4SEMANTICS | Wrocław University of Science and Technology | alphamoon

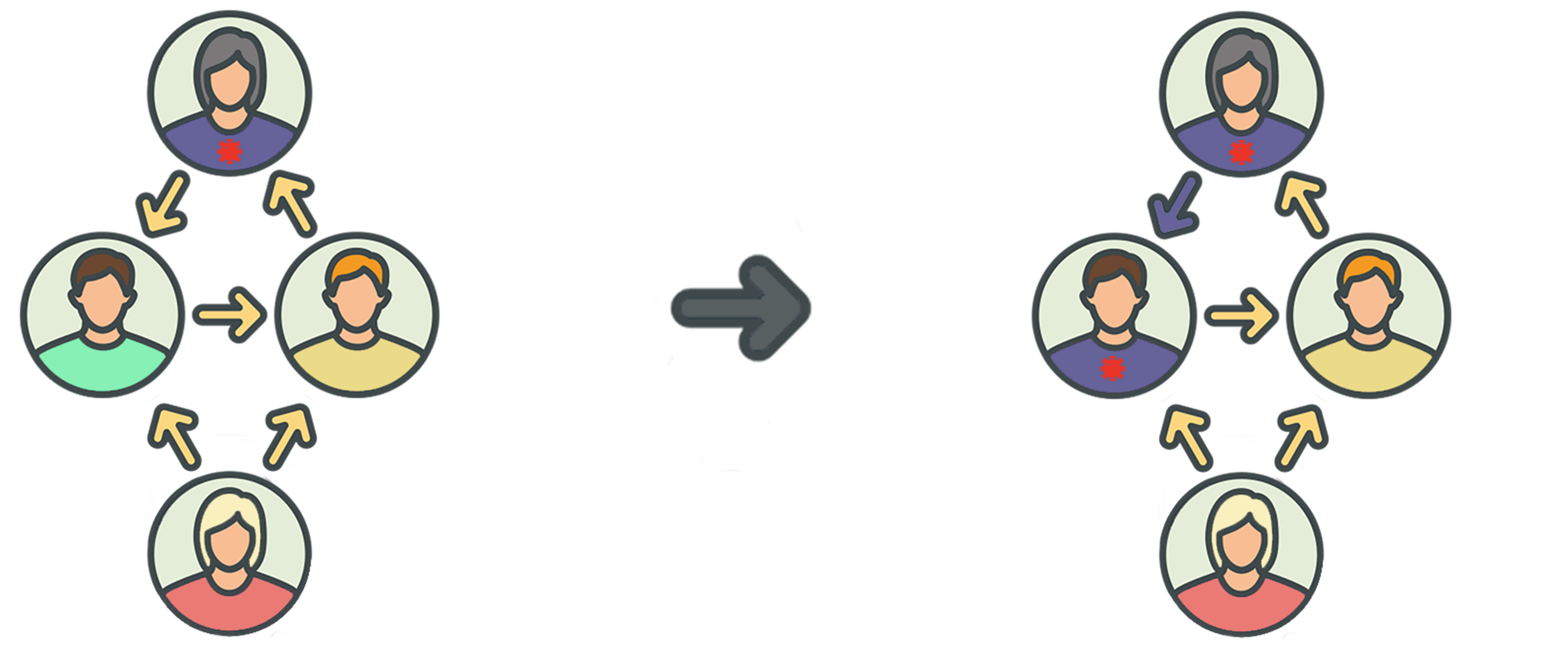
Future pandemics prevention by key nodes identification

Adam Piróg[a, b] | Mateusz Stolarski[a, c] | dr hab. Piotr Bródka[a]

Wroclaw Universityu of Science and Technology [a], 4Semantics [b], Alphamoon [c]

Introduction

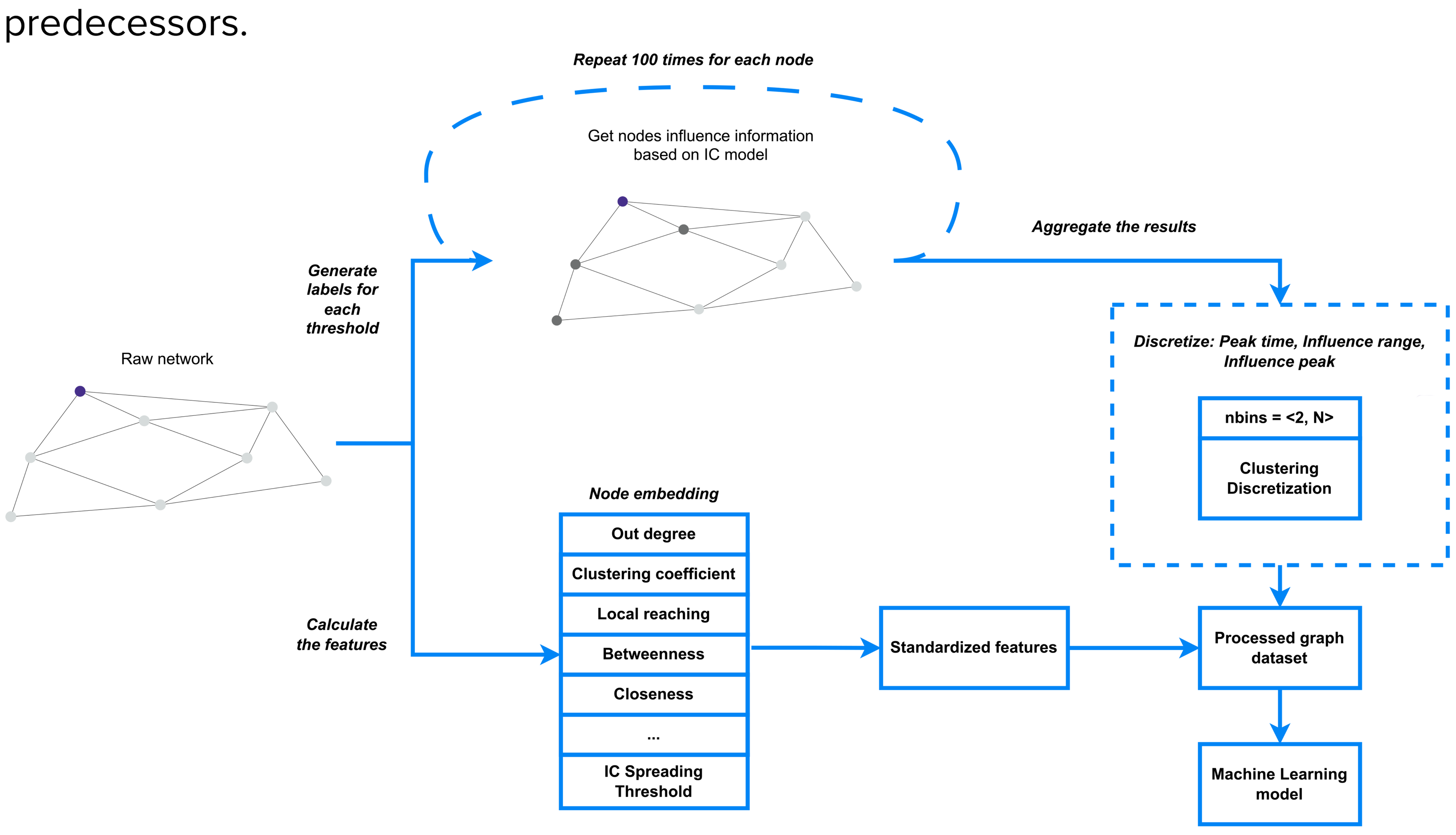
COVID-19 exposed the truth, that the world was woefully unprepared for global pandemic. This catastrophy made us realize, that with the limited medical resources we have, we need to put more focus on early forecasting and strategic planning to provide our support in the most efficient manner.



This is were Artificial Intelligence can play a key role - by properly adapting the task of key node identification in complex networks we can reliably forecast the scale of the pandemic processes.

Framework

We present an enhanced machine learning-based framework for identifying key nodes in complex networks designed to address the shortcomings of its predecessors.



1. Obtaining the labels

a) Simulating the spread with **S**usceptible, **I**nfected and **R**ecovered (SIR) model*

A complex network diffusion algorithm, commonly used to obtain ground truth for node spreading capabilities.

b) Discretizing the results using **Smart Binning** – an unsupervised approach for node labelling

Results of the diffusion model simulation require further postprocessing before feeding them into an ML model, as their fine granularity would cause multiple problems.

2. Generating node embedding

We compose the embedding vector using centrality measures, with an addition of infection rate parameter (the contagiousness). This allows our model to universally operate across different types of viral spreads.

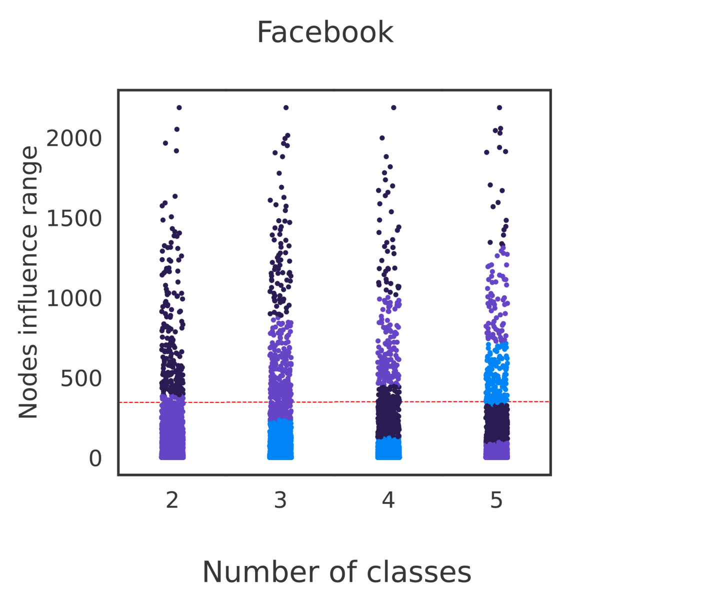
- Average neighbours degree
- Closeness
- Clustering
- Core Number
- Harmonic
- Load

3. Training machine learning models

Smart bins

Other papers [1, 2] address the granularity problem fairly poorly - authors train their models to predict an arbitrary top percent of the nodes (e.g 5%). This approach has major flaws, such as:

- Model looses the ability to classify other nodes (beyond that top percent)
- Such an arbitrary value impairs the flexibility of the method, forcing models to operate in an artificially fixed setting that may not be reflected in real networks



(Effect of smart binning vs. fixed binning. Red line denotes top 5% of the nodes)

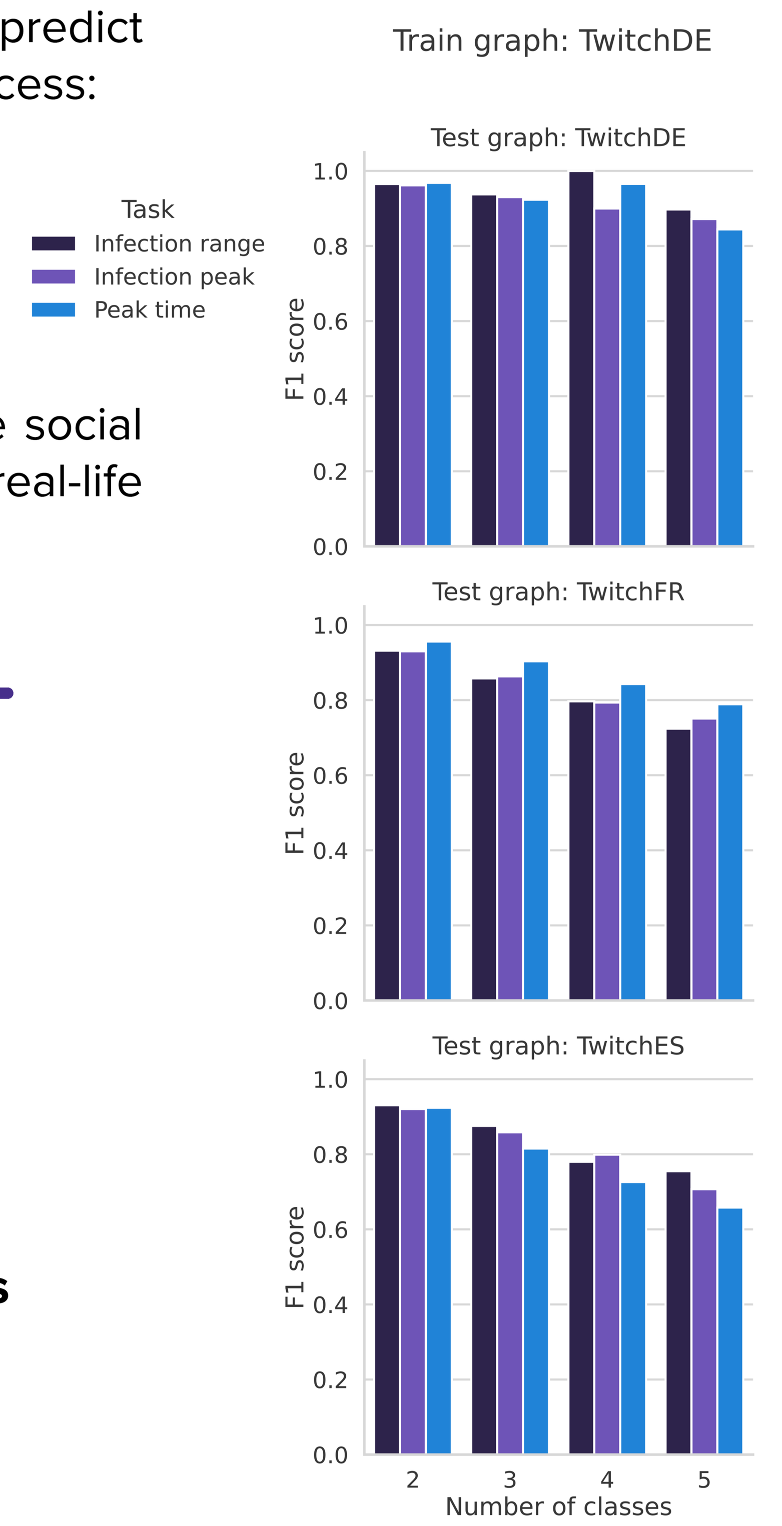
We propose a new approach – Smart bins. By utilizing unsupervised machine learning algorithms, we can achieve a flexible discretization based on actual dependencies in the data, not just an arbitrary choice of a parameter.

Performance

Our model, as the first of this kind, is able to predict 3 crucial characteristics of the spreading process:

- Infection range
- Infection peak
- Peak time

We evaluated our method on various online social networks, serving as approximations of real-life societies for our study.



Benefits

Our method vs other ML methods:

Performance increased by **~40%**
0.69 F1 ➔ 0.96 F1

Our method vs non-ML methods:

Evaluation speed increased **60 000 times**
28.2h ➔ 1.6s

Impact of smart bining

	Infection range		Infection peak		Peak time	
	Fixed bins	Smart bins	Fixed bins	Smart bins	Fixed bins	Smart bins
Facebook	0.87 ± 0.08	0.97 ± 0.001	0.89 ± 0.03	0.96 ± 0.01	0.90 ± 0.01	0.96 ± 0.01
Twitch DE	0.62 ± 0.12	0.97 ± 0.003	0.85 ± 0.05	0.96 ± 0.01	0.86 ± 0.02	0.96 ± 0.01
Twitch ES	0.70 ± 0.13	0.96 ± 0.01	0.81 ± 0.07	0.96 ± 0.01	0.84 ± 0.04	0.96 ± 0.01
Twitch FR	0.60 ± 0.09	0.96 ± 0.002	0.80 ± 0.06	0.97 ± 0.01	0.85 ± 0.02	0.95 ± 0.01

Summary

Our major contributions to the field include:

- Novel node labelling approach based on unsupervised machine learning – Smart Bins
- Improved model with the ability to predict 3 crucial characteristics of the spreading process

Our approach allowed to us to achieve state-of-the-art performance, surpassing the competition (in both performance and stability) by a significant margin.

References

*Calculations were performed using resources provided by the Wrocław Network and Supercomputing Center (http://wcss.pl/)

[1] Gouheng Zhao, Peng Jia, Cheng Huang, Anmin Zhou, and Yong Fang. 2020. A machine learning based framework for identifying influential nodes in complex networks. IEEE Access 8 (2020), 65462–65471.

[2] Doina Bucur. 2020. Top influencers can be identified universally by combining classical centralities. Scientific reports 10, 1 (2020), 1–14.