі та допомагає виявити основні моделі та структури мережі,

InFlow – інтегрований продукт, що виконує мережевий аналіз і візуалізацію мережі, але не передає файли між різними програмами, як інші інструменти. Це єдине популярне програмне забезпечення SNA / ONA, яке має доступне навчання. Пакет успішно застосовується для Team Building, знаходження спеціалістів, різноманітної робочої сили, дизайн організації, представлення дифузії інновацій, у слідчій журналістиці та управлінні знаннями. Також InFlow забезпечує легкий доступ до найпопулярніших мережевих показників.

NodeXL – це надбудова огляду мереж, виявлення та розвідки для MS Excel. Це безкоштовний та відкритий проект, який розширює знайомі таблиці та включає функції аналізу та візуалізації соціальних мереж. Пакет дозволяє відображати та аналізувати мережевий граф.

Sonamine – інтерактивний пакет аналізу та візуалізації соціальних даних, який надає змогу швидко генерувати оцінки мережі, які потім можуть бути

використані в прогнозованому моделюванні та завданнях машинного навчання. Sonamine – це масштабований програмний засіб, для мережевого аналізу рівня підприємства. Це пакетне середовище для запуску різних оцінок і вилучення графів.

VisuaLyzer – програмний пакет призначений для графічного відображення малих і середніх соціальних мереж. Дослідники можуть імпортувати свої дані з edgelist UCINET або edgearray, GraphML та інші формати в графічну мережу вузлів та послань, що їх зв'язують. VisuaLyzer надає дослідникам низку функцій аналізу мережі, таких як обчислення клік, розділів, спільнот, найкоротших шляхів, найближчих сусідів, ролей та посад.

SocNetV – це крос-платформенна, зручна для користування програма для аналізу та візуалізації соціальних мереж. За допомогою SocNetV можна моделювати соціальні графи, редагувати акторів та зв'язки, аналізу-

вати графіку та властивості соціальної мережі, створювати HTML-звіти та вставляти макети візуалізації у мережу. Підтримує стандартні графіки та показники мережної згуртованості, такі як щільність, діаметр, геодезичні та відстані, зв'язність, ексцентричність, коефіцієнт кластеризації, взаємність тощо.

Gephi – це провідне програмне забезпечення для візуалізації та дослідження всіх видів графів та мереж. Gephi має відкритий код і є безкоштовним. Аналіз досліджуваних даних: інтуїтивно-орієнтований аналіз за допомогою маніпуляцій мереж в режимі реального часу. Призначена для виявлення основних структур асоціацій між об'єктами. Має легке створення з'єднувачів соціальних даних для громадських організацій та мереж малого бізнесу.

Порівняння параметрів програм представлено у табл. 1.

Таблиця 1 – Порівняння параметрів аналітичних програмних засобів для соціальних мереж

Назва програми	Можливість візуалізації	Формат вхідних даних	Формат вихідних даних	Платформа	Ліцензія	Можливість аналізу	Застосування до великих мереж	Відкритий код	Оцінка впливовості агентів
UCINET	-	Багато форматів	Багато форматів	Windows	Умовновільна	+	-	-	-
Pajek	+	Багато форматів	Багато форматів	Мультиплатф.	Вільна	+	+	-	-
ORA	+	.csv	.csv	Windows	Вільна	+	-	-	-
Statnet	-	Багато форматів	Багато форматів	Мультиплатф.	Відкрита	+	-	-	-
iPoint	+	Тільки XML	XML, Flex	Мультиплатф.	Вільна	+	-	+	-
NetMiner	+	Багато форматів	Багато форматів	Windows	Платна	+	+	-	-
InFlow	+	.csv	Експорт даних	Windows	Платна	-	+	-	-
NodeXL	+	Багато форматів	Багато форматів	Windows	Вільна	-	-	-	-
Sonamine	-	Текстові файли	Текстові файли	Windows, Linux	Платна	+	+	-	-
VisuaLyzer	+	Списки графу	Списки графу	Мультиплатф.	Платна	-	+	+	-
SocNetV	+	Багато форматів	Багато форматів	Мультиплатф.	Вільна	+	+	+	-
Gephi	+	Багато форматів	.svg, .png, .gexf	Мультиплатф.	Відкрита	+	-	-	-

Висновки

Відповідно до даних таблиці видно, що більшість програмних засобів для аналізу є мультиплатформеними, або орієнтованими на роботу з операційною системою Windows різних версій.

Також слід зазначити, що більшість з програм, які мають широкі функціональні можливості є платними. Перевагою переважної більшості програм є великий спектр вхідних та вихідних форматів, не зважаючи на те, що найбільш популярним форматом для збереження даних про досліджувані графи є *.csv.

Зважаючи на всі переваги та недоліки розглянутих програм, можна зробити висновок, що найбільш придатними для аналізу великих графових

структур, таких як сучасні соціальні мережі є пакети програм Pajek, NetMiner, та SocNetV. Дані програмні засоби поєднують у собі функції статистичного та інтелектуального аналізу, мають широкі можливості візуалізації соціальних графів та підтримують більшість сучасних форматів вхідних та вихідних файлів. Крім того SocNetV має відкритий програмний код, що надає багато можливостей для аналізу, дослідження та модифікації існуючих алгоритмів.

Але в усьому сегменті подібних програмних засобів відсутній набір інструментів автоматичного визначення рівня впливовості агентів соціальної мережі або спільноти соціальної мережі на розповсюдження інформації та аналіз цієї самої інформації (контенту) на наявність психологічних загроз та пропаганди.

СПИСОК ЛІТЕРАТУРИ

1. Thelwall M. Big Data and Social Web Research Methods / M. Thelwall. – Wolverhampton: University of Wolverhampton, 2014. – 142 с.
2. Skold M. Social Network Visualization : дис. канд. наук з соц. комун. / Skold M. – Stockholm, 2008. – 61 с.

3. Hanneman R. Introduction to Social Network Methods / R.A. Hanneman. – Riverside: University of California, 2005. – 149 с.
4. Ferrara E. Mining and Analysis of Online Social Networks. : дис. докт. фіз.-мат. наук / Ferrara Emilio – Messina, 2012. – 176 с.
5. Suerdem A. Introduction to Social Network Analysis with UCINET / Ahmet K. Suerdem. – London: London School of Economics. – 31 с.
6. Batagelj V. Pajek Program for Analysis and Visualization of Large Networks / V. Batagelj, A. Mrvar. – Ljubljana: PdlLaTex, 2011. – 98 с.
7. Carley K. ORA: A Toolkit for Dynamic Network Analysis and Visualization / Kathleen M. Carley. – Pittsburgh: Carnegie Mellon University, 2014. – 13 с.
8. A statnet Tutorial / [S. Goodreau, M. Handcock, D. Hunter та ін.] // Journal of Statistical Software / [S. Goodreau, M. Handcock, D. Hunter та ін.], 2008. – (Volume 24, Issue 9). – С. 1–26.
9. Wild F. Learning Analytics in R with SNA, LSA, and MPIA / F. Wild. – Oxford: Springer, 2016. – 275 с.
10. Vavruska M. MVE - 2 Visualization library / M. Vavruska, M. Frank., 2005. – 15 с
11. Huisman, M., Van Duijn, M.: Software for statistical analysis of social networks/ Van Dijkum C., Blasius, J. Kleijer H., Van Hilten B. - The Sixth International Conference on Logic and Methodology, Amsterdam, The Netherlands ,2004. - 21 с.
12. Liebowitz J. Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations / Jay Liebowitz // Journal of knowledge managemen / Jay Liebowitz. – Maryland: Emerald Group Publishing Limited, 2005. – С. 76–86.
13. Hansen D. Analyzing Social Media Networks: Learning by Doing with NodeXL / D. Hansen, B. Shneiderman, M. Smith. – Maryland, 2009. – 47 с.
14. Aigner M. Analysis of social networks with focus on ram exchange in four communities in the Ethiopian highland / Marina Aigner. – Wien: University for Natural Resources and Life Sciences, 2012. – 91 с.
15. Stamelos I. Social Networks Visualizer. Software Requirements Specification Version 1.0 / Ioannis Stamelos. – Thessaloniki: Aristotle University of Thessaloniki, 2012. – 38 с.

Рецензент: д-р техн. наук, проф. О. А. Смірнов,
Центральноукраїнський національний технічний університет, Кропивницький
Received (Надійшла) 28.06.2018
Accepted for publication (Прийнята до друку) 15.09.2018

Исследование программных средств для анализа и визуализации социальных графовых структур

Д. В. Шингалов

Предметом изучения в статье является процесс анализа и визуализации социальных графов структур. **Целью** является исследование программных средств для анализа и визуализации социальных графовых структур. **Задача:** исследовать возможные платформы для построения социальных графов и их статистического анализа, сравнить их достоинства и недостатки, исследовать пригодность платформ для применения для больших наборов данных и целесообразность их использования для анализа социальных сетей. Получены следующие **результаты**. Исследованы наиболее популярные на рынке платформы для визуализации графовых структур. Осуществлен сравнительный анализ этих платформ с точки зрения наличия свободной лицензии на использование, их мультиплатформенности, многоформатности, а также возможности применения к большим социальным сетям и оценки влияния агентов, наличия открытого кода. **Выводы.** Основными преимуществами большинства программных средств является поддержка многих форматов, широкий спектр возможностей для математического и статистического анализа и условно свободная или свободная лицензия. Среди недостатков следует отметить, что не все платформы предназначены для работы в условиях больших динамических объемов данных, полученных из социальных сетей, также, к сожалению, во всех программных средствах вовсе отсутствует возможность оценки влияния агентов социальных отношений. Для решения данной проблемы необходимо создание программного обеспечения, которое бы анализировало информационные потоки и определяло степень влияния распространяемой информации, а также уровень взаимного влияния между агентами социальных сообществ.

Ключевые слова: анализ социальных сетей, визуализация графов, социальный граф, социальные сети, про-граммные средства.

Investigation of software for analysis and visualization of social graphic structures

D. Shynhalov

The **subject matter** of the article is the process of analysis and visualization of social graphs of structures. The **goal** is to explore software tools for analyzing and visualizing social graph structures. The **tasks** to be solved are: to explore modern platforms for constructing social graphs and their statistical analysis, compare their advantages and disadvantages, examine the suitability of platforms for application for large datasets and the expediency of their use for the analysis of social networks. The following **results** are obtained: The most popular platforms on visualization of graph structures are considered. A comparative analysis of these platforms has been conducted from the point of view of the availability of a free license to use, their multiplatform, multiformity, as well as the ability to apply to large social networks and assess the influence of agents, the availability of open source. **Conclusions.** The main advantages of most software are support for many formats, a wide range of possibilities for mathematical and statistical analysis and conditionally free or free license. Among the shortcomings, it should be noted that not all platforms are designed to operate in conditions of high dynamic volumes received from social networks, unfortunately, in all software tools, there is no way to assess the influence of agents of social relations. To solve this problem, it is necessary to create software that analyzes information flows and determines the degree of influence of distributed information, as well as the level of mutual influence between agents of social communities.

Keywords: analysis of social networks, visualization of graphs, social graph, social networks, programmatic means.