

Metody analizy sieci złożonych

Zadanie 3 – Ręczne tworzenie własnej sieci i jej analiza.

Zadanie 3

Celem zadania trzeciego jest poznanie możliwości tworzenia różnych sieci złożonych z danych oraz poznanie procesu analizy sieci złożonych w celu (i) definiowania pytań badawczych i (ii) wyciągnięcia wniosków na temat struktury sieci, jej właściwości, cech poszczególnych wierzchołków itd.

Zadanie jest realizowane w grupach dwuosobowych. Średni czas realizacji zadania przypadający na każdego studenta 6 godzin.

Podziału na grupy dokonuje prowadzący w sposób losowy.

Pytanie Badawcze

Do zadania każda grupa może podejść na dwa różne sposoby.

1. Definiujemy pytanie/pytania badawcze np.
 - a. Jak wyglądają moje interakcje ze znajomymi na portalu społecznościowym?
 - b. Jak dostosować strukturę organizacyjną firmy tak by lepiej odzwierciedlała rzeczywiste interakcje i zależności pomiędzy pracownikami firmy?
 - c. Gdzie są wąskie gardła komunikacyjne w naszej organizacji? Które w przypadku braku pracownika mogą poważnie zakłócić przepływ informacji w organizacji.
 - d. Gdzie są wąskie gardła w planie sieci infrastrukturalnej (wodociągi, drogi, komunikacja zbiorowa, itp.), które powodują zatory? Gdzie należałoby stworzyć nowe połączenia by przepływy w sieci były „skuteczniejsze” (np. szybsze poprzez likwidację wąskich gardeł lub skrócenie średniej długości ścieżki lub skrócenie średnicy sieci itd.).
 - e. Kogo należałoby zaszczepić w sieci (np. sieć studentów i prowadzących) mając tylko 5 szczepionek by jak najbardziej ograniczyć rozprzestrzenianie wirusa?
 - f. Kto jest w naszej organizacji największym influencerem?
 - g. itp.

Następnie w celu odpowiedzenia na to pytanie poszukujemy i zbieramy odpowiednie dane.

2. Posiadamy dane (np. przekazane przez klienta lub pozyskane przy okazji innych projektów) i zastanawiamy się jakie interesujące problemy badawcze możemy przy ich pomocy rozwiązać, na jakie interesujące pytania badawcze możemy znaleźć odpowiedź.

Wybór podejścia pozostawiony jest grupie.

Dane i sieć

Dane należy pozyskać samodzielnie, zakazane jest używanie gotowych zbiorów danych. Na podstawie zebranych danych tworzymy sieć bądź sieci. Należy pamiętać, że bardzo często istnieje więcej niż jeden sposób tworzenia sieci z danych. Kluczowe jest tutaj zdefiniowanie co w Waszym przypadku jest węzłem a co relacją, czy sieć ma być ważona czy nie, czy ma być skierowana czy nie, czy ma być wielowarstwowa lub temporalna czy wystarczy prosta statyczna sieć jednowarstwowa?

Przykłady sieci

- sieć interakcji pomiędzy osobami na portalu społecznościowym np. Facebook, Twitter – co określimy jako interakcję: polubienia, komentarze, przyjaźń, obserwowanie, używanie takich samych #tagów, itd?
- Interakcje pomiędzy osobami w grupie/na roku. Co uznamy za interakcję? Może wspólną realizację projektu naukowo-wdrożeniowego, rzeczywistą przyjaźń, bycie w tej samej grupie zajęciowej, czy należy uwzględnić prowadzących czy nie, itd.?
- Sieć interakcji między bohaterami ulubionej książki/filmu. Czy interakcja istnieje gdy dwie osoby spotykają się fizycznie w książce? A może gdy są na tej samej stronie albo w tym samym rozdziale?
- Sieć komunikacji miejskiej we małej miejscowości np. Siechnice.
- Sieć kontaktów w firmie.
- Sieć drogowa w małej miejscowości.

Uwaga: analizowane sieci nie powinny być zbyt duże. Powinniście mieć możliwość sensownego ich zwizualizowania, jak również przeprowadzenia wielu analiz bez nadmiernego narzutu związanego z czasem potrzebnym na obliczenia.

Analiza sieci złożonych

Dla posiadanej sieci przeprowadzamy jej analizę, której zakres dostosowujemy do wiedzy, którą chcemy pozyskać i do pytań, które zostały zdefiniowane. W tym celu można wykorzystać dowolne narzędzia, miary, modele i techniki poznane do tej pory.

Przykładowe analizy, które można wykonać:

- Podstawowa analiza sieci – rozkład stopni wężła, rozkład najkrótszych ścieżek, rozkład współczynnika grupowania, średnica sieci, liczba komponentów, kluczowe węzły w sieci z punktu widzenia różnych miar, do jakiego modelu sieci jest zbliżona nasza sieć (losowa, bezskalowa itd.).
- Analiza grup w sieci, jakie grupy występują w naszej sieci, ile ich jest, jak są połączone, jakiej są wielkości, z czego może wynikać dany podział, jak wygląda podział na grupy z punktu widzenia różnych algorytmów do ekstrakcji grup.
- Analiza rozprzestrzeniania w sieci – jak rozprzestrzenia się informacja, wpływ, wirus w Waszej sieci w zależności od różnych warunków np. jak wygląda proces gdy zaczniemy od wierzchołka X, Y, Z, jak wygląda proces gdy „zaszczepimy” węzły A, B, C. Kogo musimy, zaszczerpić/odizolować by w przypadku wybuchu epidemii w jednej grupie społecznej nie przedostała się ona do innych, co jeżeli mamy ograniczoną liczbę szczepionek albo szczepionki przychodzą partiami i danego dnia możemy zaszczerpić ograniczoną liczbę osób itp.
- Analiza wizualna sieci, różne layouty, różne sposoby kolorowania, kształtu, wielkości wierzchołków w zależności do wartości poszczególnych miar, w zależności od podziału na grupy, wizualizacja tylko wycinka sieci (np. jednej grupy) itd.

Prezentacja wyników na kolejnych zajęciach

Po zrealizowaniu zadania należy przygotować się do prezentacji wyników, forma prezentacji dowolna (prezentacja multimedialna, notatnik python-owy/R-owy, raport, itp). W czasie prezentacji należy przedstawić:

- problem/pytania badawcze, motywację do podjęcia takiego a nie innego tematu,
- zbiór danych, sposób pozyskania danych, sposób ekstrakcji sieci z danych,
- sieć (sieci) i jej charakterystyki, właściwości,
- wyniki przeprowadzonej analizy i wyciągnięte wnioski.

Prezentację wraz z odpowiedzią na ewentualne uwagi i pytania, które mogą się pojawić w trakcie i po prezentacji na które nie zdążycie odpowiedzieć należy przestać prowadzącemu w dniu prezentacji.

Czas prezentacji na grupę 10 minut (8 minut prezentacja + 2 minuty pytania).