

Metody analizy sieci złożonych

Zadanie nr 1 – zapoznanie się z podstawowymi narzędziami do analizy sieci złożonych oraz z podstawowymi miarami sieciowymi.

Zadanie 1

Zadanie numer 1 realizowane jest indywidualnie w trakcie pierwszych zajęć. Średni czas realizacji zadania: 75 minut.

Dane

Z repozytorium Konect <http://konect.cc/networks/> pobierz następujące sieci

- S1. Sieć Les Miserables http://konect.cc/networks/moreno_lesmis/ - Jest to nieskierowana sieć zawierająca współwystępowanie bohaterów powieści Victora Hugo „Nędznicy”. Węzeł reprezentuje postać, a krawędź między dwoma węzłami wskazuje, że te dwie postacie pojawiły się w tym samym rozdziale książki. Waga każdej krawędzi wskazuje, jak często takie współwystępowanie miało miejsce.
- S2. A Song of Ice and Fire <http://konect.cc/networks/asoiaf/> - Jest to fikcyjna sieć społeczna stworzona na podstawie powieści „Pieśń lodu i ognia” George'a R.R. Martina. Krawędź oznacza, że dwie postacie są wymienione w odstępie piętnastu słów; krotności krawędzi oznaczają liczbę takich wystąpień. Zbiór danych obejmuje książki od pierwszej do piątej.

Alternatywnie studenci mogą pobrać dwie dowolne inne sieci z dowolnego repozytorium pod warunkiem, że będą miały podobny rozmiar czyli dla sieci S1, 85+-15 wierzchołków a dla sieci S2 ponad 1000 wierzchołków.

Przykładowe repozytoria zbierające dane o sieciach

1. Konect <http://konect.cc/networks/>
2. Network Repository <https://networkrepository.com/>
3. Stanford Large Network Dataset Collection <http://snap.stanford.edu/data/>
4. The Colorado Index of Complex Networks <https://icon.colorado.edu/#/>

Gephi

Gephi to narzędzie dla naukowców i danologów, którzy chcą analizować sieci złożone. Użytkownik przy użyciu programu Gephi może w łatwy sposób wchodzić w interakcję z siecią, manipulować wizualizacją, reprezentacją wierzchołków i krawędzi, obliczać najważniejsze miary dla sieci, krawędzi i wierzchołka, czy identyfikować grupy [<https://gephi.org/features/>].

Oficjalny przewodnik użytkownika można znaleźć pod adresem <https://gephi.org/users/>

1. Pobierz, zainstaluj i uruchom oprogramowanie Gephi (w wersji 0.10 lub nowszej) <https://gephi.org/> (oprogramowanie jest już zainstalowane na komputerach w laboratorium)
2. Dla każdej z Sieci (S1 i S2) przeprowadź podstawową analizę sieci mierząc (orientacyjnie) i zapisując czas jaki był potrzebny na obliczenie każdego elementu:
 - a. Stopień węzła (degree centrality),

- b. Przewodnictwo (betweenness centrality),
 - c. Bliskość (closeness centrality),
 - d. Współczynnik grupowania (clustering coefficient),
 - e. Pagerank,
 - f. Średnia najkrótsze ścieżka sieci (average shortest path),
 - g. Średnicę sieci (diameter),
 - h. Liczbę komponentów (number of connected components),
 - i. Gęstość sieci (network density)
- 3. Narysuj rozkłady
 - a. Stopni węzła
 - b. Współczynnika grupowania
 - c. Najkrótszych ścieżek
- 4. Zwizualizuj sieć i zbadaj możliwości wizualizacji w narzędziu, wypróbuj minimum 2-3 różne układy (layouts) i zastanów się czym się różnią i co dzięki nim możesz zauważyć.
- 5. Po ukończeniu wszystkich punktów należy zaprosić do swojego stanowiska prowadzącego i skrótkowo zaprezentować zrealizowane prace.

Cytoscape

Cytoscape to oprogramowanie typu open source do wizualizacji sieci biologicznych. Chociaż Cytoscape został pierwotnie zaprojektowany do badań biologicznych, obecnie jest ogólną platformą do kompleksowej analizy i wizualizacji sieci złożonych. Cytoscape zapewnia podstawowy zestaw funkcji do integracji, analizy i wizualizacji danych. Dodatkowe funkcje są dostępne jako aplikacje (Apps), we wcześniejszych wersjach nazywane wtyczkami. Dostępne są aplikacje do analiz profilowania sieciowego i molekularnego, nowych układów, obsługi dodatkowych formatów plików, skryptów i połączenia z bazami danych. Mogą być opracowywane przez każdego, kto korzysta z otwartego interfejsu API Cytoscape opartego na technologii Java. Większość aplikacji jest dostępna bezpłatnie w Cytoscape App Store [https://cytoscape.org/what_is_cytoscape.html].

Oprócz aplikacji desktopowej, Cytoscape udostępnia także silnik do wizualizacji sieci na stronach internetowych cytoscape.js.

Oficjalny przewodnik użytkownika znajduje się pod adresem <http://manual.cytoscape.org/en/stable/>

1. Pobierz, zainstaluj i uruchom oprogramowanie Cytoscape (w wersji 3.9.1 lub nowszej) <https://cytoscape.org/> (oprogramowanie jest już zainstalowane na komputerach w laboratorium)
2. Dla każdej z Sieci (S1 i S2) przeprowadź podstawową analizę sieci mierząc (orientacyjnie) i zapisując czas jaki był potrzebny na obliczenie każdego elementu (UWAGA: obliczenie niektórych miar może wymagać zainstalowania dodatkowych aplikacji (Apps)):
 - a. Stopień węzła (degree centrality),
 - b. Przewodnictwo (betweenness centrality),
 - c. Bliskość (closeness centrality),
 - d. Współczynnik grupowania (clustering coefficient),
 - e. Średnie najkrótsze ścieżki (average shortest paths),
 - f. Średnicę sieci (diameter),
 - g. Liczbę komponentów (number of connected components),
 - h. Gęstość sieci (network density)
3. Narysuj rozkłady

- a. Stopni węzła
 - b. Współczynnika grupowania
 - c. Najkrótszych ścieżek
4. Zwizualizuj sieć i zbadaj możliwości wizualizacji w narzędziu, wypróbuj minimum 2-3 różne układy (layouts) i zastanów się czym się różnią. Należy użyć innych układów niż w przypadku prac z programem Gephi.
5. Po ukończeniu wszystkich punktów należy zaprosić do swojego stanowiska prowadzącego i skrótowo zaprezentować zrealizowane prace.

Dyskusja po zadaniu 1

Po zrealizowaniu zadania należy przygotować się do dyskusji identyfikując minimum trzy zalety i trzy wady dla każdego narzędzia (Gephi i Cytoscape). Czas trwania dyskusji około 15 minut.

[Uwaga dla prowadzącego w tym miejscu należy studentów zachęcić do dyskusji na temat tego co im się podobało a co nie w używaniu tych aplikacji, do czego ich zdaniem mogą się nadać a do czego nie. Na końcu, należy podsumować kluczowe wady i zalety aplikacji, szczególnie w kontekście bibliotek, które będą używane na kolejnych zajęciach]