УДК 519.174

Т.Г. Білова, І.О. Побіженко

*Харківська державна академія культури, Харків*

# МЕТОД ОЦІНКИ СТУПЕНЮ СТРУКТУРНОЇ БЛИЗЬКОСТІ ЗВ'ЯЗНИХ НЕОРІЄНТОВАНИХ ГРАФІВ

*Розглянуто задачу пошуку подібної структури в термінах теорії графів. Класифіковано основні випадки часткового ізоморфізму двох зв'язаних неорієнтованих графів. Визначена покрокова процедура диференціації вершин графів для пошуку найбільшого ізоморфного підграфу. Розроблено метод оцінки ступеню структурної близькості, що заснований на пошуку найбільшого спільного підграфу та визначенні функцій відстаней між графами.*

***Ключові слова**: зв'язний неорієнтований граф, ізоморфізм, інваріант графа, найбільший спільний підграф, диференціація вершин графу, метричні характеристики.*

## Вступ

**Постановка задачі та аналіз досліджень.** Проблема встановлення зв'язків між структурою об'єкту та його властивостями є однією з основних напрямків прикладної теорії графів, зокрема її методи активно використовуються для моделювання залежностей «структура-властивість» в комп'ютерній хімії, енергетиці, геоінформаційних системах та ін. Серед задач, що вирішуються в рамках математичного моделювання, важливе місце займає проблема пошуку подібних структур.

Можна визначити два основні напрямки визначення ступеню структурної близькості двох зв'язних неорієнтованих графів:

– методи, що зводяться до знаходження максимальної загальної частини графів та розрахунку функцій відстаней [3; 4];

– методи представлення графів у вигляді множин фрагментів малого розміру, близькість визначається як функція числа співпадаючих фрагментів та кількості фрагментів, що відрізняються [2].

Відомі підходи к оцінюванню структур графів обмежуються використанням набору непрямих ознак і не розглядають можливостей вирішення проблеми, що базується на прямих ознаках, які пов'язані з ізоморфізмом. При використанні цих методів найбільш складною є процедура встановлення часткового ізоморфізму, особливо коли в графах, що порівнюються, існує декілька однаково позначених вершин або ребер.

**Метою роботи** є розробка методу оцінки ступеню структурної близькості, що заснований на пошуку спільного підграфу та визначенні функцій відстаней між графами.

В рамках поставленої мети було сформульовано наступні завдання: визначити задачу пошуку подібної структури в термінах теорії графів; класифікувати основні випадки часткового ізоморфізму графів; визначити міру оцінки подоби для двох графів; оцінити ефективність та адекватність розробленого методу.

## Основна частина

**Визначення основних термінів та понять пошуку подібної структури.** Визначення ступеню подоби графових структур – складна та нетривіальна задача, яка актуальна, наприклад, в комп'ютерній хімії. Розглянемо задачу пошуку ідентичної структури в термінах теорії графів.

Зв'язний неорієнтований граф – це упорядкована пара G(V, U), де $V = \{v_1, v_2, \ldots v_n\}$ – скінченна множина вершин, а множина ребер $U \subseteq V \times V$ – бінарне відношення на V, для якого виконуються наступні умови:

1) $\forall a, b \in V\ ((a, b) \in U \Rightarrow (b, a) \in U)$ – симетричність відношення;

2) $\forall a \in V\ ((a, a) \notin U)$ – антірефлексивність відношення.

Графу G, у якого $|V| = n$, ставиться у відповідність матриця суміжності A по наступному правилу:

$$A(i, j) = 1 \Leftrightarrow (v_i, v_j) \in U;\ A(i, j) = 0 \Leftrightarrow (v_i, v_j) \notin U.$$

Два помічених графа називають ізоморфними, якщо між множинами їх вершин існує взаємно однозначне відображення, що зберігає мітки вершин та їх суміжність по однаково поміченим ребрам. Для визначення ізоморфізму графів G та H, що містять по n вершин кожний, необхідно перебрати n! варіантів суміщення вершин. Існують алгоритми, що значно зменшують число кроків процедури порівняння [3–5], але вони потребують доробки для практичного застосування.

Повний ізоморфізм – лише окремий випадок повної схожості структур. Для визначення ступеню структурної близькості доцільно використовувати поняття частини та підграфу графу.

Граф G'(V', U') є частиною (або фрагментом) графу G (G'⊆G), якщо V'⊆V та U'⊆ U, тобто множини його вершин та ребер є підмножинами відповідних множин графа G. Підграф графа G – частина графа, в якому множина ребер U' складається з усіх ребер, що з'єднують у G вершини з V'.

Граф F є спільним (або ізоморфним) підграфом графів G та H, якщо в них існують підграфи G' та H', які ізоморфні графу F. Частковий спільний підграф є таким, що не розширюється (тобто найбільшим або максимальним за включенням), якщо в G та H не існує іншого часткового спільного підграфа, що строго його включає.

Спільний підграф двох графів є найбільшим, якщо цей підграф містить максимальну кількість вершин або ребер серед інших спільних підграфів. Нехай $Q_{G,H}^{V}$ – перетин графів G та H, найбільша по кількості вершин спільна частина, а $Q_{G,H}^{U}$ – перекриття графів G та H, найбільша по кількості ребер частина.

Спільні підграфи можуть бути різних типів. Два графи можуть мати декілька не ізоморфних спільних підграфів, але в силу визначення всі їх перетини мають однакову кількість вершин, а усі перекриття – однакову кількість ребер. При цьому перетини можуть мати різне число ребер, а перекриття – різне число вершин.

Найбільший перетин (перекриття) пари графів може не містити їх найбільший зв'язний перетин (перекриття). Перетин та перекриття пари графів можуть не містити деякий їх спільний підграф.

Оцінки взаємного положення заданих підграфів $G_1$ та $G_2$ в графі G базуються на понятті відстаней між його вершинами [2]. Відстані задаються матрицею $D(G)=\{d_{ij}\}$, кожний елемент якої – довжина найкоротшого простого ланцюгу, що з'єднує дві відповідні вершини.

Відстань від вершини i до множини вершин $V_1 \subseteq V$ графу G – величина мінімальної відстані від i до будь-якої вершини з $V_1$: $d_{iV_1}=\min_j d_{ij},\ j\in V_1$.

Функція $d(G_1,G_2)$ є мінімальною відстанню в графі між множиною вершин підграфів:

$$d(G_1,G_2)=\min_i d_{iV_2}=\min_i \min_j d_{ij},\ i\in V_1,\ j\in V_2.$$

Функція $D(G_1,G_2)$ є сумою двох величин, перша з яких є сумою відстаней от кожної вершини $G_1$ до множини вершин $V_2$, а друга – сумою відстаней від кожної вершини $G_2$ до множини $V_1$:

$$D(G_1,G_2)=\sum_i d_{iV_2}+\sum_j d_{jV_1},\ i\in V_1,\ j\in V_2.$$

Так як $d_{i,V_1\cap V_2}=0$ при $i\in V_1\cap V_2$, функцію $D(G_1,G_2)$ можна представити у вигляді

$$D(G_1,G_2)=\sum_i d_{iV_2}+\sum_j d_{jV_1},\ i\in \bar{V}_1,\ j\in \bar{V}_2,$$
$$\bar{V}_1=V_1\setminus(V_1\cap V_2),\ \bar{V}_2=V_2\setminus(V_1\cap V_2).$$

Для практичного застосування більш ефективною є функція $D'(G_1,G_2)$:

$$D'(G_1,G_2)=\sum_i\sum_j d_{ij}-\sum_k\sum_l d_{kl},$$
$$i\in V_1,\ j\in V_2,\ k,l\in V_1\cap V_2.$$

Тобто пошук подобної структури зводиться до визначення найбільшого ізоморфного підграфу та оцінювання на його основі міри близькості двох графів. При знаходженні спільних підграфів задаються умови еквівалентності міток. Для кількісної величини задаються інтервали її значень, всередині яких усі значення вважаються рівними, для якісної – списки еквівалентних імен.

В [1; 2] досліджуються метричні властивості функцій відстаней між молекулярними графами та за їх допомогою даються оцінки положення підграфів в них. В [5; 6] наводяться ефективні методи визначення графового ізоморфізма, основані на диференціації вершин графів об'єктів різної природи, що значно скорочують кількість кроків процедури порівняння. Ці методи потребують уточнення методів пошуку для їх практичного використання на графах та доповнення їх функціями відстаней, що відповідають метричним властивостям.

**Метод оцінки ступеню структурної близькості графів.** Розглянемо процедуру пошуку подібної структури на множині зв'язних неорієнтованих графів $Q^{(n)}$. Нехай $Q^{\cap}$ – множина їх найбільших спільних підграфів:

$$Q^{\cap}=\left\{Q_{G,H}\ \middle|\ \forall G,H\in Q^{(n)},\ Q_{G,H}=Q_{H,G},\ G\neq H\right\}.$$

Відношення $Q^{\cap}$ має властивості рефлексивності та симетричності, а також транзитивності (якщо $Q_{G,F}=Q_{F,H}$, то $Q_{G,H}=Q_{G,F}=Q_{F,H}=Q_{G,H}$ ). Тобто відношення $Q^{\cap}$ задає на множині $Q^{(n)}$ відношення строгого порядку. Множина $Q^{(n)}$ є частково упорядкованою відносно відношення $Q^{\cap}$. Якщо $\left|Q^{(n)}\right|=n$, то $\left|Q^{\cap}\right|=(n^2-n)/2$.

Множині графів $Q^{(n)}$ відповідає множина вершин $V^Q=\{V_i\}$ та ребер $U^Q=\{U_j\}$, а множині $Q^{\cap}$ їх спільних підграфів – множини вершин та ребер $V^{\cap}=\left\{V_{g,h}\ \middle|\ V_{g,h}\in V_i\right\}$ та $U^{\cap}=\left\{U_{g,h}\ \middle|\ U_{g,h}\in U_j\right\}$.

Розглянемо два випадки часткового ізоморфізму:

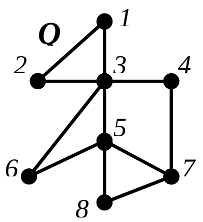
1. $G\subset H,\ Q_{G,H}=G$ – ізоморфне вкладання графа G в граф H.

2. $G \not\subset H, H \not\subset G, Q_{G,H} = G \cap H$ – графи G та H мають найбільший спільний підграф.

Для диференціації вершин графів застосуємо метод, запропонований в [5; 6], що базується на побудові упорядкованих ступеневої та маршрутної характеристик.

S(Q) – ступеневий інваріант графу Q, розраховується для кожної вершини графа як кількість ребер, що інцидентні даній вершині. $\tilde{S}(Q)$ – упорядкована ступенева характеристика, вектор упорядкованих за зростанням ступеню вершин графу.

На рис. 1 наведено приклад побудови упорядкованої ступеневої характеристики для зв'язного неорієнтованого графу Q. На основі $\tilde{S}(Q)$ вершини розбиваються на чотири групи: {1, 2, 4, 6, 8}, {7}, {5} та {3}.




| A(Q) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | S(Q) | $\tilde{S}$(Q) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 |
| 3 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 5 | 4 |
| 4 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 1 |
| 5 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 4 | 3 |
| 6 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 1 |
| 7 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 3 | 2 |
| 8 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 1 |

Рис. 1. Диференціація вершин графу на основі упорядкованої ступеневої характеристики

$M^1(Q)$ – маршрутний інваріант графу, відображує для кожної вершини відповідну упорядковану строку матриці $A^1(Q), a_{ij} \neq 0$. Елементи матриці $A^1(Q)$ характеризують кількість маршрутів довжиною *l* між відповідною парою вершин. $\tilde{M}^1(Q)$ – вектор упорядкованих за зростанням маршрутних характеристик вершин графу.

На рис. 2 наведено приклад побудови упорядкованої маршрутної характеристики другого ступеню для графа Q, зображеного на рис. 1. Згідно з нею вершини графа розбиваються на чотири групи: {7}, {1, 2, 4, 8}, {3} та {5, 6}.

| $A^2$(Q) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | $M^2$(Q) | $\tilde{M}^2$(Q) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 11111 | 2 |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 11111 | 2 |
| 3 | 1 | 1 | 0 | 0 | 1 | 1 | 2 | 1 | 111112 | 3 |
| 4 | 1 | 1 | 0 | 0 | 2 | 1 | 0 | 1 | 11112 | 2 |
| 5 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 1 | 1111112 | 4 |
| 6 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1111111 | 4 |
| 7 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 1112 | 1 |
| 8 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 11111 | 2 |

Рис. 2. Диференціація вершин графу на основі упорядкованої маршрутної характеристики

Для визначення ступеню структурної близькості графів задамо функцію відстаней. Абсолютні відстані залежать від розмірів двох графів та їх найбільших підграфів:

$$r^v_{g,h} = v_g + v_h - 2v_{g,h}, \quad r^u_{g,h} = u_g + u_h - 2u_{g,h}, \quad r_{g,h} = (r^v_{g,h} + r^u_{g,h})/2, \tag{1}$$

де $v_g, v_h, v_{g,h}$ – кількість вершин графів G, H та їх найбільшого спільного підграфу $Q_{G,H}$ відповідно;

$u_g, u_h, u_{g,h}$ – кількість ребер графів G, H та їх найбільшого спільного підграфу $Q_{G,H}$ відповідно.

Відносні (або нормовані) відстані розраховуються за формулами

$$R'^v_{g,h} = \frac{r^v_{g,h}}{v_g + v_h - v_{g,h}}, \quad R'^u_{g,h} = \frac{r^u_{g,h}}{u_g + u_h - u_{g,h}}, \quad R'_{g,h} = (R'^v_{g,h} + R'^u_{g,h})/2. \tag{2}$$

Для практичного застосування функції відстаней (1, 2) повинні задовольняти умовам метрики:

1) $r_{g,h} = 0$ або $R'_{g,h} = 0$ тоді і тільки тоді, коли G=H, тобто два графи ізоморфні;

2) $r_{g,h} = r_{h,g}$ та $R'_{g,h} = R'_{h,g}$ – функції відстаней симетричні;

3) $r_{g,f} + r_{f,h} \geq r_{g,h}$, $R'_{g,f} + R'_{f,h} \geq R'_{g,h}$ – виконується нерівність трикутника.

Якщо функції відстані не задовольняють цім умовам, можливі наступні протиріччя:

1) $r_{g,h} = 0$ або $R'_{g,h} \neq R'_{h,g}$, при цьому G≠H – графи, що не співпадають по структурі, будуть розглядатися як повністю ізоморфні;

2) $r_{g,h} \neq r_{h,g}$ або $R'_{g,h} \neq R'_{h,g}$, то міра подоби двох графів буде залежати від порядку обчислень;

3) $r_{g,f} + r_{f,h} < r_{g,h}$ або $R'_{g,f} + R'_{f,h} < R'_{g,h}$, то оцінка ступеню подоби двох графів буде залежати від присутності якогось третього графу.

Відстань $r^v_{g,h}$ в деяких випадках недоцільно використовувати для оцінки структурної подоби графів. Наприклад, для структурно різних графів G та H, що не містять спільного підграфу $(v_{g,h} = 0)$, відстань $r^v_{g,h}$ менше відстані $r^v_{g,f}$ між графами G та F, що мають спільний підграф $(v_{g,f} > 0)$, у випадку, якщо $v_h < v_f - 2v_{g,f}$. При цьому для нормованих відстаней нерівність $R'^v_{g,h} = 1 > R'^v_{g,f} = 1 - v_{g,f}/s_{g,h}$ виконується. Тому для оцінки структурної подоби графів доцільно використовувати нормовану відстань $R'^v_{g,h}$.

На основі розрахованих нормованих відстаней (2) будується матриця відстаней (структурних відмінностей) між графами $R_{ij} = \{R'_{i,j}\}$. Пошук найбільш структурно близького графу зводиться до пошуку у відповідному рядку мінімального елементу або елементів, для яких функція відстані не перевищує наперед задане граничне значення.

Тобто процедура оцінки ступеню структурної подоби двох графів містить наступні кроки:

$$G \cap H \Rightarrow \tilde{S}(G) \cap \tilde{S}(H),\ \tilde{M}^2(G) \cap \tilde{M}^2(H) \Rightarrow$$
$$\Rightarrow Q_{G,H}\left\{(v_{g,h}, u_{g,h})\right\} \Rightarrow r_{g,h}, R'_{g,h} \Rightarrow R.$$

## Висновки

Запропонований метод оцінки ступеню структурної близькості графів базується на пошуку найбільшого спільного підграфу та поєднує в собі переваги методів диференціації вершин та обчислення функцій відстаней. Адекватність методу підтверджується використанням функцій відстаней, що відповідають метричним властивостям. Ефективність моделі визначається процедурою диференціації вершин графу на основі упорядкованих ступеневих та маршрутних характеристик, що на порядок знижує кількість необхідних обчислень при перебиранні вершин.

Використання запропонованого методу підвищить ефективність обробки та пошуку інформації в графових моделях. Зокрема, в комп'ютерній хімії метод оцінки структурної близькості дозволить виділити спільні структурні фрагменти молекулярних графів, що визначають певні властивості хімічних сполук.

Запропонований метод при оцінці структурної близькості враховує пошук лише найбільшого зв'язного спільного підграфу для двох графів. Розвиток моделі в частині пошуку декількох незв'язаних ізоморфних підграфів з визначенням міри подоби за рахунок аналізу кількох спільних фрагментів та оцінки їх відносного розташування може стати перспективною темою подальших досліджень по цьому напрямку.

## Список літератури


*1. Макаров Л.И. Метрические свойства функций расстояний между молекулярными графами [Текст] / Л.И. Макаров // Журнал структурной химии. – 2007. – Том 48, № 2. – С. 223-229.*

*2. Макаров Л.И. Оценки положения подграфов в молекулярных графах и особенности их общих подграфов [Текст] / Л.И. Макаров // Журнал структурной химии. – 2005. – Том 46, № 4. – С. 759-763.*

*3. Москаленко С.В. Оптимизация алгоритмов идентификации графового изоморфизма [Текст] / С.В. Москаленко, Ю.А. Гатчин // Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики. – 2010. – № 2 (66). – С. 98-103.*

*4. Назаров М.И. Об альтернативном способе задания конечных графов [Текст] / М.И. Назаров // Прикладная дискретная математика. – 2015. – № 3 (29). – С. 37-45.*

*5. Погребной А.В. Метод дифференциации вершин графа и решение проблемы изоморфизма [Текст] / А.В. Погребной, В.К. Погребной // Известия Томского политехнического университета. – 2015. – Том 326, № 6. – С. 37-45.*

*6. Погребной А.В. Полный инвариант графа и алгоритм его вычисления [Текст] / А.В. Погребной // Известия Томского политехнического университета. – 2014. – Том 325, № 5. – С. 110-122.*


*Надійшла до редколегії 15.12.2016*

**Рецензент:** д-р техн. наук, проф. Г.Г. Асєєв, Харківська державна академія культури, Харків.

### МЕТОД ОЦЕНКИ СТЕПЕНИ СТРУКТУРНОЙ БЛИЗОСТИ СВЯЗНЫХ НЕОРИЕНТИРОВАННЫХ ГРАФОВ

Т.Г. Белова, И.А. Побиженко

*Рассмотрена задача поиска подобной структуры в терминах теории графов. Классифицированы основные случаи частичного изоморфизма двух связанных неориентированных графов. Определена пошаговая процедура дифференциации вершин графов для поиска наибольшего изоморфного подграфа. Разработан метод оценки степени структурной близости, основанный на поиске наибольшего общего подграфа и определении функций расстояний между графами.*

***Ключевые слова:** связный неориентированный граф, изоморфизм, инвариант графа, наибольший общий подграф, дифференциация вершин графа, метрические характеристики.*

### METHOD FOR ASSESSING THE DEGREE OF STRUCTURAL SIMILARITY CONNECTED UNDIRECTED GRAPHS

T.G. Belova, I.O. Pobizhenko

*We consider the problem of searching for similar patterns in terms of graph theory. It classifies the main cases of partial isomorphism of two connected undirected graphs. Determined step by step procedure of differentiation of graph vertices to find the highest isomorphic subgraph. A method for assessing the structural similarity based on the search for a common sub-graph and determining the functions of the distance between the graphs.*

***Keywords:** connected undirected graph, isomorphism, graph invariant, the greatest common subgraph, differentiation of vertices, metric characteristics.*