GUS | MET2023

Trollowanie w sieci. Testowanie podatności różnych struktur sieciowych na dezinformację

Dr hab.Tomasz Kopczewski
Wydział Nauk Ekonomicznych
Uniwersytetu Warszawskiego

University of Warsaw
Faculty of Economic Sciences

Julian Kocerka
Wydział Nauk Ekonomicznych
Uniwersytetu Warszawskiego

Poznaj siebie — nauka ekonomii przez storytelling i Data Science

Rdzeń metody – opowieść o człowieku

My (ekonomiści) opisujemy rzeczywistość za pomocą notacji matematycznej – tworzymy modele. Ale ... wszystkie modele ekonomiczne są opowieścią o człowieku / społeczeństwie → Powinniśmy mieć powód, aby opowiadać, te historie lub jest to tylko czysta retoryka.

Główna idea tworzenia ciekawości naukowej przez storytelling:

Po pierwsze

Istnieje wiele historii, w których ktoś opowiada o nas (podręcznik, artykuły, ...). Powinniśmy poznać te opowieści i skonfrontować je z naszą wiedzą o sobie.

Po drugie

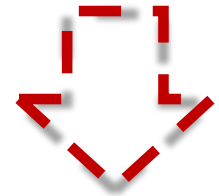
Powinniśmy nauczyć się radzić sobie z konsekwencjami tych opowieści w naszym życiu.

Rdzeń metody – opowieść o człowieku

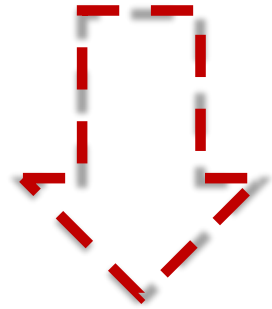
Na podstawie wieloletnich doświadczeń powstała metoda *Know Thyself*. W metodzie tej studenci poznają ekonomię jako zbiór opowieści o nich samych, które kryją się w podręcznikach. Celem tej metody jest rozszyfrowanie tych opowieści. Każdy ze studentów uzyskuje też dostosowany do własnych potrzeb sposób narracji. Dla humanistów są to odwołania do literatury i sztuki, dla studentów STEM jest to odwołanie do opisu statystycznego i analizy danych eksperymentalnych. Dla nauk społecznych jest to wykorzystanie eksperymentów ekonomicznych i licznych powiązań ekonomii z psychologią i socjologią, ….

Poznaj siebie – flipped classroom na zajęciach z mikroekonomii

Przed zajęciami studenci biorą udział w badaniu / eksperymencie on-line, w którym przyjmują „bierną rolę”.



Po badaniu zachęcam studentów do próby odpowiedzi na pytania: czy wasze decyzje lub odpowiedzi były zgodne z teorią ekonomii. W ten sposób kreuje *information gap*, poczucie deprywacji - niezaspokojeni potrzeby wiedzy na temat podjętych decyzji → *science curiosity*



Ciekawość nauki jest zaspokojona podczas wykładów -->
wyniki badania wykorzystane są do przedstawienia modeli teoretycznych podczas zajęć w formie raportu.

Raport

Na wykładzie, przedstawiam teorię i konfrontuje ją bezpośrednio z wynikami badań ad hoc. Przedstawiam wyniki zbiorcze i indywidualne (zanonimizowane).

W porównaniu z tradycyjnymi materiałami do nauki, zbieranie i analiza danych oraz przygotowywanie prezentacji na podstawie tych danych jest bardzo czasochłonne. Jednak narzędzia IT mogą wspierać ten proces przygotowawczy.

Raw data

RStudio

Markdown

self-publishing system

R-CRAN

Raport

Tax ...riment by TK | 0.Description | 1.B-law | 3.Test experimental results | Source Code

Nickname

niebieski21

Test

Test

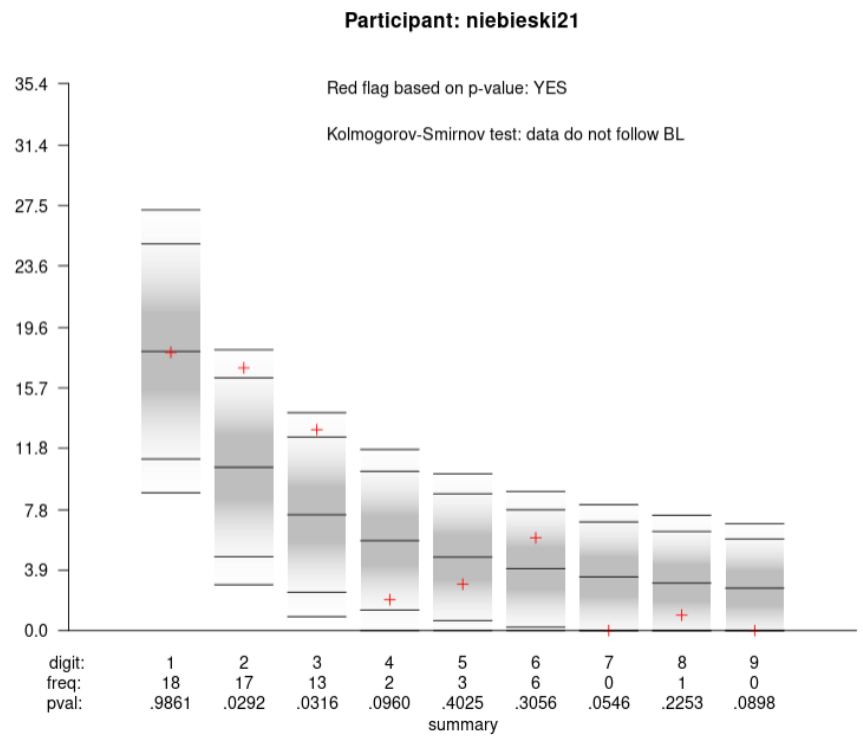
We used a simple one-sample Kolmogorov–Smirnov test to compare experimental samples with Benford's probability distribution. We tested only the first digits' appearance.

The anti-fraud algorithm, which is used by tax authorities, is more straightforward. If only the frequency of appearance of 1 is less than 0.15, the tax declaration will be checked carefully :)

First proposed by Varian in 1972. Now, Benford's Law is legally admissible as evidence in the US in criminal cases at the federal, state and local levels.

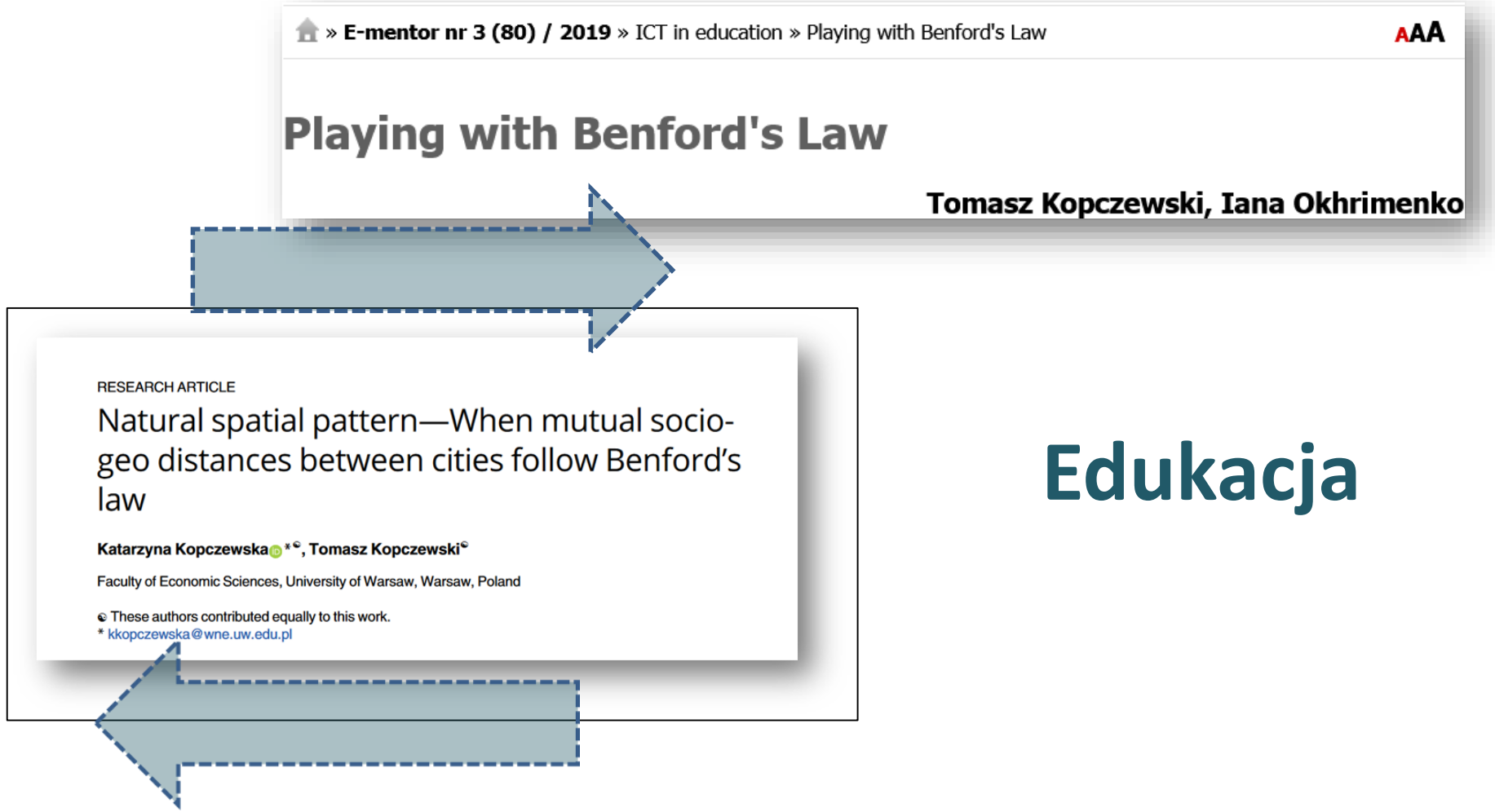
You can use it to detect any fraud: financial/accounting fraud, scientific fraud and data fabrication (you can check the quality of your datasets)

Graph test



Dwukierunkowy przepływ wiedzy

Teoria / badania empiryczne



Edukacja

Poszukiwanie prostoty i odpowiedzi na pytanie: dlaczego?

Troll w sieci

Opowieść o człowieku w podręcznikach mikroekonomii nie zawiera opowieści o dezinformacji, fake news, …, Homo Oeconomicus ma pełną informację i pełne zdolności poznawcze podjęcia najlepszej dla niego decyzji – nie popełnia błędów poznawczych, decyzje HO są niezależne od innych.

→ Poszukiwanie prostego modelu – nie ma takiego i trzeba było go stworzyć

Model DeGroota

→ **Jest to prosty model uczenia społecznego, który stanowi bazę do tworzenia skomplikowanych modeli interakcji społecznych i znajdywania konsensusu lub polaryzacji społecznej.**

→ **Jest to ciekawy przykład zastosowania łańcuchów Markowa do analizy zachowań społecznych.**

$$x = (x_1, x_2, \dots, x_n) \in [0,1]^n$$

Wektor opinii / przekonań
0 – nie wierzę w szczepionki
1 - wierzę w szczepionki

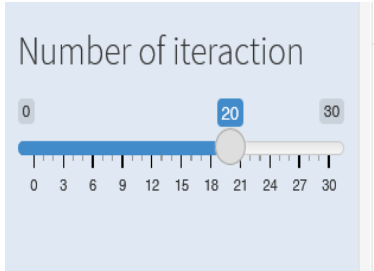
$$W = \begin{bmatrix} w_{1,1} & \cdots & w_{1,n} \\ \vdots & \ddots & \vdots \\ w_{n,1} & \cdots & w_{n,n} \end{bmatrix}$$

Macierz interakcji społecznych, kto kogo słucha.

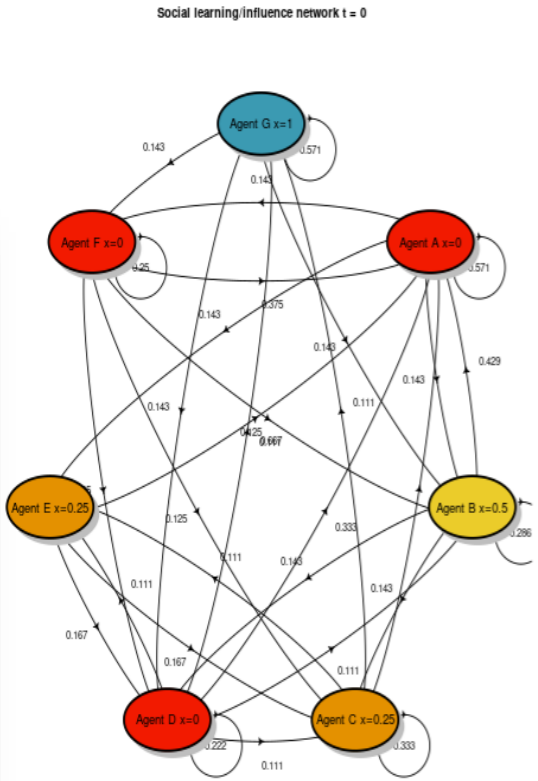
Model DeGroota

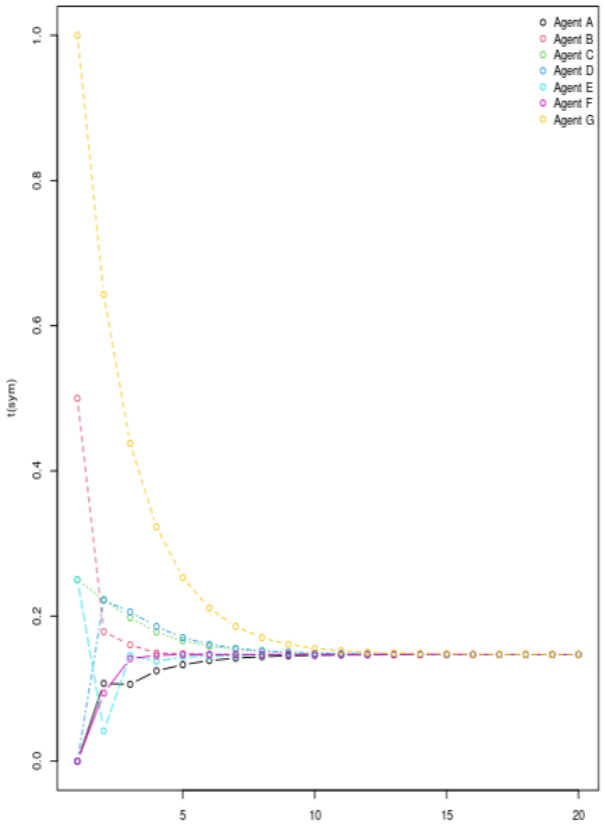
Model dąży do stanu stacjonarnego.

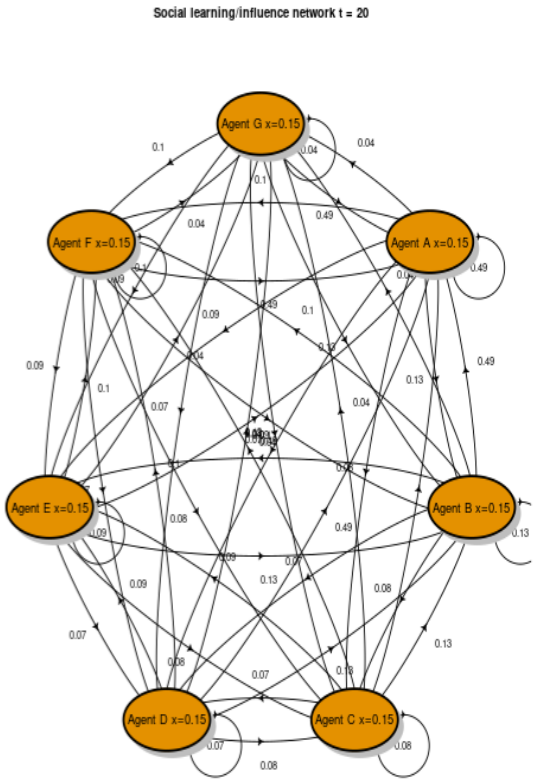
$$x(t+1) = W^t x(1) = W^{t+1} x(0)$$



DeGroot model of social learning







Model DeGroota z trollem

Przyjęliśmy, że jeden agent wierzy tylko sobie, nikt nie ma na niego wpływu oraz ma on wpływ przynajmniej na jednego agenta. Nazwaliśmy tego agenta trollem algorytmicznym.

$$W = \begin{bmatrix} 1/3 & 1/6 & 1/3 & 1/6 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Model DeGroota z trollem

Wprowadzenie trolla uchyliło podstawowe założenie modelu DeGroota: macierz W nie jest silnie spójnym grafem (*non-strongly connected*). Zanalizowaliśmy wpływ uchylenia tego założenia na zachowanie się modelu.

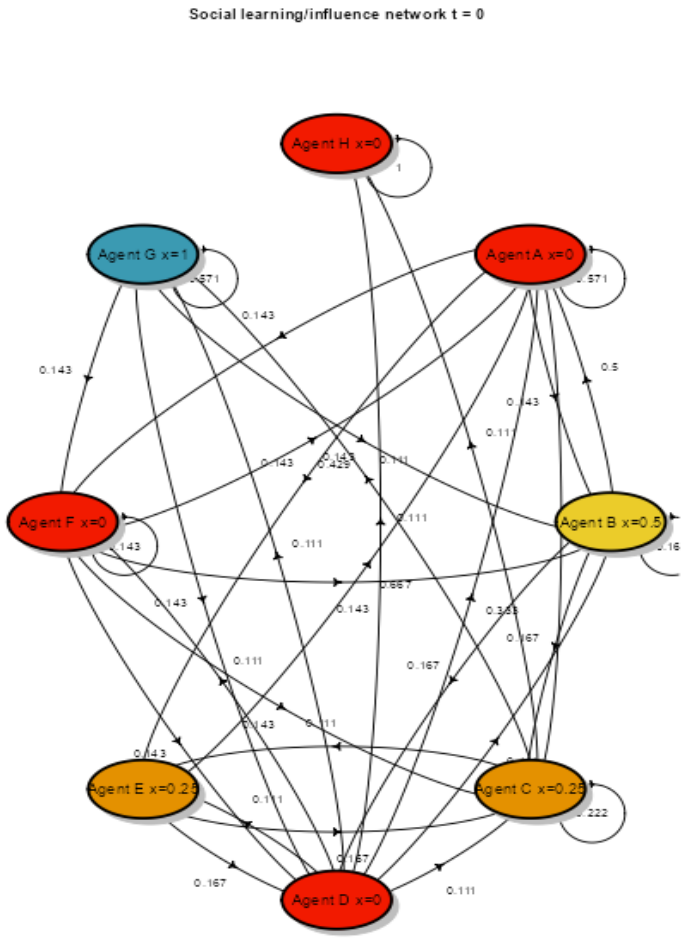
$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} W^t x(0) = \begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} x(0)$$

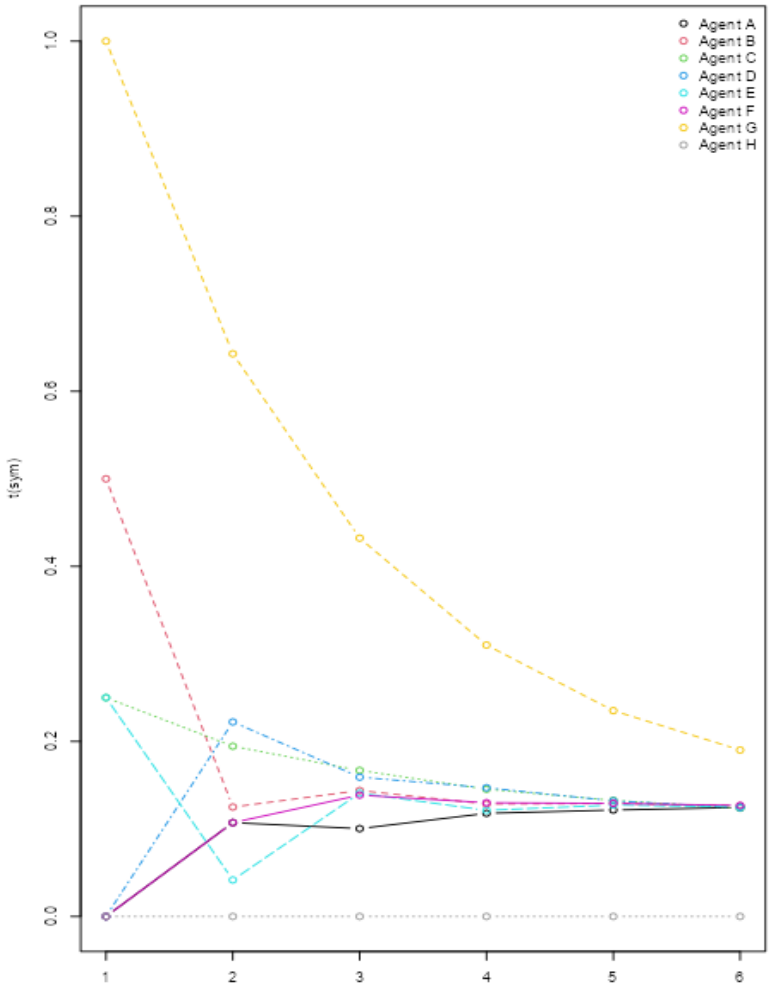
Nie jest to ciekawe rozwiązanie algebraiczne, ale ciekawie zachowuje się system:

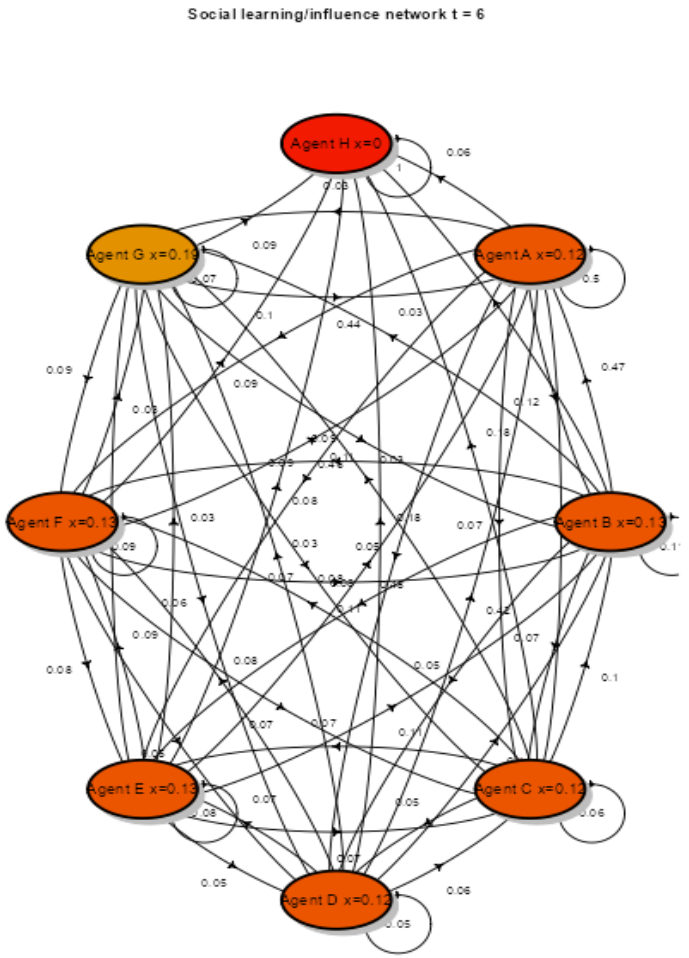
- **Pierwsza faza łączenia agentów**
- **Druga faza degeneracji grafu**

Model DeGroota z trollem

- Faza łączenia – graf staje się pełny

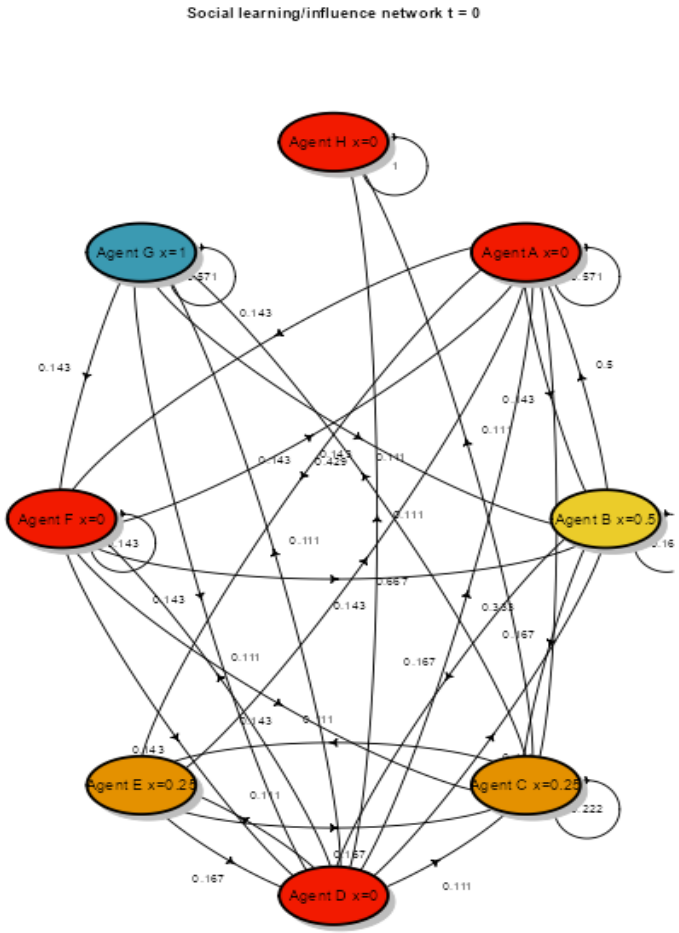


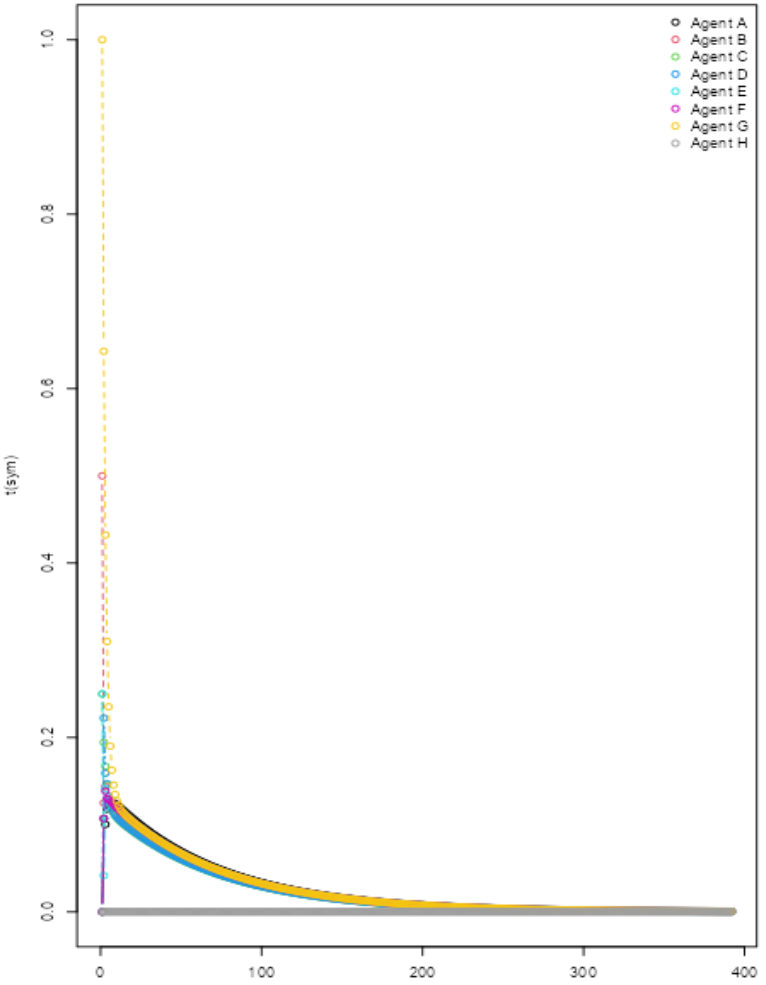


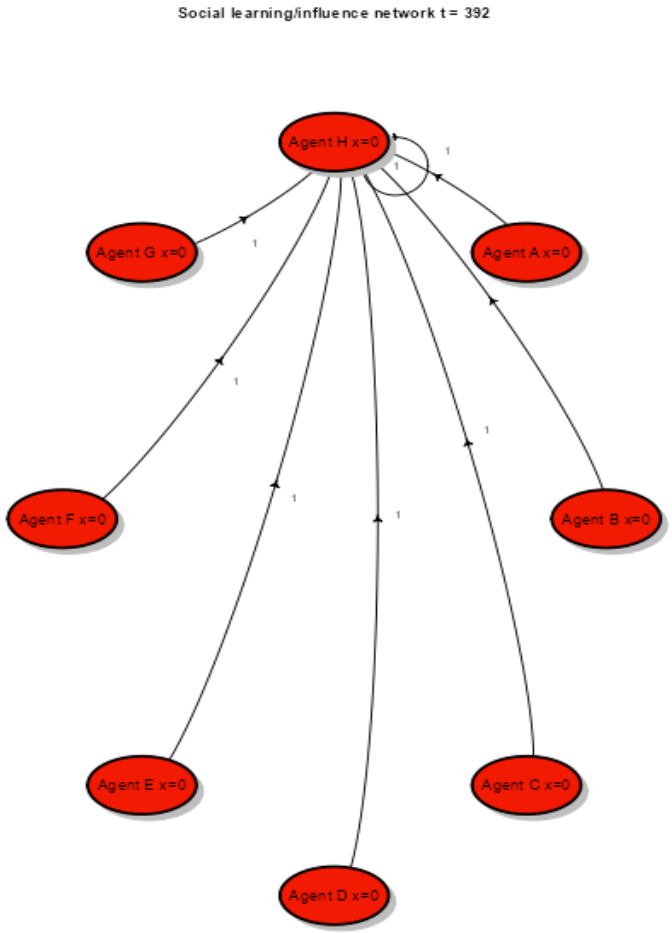


Model DeGroota z trollem

- Faza degeneracji grafu





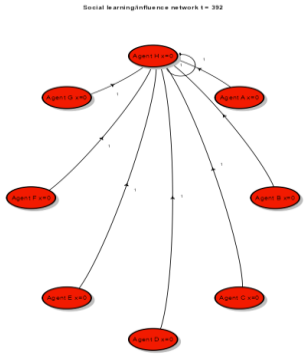


Model DeGroota z trollem

Nie ma znaczenia czy troll wierzy (1) czy nie (0) - system i tak zmierza do stanowiska trolla.

Model dobrze też pokazuje dwa zjawiska:

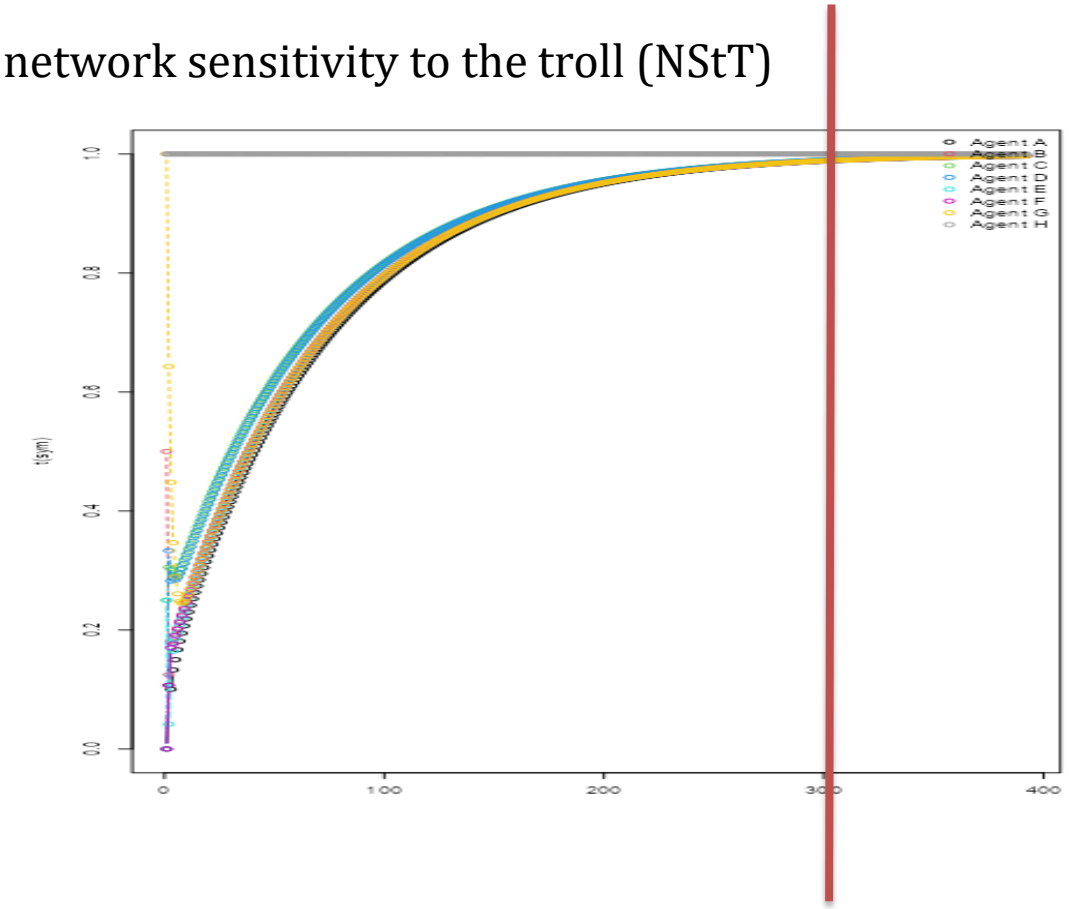
- **Efekt iluzorycznej prawdy (efekt powtórzenia) - to tendencja do wierzenia, że fałszywe informacje są poprawne po wielokrotnym powtórzeniu (Hasher et. al, 1977)**
- **Taktyka salami – powolna zmiana, która jest niezauważalna - jest akceptowalna społecznie (Enh, 2010).**



Mara wrażliwości sieci na trolla

Ilość iteracji potrzebnych do przyjęcia przez wszystkich stanowiska trolla z zadanym przybliżeniem (0.005).

Ilość iteracji będzie zależała od struktury sieci i miejsca trolla w tej sieci.



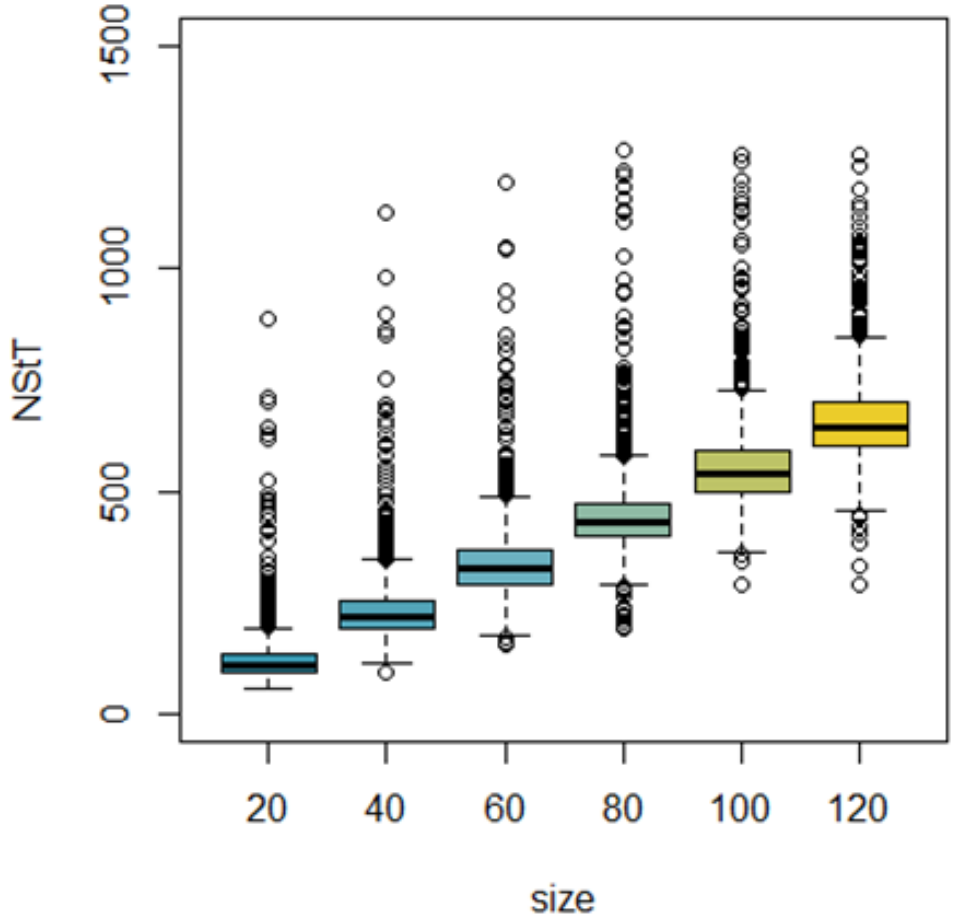
Symulacja MC

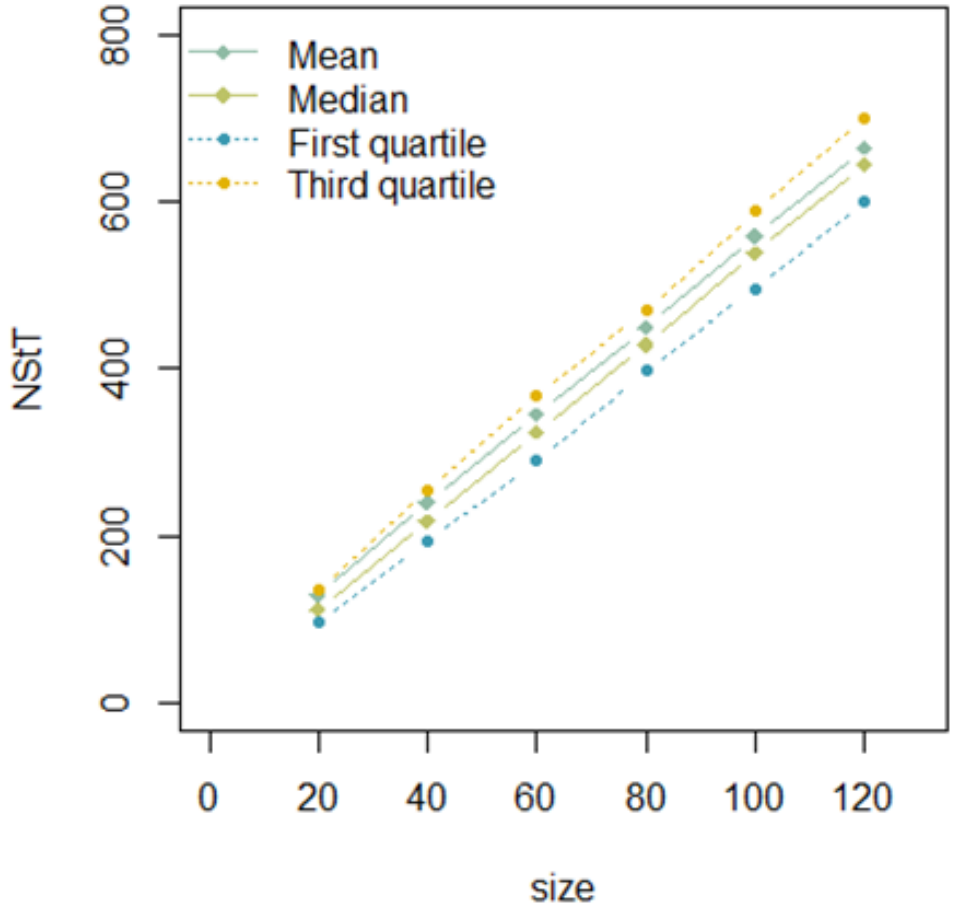
Można przeprowadzić symulację MC i określić jak statystyki /charakterystyki sieci i węzła, w którym jest troll wpływają na szybkość konwergencji do stanowiska trolla.

1) **Generowanie losowej macierzy W z losowo umieszczonym trollem (Erdős-Rényiof G_(n,p) + troll)**
2) **Obliczenie charakterystyk sieci oraz policzenie charakterystyk wierzchołka zajętego przez trolla**
3) **Estymacja miary *network sensitivity to the troll* (NStT)**
4) **Powtórzenie M razy punktu 1- 3**

Symulacja MC

Pierwsza (zła) wiadomość: skalowalność linowa wyników dla grafów o różnej licznie wierzchołków.



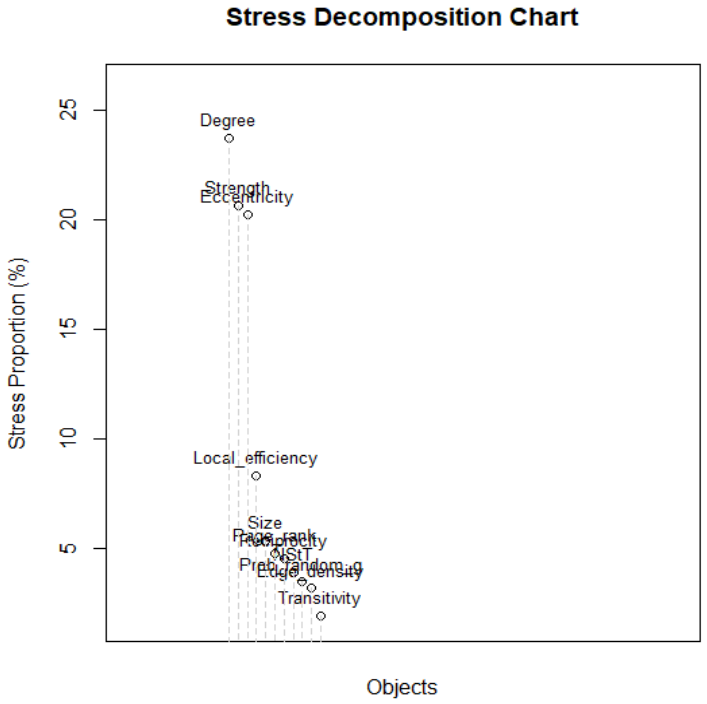


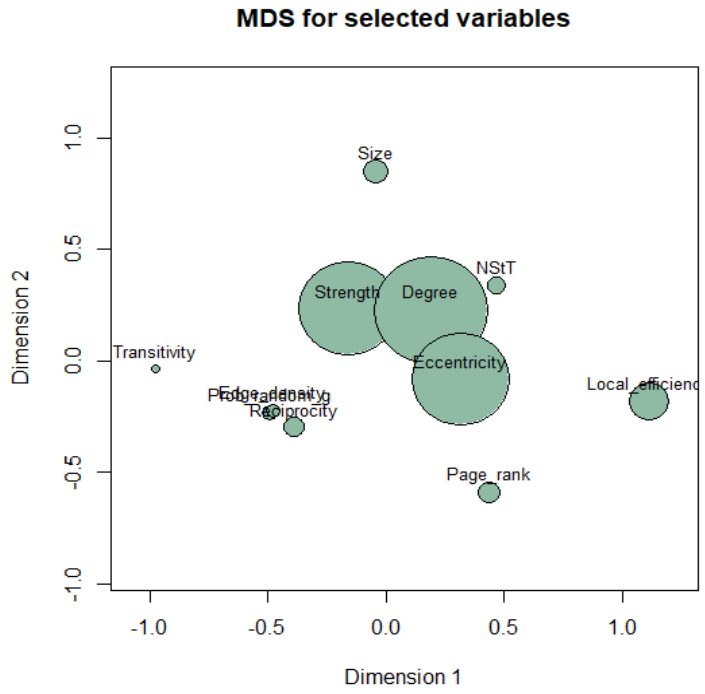
Symulacja MC

Druga (dobra/zła?) wiadomość: charakterystyki sieci oraz charakterystyki węzła częściowo grupują się, ale ich ładunek informacyjny jest inny.

Edge density, reciprocity Transitivity opisują ogólnie siec /graf

Degree, strength, page rank, local efficiency eccentricity opisują znaczenie danego węzła w sieci

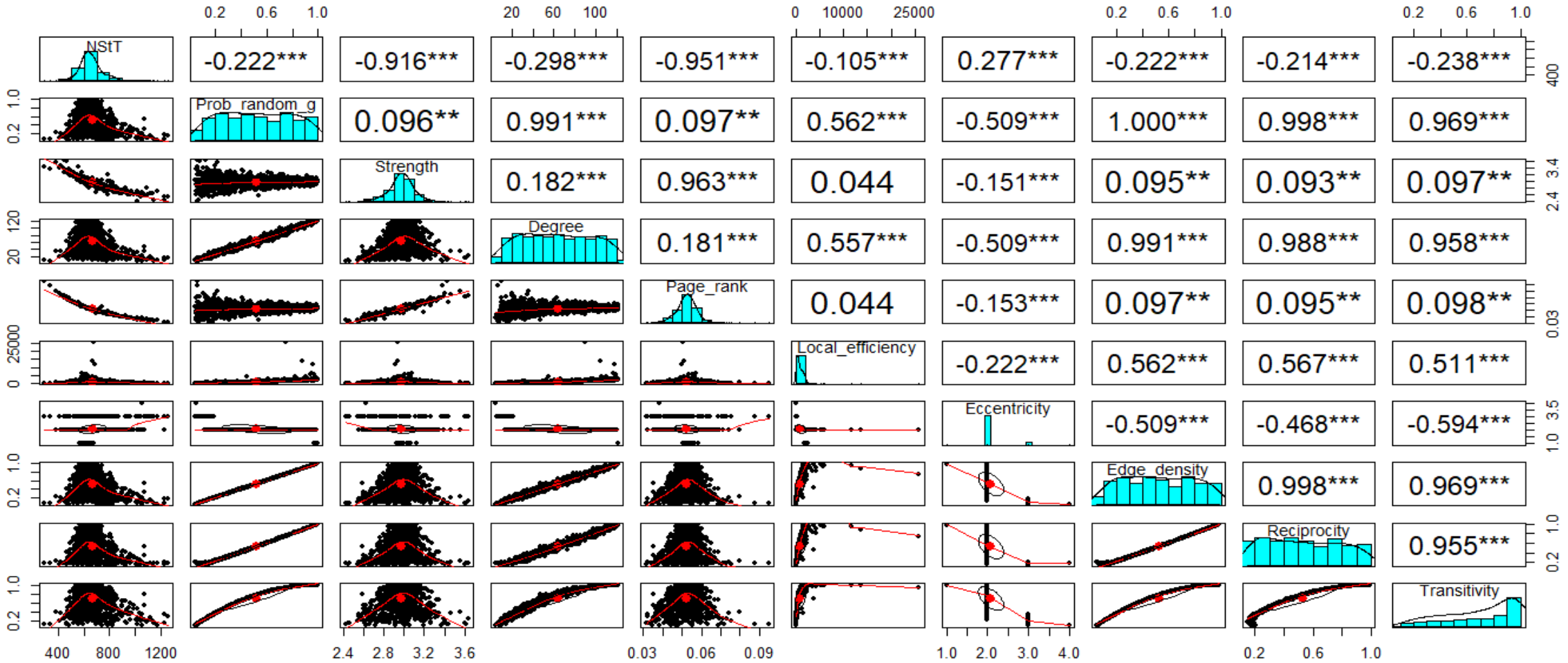




Trzecia (zła) wiadomość: trudno znaleźć charakterystyki sieci jako ogólnie dostępne dane. Znany wskaźnik Page Rank został wycofany przez Googla, ze względu na manipulacje rankingiem stron WWW.

Symulacja MC

Czwarta (dobra) wiadomość: nie trzeba udowadniać na siłę istnienia zależności między NStT a miarami sieci i charakterystykami węzła



Symulacja MC

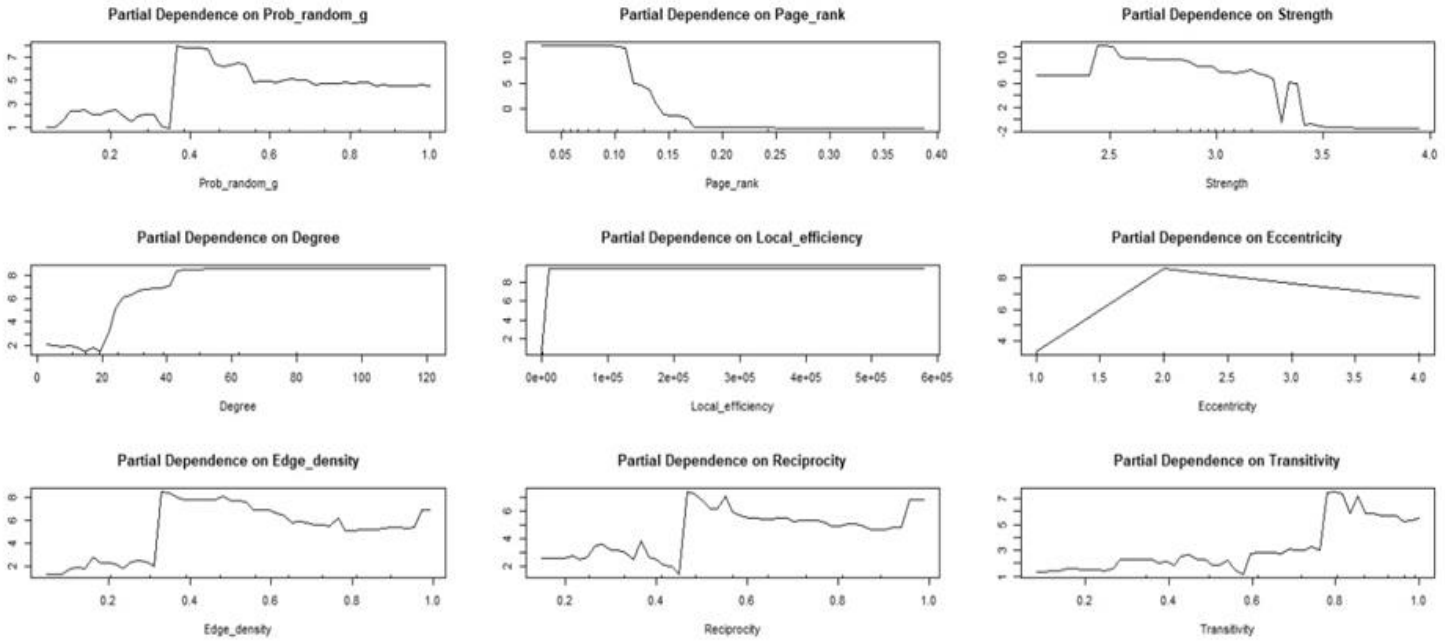
Piąta (dobra) wiadomość: łatwość prognozowania / model Random Forest - klasyfikacja 25% najmniej odpornych na trolla sieci.

```
Type of random forest: classification
Number of trees: 500
No. of variables tried at each split: 3

OOB estimate of error rate: 0.74%
Confusion matrix:
```

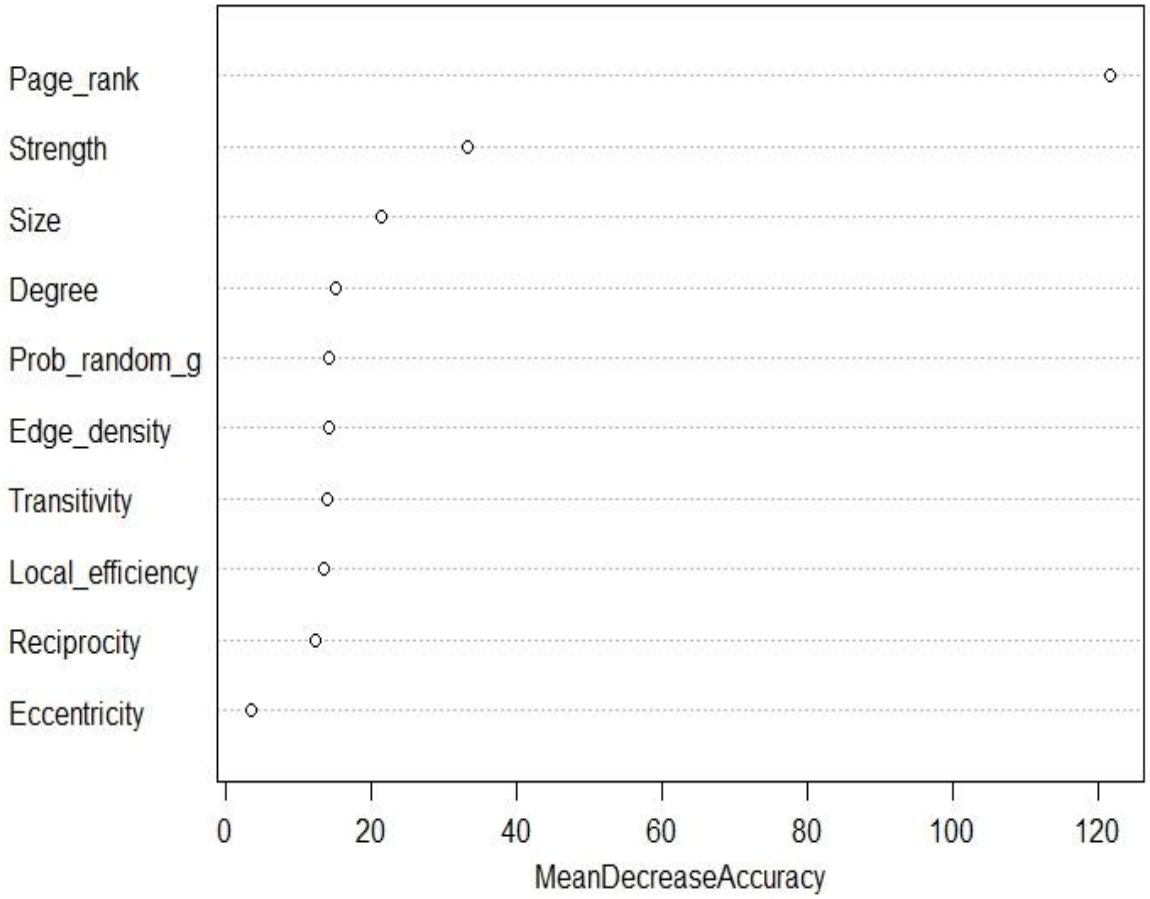
	0	1	class.error
0	4477	10	0.002228661
1	34	1449	0.022926500

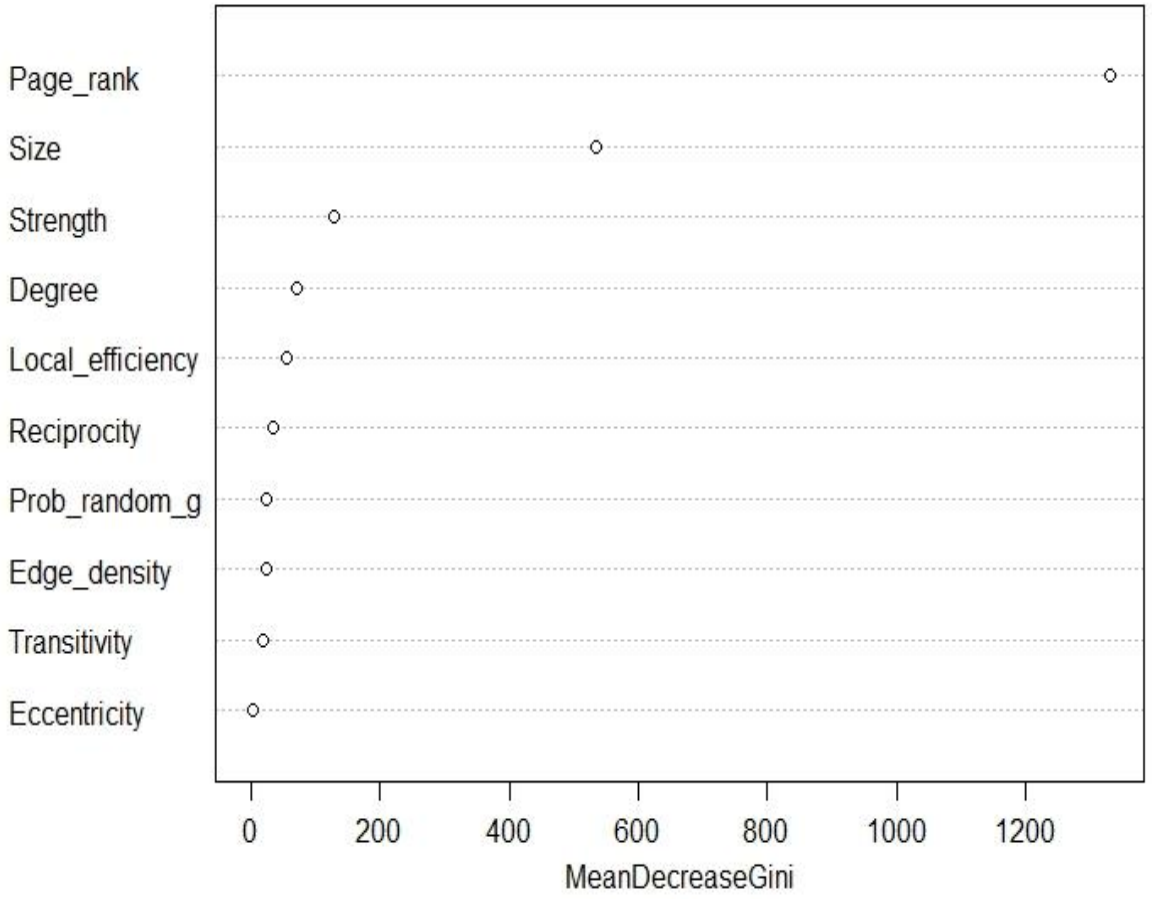
Nieliniowe i częściowe zależności



Symulacja MC

Sósta (zła/dobra??) wiadomość: w prognozowaniu charakterystyki sieci mają mniejsze znaczenie niż charakterystyki położenie trolla.





Rozszerzenie badania - model prognostyczny wrażliwości sieci na trolla

Poprawienie/urealnienie modelu symulacyjnego. Na podstawie uzyskanych wyników wytrenowanie modelu prognostycznego klasyfikacyjnego (Random Forest). Podstawienie do modelu wartości rzeczywistych w celu określenia wrażliwości sieci na trolling.

Pytanie filozoficzne: co się stanie, gdy Chat GTP będzie trollem algorytmicznym?

- Rozwiązanie problemu nadmiaru informacji (ekonomia uwagi Herberta Simona)

- Zanik wiedzy kolektywnej wynikającej z zaniku różnorodności (*Wisdom of Crowds*).

Dziękujemy za uwagę.

Na stronie projektu będą pojawiać się nowe wersje materiałów.