

Metody analizy sieci złożonych

Zadanie nr 5 – praktyczna analiza sieci złożonych

1 Zadanie 5

Celem zadania czwartego jest wykorzystanie analizy sieci złożonych w rzeczywistych projektach o charakterze naukowo wdrożeniowym.

Zadanie jest realizowane w grupach 3-4 osobowych. Czas realizacji zadania przypadający na każdego studenta w grupie: 18 godzin. Wielkość grupy powinna korespondować z trudnością projektu. Studenci dobierają się w grupy samodzielnie. Alternatywnie, podziału na grupy dokonuje prowadzący.

W ramach realizacji zadania można wykorzystać inne techniki poznane w trakcie innych przedmiotów, należy jednak pamiętać, że oceniane w głównej mierze będzie wykorzystanie **technik analizy sieci złożonych**.

1.1 Pytanie Badawcze

Do zadania każda grupa może podejść na dwa różne sposoby. Wybór podejścia pozostawiony jest grupie.

1. Definiujemy pytanie/pytania badawcze, a następnie w celu odpowiedzenia na to pytanie poszukujemy i zbieramy odpowiednie dane:
 - (a) Jak wyglądają moje interakcje na portalu społecznościowym? Analiza mediów społecznościowych Facebook, Twitter, Instagram - grupy ludzi, rozprzestrzenianie informacji np. #coronavirus.
 - (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 aby przepływy w sieci były „skuteczniejsze” (np. szybsze poprzez likwidację wąskich gardeł lub skrócenie średniej długości ścieżki lub skrócenie średnicy sieci itd.).
 - (e) Kto jest w naszej organizacji największym influencerem?

- (f) Jak wygląda współpraca pomiędzy naukowcami (np. na W04N lub PWr), analiza sieci cytowań/współautorstwa pracowników PWr - kto z kim współpracuje, kto prowadzi badania interdyscyplinarne, czy są grupy wzajemnie się cytujących osób. (Google scholar, ORCID, <https://dona.pwr.edu.pl/szukaj/>)
 - (g) Jak zrobić sobie kanapkę z tego co mam w lodówce - Analiza składników współwystępujących w przepisach
 - (h) Jak rozprzestrzeniania się (korona)wirus
 - i. Rzeczywisty przepływ, propagacja wirusa, analiza środków prewencyjnych, sugerowanie środków prewencyjnych,
 - ii. Analiza mediów społecznościowych pod kątem #tagów, propagacji informacji
 - iii. Kogo należałoby zaszczepić w sieci (np. sieć studentów i prowadzących) mając tylko 5 szczepionek by jak najbardziej ograniczyć rozprzestrzenianie wirusa?
 - (i) Jak balansować zapelnieniem stacji roweru miejskiego – analiza przepływu rowerów miejskich, hulajnóg itp.
 - (j) Jak wykryć nielegalne transakcje w sieci kryptowalut - Analiza sieci blockchain, struktura, transakcje, podejrzone transakcje, przepływy w kontekście tego co się dzieje na świecie <https://dru.readthedocs.io/en/master/>
 - (k) Jak wykryć nepotyzm – analiza sieci powiązań pomiędzy posłami, firmami, stowarzyszeniami itd.
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.

1.2 Dane i sieć

Dane można pozyskać samodzielnie, można też korzystać z gotowych zbiorów danych lub agregować istniejące zbiory danych i dane własne. 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 dużej miejscowości np. Wrocław.
- Sieć kontaktów w firmie.
- Sieć drogowa w dużej miejscowości.

1.3 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 węzła, 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, zaszczyć/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 zaszczyć ograniczoną liczbę osób itp.
- Analiza wizualna sieci, różne layouty, różne sposoby kolorowania, kształtu, wielkości wierzchołków w zależności od wartości poszczególnych miar, w zależności od podziału na grupy, wizualizacja tylko wycinka sieci (np. jednej grupy) itd.

2 Etapy

Na początku każdych zajęć każda grupa prezentuje postępy zgodnie z danym etapem. Następnie na forum całej grupy lub indywidualnie z prowadzącym prowadzona jest dyskusja na temat zaprezentowanych postępów, przedstawionych pomysłów, czy też ewentualnych problemów. Poniżej plan prac:

Lab. 5.	Wprowadzenie do zadania i podział na grupy.
Lab. 6.	Definiowanie problemu (biznesowego, naukowego, społecznego) do którego można wykorzystać analizę sieci oraz identyfikacja źródeł danych sieciowych umożliwiających tą analizę. Zbieranie danych sieciowych, budowa sieci, podstawowa analiza sieci.
Lab. 7.	Pogłębiona analiza sieci, definiowanie systemu/warstwy prezentacji przedstawiającej odpowiedź na zdefiniowany wcześniej problem. Dalsza analiza sieci, prace nad prototypem systemu/warstwą prezentacji i prezentacją. (Ewentualne rozpoczęcie prezentacji z zadania na drugiej połowie laboratorium).
Lab. 8.	Prezentacja opracowanego rozwiązania, odkryć i wniosków z analiz.

3 Prezentacja wyników

Po zrealizowaniu zadania należy przygotować się do prezentacji wyników:

- 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,
- wyniki przeprowadzonej analizy i wyciągnięte wnioski,
- prototyp systemu/warstwę prezentacji.

Czas prezentacji będzie dostosowany do liczby grup, ale nie powinien być dłuższy niż 15 minut. **Prezentacje należy przygotować w postaci plakatu konferencyjnego formatu A0.** Po przedstawieniu postera (w formie elektronicznej - nie trzeba drukować!) należy zamieścić na e-portalu jako plik PDF. Oceniane elementy:

1. Sposób wykorzystania narzędzi/technik analizy sieci złożonych: $[0 - 1, 5]$,
2. Uzyskane wyniki – analizy, obserwacje, wnioski: $[0 - 1]$,
3. Jakość prezentacji i plakatu: $[0 - 0, 5]$,
4. Pomysł na wykorzystanie wyników - prototyp systemu/warstwa prezentacji: $[0 - 0, 5]$.