

Nowoczesne narzędzia analizy sieci złożonych

W dzisiejszych czasach zarówno naukowcy, jak i pracownicy przedsiębiorstw różnych branż mierzą się z przetwarzaniem ogromnych ilości danych. Często są to dane, które naturalnie formują sieć (tzw. sieć danych rzeczywistych). Można tu wymienić duże zbiory danych pochodzące z sieci społecznościowych (takich jak Facebook, Twitter, czy Pinterest), sieci biologicznych (np. sieci połączeń białkowych, czy połączeń neuronów w mózgu), sieci telekomunikacyjnych, sieci energetycznych, sieci transportowych (np. połączeń lotniczych), itp.

Modelami matematycznymi reprezentującymi dane o takim charakterze są duże grafy, tzw. sieci złożone (ang. complex networks) - struktury składające się z wierzchołków i krawędzi. W zależności od tego, co reprezentuje sieć, wierzchołki i krawędzie mogą mieć różne znaczenie, np. (i) w mediach społecznościowych wierzchołkami są osoby/podmioty w nich zarejestrowane, a krawędź pomiędzy dwoma wierzchołkami symbolizuje relację znajomości lub obserwowania się, (ii) w sieciach biologicznych wierzchołki odpowiadają molekułom białkowym, a krawędzie różnego rodzaju odpowiadają poszczególnym typom interakcji pomiędzy molekułami. Dynamiczna natura danych w sieciach rzeczywistych (np. przyływy i odpływy użytkowników w sieciach społecznościowych, czy pojawiające się i znikające połączenia pomiędzy neuronami) jest odzwierciedlona w modelach grafowych poprzez dodawanie lub usuwanie wierzchołków i krawędzi.

Duże wielkości danych rzeczywistych oraz ich dynamiczna natura stanowią ogromne wyzwanie zarówno dla analityków, jak i algorytmików. Zgromadzenie takich danych na jednym nośniku i poddanie ich statycznej analizie jest często niemożliwe. Podobnie jak niemożliwym jest zastosowanie do takich danych standardowych algorytmów, gdyż wymagałyby one wykorzystania całkowitej wiedzy na temat badanej sieci. Ze względu na ograniczenia pamięciowe nie sprawdzają się także metodologie typu 'online' zakładające, że strukturę sieci poznajemy z czasem. Niezbędne okazuje się tu stosowanie niestandardowych technik przetwarzania i analizowania danych, np. metod wywodzących się z tzw. analityki Big Data. Przykładem może być tworzenie szkicu danych (ang. data sketch) - małego zbioru pozwalającego odtwarzać najistotniejsze informacje o sieci rzeczywistej.

Celem projektu jest konstruowanie nowych oraz udoskonalanie istniejących narzędzi pracy z dużymi sieciami danych rzeczywistych. Problemy badawcze zostały podzielone na trzy grupy.

1. Konstrukcja matematycznych modeli dobrze opisujących współczesne sieci rzeczywiste.
2. Identyfikacja i analiza istotnych z punktu widzenia algorytmiki parametrów sieci złożonych.
3. Konstrukcja efektywnych algorytmów pozwalających oszacować wybrane parametry sieci rzeczywistych z wykorzystaniem technik analityki Big Data.

Trafnie skonstruowane modele matematyczne pozwolą lepiej zrozumieć zjawiska zachodzące w otaczających nas systemach, a także dobrze przewidywać ich przyszłe zachowania. Posiadanie metod precyzyjnie szacujących wybrane parametry sieci rzeczywistych da możliwość konstruowania algorytmów odpowiednio dostosowanych do danych, zatem np. ulepszać systemy rekomendacyjne, celnie identyfikować użytkowników o podobnych profilach, skuteczniej filtrować spam, czy wykrywać funkcjonujące w sieciach uciążliwe boty oraz oszustów.