

Projekt 1

Analiza statystyczna grafu przy użyciu standardowych narzędzi – 20 punktów

Projekt składa się z dwóch zadań.

A) Za pomocą Pajeka należy dokonać analizy grafu „małego”, tj. o rozmiarze umożliwiającym wykorzystanie standardowych algorytmów wizualizacji i grupowania.

B) Za pomocą networkx należy zbadać występowanie w większym grafie wybranych zjawisk charakterystycznych dla sieci złożonych.

W obu zadaniach należy działać na wyodrębnionej największej składowej spójnej grafu nieskierowanego, bez wag, pętli i duplikatów krawędzi. W nawiasach podano punktację za każdy etap zadania.

Zadanie A. Wczytaj w Pajeku sieć o numerze będącym resztą z dzielenia Twojego numeru indeksu przez 7, a następnie

- zbadaj, jaki jest rząd i rozmiar całej sieci, a następnie wyodrębnij największą składową spójną, zbadaj jej rząd i rozmiar (1 pkt);
- wykreśl największą składową spójną i skomentuj wynik (1);
- przeprowadź grupowanie metodą Warda z metryką dI (odległość dwóch węzłów to liczba sąsiadów połączonych tylko z jednym z nich) (1);
- wykreśl dendrogram (1 pkt) i zaproponuj cięcie (1);
- wykreśl wyodrębnione grupy (2).

Zbiory danych do zadania A:

0. [Interakcje pomiędzy pracownikami małej firmy, zanotowane przez obserwatora](#)
1. [Stacje metra londyńskiego i łączące je linie](#)
2. [Deklarowane związki pomiędzy więźniami](#)
3. [Związki pomiędzy naukowcami w określonej dziedzinie](#)
4. [Sieć energetyczna Włoch](#)
5. [Sieć interakcji pomiędzy delfinami](#)
6. [Współwystępowanie bohaterów „Nędzników” w tekście powieści](#)

Zadanie B. Wczytaj w networkx sieć o numerze będącym resztą z dzielenia Twojego numeru indeksu przez 6, a następnie

- zbadaj jaki jest rząd i rozmiar całej sieci: pierwotnej oraz po usunięciu pętli i duplikatów krawędzi (1);
- wyodrębnij największą składową spójną, zbadaj jej rząd i rozmiar (1);
- wyznacz aproksymacje średniej długości ścieżki, operując na próbie losowej 100, 1000 i 10 tys. par wierzchołków (2);
- wyznacz liczbę rdzeni o największym możliwym rzędzie, o drugim możliwie największym rzędzie o trzecim możliwie największym rzędzie; jakie to są rzędy? (3);
- wykreśl rozkład stopni wierzchołków (1);
- wyznacz wykładnik rozkładu potęgowego metodą regresji dla dopełnienia dystrybucji rozkładu stopni, dla przedziałów rozłożonych logarytmicznie (3);
- wyznacz wykres Hilla (3).

Zbiory danych do zadania B:

0. [Sieć spontanicznych skojarzeń słownych](#)
1. [Powiązania użytkowników Twittera](#)
2. [Sieć połączeń lotniczych](#)
3. [Sieć stron www](#)
4. [Sieć współpracy programistów na Githubie](#)
5. [Powiązania użytkowników Wykopu](#)

W sprawozdaniu z projektu (nie *notebooku!*) przedstaw i omów wszystkie otrzymane wyniki oraz uzasadnij podjęte decyzje. Sprawozdanie wraz ze skomentowanym kodem źródłowym obliczeń lub opisem wykonanych operacji prześlij mailem do obu prowadzących **do 25 listopada** (włącznie, tj. do północy – decyduje data przyjęcia maila przez serwer uczelniany). Za każdy dzień opóźnienia naliczana jest kara: **0,5 punktu**. Kara uwzględniana jest przy obliczaniu ostatecznej oceny, ale nie ma ona wpływu na zaliczenie projektu 1.