



UNIwersytet
Warszawski

Biuro Prasowe

22.03.2017

ROZCHODZENIE SIĘ PLOTEK

Zespół algorytmików z Wydziału Matematyki, Informatyki i Mechaniki UW stara się opisać zjawiska zachodzące na Twitterze: jaka jest szansa, że plotka opublikowana przez jedną osobę trafi do ponad tysiąca innych, w jakim tempie plotka będzie się rozchodzić, jakie cechy musi mieć sieć społecznościowa, żeby plotka rozchodziła się w taki sam sposób jak na Twitterze?

Odpowiedzi na te pytania znalazły się w pracy "Why Do Cascade Sizes Follow a Power-Law?", autorstwa prof. Piotra Sankowskiego oraz jego zespołu, która została przyjęta na konferencję International World Wide Web Conference (odbędzie się 3-7 kwietnia 2017 w Perth w Australii).

Liczba osób, do których trafi informacja, może być szacowana przy użyciu prawdopodobieństwa. Szansa na to, że plotka dotrze do wielu osób jest w rzeczywistości bardzo niewielka. Z danych empirycznych wynika, że jedynie 0.1 promila wszystkich plotek na Twitterze dociera do ponad 2500 osób.

Dotychczasowe modele, którymi naukowcy zajmują się od ponad 10 lat, nie opisywały dobrze stanu faktycznego.

W pracy "Why Do Cascade Sizes Follow a Power-Law?" zaproponowano nowy model, w którym rozkład plotek jest linią prostą (zbliżoną do prawdziwych danych), czyli szansę na to, że plotka dotrze do określonej liczby osób, opisuje tzw. „rozkład potęgowy”.

W pracy zaproponowano nowy sposób generowania kaskad, taki że rozchodzenie się w nich plotek ma takie same cechy (rozkład prawdopodobieństwa) co rozchodzenie się plotek na Twitterze. Jest to zupełnie nowy sposób weryfikowania tego typu modeli. Do tej pory należało przeprowadzić ogromną liczbę symulacji i porównać ich wyniki z prawdziwymi danymi. Korzystając z nowej metody, poprawność modelu może być sprawdzana szybciej przy użyciu mniejszej ilości danych.

Jednym z zastosowań wyników pracy "Why Do Cascade Sizes Follow a Power-Law?" jest testowanie wydajności. Do tej pory, aby sprawdzić czy duży system informatyczny przetrwa wysokie obciążenie użytkowników, trzeba było przeprowadzić symulacje przy użyciu ogromnej ilości danych. Teraz można wykonać taki test mniejszym kosztem (energii, czasu, danych). – Nasz model pozwoli tanio sprawdzić czy narzędzia wytrzymają obciążenie – komentują autorzy pracy.