

# Graph Neural Networks (Scarselli, 2009)

Aleksy Barcz

19 maja 2013

# Cechy GNN

- ▶ Klasyfikator dowolnych grafów (niepozycyjne, cykle)
- ▶ Klasyfikacja węzłów i klasyfikacja grafów
- ▶ Oparty na FNN
- ▶ Proste rozwiązanie problemu cykli

# Cel implementacji

- ▶ Sprawdzenie działania GNN i identyfikacja kluczowych parametrów
- ▶ Utworzenie wygodnego i uniwersalnego środowiska
- ▶ Wybrana platforma implementacji : Octave

# Format reprezentacji grafu

- ▶ *nodes.csv* (etykiety węzłów):  
każdy wiersz =  $l_n$ ; ( $|l_n| \geq 1$ )
- ▶ *edges.csv* (etykiety krawędzi  $x_u \Rightarrow x_n$ ):  
każdy wiersz =  $id_u, id_n, l_{un}$ ;  
( $id_n$  : indeks węzła  $n$  w pliku *nodes.csv*)  
( $l_{un}$  : etykieta krawędzi,  $|l_{un}| \geq 0$ )
- ▶ *output.csv* (oczekiwane wyjścia węzłów lub grafu):  
każdy wiersz =  $o_n$ ; ( $|o_n| \geq 1$ )  
lub  
pojedynczy wiersz = klasa grafu

# Koncepcja GNN

- ▶ Pojedyncza sieć dla wszystkich grafów należących do problemu
- ▶ Dla każdego wężła automatycznie budowana reprezentacja:  
 $x_n = f(\dots)$
- ▶ Klasyfikacja wężła:  $o_n = g(x_n)$
- ▶ Dla zagadnienia klasyfikacji grafów, wybieramy wierzchołek reprezentujący graf

# Składowe GNN

- ▶ Dwie jednostki obliczeniowe :  $f_w$  i  $g_w$
- ▶ Wszystkie instancje  $f_w$  współdzielą wagi
- ▶ Wszystkie instancje  $g_w$  współdzielą wagi
- ▶ Instancje  $f_w$  połączone w metasieć odwzorowującą połączenia w grafie

# Składowe GNN - grafy niepozycyjne

- ▶ Stan węzła:

$$x_n = f_w(\dots) = \sum_{u: u \Rightarrow n} h_w(l_n, l_{nu}, x_u)$$

- ▶ Wyjście węzła:

$$o_n = g_w(x_n)$$

- ▶  $h_w$  : FNN

(wejścia, warstwa ukryta *tanh*, warstwa wyjściowa *tanh*)

- ▶  $g_w$  : FNN

(wejścia, warstwa ukryta *tanh*, warstwa wyjściowa dowolna)

# Postać globalna

- ▶  $X = x_1, x_2, \dots, x_m$  : stan - reprezentacja wszystkich węzłów grafu, zbudowana przez  $F_w$
- ▶  $F_w(X)$  : globalna funkcja przejścia,  $F_w(X) = X$
- ▶  $F_w$  jest kontrakcją, a wyznaczany  $X$  - jej punktem stałym
- ▶  $G_w(X)$  : globalna funkcja wyjścia



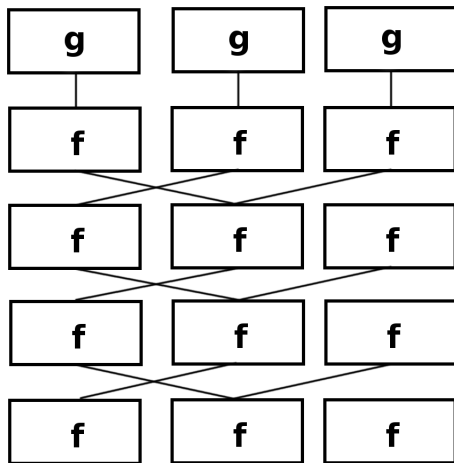
# Własności $F_w$ (tw. Banacha)

- ▶ kontrakcja :  $\|F_w(X_1) - F_w(X_2)\| \leq \|X_1 - X_2\|$
- ▶ posiada dokładnie jeden punkt stały
- ▶ zbieżność  $F_w$  niezależna od punktu początkowego
- ▶ zbieżność wykładnicza - potrzebna bardzo mała ilość powtórzeń (ok. 5 do 15)
- ▶ jak zapewnić by  $F_w$  była kontrakcją?

# Schemat działania

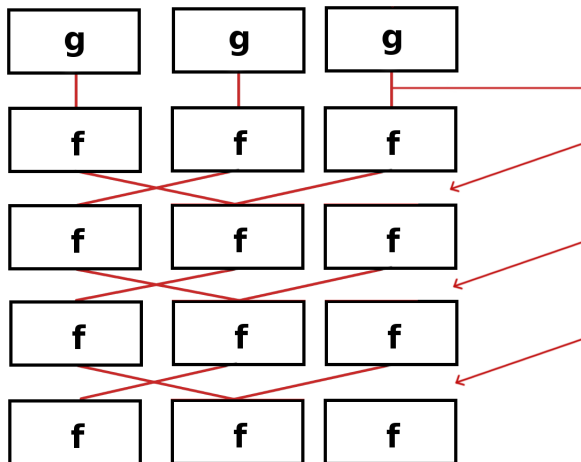
1. Losowa inicjalizacja stanu  $X$
2. do osiągnięcia kryterium stopu:
  - ▶ FORWARD : obliczenie  $X = F_w(X)$  aż do osiągnięcia punktu stałego
  - ▶ BACKWARD : obliczenie  $G_w(X)$  i propagacja wsteczna błędów
  - ▶ aktualizacja wag  $f_w$  i  $g_w$

## Forward - budowanie stanu



- ▶ BPTT
- ▶ *minStateDiff*

## Backward - propagacja wsteczna błędu



- ▶ Almeida-Pineda *minErrorDiff*
- ▶ Kara za brak kontrakcji *contracionConstant*
- ▶ RPROP *deltaMax*

# Zbiór danych

- ▶ Wykrywanie podgrafu
- ▶ Podobny do zbioru Scarselli
- ▶ 6 węzłów, 3 węzły podgrafu,  $p_{edge} = 0.8$  (zamiast 0.2)
- ▶ Etykiety węzłów 0..10
- ▶ Szum Gaussowski średnia=0 std=0.25
- ▶ 20 grafów

# Parametry modelu - stałe

FNN - zgodne z zaleceniami:

- ▶ liczba neuronów ukrytych  $h_w$  : 5
- ▶ liczba neuronów ukrytych  $g_w$  : 5
- ▶ rozmiar stanu  $x_n$  : 5

RPROP - zgodne z zaleceniami:

- ▶ *initialDelta* : 0.1
- ▶ *minDelta* : 10e-6
- ▶ *maxDelta* : 1.0 (zalecana wartość alternatywna)
- ▶ *factor+* : 1.2
- ▶ *factor-* : 0.5

# Parametry modelu - modyfikowane

Zależne od problemu:

- ▶ *minStateDiff* :  $10e-8$  ..  $10e-5$
- ▶ *minErrorDiff* : jw.
- ▶ *contractionConstant* :  $0.5$  ..  $1.2$

# Eksperymenty

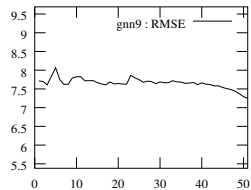
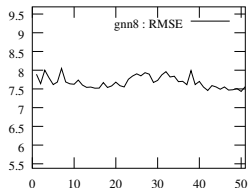
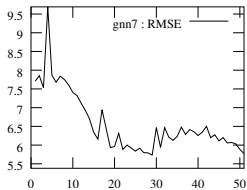
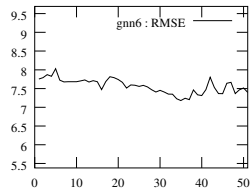
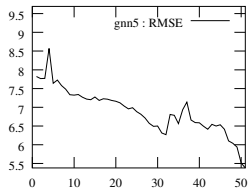
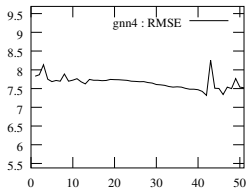
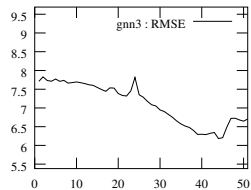
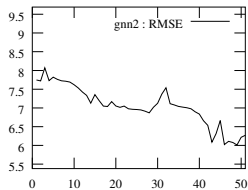
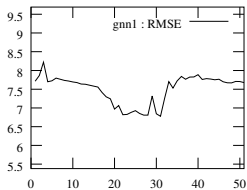
## Legenda:

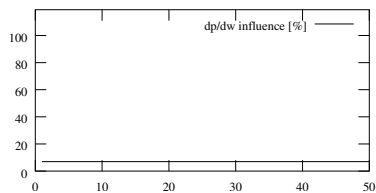
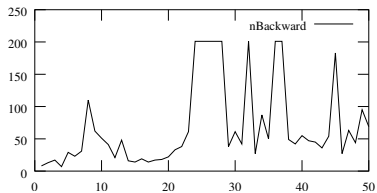
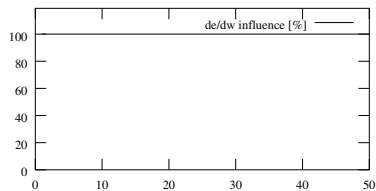
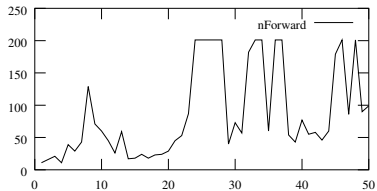
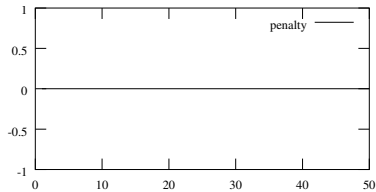
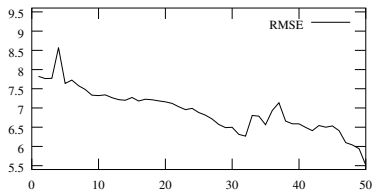
- ▶ *nForward* : ilość kroków procedury budowania stanu (przerywana po przekroczeniu 200)
- ▶ *nBackward* : ilość kroków akumulacji błędu (Almeida-Pineda, przerywane po przekroczeniu 200)
- ▶ *penalty* : czy na którąkolwiek z wag została nałożona kara kontrakcji?
- ▶ *de/dw influence* : procent korekt wag zgodnych znakiem z gradientem błędu
- ▶ *dp/dw influence* : procent korekt wag zgodnych znakiem z pochodną kary kontrakcji

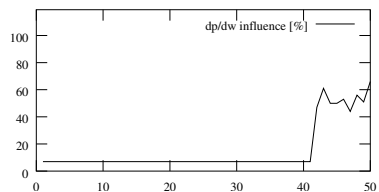
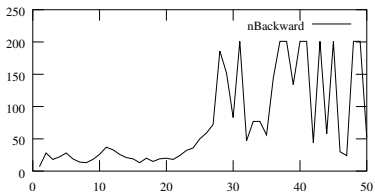
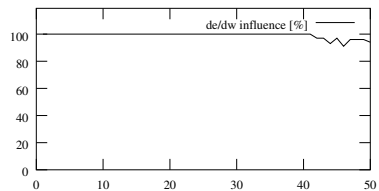
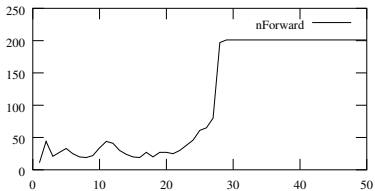
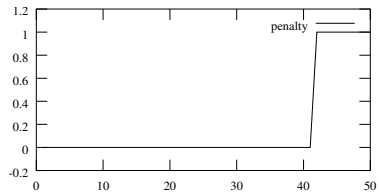
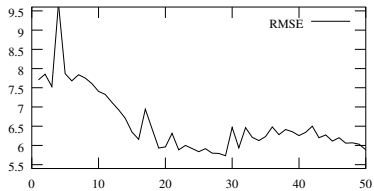


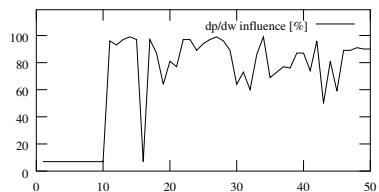
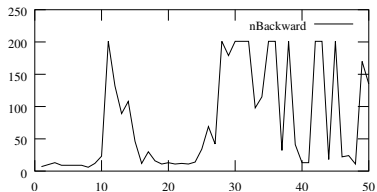
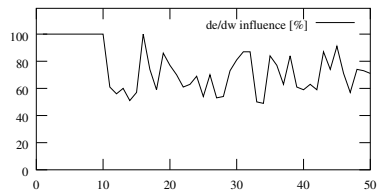
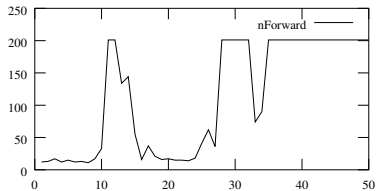
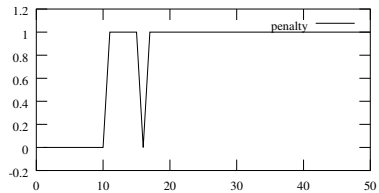
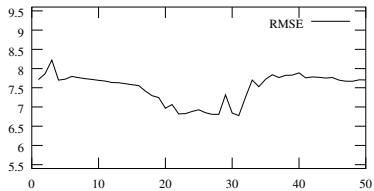
# Eksperyment 1 : wpływ początkowych wartości wag

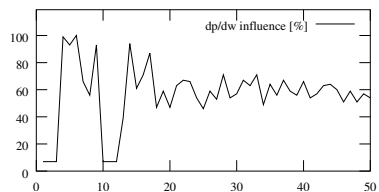
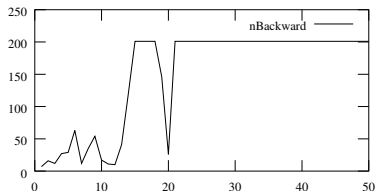
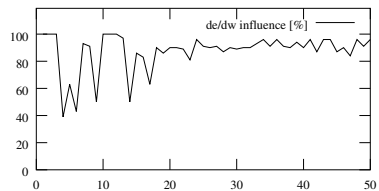
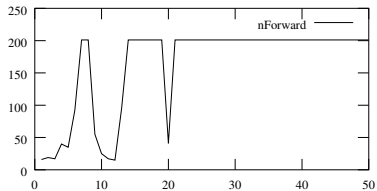
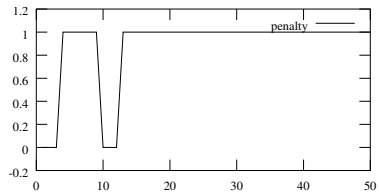
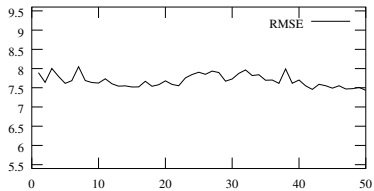
- ▶ 10 grafów
- ▶ 9 sieci GNN o losowych wagach
- ▶  $\text{contractionConstant} = 0.9$
- ▶  $\text{minStateDiff} = 10\text{e-}08$
- ▶  $\text{minErrorDiff} = \text{minStateDiff}$





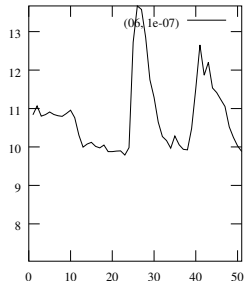
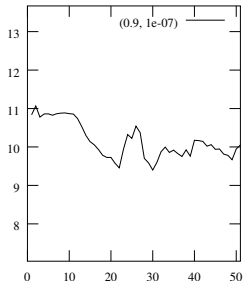
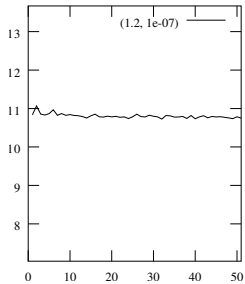
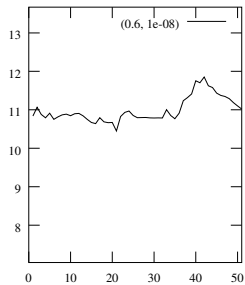
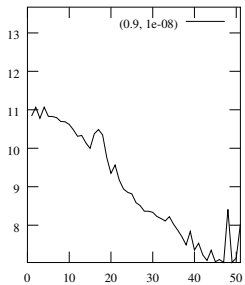
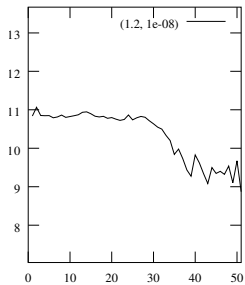




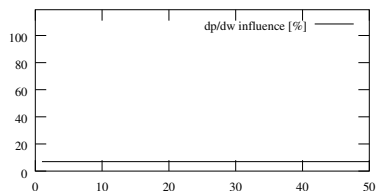
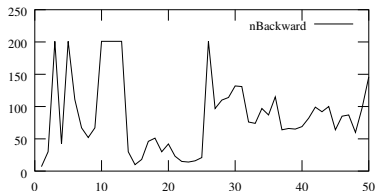
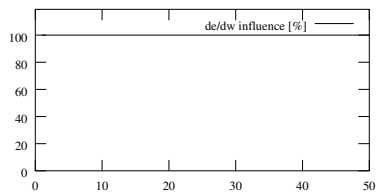
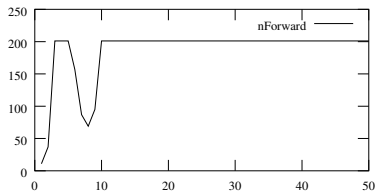
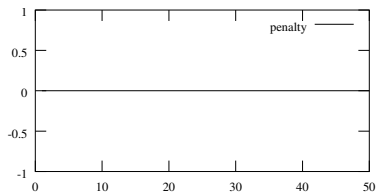
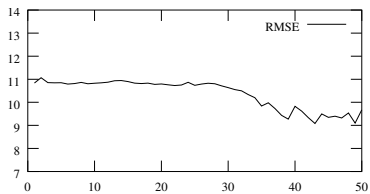


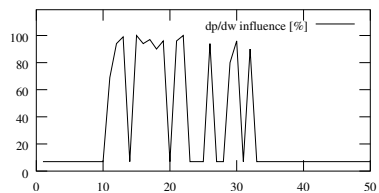
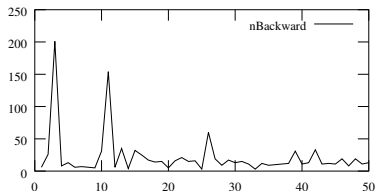
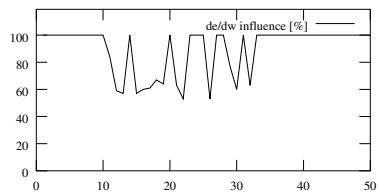
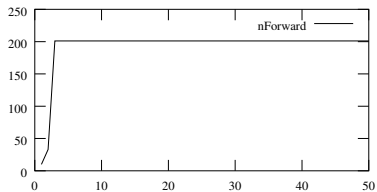
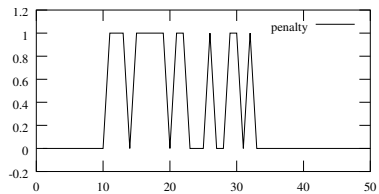
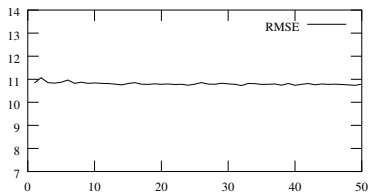
## Eksperyment 2 : wpływ parametrów GNN

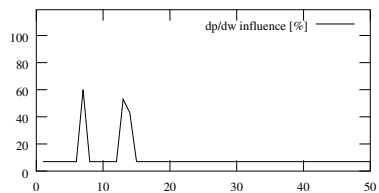
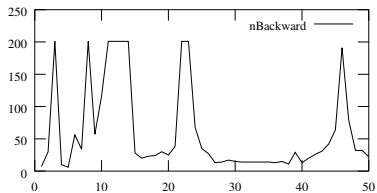
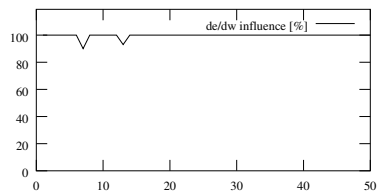
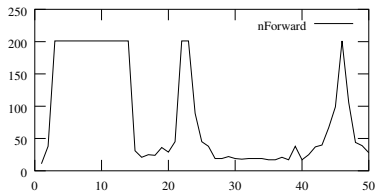
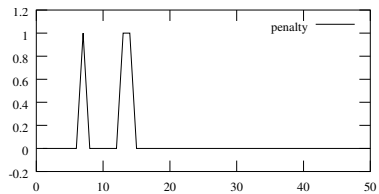
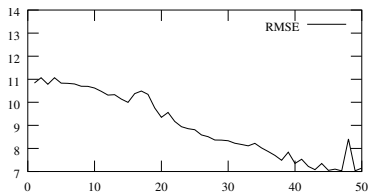
- ▶ 20 grafów
- ▶ jeden zestaw początkowych wartości wag
- ▶ *contractionConstant* = 1.2, 0.9, 0.6
- ▶ *minStateDiff* = 10e-08, 10e-07
- ▶ *minErrorDiff* = *minStateDiff*
- ▶ wybrana sieć nr 7 z poprzedniego eksperymentu

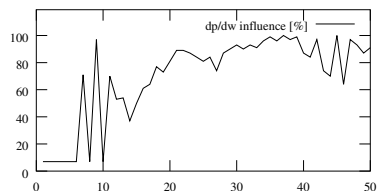
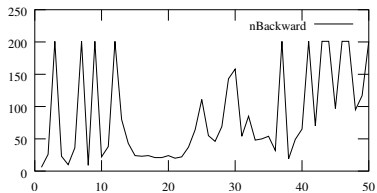
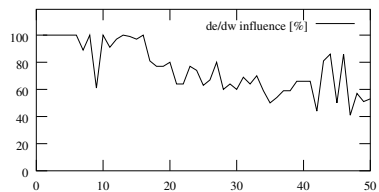
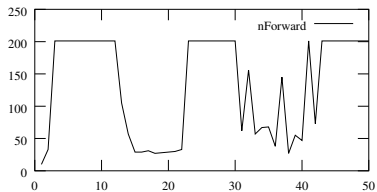
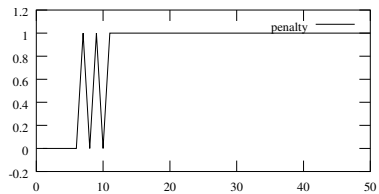
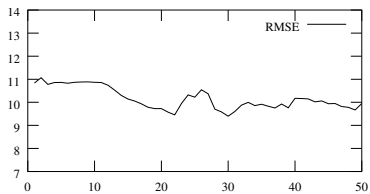


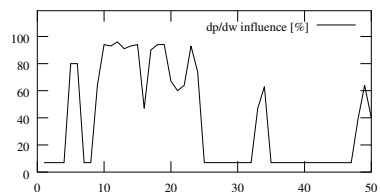
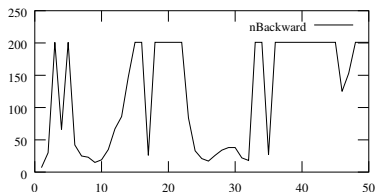
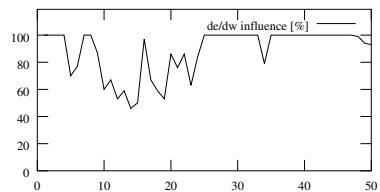
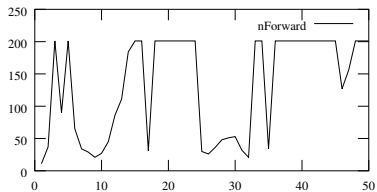
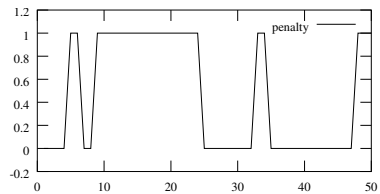
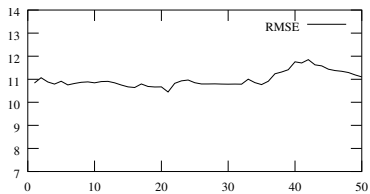


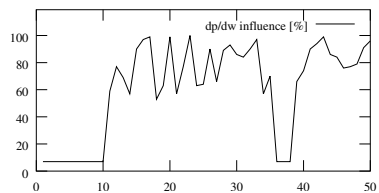
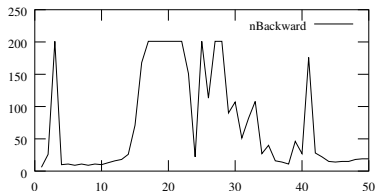
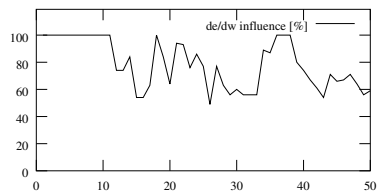
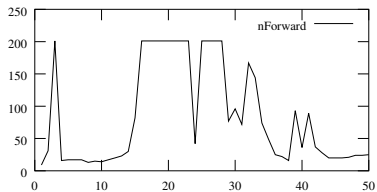
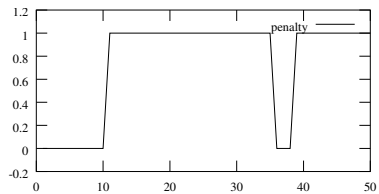
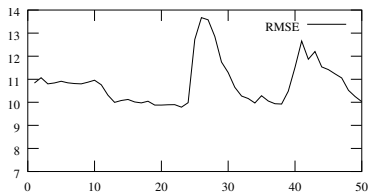












# Klasyfikacja węzłów - wyniki

	accuracy	precision	recall
TR - średnia	79%	83%	79%
TR - std	8%	4%	14%
TST - średnia	75%	76%	81%
TST - std	8%	10%	14%

Tabela: 5-krotna walidacja krzyżowa, najlepsza sieć z 10ciu, 200 iteracji

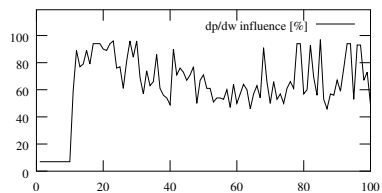
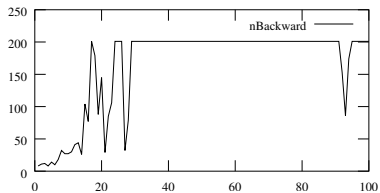
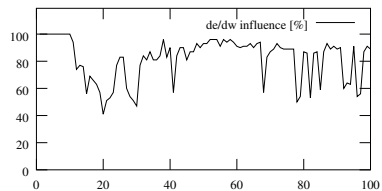
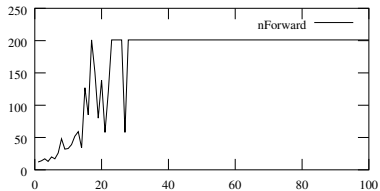
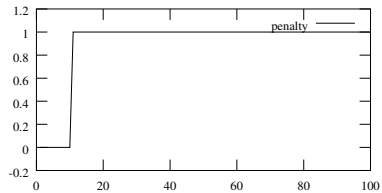
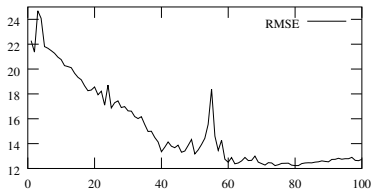
# Klasyfikacja węzłów - 100 grafów

	accuracy	precision	recall
GNN - TR	90%	88%	95%
GNN - TST	87%	84%	95%
FNN - TR	96%	94%	100%
FNN - TST	97%	95%	100%

**Tabela:** Klasyfikacja węzłów 100 iteracji

- ▶ zbiór uczący : 80 grafów
- ▶ zbiór testowy : 20 grafów
- ▶ wykorzystana GNN nr 5 z eksperymentu 1 (najlepsza)
- ▶ wykorzystana do porównania FNN z 5 neuronami ukrytymi





# Wnioski

- ▶ Najlepszy efekt przy zachowaniu kontrakcji
- ▶ Brak zbieżności kontrakcji – brak poprawy
- ▶ Balansowanie na granicy zbieżności / ograniczona ilość kroków obliczenia stanu i błędu dopuszczalne
- ▶ Należy unikać nakładania niepotrzebnie kary kontrakcji
- ▶ Konieczna duża dokładność w sprawdzaniu punktu stałego
- ▶ Duża wrażliwość modelu na parametry
- ▶ Można dobrać zestaw parametrów (i początkowych wag) na podzbiorze danych