



Politechnika  
Wrocławska

Adam Piróg | Mateusz Stolarski | dr hab. inż. Piotr Bródka

# Znajdź swojego influencera

Czyli jak szybko zidentyfikować kluczową osobę w sieci

## Problem badawczy

Identyfikacja wpływowych węzłów jest jednym z kluczowych zadań nauki o sieciach, mających jednocześnie praktyczne zastosowanie w wielu problemach biznesowych.



Chcąc przeprowadzić efektywną kampanie reklamową i skutecznie rozpowszechnić informację o naszym produkcie, trzeba zdecydować kogo zatrudnić do współpracy reklamowej, aby uzyskać największe zasięgi.

Rozwiązania aktualnie stosowane w tym obszarze opierają się na wielokrotnych symulacjach, co gwarantuje wysoką dokładność, ale jest bardzo kosztowne obliczeniowo.

W ramach naszej pracy wykorzystujemy techniki uczenia maszynowego, do stworzenia systemu oceniającego daną osobę w sieci pod trzema kluczowymi względami:

- Całkowitego zasięgu, jaki jest w stanie wygenerować - **całkowity zasięg reklamy**
- Liczby innych użytkowników, których zachęci do aktywnego reklamowania produktu - **maksymalna liczba reklamujących**
- Szybkości, z jaką rozprzestrzeni informację o produkcie - **szybkość rozprzestrzeniania**

## Metodologia

### Zbiór danych

Sieci społecznościowe z serwisów: <sup>1</sup>

- Twitch (DE) - 9 tys. wierzchołków
- Facebook - 22 tys. wierzchołków
- Github - 37 tys. wierzchołków

Każdy z badanych grafów został podzielony na zbiór uczący (80% grafu) oraz testowy (20% grafu).

### Model symulacji

Model niezależnych kaskad - w danej iteracji aktywny wierzchołek ma dokładnie jedną szansę by aktywować każdego ze swoich sąsiadów z prawdopodobieństwem *beta*.

### Modele sztucznej inteligencji

- Support Vector Machine
- K-Nearest Neighbors
- Random Forest

### Dane wejściowe modelu

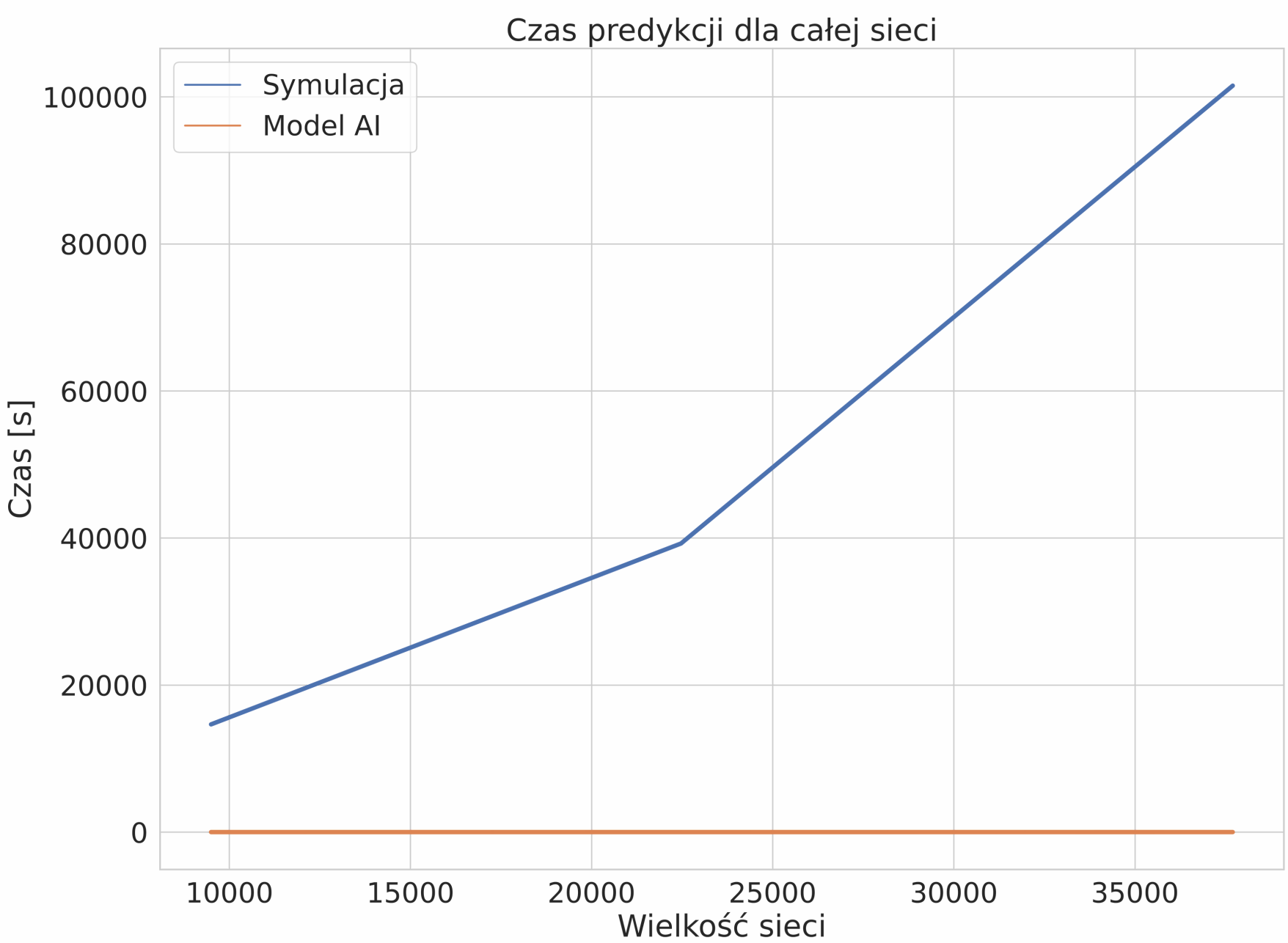
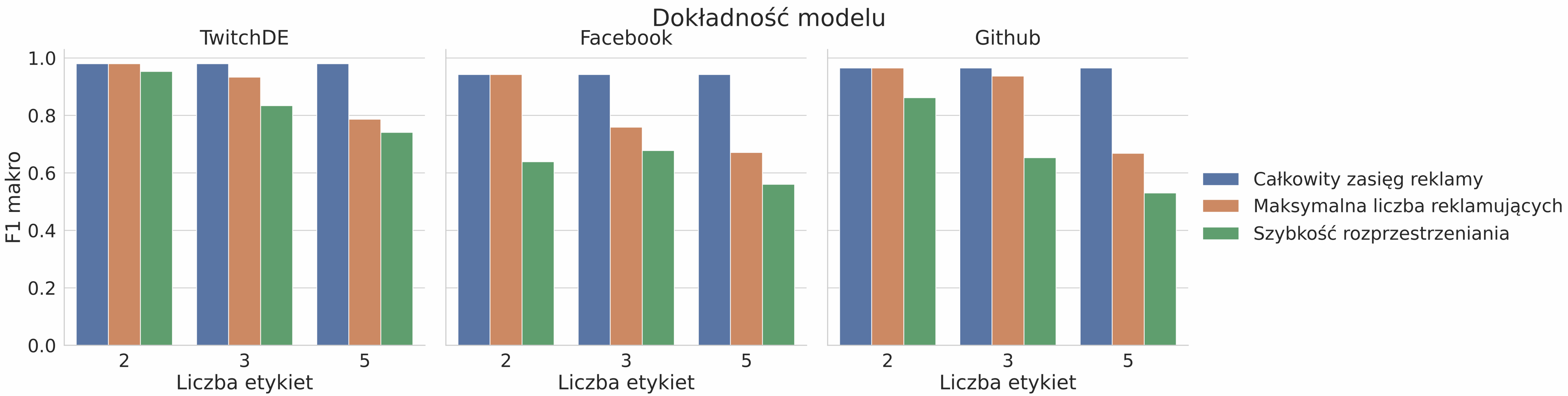
Miary centralności każdego wierzchołka:

- degree
- closeness
- load
- pagerank
- betweenness
- eigenvector
- clustering coefficient

### Etykiety

- Symulacja z użyciem modelu niezależnych kaskad
- Dyskretyzacja wyników symulacji

## Rezultaty



**1.6s**  
**model AI**  
**28.2h**  
**symulacja**

## Podsumowanie

- Udało się stworzyć model AI, potrafiący ocenić potencjał influencera z dokładnością ~80%
- Dzięki technikom uczenia maszynowego, udało się przyspieszyć proces oceny ponad 60 000 razy
- Projekt ma duży potencjał do dalszego rozwoju:
  - Zbadanie umiejętności generalizacji dla silnie zróżnicowanych sieci
  - Analiza wpływu poszczególnych miar centralności na uzyskiwane wyniki
  - Wykorzystanie cech indywidualnych wierzchołka



## Referencje

1. Benedek Rozemberczki, Carl Allen, Rik Sarkar, Multi-Scale attributed node embedding, Journal of Complex Networks, Volume 9, Issue 2, April 2021, cnab014, <https://doi.org/10.1093/comnet/cnab014>  
2. G. Zhao, P. Jia, C. Huang, A. Zhou and Y. Fang, "A Machine Learning Based Framework for Identifying Influential Nodes in Complex Networks," in IEEE Access, vol. 8, pp. 65462-65471, 2020, doi: 10.1109/ACCESS.2020.2984286.  
3. Icon made by Freepik, noomtah, photo3idea\_studio from www.flaticon.com