

Procesy rozprzestrzeniania oraz maksymalizacja wpływu w sieciach wielowarstwowych

Streszczenie

mgr inż. Michał Jerzy Czuba

Sieci stanowią przystępną metodę reprezentacji wielu systemów złożonych. Węzły społeczne, interakcje pomiędzy białkami, szlaki transportowe czy też systemy zaopatrzenia w wodę — wszystkie one mogą być modelowane przy użyciu formalizmów bazujących na grafach. To właśnie między innymi dlatego w ostatnich latach opracowano liczne metody modelowania tego typu struktur. W szczególności, dużą popularność zyskały sieci wielowarstwowe, które umożliwiają jednoczesną reprezentację wielu typów relacji. Koncepcyjnie, sieć wielowarstwową można postrzegać jako zbiór grafów, z których każdy rozpięty jest na tym samym zbiorze węzłów (lub jego podzbiorze), a same węzły stanowią reprezentację aktorów systemu względem danej relacji. Taka struktura może zostać uzyskana np. poprzez agregację użytkowników z kilku portali społecznościowych wraz z relacjami znajomości pomiędzy nimi.

Jednym z najważniejszych aspektów sieci jest ich dynamika. Może ona przyjmować różne formy, obejmujące zarówno zmiany strukturalne, jak i fluktuacje stanów aktorów budujących ją. Ten drugi rodzaj zjawisk badany jest zazwyczaj w ramach procesów rozprzestrzeniania i stanowi podstawę wielu obszarów badań specyficznych dla danych dziedziny, jak np.: modelowanie epidemii, dynamika opinii czy dyfuzja informacji. Kluczowym zagadnieniem w tym kontekście jest znalezienie sposobu na możliwie najefektywniejszą propagację danego zjawiska w sieci (w wymiarze praktycznym może być to reklama, opinia, wirus itp.). Problem ten tradycyjnie rozpatrywany jest jako zadanie maksymalizacji wpływu. W swojej kanonicznej postaci polega ono na identyfikacji takiego zbioru agentów, który jeśli użyty do zainicjowania propagacji, zapewni jej maksymalny zasięg. Z biegiem lat problem ten bywał jednak wielokrotnie przeformułowywany, a do jego rozwiązania wykorzystywano różne paradygmaty.

Celem niniejszej rozprawy jest ujęcie problemu modelowania zjawisk rozprzestrzeniania w sieciach wielowarstwowych oraz analiza różnych podejść do rozwiązywania zadania maksymalizacji wpływu w tych strukturach. Z uwagi na to, że praca została zrealizowana w dyscyplinie Informatyki Technicznej i Telekomunikacji, jej równoległym wynikiem są opublikowane zbiory danych, algorytmy, metody, ich implementacje oraz środowiska symulacyjne. W ramach rozprawy zidentyfikowano siedem luk badawczych, a te następnie przekształcono w odpowiadające im cele oraz prerekwizyty, których uprzednie wykonanie pozwala na realizację celów. Sam cel główny natomiast sformułowano jako *opracowanie i ewaluację podejść do maksymalizacji wpływu w sieciach wielowarstwowych poprzez wykorzystanie wybranych paradygmatów dostarczających komplementarne perspektywy na systemy złożone: pojęcia centralności w sieci, kontroli/sterowania w sieciach, uczenia maszynowego na grafach oraz strukturalnych interwencji w analizowanym systemie*. Przedstawione badania realizują ten cel poprzez zbiór siedmiu tematycznie powiązanych publikacji wraz ze wstępem merytorycznym, ukazującym wkład każdej z nich w rozpatrywany problem.

Pierwszy prerekwizyt: *identyfikacja metod modelowania zjawisk rozprzestrzeniania w sieciach wielowarstwowych oraz wybór jednego mechanizmu, w oparciu o który zrealizowane zostaną wszystkie cele badawcze*, osiągnięto poprzez opracowanie środowiska symulacyjnego (*network-diffusion*), które pozwala na reprezentowanie dowolnej konfiguracji współlistniejących zjawisk za pomocą n -wymiarowej siatki stanów. Ponadto, niedawno zaproponowane podejście do modelowania dyfuzji w sieci z perspektywy aktora (tj. funkcje protokołów określające sposób agregacji aktywacji uzyskanych na poziomie poszczególnych warstw) zostało zaadaptowane do tematu pracy i rozszerzone z modelu proggu liniowego na model niezależnych kaskad.

Drugie z wymagań wstępnych — *opracowanie miar opisujących efektywność rozprzestrzeniania w sieciach wielowarstwowych, umożliwiających porównanie dyfuzji inicjowanych w różnych*

warunkach — doprowadziło (zgodnie ze swym brzmieniem) do zaproponowania zestawu nowych metryk pozwalających na porównywanie dyfuzji zarówno z perspektywy statycznej, jak i dynamicznej. Opublikowano również autorską metodę ewaluacji pozwalającą abstrahować od konkretnych warunków rozprzestrzeniania. Ponadto sformułowano i przebadano metodę estymacji zdolności do propagacji pojedynczych aktorów.

Trzeci prerekwizyt określono jako: *pozyskanie wielorelacyjnych danych mogących stanowić strukturalne zaplecze eksperymentów nad maksymalizacją wpływu oraz opracowanie metod umożliwiających ich uzupełnianie w czasie rzeczywistym*. Został on zrealizowany poprzez opublikowanie nowej sieci opisującej fizyczne interakcje w izolowanej grupie osób. Reprezentatywny zbiór danych wykorzystanych we wszystkich symulacjach zapewnił natomiast spójność badań. Ponadto, zaproponowano nowy syntetyczny generator sieci wielowarstwowych o rozkładzie potęgowym, mABCD, wyróżniający się wysoką wydajnością oraz zdolnością do modelowania złożonych zależności pomiędzy warstwami.

Po zrealizowaniu prerekwizytów przeprowadzono prace ukierunkowane na osiągnięcie głównego celu rozprawy. W ramach *ewaluacji metod rankingowych w zadaniu maksymalizacji wpływu w sieciach wielowarstwowych oraz ocenie różnic w ich skuteczności* wykonano kompleksowe badanie najbardziej podstawowego podejścia z rozpatrywanych. Oceniono szesnaście heurystyk, w tym trzy rozszerzone do scenariusza wielowarstwowego na bazie centralności pierwotnie zdefiniowanych dla sieci jednowarstwowych. Zaproponowano również nową metodę: *Neighbourhood Size Discount*. Główny wniosek wskazuje, że wszystkie analizowane parametry, zarówno te dotyczące modelu rozprzestrzeniania, typu sieci, wielkości zbioru początkowego oraz metody jego wyboru, łącznie determinują ogólną efektywność maksymalizacji wpływu.

Kolejny cel stanowiło *zaproponowanie metody wyboru aktorów początkowych wykorzystującej minimalny zbiór dominujący w zadaniu maksymalizacji wpływu w sieciach wielowarstwowych oraz weryfikację jej skuteczności*. Realizacja tego celu doprowadziła do rozszerzenia pojęcia minimalnego zbioru dominującego — zbioru agentów umożliwiających kontrolę nad całym systemem — na sieci wielowarstwowe. Następnie, z wykorzystaniem najskuteczniejszych metod rankingowych, sprawdzono czy inicjowanie rozprzestrzeniania z poziomu zbioru dominującego zwiększa jego zasięg. Wyniki wykazały znaczną, lecz ograniczoną do specyficznych warunków, przewagę tego podejścia.

Badania w obszarze uczenia maszynowego miały na celu *opracowanie indukcyjnej grafowej sieci neuronowej do predykcji najefektywniejszych inicjatorów dyfuzji w danej sieci wielowarstwowej oraz walidację jej działania w porównaniu z konkurencyjnymi podejściami*. W tym celu utworzono zbiór danych ponad dwustu sieci wielowarstwowych z etykietami określającymi potencjał rozprzestrzeniania każdego aktora. Następnie opracowano grafową sieć neuronową *TopSpreadersNetwork*, zdolną do przewidywania tych wartości. Kluczowym założeniem projektowym było umożliwienie wnioskowania na podstawie wcześniej niewidzianych danych, co stanowi rzadkość w obszarze rozpatrywanego problemu. Badania porównawcze wykazały zdolność opracowanej metody do generalizacji oraz przewagę nad rozwiązaniami konkurencyjnymi.

Ostatnie zadanie polegało na *zaprojektowaniu metody manipulowania strukturą sieci wielowarstwowej w skali makro, umożliwiającej porównanie efektywności rozprzestrzeniania pomiędzy siecią oryginalną i zmodyfikowaną oraz identyfikację własności strukturalnych sprzyjających maksymalizacji wpływu*. W tym celu przeprowadzono studium koncepcyjne analizując wpływ parametrów modelu mABCD na możliwości propagacji w sieci. Opracowano również metodę osadzania rzeczywistej sieci w przestrzeni konfiguracji mABCD, co może mieć szczególną wartość w kontekście tzw. *cyfrowych bliźniaków* systemów złożonych. Uzyskane wyniki wskazują między innymi na silny wpływ poziomu szumu międzyspołecznościowego na efektywność rozprzestrzeniania.

Podsumowując, rozprawa przedstawia cztery perspektywy dotyczące problemu maksymalizacji wpływu w sieciach wielowarstwowych. Odnosi się również do szeregu prerekwizytów (takich jak modelowanie zjawisk rozprzestrzeniania) których rozwiązanie umożliwiło przeprowadzenie właściwych badań. Uzyskane rezultaty pokazują, że problem w analizowanym typie sieci cechuje się większą złożonością niż jego kanoniczna postać. Wkład pracy obejmuje rozwój badań nad sieciami wielowarstwowymi, w tym nowe miary oceny rozprzestrzeniania, centralności, podejścia do modelowania procesów rozprzestrzeniania, metody identyfikacji najbardziej wpływowych aktorów, zbiory danych oraz syntetyczny generator grafów. Nadrzędny cel został osiągnięty, choć z zastrzeżeniami wynikającymi z ograniczeń zaproponowanych metod.