

**Міністерство освіти і науки України**  
**Львівський національний університет імені Івана Франка**

Факультет електроніки та комп'ютерних технологій

**Звіт**

Про виконання лабораторної роботи №9

**3 курсу «Системи опрацювання даних»**

«Дані для побудови графів, параметри графів та візуалізація графів»

Виконав:

Студент групи Фес-21

Осадчук Дмитро

Львів-2025

## Хід роботи:

### Мета роботи

Зобразити дані 1 та дані 2 у вигляді графу та обчислити його параметри

#### Завдання 1

1. Зчитати дані для графу 1 – дані 1
2. Ознайомитись з описом **даних 1**  
<http://www.orgnet.com/divided2.html>
3. Побудувати граф.
4. Визначити параметри графу.
5. За допомогою кольору, товщини, розміру вершин і ребер графа відобразити параметри графу на графіку.
6. Пояснити яку загальну інформацію про покупців книг можемо з нього отримати.



#### Завдання 2

1. Зчитати дані для графу 2 – дані 2
2. Ознайомитись з описом **даних 2**

Ці дані представляли вибір партнера за обіднім столом. Кожна з вершин у наборі даних представляє одну дівчину. Ребра в цій мережі представляють перший і другий вибір дівчат, коли їх запитують, які інші дівчата їм найбільше підходять, як їхні партнери по обіді. Ребра в цих даних мають цінний атрибут, який представляє перший (1) і другий (2) варіанти.

3. Побудувати напрямлений граф для даних 2 фрагмент якого показано на рис



4. Визначити ім'я дівчини, яка на думку дівчат, є найкращою сусідкою за столом.
5. Обчислити та візуалізувати параметри графа

## Хід роботи:

```
import networkx as nx
import matplotlib.pyplot as plt
import pandas as pd
✓ [4] 3s 53ms

edges = pd.read_csv('political-books-edges.csv')
nodes = pd.read_csv('political-books-nodes.csv')

G_books = nx.Graph()

# Додавання вершин
for _, row in nodes.iterrows():
    G_books.add_node(row['Id'], label=row['Label'], ideology=row['political_ideology'])

# Додавання ребер
for _, row in edges.iterrows():
    G_books.add_edge(row['Source'], row['Target'], weight=row['Weight'])

print(f"Кількість вершин: {G_books.number_of_nodes()}")
print(f"Кількість ребер: {G_books.number_of_edges()}")
✓ [7] 94ms

Кількість вершин: 105
Кількість ребер: 441
```

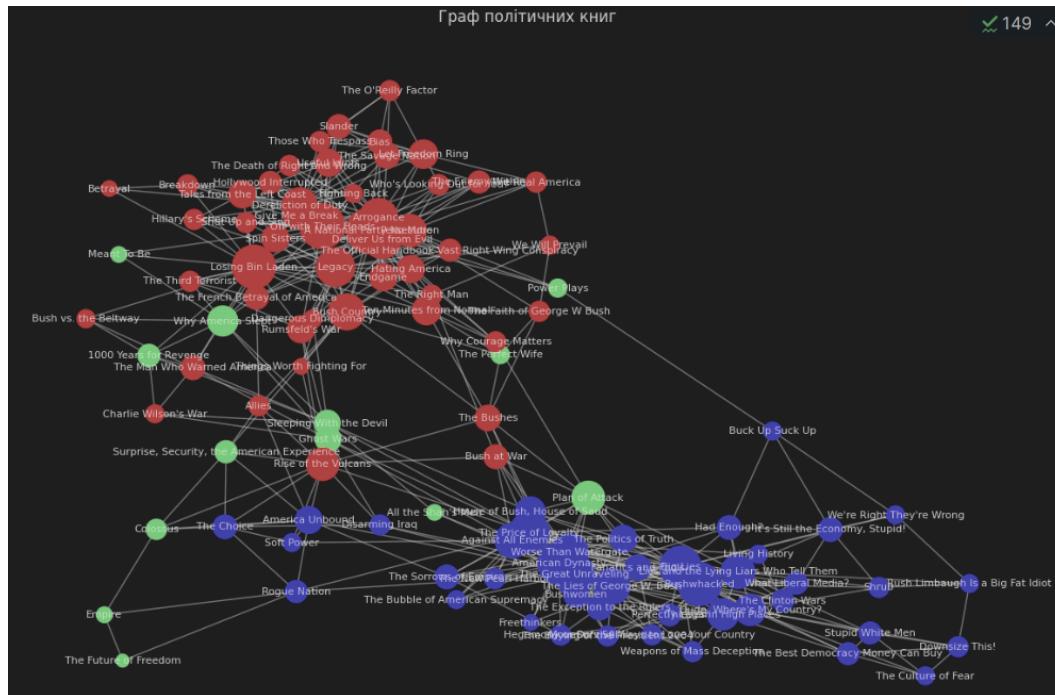
Імпорт потрібних для робіт бібліотек, зчитування CSV файлів. Виведення кількості ребер та вершин графів.

```
color_map = {
    'conservative': 'red',
    'liberal': 'blue',
    'neutral': 'green'
}
colors = [color_map[G_books.nodes[n]['ideology']] for n in G_books.nodes()]

# Визначення розмірів вершин за ступенем
degrees = dict(G_books.degree())
node_sizes = [v * 50 for v in degrees.values()]

plt.figure(figsize=(15, 10))
pos = nx.spring_layout(G_books, seed=42)
nx.draw_networkx_nodes(G_books, pos, node_size=node_sizes, node_color=colors)
nx.draw_networkx_edges(G_books, pos, alpha=0.5)
nx.draw_networkx_labels(G_books, pos, labels={n: G_books.nodes[n]['label'] for n in G_books.nodes()},
    font_size=8)
plt.title("Граф політичних книг")
plt.axis('off')
plt.show()
```

## Візуалізація графу:



Граф показує мережу політичних книг на основі даних про спільні покупки.

- 1) Червоні (консервативні) вузли утворюють щільний кластер - і це вказує на сильні зв'язки всередині групи
- 2) Центральні зв'язки (нейтральні книги) мають високий ступінь центральності - це вказує на їх популярність у мережі. Вузли, які розташовані ближче до центру слугують ланками між ідеологічними групами.
- 3) Сині (ліберальні) вузли більш розсіяні, проте мають багато багато зв'язків всередині групи.

Топ-5 вершин за ступенем центральності:

```
[(8, 0.2403846153846154), (12, 0.2403846153846154), (3, 0.22115384615384617), (84, 0.22115384615384617), (72, 0.21153846153846156)]
```

Топ-5 вершин за betweenness центральністю:

```
[(30, 0.13947827864287202), (49, 0.10364920953531946), (9, 0.09839490722763909), (12, 0.09526168061799924), (72, 0.09093422527408519)]
```

Середній коефіцієнт кластеризації:

```
0.4875267912317313
```

**(8, 0.2403846153846154):** Книга 8 має 25 зв'язків ( $0.2404 \times 104 \approx 25$ ). Це означає, що вона пов'язана з 25 іншими книгами.

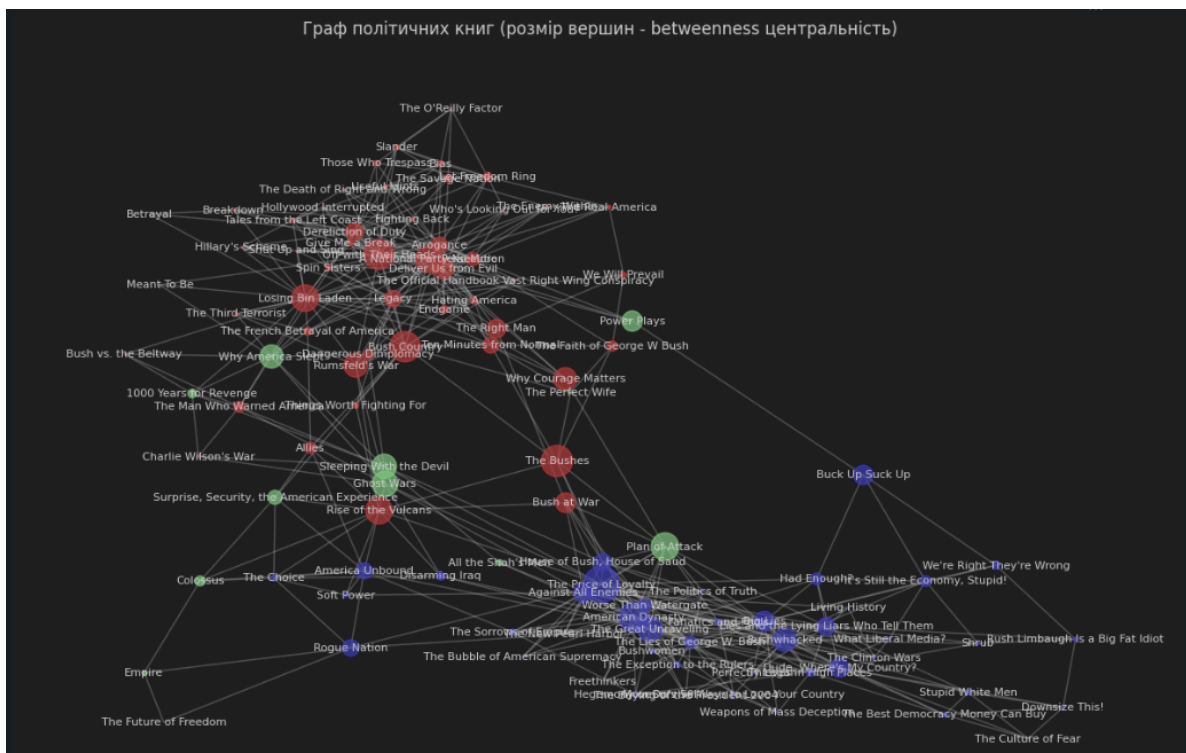
**(12, 0.2403846153846154):** Книга 12 також має 25 зв'язків.

**(3, 0.22115384615384617):** Книга 3 має 23 зв'язки ( $0.2212 \times 104 \approx 23$ ).

**(84, 0.22115384615384617):** Книга 84 має 23 зв'язки.

(72, 0.21153846153846156): Книга 72 має 22 зв'язки ( $0.2115 \times 104 \approx 22$ )

```
1 plt.figure(figsize=(15, 10))
2 node_sizes = [v * 5000 for v in betweenness centrality.values()] node_sizes betweenness_cent
3 nx.draw_networkx_nodes(G_books, pos, node_size=node_sizes, node_color=colors, alpha=0.7)
4 nx.draw_networkx_edges(G_books, pos, alpha=0.3)
5 nx.draw_networkx_labels(G_books, pos, labels={n: G_books.nodes[n]['label'] for n in G_books.nodes()},
6 font_size=8)
7 plt.title("Граф політичних книг (розмір вершин - betweenness центральність)")
8 plt.axis('off')
9 plt.show()
✓ [10] 882ms
```



Betweenness (міжцентральність) — це метрика в аналізі мереж, яка вимірює, наскільки вузол є "мостом" між іншими вузлами графу.

Цей графік показує мережу політичних книг з акцентом на потрібну метрику.

Нейтральні книги (зелені) мають вищу метрику, адже вони є привабливими для двох типів читачів.

Консервативні книги утворюють щільне ядро, що свідчить про згуртовану спільноту читачів із подібними поглядами. Ліберальні книги більш ізольовані, що може вказувати на різноманітність інтересів у цій групі.

Граф показує зв'язки між книгами, які часто купують разом

- Консервативні книги (червоні) утворюють щільне ядро
- Ліберальні книги (сині) мають менше зв'язків між собою

- Книги з високою betweenness центральністю є 'мостами' між різними ідеологічними групами
- Центральні книги можуть бути популярними або базовими для різних ідеологій

```

1 nodes_dining = pd.read_csv('Dining-table_partners [Nodes].csv')
2 edges_dining = pd.read_csv('Dining-table_partners [Edges].csv')
3
4 G_dining = nx.DiGraph()
5
6 # Додавання вершин
7 for _, row in nodes_dining.iterrows():
8     G_dining.add_node(row['Id'], label=row['Label'])
9
10 # Додавання ребер
11 for _, row in edges_dining.iterrows():
12     G_dining.add_edge(row['Source'], row['Target'], weight=row['Weight'])
13
14 print(f"Кількість вершин: {G_dining.number_of_nodes()}")
15 print(f"Кількість ребер: {G_dining.number_of_edges()}")
16
17 ✓ [15] 31ms
18
19 Кількість вершин: 26
20 Кількість ребер: 52

```

```

plt.figure(figsize=(15, 10))
pos = nx.spring_layout(G_dining, seed=42)

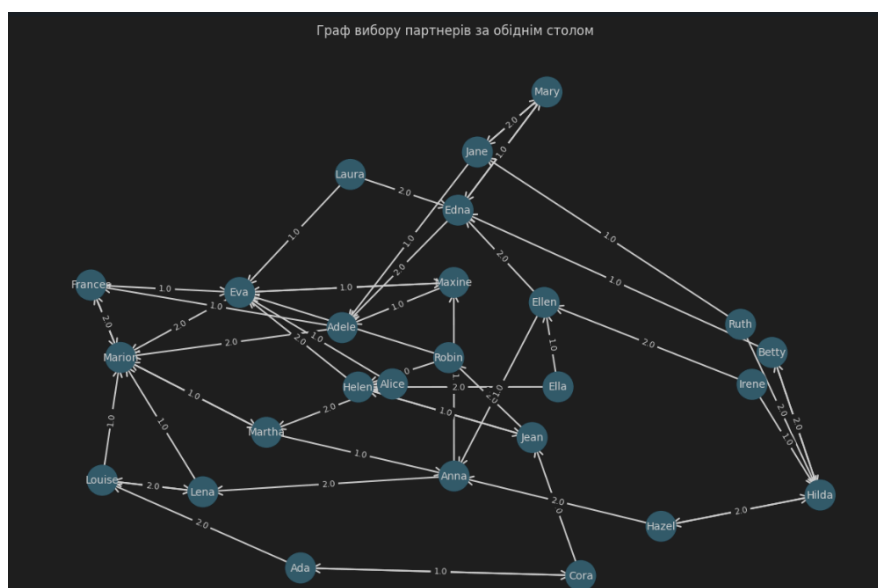
colors = ['skyblue' for _ in G_dining.nodes()]

nx.draw_networkx_nodes(G_dining, pos, node_color=colors, node_size=800)
nx.draw_networkx_edges(G_dining, pos, arrowstyle='->', arrowsize=20, width=1.5)
nx.draw_networkx_labels(G_dining, pos, labels={n: G_dining.nodes[n]['label'] for n in G_dining.nodes()}, font_size=10)

# Додавання ваг ребер
edge_labels = {(u, v): d['weight'] for u, v, d in G_dining.edges(data=True)}
nx.draw_networkx_edge_labels(G_dining, pos, edge_labels=edge_labels, font_size=8)

plt.title("Граф вибору партнерів за обіднім столом")
plt.axis('off')
plt.show()

```



Цей граф показує зв'язки між партнерами за обіднім столом. Кожен вузол - людина, ребра вказують та силу зв'язку між партнерами.

Центральні вузли мають багато зв'язків. Вони розташовані ближче до центру. Решта вузлів це люди з меншою кількістю зв'язків між собою.

Активними учасниками є Анна (має найбільше зв'язків з Maxine, Robin, Alice). Являється центральною людиною в мережі.

Утворені певні кластери людей:

- 1) Центральний (мають багато зв'язків між собою)
- 2) Периферійний (це люди, які мають мало зв'язків із центральною групою)

Можна інтерпретувати як соціальну активність людей.

Топ-5 вершин за PageRank (важливість вузла):

[(9, 0.23285988006162917), (14, 0.1487492905134213), (15, 0.09289623728330257), (6, 0.07501577365999641), (11, 0.05108158085903622)]

Топ-5 вершин за betweenness центральністю:

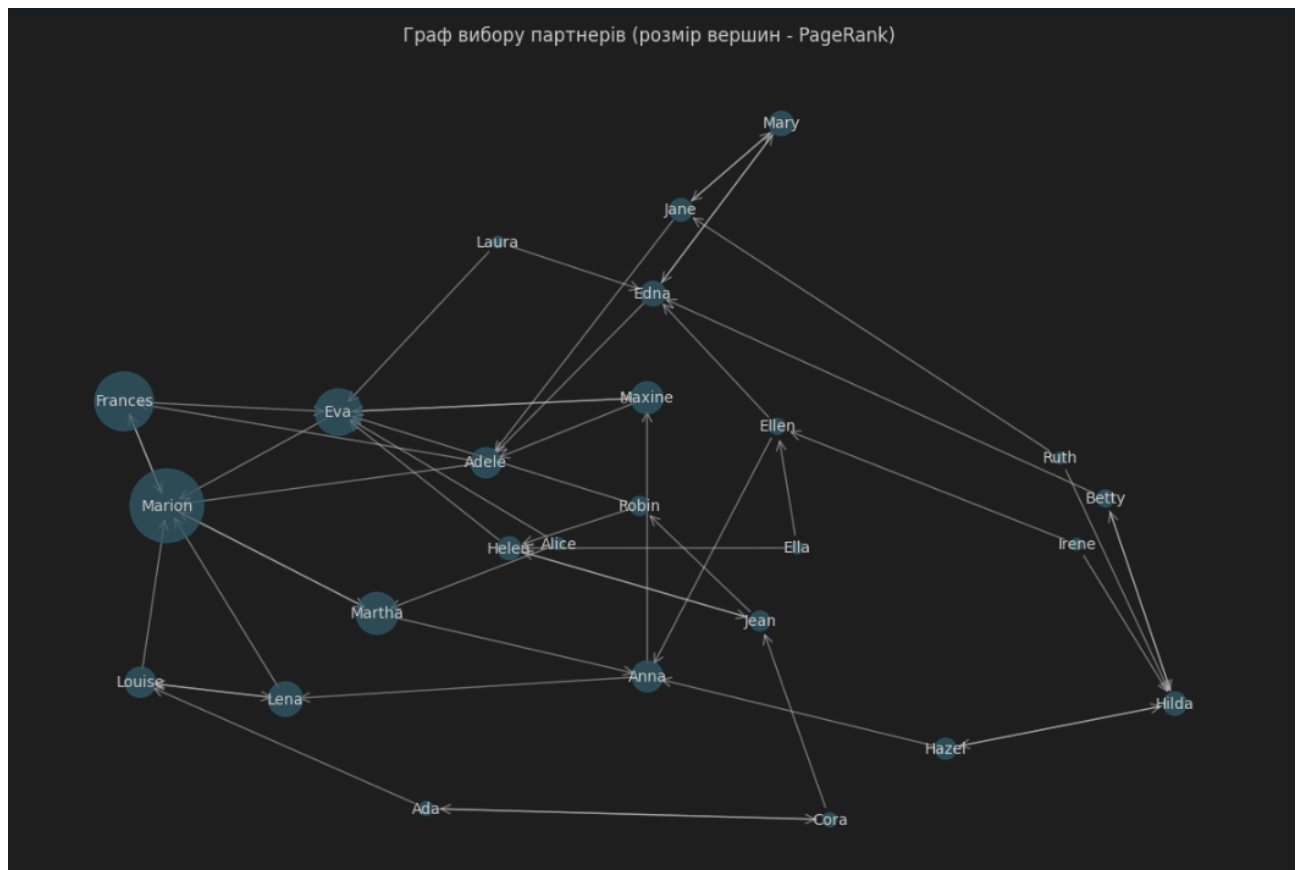
[(9, 0.1330555555555556), (20, 0.10916666666666668), (15, 0.08583333333333334), (6, 0.07916666666666668), (18, 0.07250000000000001)]

Середній коефіцієнт кластеризації:

0.11776556776556776

Hazel і Jean мають високий PageRank, тому що вони пов'язані з центральними особами (наприклад, Anna). PageRank враховує не лише кількість зв'язків, а й те, наскільки важливі їхні сусіди.

Ellen і Hilda, хоча мають менше зв'язків, опосередковано пов'язані з ключовими учасниками, що підвищує їхню важливість.



Аналіз графу вибору партнерів:

- Найпопулярніша дівчина: Marion (отримала 10.0 виборів)
- Вершини з високим PageRank є 'центром уваги' в мережі
- Вершини з високою betweenness центральністю є важливими 'посередниками'
- Кластеризація показує, наскільки друзі дівчат пов'язані між собою

**Висновок:** в ході виконання ЛР-9 я навчився працювати з графами, використовувати їх для різноманітних візуалізацій даних, проводити дослідження на їх основі та інтерпритовувати результат цих досліджень потрібним чином.