



# Struktury bezskalowe w rekurencyjnych sieciach neuronowych

Filip Piękniewski

Wydział Matematyki i Informatyki  
UMK

Prezentacja na Seminarium Probabilistyczne KTHiAS dostępna na  
[http://www.mat.uni.torun.pl/~philip/sem\\_2007.pdf](http://www.mat.uni.torun.pl/~philip/sem_2007.pdf)

6 grudnia 2007



1

## Wstęp

- Grafy losowe, model Erdős'a-Rényi
- Model Small World Network (Watts-Strogatz)
- Sieci bezskalowe, model Barabási-Albert
- Rozkład Pareto

2

## Sieci neuronowe

- Perceptron
- Sieci Feed Forward Network
- Sieci rekurencyjne
- Model Hopfielda
- Sieci impulsujące
- Dynamiczne sieci impulsujące
- Model Eugene Izhikevicha

3

## Model przepływowy

- Grafy przepływu impulsów
- Model matematyczny i dowód bezskalowości
- Warianty modelu przepływowego
- Wyniki numeryczne
- Bezskalowość w sieciach dynamicznych?

4

## Podsumowanie

- Co dalej?
- Podsumowanie



# Grafy losowe, model Erdős'a-Rényi

- Ciekawym z punktu widzenia informatyki jest to jak wygląda przeciętny graf?
- Sformułowanie to wymaga określenia czym jest "przeciętny" graf!
- Oznacza to zdefiniowanie przestrzeni probabilistycznej, której elementami będą grafy, oraz zadania pewnej miary na tej przestrzeni.
- Jak to zrobić?



# Grafy losowe, model Erdős'a-Rényi

- Łatwo zdefiniować przestrzeń wszystkich grafów  $n$  wierzchołkowych i zadać na niej miarę jednostajną (osobną kwestią jest problem losowania z tego rozkładu)
- Wszystkich grafów jest jednak nieskończenie wiele, podobnie jak liczb rzeczywistych, nie da się zatem zadawać na nich miary poprzez przypisanie każdemu grafowi skończonego prawdopodobieństwa...
- Możliwe są inne podejścia do stworzenia modelu grafu losowego, każdy jednak model może być inny bo...
- ... do końca nie wiemy czym jest graf losowy...



# Grafy losowe, model Erdős'a-Rényi

- Podstawowym modelem grafu losowego był od lat model Erdős'a-Rényi  
[Erdős and Rényi(1959), Erdős and Rényi(1960)]
- Zakłada on następującą drogę wylosowania grafu:
  - Ustalamy pewną liczbą  $p \in (0, 1)$ , wybieramy liczbą naturalną  $n$  (losowo)
  - Pomiedzy każdą z  $n^2$  par wierzchołków dodajemy krawędź prawdopodobieństwem  $p$
- Model E-R jest przyzwoicie losowy, a jednocześnie na tyle prosty, iż dało się go łatwo zbadać



# Grafy losowe, model Erdős'a-Rényi

- Łatwo sprawdzić, iż histogram stopni wierzchołków w modelu E-R ma rozkład dwumianowy:

$$P(\deg(v) = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1)$$

który zbiega do rozkładu Poissona przy  $n \rightarrow \infty$

- Warto też odnotować przejścia fazowe w grafach E-R. Dla  $p < \frac{(1-\epsilon) \ln n}{n}$  graf prawie na pewno nie będzie spójny. Natomiast dla  $p > \frac{(1-\epsilon) \ln n}{n}$  praktycznie z pewnością będzie spójny.



- Model E-R ma jednak pewne wady które powodują, że nie nadaje się do symulowania grafów empirycznych
- Po pierwsze, grafy E-R nie przejawiają struktury lokalnej, nie posiadają dobrze zdefiniowanych ośrodków i skupisk
- Grafy empiryczne nie przejawiają rozkładu dwumianowego w histogramach stopni wierzchołków...



## Model Small World Network (Watts-Strogatz)

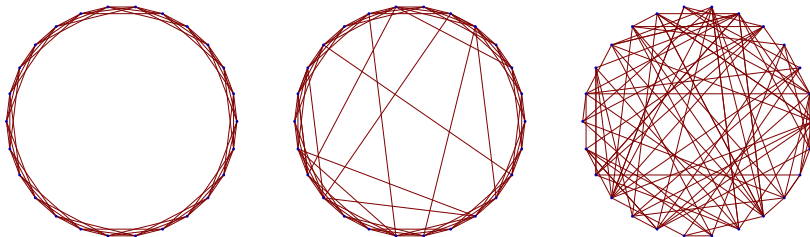
- Grafy empiryczne mają często dość dziwne cechy. Pomimo iż znaczna większość połączeń jest lokalnych (graf dziedziczy pewną strukturę odległości), to średnia długość drogi liczona w ilości krawędzi między dowolną parą wierzchołków jest bardzo niewielka, znacznie mniejsza niż w pełni lokalnym grafie
- Własność ta została zaobserwowana przez socjologów w strukturze znajomości w społeczeństwie ludzkim, nosi nazwę fenomenu sześciu uścisków dłoni
- Okazuje się, że przeciętnie pomiędzy dowolnymi dwoma osobami na świecie istnieje łańcuch sześciu osób które uściśnęły sobie dłonie



## Model Small World Network (Watts-Strogatz)

- W 1998 roku Watts i Strogatz [Watts and Strogatz(1998)] zauważyli, iż pewien bardzo prosty proces losowy prowadzi do powstania dobrze zlokalizowanych sieci z niewielką średnią odległością między wierzchołkami
- Model sprowadza się do następującego algorytmu:
  - Zaczynij od zlokalizowanej sieci, np. kraty lub pierścienia w którym każdy wierzchołek połączony jest z kilkoma najbliższymi sąsiadami
  - Następnie losowo pozamieniaj wierzchołki startowe pewnej ilości krawędzi
- Jak się okazuje, wystarczy zaingerować w kilka procent krawędzi, aby średnia odległość dramatycznie spadła

# Model Small World Network (Watts-Strogatz)



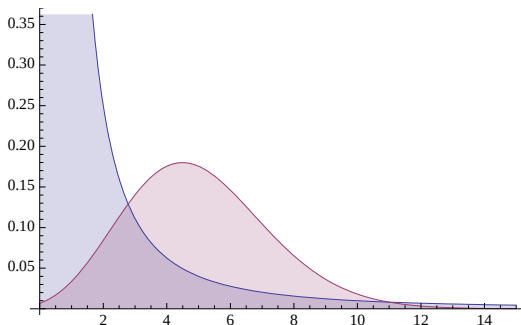
**Rysunek:** Od grafu zlokalizowanego (po lewo) do grafu losowego (pravo).  
Sieci małego świata są blisko grafów lokalnych, jednak niewielka ilość losowych połączeń istotnie skraca średnią drogę między wierzchołkami.



## Sieci bezskalowe, model Barabási-Albert

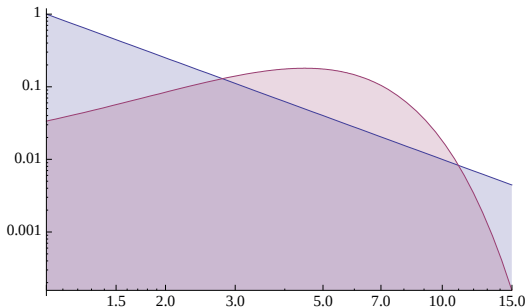
- Model W-S interpoluje między grafami z lokalną strukturą połączeń a losowymi grafami E-R
- Wiele grafów empirycznych, pozornie losowych jest raczej bliżej grafów zlokalizowanych niż E-R
- Sama lokalność i niewielka średnia odległość to jednak nie wszystko
- Wiele empirycznych grafów posiada bardzo specyficzny rozkład stopni wierzchołków, nie przypominający rozkładu dwumianowego

# Sieci bezskalowe, model Barabási-Albert



**Rysunek:** Rozkład stopni wierzchołków spotykany w grafach empirycznych (niebieski) oraz ten spodziewany w dużych grafach losowych w sensie Erdős'a-Rényi (różowy)

# Sieci bezskalowe, model Barabási-Albert



**Rysunek:** Rozkład stopni wierzchołków spotykany w grafach empirycznych (niebieski) oraz ten spodziewany w dużych grafach losowych w sensie Erdős'a-Rényi (różowy), wykres logarytmiczny.



## Sieci bezskalowe, model Barabási-Albert

- Grafy empiryczne okazują się posiadać rozkład stopni wierzchołków raczej przypominający rozkład Pareto, czyli  $\approx k^{-\gamma}$  dla pewnego  $\gamma > 2$ .
- Właściwość ta, nosi nazwę bezskalowości, gdyż zarówno małe jak i duże grafy tego typu parametryzują się za pomocą pewnego  $\gamma$ . Oznacza to, że rozkład stopni wierzchołków zasadniczo nie różni się dla małych i dla dużych instancji.
- Skoro tak, to oznacza, iż istnieje pewien proces przyrostu, który z małego grafu bezskalowego tworzy duży, jednocześnie zachowując parametr prawa potęgowego.
- Przez ładnych parę lat nie było wiadomo co to za mechanizm.



## Sieci bezskalowe, model Barabási-Albert

- W roku 1999 Albert-László Barabási oraz Reka Albert zaproponowali model [Barabási and Albert(1999)], który w naturalny sposób produkuje sieć bezskalową.
- Od czasu tej publikacji posypała się lawina prac z analizami rozmaitych grafów, w których obserwuje się strukturę bezskalową:
  - WWW [Albert et al.(1999)Albert, Jeong, and Barabási],
  - Sieć kolaboracji naukowej [Barabási et al.(2002)Barabási, Jeong, Néda, Ravasz, Schubert, and Vicsek]
  - Sieć cytowań [Redner(1998)],
  - Sieci ekologiczne [Montoya and V.(2002)],
  - Sieci lingwistyczne [i Cancho and Solé(2001)]
  - Sieci metabolizmu [Jeong et al.(2000)Jeong, Tombor, Albert, Oltvai, and Barabási]
  - I wiele innych [Albert and Barabási(2002b)].



## Sieci bezskalowe, model Barabási-Albert

- Model B-A jest bardzo prosty:
  - Założmy, że mamy już pewien graf bezskalowy (na początku jeden wierzchołek)
  - Dodajemy nowy wierzchołek i łączymy go z już obecnymi w ten sposób, iż preferowane są te wierzchołki które w poprzednim grafie miały wysokie stopnie. Losujemy zatem pewną ilość wierzchołków z którymi nowy zostanie połączony, z rozkładu proporcjonalnego do obecnego histogramu stopni wierzchołków.
- Można pokazać, że taki model zbiega do grafu bezskalowego z  $\gamma = 3$

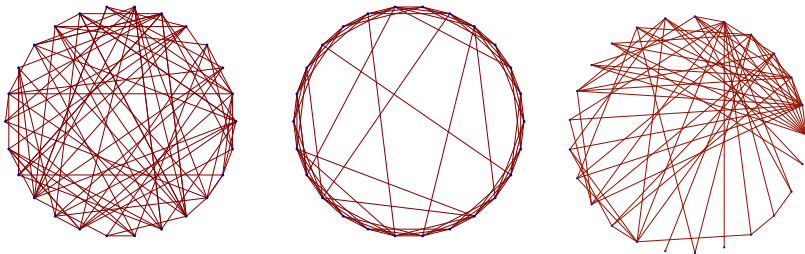


## Sieci bezskalowe, model Barabási-Albert

Sieci bezskalowe mają dość ciekawe cechy:

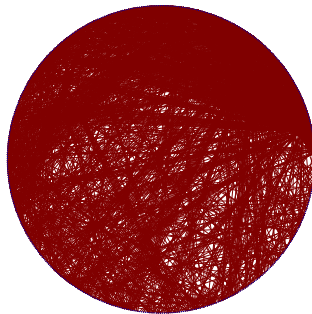
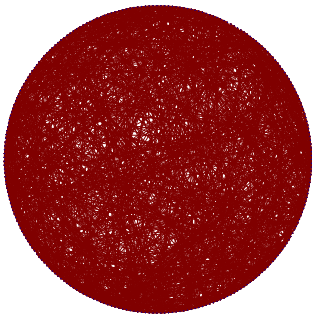
- Istnieje niewielka ilość węzłów o wysokich stopniach (hubów) oraz cała masa węzłów o niskich stopniach
- Są dobrze sklasteryzowane, zatem istnieje pewna struktura lokalna
- Przejawiają cechy sieci małego świata, ze względu na istnienie hubów, średnia droga między wierzchołkami jest bardzo niewielka
- Podobnie jak w życiu, gdy masz dużo łatwo ci zdobyć więcej. Gdy twoja praca jest już cytowana, łatwiej zdobyć dalsze cytowania...

# Sieci bezskalowe, model Barabási-Albert



**Rysunek:** Grafy: losowy E-R, małego świata W-S, bezskalowy B-A. W przypadku tego ostatniego rzuca się w oczy węzeł skupiający wiele połączeń, i spora liczba dość osamotnionych węzłów.

## Sieci bezskalowe, model Barabási-Albert



Rysunek: Większa sieć losowa (lewo) i bezskalowa (prawo).



## Rozkład Pareto

- Rozkłady bezskalowe ( $\gamma = k + 1$ ) to tak zwane rozkłady Pareto o następujących własnościach:

$$P(X > x) = \left( \frac{x}{x_m} \right)^{-k} \quad (2)$$

gęstości

$$f(x, k, x_m) = k \frac{x_m^k}{x^{k+1}} \quad (3)$$

Wartości oczekiwanej

$$E(X) = \frac{kx_m}{k-1} \quad (4)$$



## Rozkład Pareto

- Rozkłady bezskalowe ( $\gamma = k + 1$ ) to tak zwane rozkłady Pareto o następujących własnościach:

$$P(X > x) = \left( \frac{x}{x_m} \right)^{-k} \quad (2)$$

gęstości

$$f(x, k, x_m) = k \frac{x_m^k}{x^{k+1}} \quad (3)$$

Wartości oczekiwanej

$$E(X) = \frac{kx_m}{k-1} \quad (4)$$



## Rozkład Pareto

- Rozkłady bezskalowe ( $\gamma = k + 1$ ) to tak zwane rozkłady Pareto o następujących własnościach:

$$P(X > x) = \left(\frac{x}{x_m}\right)^{-k} \quad (2)$$

gęstości

$$f(x, k, x_m) = k \frac{x_m^k}{x^{k+1}} \quad (3)$$

Wartości oczekiwanej

$$E(X) = \frac{kx_m}{k-1} \quad (4)$$



## Rozkład Pareto

- Rozkłady bezskalowe ( $\gamma = k + 1$ ) to tak zwane rozkłady Pareto o następujących własnościach:

$$P(X > x) = \left(\frac{x}{x_m}\right)^{-k} \quad (2)$$

gęstości

$$f(x, k, x_m) = k \frac{x_m^k}{x^{k+1}} \quad (3)$$

Wartości oczekiwanej

$$E(X) = \frac{kx_m}{k-1} \quad (4)$$

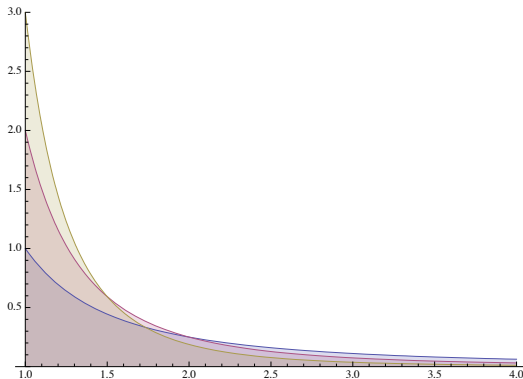


Spis treści  
Wstęp  
Sieci neuronowe  
Model przepływowy  
Podsumowanie  
Bibliografia

Grafy losowe, model Erdős'a-Rényi  
Model Small World Network (Watts-Strogatz)  
Sieci bezskalowe, model Barabási-Albert  
Rozkład Pareto



# Rozkład Pareto



Rysunek: Rozkład Pareto dla  $k = 1, 2, 3$ .



# Sieci Neuronowe

- Sieci neuronowe to dział informatyki rozwijający się od lat 50 nieco z boku głównego nurtu
- W dziedzinie tej istnieją dwa równoważne cele badawcze: zrozumieć w jaki sposób układy nerwowe zwierząt i człowieka prowadzą obliczenia oraz jak tą samą funkcjonalność przenieść na maszynę Turinga.
- Inspiracją wielu koncepcji w sieciach neuronowych są nierzadko brutalnie uproszczone modele działania neuronów. Często nawet jeśli model pojedynczej jednostki jest bardzo prosty, układ wielu takich jednostek posiada ciekawe możliwości.



# Sieci Neuronowe

- Z początku sieci neuronowe rozwijały się dość wolno, modele neuronu przypominały nieco bramki logiczne, wydawało się jednak, iż uzyskanie "ludzkiej" inteligencji na komputerze jest na wyciągnięcie ręki... później badania przyspieszyły zaś entuzjizm opadł...
- W latach 70 i 80 XX wieku, w czasie burzliwego rozwoju informatyki wokół sieci neuronowych pojawiły się działy takie jak *machine learning*, *fuzzy logic*, *data mining*. Dzisiaj można całą tą dziedzinę nazwać *soft computing*.
- W pobliżu sieci neuronowych pojawiła się nowa dziedzina *neuroscience*, która za przedmiot swoich badań objęła neurony biologiczne



# Neurony

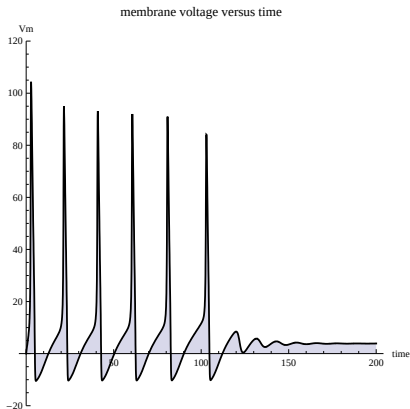
- Biologiczne neurony pomimo mikroskopijnych rozmiarów są skomplikowanymi urządzeniami elektrochemicznymi.
- Każde przekazanie sygnału elektrochemicznego wiąże się z falą skomplikowanych przemian błony komórkowej.
- W każdym punkcie błony przemiany te są opisywane skomplikowanym równaniem różniczkowym Hodgkina-Huxleya [Hodgkin and Huxley(1952)].
- W największym uproszczeniu można przyjąć, iż neuron przyjmuje sygnały z dendrytów i sumuje je. Jeśli zsumowany sygnał przekroczy pewien próg, neuron wzbudza się, wysyłając sygnał dalej. To brutalne uproszczenie stało się podstawą działania perceptronu.



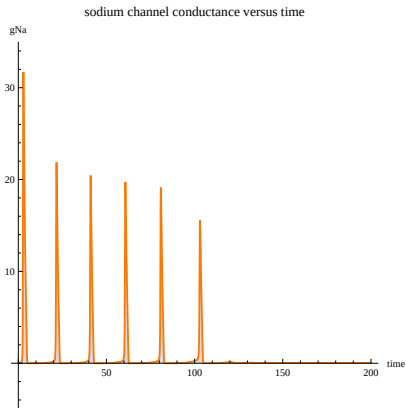
# Neurony

Równania Hodgkina-Huxleya:

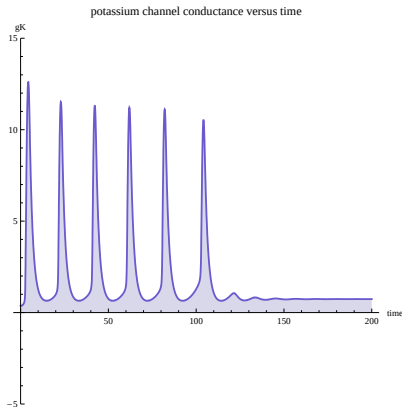
$$\begin{aligned}C_m \frac{dV}{dt} &= -g_L(V - V_L) - g_{Na}m^3h(V - V_{Na}) - g_Kn^4(V - V_K) \\ \frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m \\ \frac{dh}{dt} &= \alpha_h(V)(1 - h) - \beta_h(V)h \\ \frac{dn}{dt} &= \alpha_n(V)(1 - n) - \beta_n(V)n\end{aligned}$$



**Rysunek:** Przykładowy przebieg zmian potencjału na membranie neuronowej w odpowiedzi na sygnał 6.2 milivolta



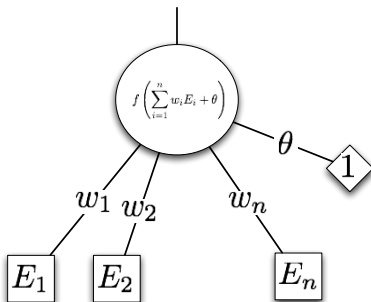
**Rysunek:** Przykładowy przebieg zmian przewodnictwa jonów sodu na membranie neuronowej w odpowiedzi na sygnał 6.2 milivolta



**Rysunek:** Przykładowy przebieg zmian przewodnictwa jonów potasu na membranie neuronowej w odpowiedzi na sygnał 6.2 milivolta



# Perceptron



**Rysunek:** Schemat perceptronu, za  $f$  można wziąć funkcję progową lub  $\tanh$ .

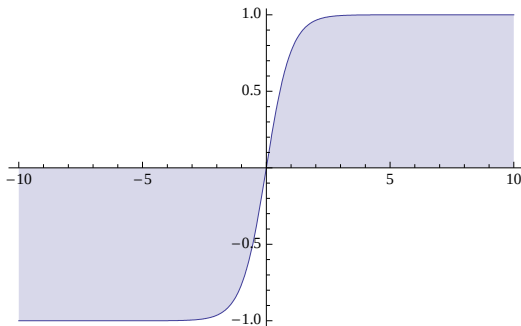


# Perceptron

- Pojedynczy perceptron potrafi (po odpowiednim dostosowaniu wag), rozwiązywać tak zwane problemy liniowo separowalne.
- Działanie tego prostego mechanizmu sprowadza się na wyznaczeniu płaszczyzny w przestrzeni wejść parametryzowanej wagami i progiem.
- Dla punktów wejścia znajdujących się z jednej strony tej płaszczyzny perceptron zwraca 1 (sygnał wzbudzenia), dla pozostałych -1 (sygnał braku wzbudzenia).
- Granice decyzji można nieco "rozmyć" poprzez zastosowanie sigmoidalnej funkcji aktywacji (zamiast progowej)

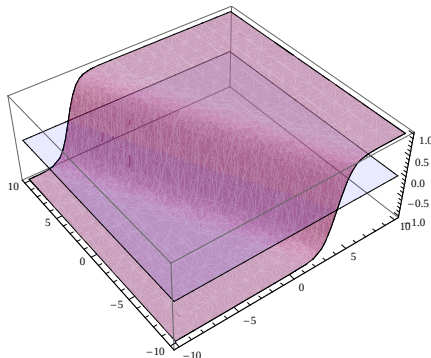


# Perceptron



Rysunek: Funkcja sigmoidalna, w tym przypadku tangens hiperboliczny.

# Perceptron



**Rysunek:** Geometryczne przedstawienie zadania realizowanego przez perceptron.

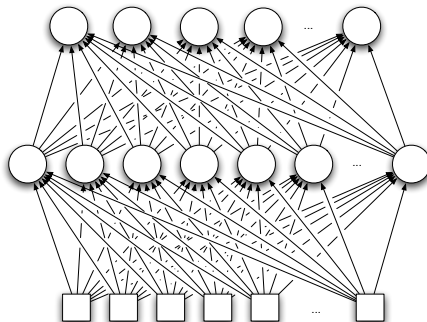


# Sieci Feed Forward Network

- Problem liniowo separowalny nie jest niczym ciekawym...
- Pojedynczy perceptron nie jest zbyt "inteligentny", ale być może siła w ilości?
- Okazuje się, że już sieć złożona z trzech warstw perceptronów jest w stanie rozwiązać po odpowiednim doborze wag i progów dowolny problem klasyfikacji.
- Jak jednak dobierać wagi i progi? To zagadnienie nosi nazwę "problemu uczenia" i wymaga zastosowania "algorytmu uczenia sieci"
- Problem w tym, że do połowy lat 70 XX wieku taki algorytm dla MLP (multi layer perceptron) nie był znany...



# Sieci Feed Forward Network



**Rysunek:** Schemat sieci MLP (Multi Layer Perceptron). Informacja w tej sieci biegnie z dołu do góry zawsze w jednym kierunku, stąd "Feed forward network".



# Sieci Feed Forward Network

- Sposobem na rozwiązanie trudnych problemów jest często wyrażenie ich w terminach optymalizacji.
- Dana sieć neuronowa to ostatecznie funkcja zależna od parametrów (wag i progów) która przykładowi przypisuje wartość aktywacji.
- Zbiór przykładów zawiera także docelowe aktywacje (dlatego nazwany jest zbiorem treningowym). Zbiór treningowy powinien być niesprzeczny.
- Przez  $O_W(e)$  oznaczmy aktywację sieci z macierzą wag  $W$  (wektor progów  $\theta$  można uwikłać w macierz  $W$  łącząc każdą jednostkę ze sztucznym wejściem stale równym jeden) na przykładzie  $e$ .



## Algorytm wstecznej propagacji

- Mając pewien zbiór przykładów  $(e_i, t_i)$  gdzie  $e_i$  jest wejściem do sieci a  $t_i$  oczekiwanym wyjściem możemy wyrazić problem uczenia sieci w terminach minimalizacji błędu średniokwadratowego:

$$\mathbb{E} = \sum_{i=1}^n ||O_W(e_i) - t_i|| \quad (5)$$

- Następnie aplikujemy algorytm spadku gradientowego na funkcji  $\mathbb{E}$ :

$$W := W - \eta \left[ \frac{\partial \mathbb{E}}{\partial w_{i,j}} \right]_{i=1 \dots n, j=1 \dots m} \quad (6)$$



## Algorytm wstecznej propagacji

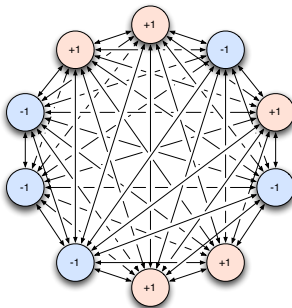
- Od tej pory cały problem sprowadza się do efektywnego policzenia pochodnej względem wag...
- W ogólności przy wielu tysiącach zmiennych policzenie gradientu jest bardzo trudne i niestabilne numerycznie...
- Jeśli sieć ma charakter jednokierunkowy, można jednak bardzo łatwo wyliczyć pochodną, poprzez proces w którym w odpowiedni sposób przesyłamy błąd jako aktywację do tyłu
- Algorytm ten nosi nazwę "algorytmu wstecznej propagacji" i jest bez wątpienia perełką
- Wsteczną propagację zawdzięczamy Paulowi Werbosowi [Werbos(1974)], który zaproponował tą metodę w swojej pracy doktorskiej w 1974 roku. Algorytm został zapomniany i na nowo odkryty w latach 80 przez Rumelharta, Hintona i Williamsa [Rumelhart et al.(1986)Rumelhart, Hinton, and Williams].



## Sieci rekurencyjne

- Uczenie algorytmem wstecznej propagacji o ile w sieci nie ma połączeń zwrotnych (rekurencji)
- Jest to duże ograniczenie, gdyż biologiczne sieci są często prawie całkowicie rekurencyjne...
- Ta sytuacja wymaga trochę innego podejścia, zahaczającego o mechanikę statystyczną ...

# Sieci rekurencyjne



**Rysunek:** Przykład sieci w pełni rekurencyjnej. Właściwie trudno nawet określić które jednostki są wejściowe a które wyjściowe. Można ustalić, że cała sieć jest wejściem i wyjściem, wtedy wynikiem działania sieci może być pewien stan równowagi.



## Sieci rekurencyjne

- W roku 1982 J. Hopfield [Hopfield(1982)] zaproponował prosty model rekurencyjnej sieci neuronowej wraz ze skutecznym algorytmem uczenia opartym o tak zwaną regułę Hebba.
- Reguła Hebba [Hebb(1949)] była używana już wcześniej w tak zwanym uczeniu bez nadzoru
- Reguła ta sprowadza się do obserwacji, iż w biologicznych neuronach synapsy często używane do wymiany sygnałów się rozrastają, te zaś które nie są używane zmniejszają się.
- W języku sztucznych sieci neuronowych tłumaczy się to na wzmacnianie połączeń pomiędzy skorelowanymi jednostkami



# Model Hopfielda

- Sieć składa się z  $n$  jednostek binarnych, każda jednostka przyjmuje wartości ze zbioru  $\{-1, 1\}$
- Wszystkie jednostki są połączone, wagi w układzie zadaje macierz  $W = [w_{i,j}]$ , zakładamy że  $w_{i,i} = 0$  oraz  $w_{i,j} = w_{j,i}$
- Z siecią stowarzyszona jest funkcja energetyczna:

$$\mathbb{E} = -\frac{1}{2} \sum_{i < j} w_{i,j} \sigma_i \sigma_j + \sum_i \theta_i \sigma_i \quad (7)$$

Pierwszy czynnik sumy zależy od wewnętrznych interakcji sieci, drugi zwany jest polem zewnętrznym.



## Model Hopfielda

- W trakcie swojej ewolucji sieć startuje ze stanu początkowego a następnie w każdym kroku:
  - wybieramy losowo jednostkę  $\sigma_i$
  - obliczamy pole lokalne dla tej jednostki czyli

$$\Sigma_i = \sum_j w_{j,i} \sigma_j \quad (8)$$

- Sprawdzamy czy  $\Sigma_i > \theta_i$ . Jeśli tak, ustawiamy  $\sigma_i = 1$ , gdy  $\Sigma_i < \theta_i$  ustawiamy  $\sigma_i = -1$ , zaś gdy  $\Sigma_i = \theta_i$  pozostawiamy sieć bez zmian.
- Powyższa dynamika w wersji asynchronicznej (naraz uaktualniamy tylko jedną jednostkę) nigdy nie zwiększa globalnej energii układu, a zatem zbiega do pewnego minimum.



## Model Hopfielda

Oznaczmy przez  $\bar{\sigma}$  stan wyjściowy sieci, a przez  $\bar{\sigma}'$  stan po dokonaniu zmiany przez dynamikę. Załóżmy, że zmieniony został neuron  $i$ :

$$\mathbb{E}(\bar{\sigma}') - \mathbb{E}(\bar{\sigma}) = -\frac{1}{2} \sum_j w_{i,j} \sigma_j (\sigma'_i - \sigma_i) - \theta_i (\sigma'_i - \sigma_i) \leq -(\sigma'_i - \sigma_i) (\Sigma_i - \theta_i) \leq 0 \quad (9)$$

Jeśli założymy, że przy  $\Sigma_i = \theta_i$  nie następuje zmiana sieci, to ostatnia nierówność jest ostra. Ponieważ energia jest funkcją ograniczoną, zatem w trakcie dynamiki asynchronicznej osiągany jest pewne minimum (lokalne). Zatem problem uczenia sieci, sprowadza się do odpowiedniego zaprojektowania krajobrazu energetycznego...



## Model Hopfielda

Oznaczmy przez  $\bar{\sigma}$  stan wyjściowy sieci, a przez  $\bar{\sigma}'$  stan po dokonaniu zmiany przez dynamikę. Załóżmy, że zmieniony został neuron  $i$ :

$$\mathbb{E}(\bar{\sigma}') - \mathbb{E}(\bar{\sigma}) = -\frac{1}{2} \sum_j w_{i,j} \sigma_j (\sigma'_i - \sigma_i) - \theta_i (\sigma'_i - \sigma_i) \leq -(\sigma'_i - \sigma_i) (\Sigma_i - \theta_i) \leq 0 \quad (9)$$

Jeśli założymy, że przy  $\Sigma_i = \theta_i$  nie następuje zmiana sieci, to ostatnia nierówność jest ostra. Ponieważ energia jest funkcją ograniczoną, zatem w trakcie dynamiki asynchronicznej osiągany jest pewne minimum (lokalne). Zatem problem uczenia sieci, sprowadza się do odpowiedniego zaprojektowania krajobrazu energetycznego...



## Model Hopfielda

Oznaczmy przez  $\bar{\sigma}$  stan wyjściowy sieci, a przez  $\bar{\sigma}'$  stan po dokonaniu zmiany przez dynamikę. Załóżmy, że zmieniony został neuron  $i$ :

$$\mathbb{E}(\bar{\sigma}') - \mathbb{E}(\bar{\sigma}) = -\frac{1}{2} \sum_j w_{i,j} \sigma_j (\sigma'_i - \sigma_i) - \theta_i (\sigma'_i - \sigma_i) \leq -(\sigma'_i - \sigma_i)(\Sigma_i - \theta_i) \leq 0 \quad (9)$$

Jeśli założymy, że przy  $\Sigma_i = \theta_i$  nie następuje zmiana sieci, to ostatnia nierówność jest ostra. Ponieważ energia jest funkcją ograniczoną, zatem w trakcie dynamiki asynchronicznej osiągany jest pewne minimum (lokalne). Zatem problem uczenia sieci, sprowadza się do odpowiedniego zaprojektowania krajobrazu energetycznego...



# Model Hopfielda

- Ukształtowanie krajobrazu energetycznego sprowadza się do ustalenia macierzy wag
- Załóżmy, iż chcemy, aby minima energetyczne przypadwały w określonych konfiguracjach wzorcowych  $\mathbb{I}^\mu = \{\zeta_i^\mu\}$  sieci
- Ustalmy miarę podobieństwa konfiguracji sieci  $\bar{\sigma}$  do wzorca  $\mu$ :

$$\mathbb{M}^\mu(\bar{\sigma}) = \frac{1}{N} \sum_i \sigma_i \zeta_i^\mu = \frac{1}{N} \langle \bar{\sigma}, \mathbb{I}^\mu \rangle \quad (10)$$

- Ustalmy funkcję energetyczną jako błąd średniokwadratowy odtwarzania wzorca:

$$\mathbb{E}(\bar{\sigma}) = \frac{1}{2} N \sum_\mu (\mathbb{M}^\mu(\bar{\sigma}))^2 \quad (11)$$



# Model Hopfielda

Wtedy:

$$\mathbb{E}(\bar{\sigma}) = \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j - \frac{1}{2} P \equiv \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j \quad (12)$$

Ustalając:

$$\frac{1}{N} \sum_{\mu} \zeta_i^{\mu} \zeta_j^{\mu} = w_{i,j} \quad (13)$$

Mamy:

$$\mathbb{E} = \frac{-1}{2} \sum_{i,j} \sigma_i \sigma_j w_{i,j} \quad (14)$$

Zatem wagi zdefiniowane regułą Hebba ustalają dobry krajobraz energetyczny.



# Model Hopfielda

Wtedy:

$$\mathbb{E}(\bar{\sigma}) = \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j - \frac{1}{2} P \equiv \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j \quad (12)$$

Ustalając:

$$\frac{1}{N} \sum_{\mu} \zeta_i^{\mu} \zeta_j^{\mu} = w_{i,j} \quad (13)$$

Mamy:

$$\mathbb{E} = \frac{-1}{2} \sum_{i,j} \sigma_i \sigma_j w_{i,j} \quad (14)$$

Zatem wagi zdefiniowane regułą Hebba ustalają dobry krajobraz energetyczny.



# Model Hopfielda

Wtedy:

$$\mathbb{E}(\bar{\sigma}) = \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j - \frac{1}{2} P \equiv \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j \quad (12)$$

Ustalając:

$$\frac{1}{N} \sum_{\mu} \zeta_i^{\mu} \zeta_j^{\mu} = w_{i,j} \quad (13)$$

Mamy:

$$\mathbb{E} = \frac{-1}{2} \sum_{i,j} \sigma_i \sigma_j w_{i,j} \quad (14)$$

Zatem wagi zdefiniowane regułą Hebba ustalają dobry krajobraz energetyczny.



# Model Hopfielda

Wtedy:

$$\mathbb{E}(\bar{\sigma}) = \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j - \frac{1}{2} P \equiv \frac{-1}{2N} \sum_{\mu} \sum_{i,j} \sigma_i \zeta_i^{\mu} \zeta_j^{\mu} \sigma_j \quad (12)$$

Ustalając:

$$\frac{1}{N} \sum_{\mu} \zeta_i^{\mu} \zeta_j^{\mu} = w_{i,j} \quad (13)$$

Mamy:

$$\mathbb{E} = \frac{-1}{2} \sum_{i,j} \sigma_i \sigma_j w_{i,j} \quad (14)$$

Zatem wagi zdefiniowane regułą Hebba ustalają dobry krajobraz energetyczny.



# Model Hopfielda

- Sieci Hopfielda mają tendencję do popadania w minima lokalne, aby tego uniknąć stosuje się wersję stochastyczną dynamiki (maszyna Boltzmanna [Ackley et al.(1985)Ackley, Hinton, and Sejnowski])
- Model Hopfielda znajduje zastosowanie w rozpoznawaniu wzorców, a także jako model prostych układów nerwowych dla robotów, biotów, systemów agentowych
- Prawdziwe neurony nie działają w sposób asynchroniczny, faktycznie zaś przesyłają impulsy których przesunięcia czasowe mogą mieć istotne znaczenie dla przetwarzanego sygnału.
- Dotychczas wymienione modele neuronowe zupełnie ignorują to zjawisko...



## Sieci impulsujące

- Trzecią generacją modeli neuronowych są tak zwane sieci impulsujące (*spiking neural network*) opisane obszernie w [Gerstner and Kistler(2002)].
- Zagadnienie sprowadza się do mniej lub bardziej dokładnego modelowania dynamiki błony komórkowej z uwzględnieniem opóźnień i innych czynników wpływających na impulsy
- Najprostszą wersją neuronu impulsującego jest tzw. *integrate and fire*. Neuron całkuje po czasie nadchodzący sygnał, gdy całko osiągnie pewien próg, wysyła własny sygnał, zerując zliczaną całkę (ewentualnie wchodząc w okres refrakcji poprzez przyjęcie ujemnej wartości całki).
- Minimalnym ulepszeniem jest model *leaky integrate and fire* w którym część zbieranego potencjału umyka



## Dynamiczne sieci impulsujące

- Model *integrate and fire* nie przejawia żadnych oscylacji, tymczasem o neuronach wiadomo, iż ich odpowiedź impulsowa nie zależy jedynie od mocy sygnału, ale także od częstotliwości i jej synchronizacji z wewnętrznym *rytmem* komórki.
- Tego typu przekształcenie nie jest realizowalne za pomocą prostego splotu czy podobnych mechanizmów, wymaga zastosowania układu równań różniczkowych o odpowiednio bogatej przestrzeni fazowej
- Punktem wyjścia może być równanie Hodgkina-Huxleya, które pasuje do wyników empirycznych, jednak jest ono dość skomplikowane i (prawdopodobnie) nadmiernie sparametryzowane

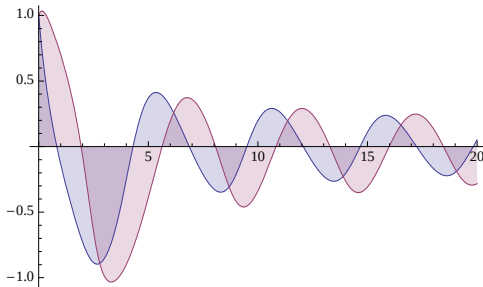
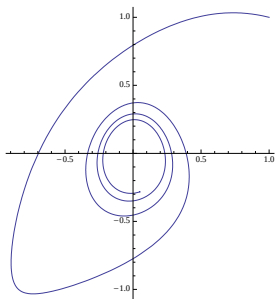


## Dynamiczne sieci impulsujące

- Z pewnością niezbędnym minimum do wygenerowania oscylacji jest układ dwóch równań różniczkowych, jedno określające odpowiednik potencjału membrany, oraz drugi składnik wolny, odpowiadający przewodności kanałów jonowych
- Przestrzenią fazową takiego układu jest płaszczyzna, stan komórki zatacza na tej płaszczyźnie cykle
- Nadchodzący sygnał wejściowy na bieżąco wpływa na trajektorie układu



## Dynamiczne sieci impulsujące



**Rysunek:** Trajektorie w przestrzeni fazowej pewnego dwu parametrowego układu (lewo), wykres zmian poszczególnych parametrów układu (prawo).



## Dynamiczne sieci impulsujące

- Układy dynamiczne, zależnie od parametrów posiadają pewne ulubione częstotliwości wzbudzające i inne wygaszające...
- Nawet niewielkie pobudzenie trafiające w neuron w odpowiednim stanie może wywołać silne pobudzenie ujawniające się szeregiem impulsów
- Silne wzbudzenie trafiające w nieodpowiednim momencie może minąć bez echa...



## Model FitzHugh-Nagumo

- Równania Hodgkina-Huxleya są skomplikowane i mają dużo zmiennych. Już w latach 60-tych XX wieku trwały prace nad stworzeniem modelu, który dziedziczyłby większość cech neurodynamiki, a jednocześnie aby był prosty i minimalistyczny
- Prace te doprowadziły Richarda FitzHuga [FitzHugh(1961), Nagumo et al.(1962)Nagumo, Arimoto, and Yoshizawa] do ogólnej koncepcji uproszczonego neuronu, dla którego następnie został podany równoważny układ elektryczny przez Jin-Ichi Nagumo. Model ten ma ogólną formę:

$$\begin{aligned}\frac{dV}{dt} &= f(V) - U + I \\ \frac{dU}{dt} &= a(bV - cU)\end{aligned}\tag{15}$$

gdzie  $f$  jest wielomianem trzeciego stopnia,  $a, b, c$  są parametrami zaś  $I$  jest sygnałem wejściowym.



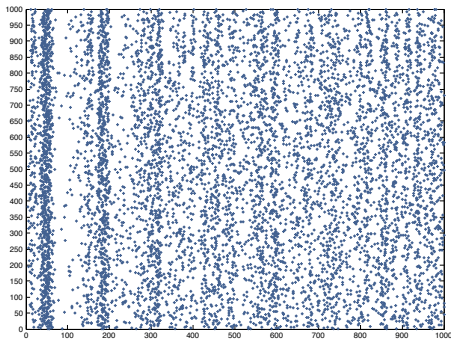
## Model Eugene Izhikevicha

- Eksperymentując z modelem Fize-Hugh-Nagumo w 2003 roku Eugene Izhikevich zaproponował jeszcze bardziej uproszczony model [Izhikevich(2003)].
- Wyraża się on następującym równaniem:

$$\begin{aligned}\frac{dV}{dt} &= 0.04V^2 + 5V + 140 - U + I \\ \frac{dU}{dt} &= a(bV - U)\end{aligned}\tag{16}$$

Przy czym gdy potencjał  $V$  przekroczy 30mV (impuls), cały układ jest resetowany i  $V := c$ ,  $U := U + d$ . Zależnie od doboru parametrów  $a, b, c, d$  układ przejawia wszystkie znane rodzaje aktywności neuronowej.

# Model Eugene Izhikevicha



**Rysunek:** Wykres aktywności 1000 losowo połączonych jednostek Izhikevicha. Widać fazy globalnej synchronizacji. Tego typu synchronizacje są źródłem fal mózgowych.



# Model przepływowy

- Jak dotychczas dwie dziedziny - grafy losowe oraz sieci neuronowe rozwijały się niezależnie
- W 2005 roku wraz z Tomaszem Schreiberem postanowiliśmy rozpocząć projekt który związałyby ze sobą te dwie dziedziny
- Nasze intuicje kierowały nas na sieci bezskalowe, gdyż mnogość ich występowania w naturze sugeruje iż są ważnym pojęciem
- Dotychczasowe badania empiryczne biologicznych sieci neuronowych nie wykazywały struktury bezskalowej w połączeniach  
[Amaral et al.(2000)Amaral, Scala, Barthelemy, and Stanley, Koch and Laurent(1999)]
- Postanowiliśmy zbadać dlaczego, a jednocześnie zaprezentować modele Hopfielda podobne, w których bezskalość spontanicznie się objawia



# Model przepływowy

- Potrzebowaliśmy modelu, który łączyłby cechy modelu Hopfielda i sieci impulsujących
- Powinien być szybki, gdyż interesujące nas zjawiska występują dopiero w instancjach mających tysiące jednostek
- Nasz model nie był tworzony z myślą o rozwiązywaniu konkretnych problemów algorytmicznych, raczej jako fundament syntezy dziedzin
- Ostatecznie powstał model, który można określić jako dyskretna sieć impulsująca, nazwaliśmy go roboczo *model przepływowy*

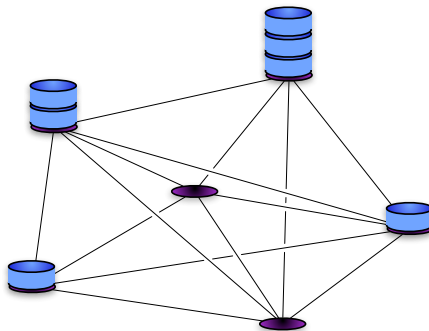


## Model przepływowy

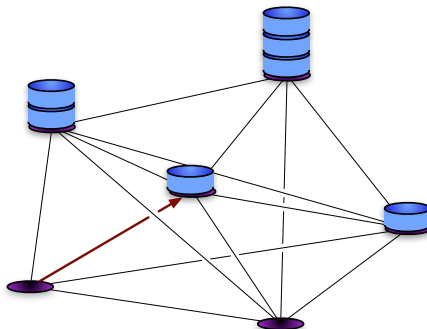
- Model składa się z  $n$  jednostek  $\sigma_i \in \mathbb{N}$ , połączonych losowo z wagami gaussowskimi.
- Model wyposażony jest funkcję energetyczną postaci:

$$\mathbb{E}(\vec{\sigma}) := \frac{1}{2} \sum_{i \neq j} w_{ij} |\sigma_i - \sigma_j| \quad (17)$$

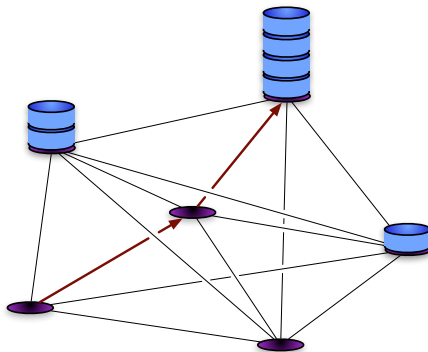
- Dynamika modelu przedstawia się następująco:
  - Losujemy dwie jednostki  $\sigma_i, \sigma_j$ . Sprawdzamy czy  $\sigma_i > 0$
  - Sprawdzamy czy przeniesienie jednostki *potencjału* z  $\sigma_i$  do  $\sigma_j$  zmniejsza energię
  - Jeśli tak, akceptujemy zmianę, w przeciwnym wypadku przywracamy oryginalny stan



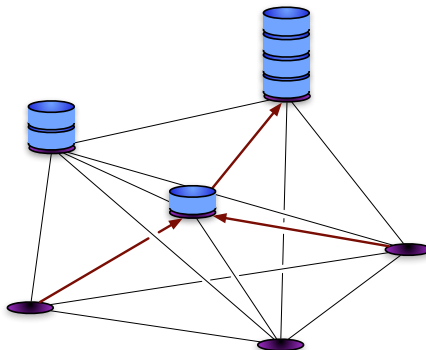
**Rysunek:** Schemat modelu przepływowego, jednostki wymieniają tokeny (niebieskie dropsy).



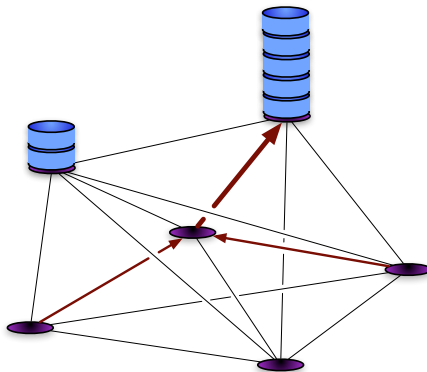
**Rysunek:** Schemat modelu przepływowego, jednostki wymieniają tokeny (niebieskie dropsy).



**Rysunek:** Schemat modelu przepływowego, jednostki wymieniają tokeny (niebieskie dropsy).



**Rysunek:** Schemat modelu przepływowego, jednostki wymieniają tokeny (niebieskie dropsy).



**Rysunek:** Schemat modelu przepływowego, jednostki wymieniają tokeny (niebieskie dropsy).



# Model przepływowy

- Model przepływowy w tym sformułowaniu nie rozwiązuje żadnego konkretnego problemu
- Dziedziczy cechy modelu Hopfielda, ma asynchroniczną dynamikę
- Jednocześnie jest to model impulsujący, przepływ jednostki potencjału można traktować jako impuls nerwowy
- Każda jednostka ma nieograniczony zbiór stanów zamiast zbioru  $\{-1, 1\}$



# Model przepływowy

- Łatwo zauważyć, że w modelu przepływowym dodatnie wagi wspomagają zgodność jednostek, ujemne zaś niezgodność
- Wydaje się zatem, że struktura optimów energetycznych tego układu jest bardzo skomplikowana
- Tak jednak nie jest, w modelu przepływowym unikatowym stanem minimalnej energii jest przypadek w którym jedna jednostka o najlepszej konfiguracji wag względem reszty przechowuje całą energię
- Aby tego dowieść potrzebujemy pewnych faktów i pomocniczych pojęć



# Model przepływowy

- Dla każdej jednostki  $\sigma_i$  zdefiniujemy *wsparcie* następująco:

$$S_i := - \sum_{j \neq i} w_{i,j} \quad (18)$$

- Wagi  $w_{i,j}$  były wylosowane z rozkładu normalnego o wariancji 1 i wartości oczekiwanej 0, zatem  $S_i$  są zmiennymi losowymi o rozkładzie normalnym z wariancją  $N - 1$ , w dodatku  $S_i$  są *praktycznie* niezależne (każdy  $S_i$  z innym współdzieli tylko jeden wyraz sumy)
- Oznaczmy przez  $S_{:k}$ ,  $k$ -tą największą wartość spośród  $S_i$



## Model przepływowy

- Z teorii wartości ekstremalnych wiadomo, iż  $S_{:k}$  może być przybliżany przez

$$S_{:k} \approx \sqrt{\frac{2N}{\log_2(N)}} \left( \log 2 \log_2 N - \log \sqrt{\log_2 N} \frac{\xi_k}{2\sqrt{\log 2}} \right) \quad (19)$$

przy czym ciąg  $0 < \xi_1 < \xi_2 < \dots$  jest wybrany z procesu punktowego Poissona, z intensywnością  $\frac{1}{\pi} \exp(-t)$ ,  $t > 0$  oraz  $\xi_k$  jest rzędu  $\log k$

- Wybierzmy niewielki ułamek  $o(N)$  wszystkich jednostek które przechowują najwięcej potencjału. Nazwijmy tę grupę elitą, zaś pozostałych tłumem.



## Model przepływowy

- Z teorii wartości ekstremalnych wiadomo, iż  $S_{:k}$  może być przybliżany przez

$$S_{:k} \approx \sqrt{\frac{2N}{\log_2(N)}} \left( \log 2 \log_2 N - \log \sqrt{\log_2 N} \frac{\xi_k}{2\sqrt{\log 2}} \right) \quad (19)$$

przy czym ciąg  $0 < \xi_1 < \xi_2 < \dots$  jest wybrany z procesu punktowego Poissona, z intensywnością  $\frac{1}{\pi} \exp(-t)$ ,  $t > 0$  oraz  $\xi_k$  jest rzędu  $\log k$

- Wybierzmy niewielki ułamek  $o(N)$  wszystkich jednostek które przechowują najwięcej potencjału. Nazwijmy tę grupę elitą, zaś pozostałych tłumem.



# Model przepływowy

- Ponieważ elita jest bardzo niewielka (jej udział zbiega do zera przy  $N \rightarrow \infty$ ), zatem energia przybiera wygodną formę:

$$\mathbb{E}(\bar{\sigma}) \approx - \sum_{i \in \text{elite}} \sigma_i S_i + \frac{1}{2} \sum_{j, l \in \text{bulk}} w_{j, l} |\sigma_j - \sigma_l| \quad (20)$$

- Wynika z tego, że za każdym razem w gdy w toku dynamiki proponowany jest przepływ z elementu tłumu  $\sigma_j$  do elitarnego elementu  $\sigma_i$ , spodziewana zmiana energii jest przybliżana przez  $-S_i$  plus czynnik związany z interakcją tłum-tłum. Generalnie nie mamy kontroli nad tym dodatkowym czynnikiem, jednak jeśli  $\sigma_i$  jest jednym z neuronów o największym wsparciu, wtedy jest prawie pewne, iż dodatkowy czynnik będzie zaniedbywalny w porównaniu z  $-S_i$ .



# Model przepływowy

- Ponieważ elita jest bardzo niewielka (jej udział zbiega do zera przy  $N \rightarrow \infty$ ), zatem energia przybiera wygodną formę:

$$\mathbb{E}(\bar{\sigma}) \approx - \sum_{i \in \text{elite}} \sigma_i S_i + \frac{1}{2} \sum_{j, l \in \text{bulk}} w_{j, l} |\sigma_j - \sigma_l| \quad (20)$$

- Wynika z tego, że za każdym razem w gdy w toku dynamiki proponowany jest przepływ z elementu tłumy  $\sigma_j$  do elitarnego elementu  $\sigma_i$ , spodziewana zmiana energii jest przybliżana przez  $-S_i$  plus czynnik związany z interakcją tłum-tłum. Generalnie nie mamy kontroli nad tym dodatkowym czynnikiem, jednak jeśli  $\sigma_i$  jest jednym z neuronów o największym wsparciu, wtedy jest prawie pewne, iż dodatkowy czynnik będzie zaniedbywalny w porównaniu z  $-S_i$ .



# Model przepływowy

- Ponieważ elita jest bardzo niewielka (jej udział zbiega do zera przy  $N \rightarrow \infty$ ), zatem energia przybiera wygodną formę:

$$\mathbb{E}(\bar{\sigma}) \approx - \sum_{i \in \text{elite}} \sigma_i S_i + \frac{1}{2} \sum_{j, l \in \text{bulk}} w_{j, l} |\sigma_j - \sigma_l| \quad (20)$$

- Wynika z tego, że za każdym razem w gdy w toku dynamiki proponowany jest przepływ z elementu tłumu  $\sigma_j$  do elitarnego elementu  $\sigma_i$ , spodziewana zmiana energii jest przybliżana przez  $-S_i$  plus czynnik związany z interakcją tłum-tłum. Generalnie nie mamy kontroli nad tym dodatkowym czynnikiem, jednak jeśli  $\sigma_i$  jest jednym z neuronów o największym wsparciu, wtedy jest prawie pewne, iż dodatkowy czynnik będzie zaniedbywalny w porównaniu z  $-S_i$ .



# Model przepływowy

- Jeśli w elicie znajdzie się element o wysokim wsparciu, wtedy przepływy w jego kierunku stają się bardzo prawdopodobne, zaś odwrotne praktycznie niemożliwe. Zatem z wielkim prawdopodobieństwem taki element pozostanie w elicie tracąc swój potencjał jedynie dla kolegów z elity o jeszcze wyższym wsparciu.
- Jeśli na początku ewolucji pewien neuron o niskim wsparciu, jego potencjał zostanie szybko wyspany, gdy tylko do elity wejdzie element o większym wsparciu.
- Ostatecznie w tłumie nie będzie już praktycznie w ogóle potencjału, wtedy Hamiltonian przyjmuje postać:

$$\mathbb{E}(\bar{\sigma}) \approx - \sum_{i \in \text{elite}} \sigma_i S_i \quad (21)$$

i wszystkie dalsze transfery będą się odbywały wewnątrz elity.



# Model przepływowy

- Ponieważ różnice energii wynikające z wag połączeń między elementami elity są zaniedbywalnie małe, w porównaniu z czynnikiem pochodzącym od wsparcia, więc na samym końcu dynamika układu przyjmuje prostą formę
- Wybierane są dwa losowe elementy elity, następnie następuje (prawie na pewno) przepływ potencjału od tego z mniejszym wsparciem do tego z większym wsparciem
- Ostatecznie na placu boju pozostaje jedna jednostka skupiająca cały potencjał



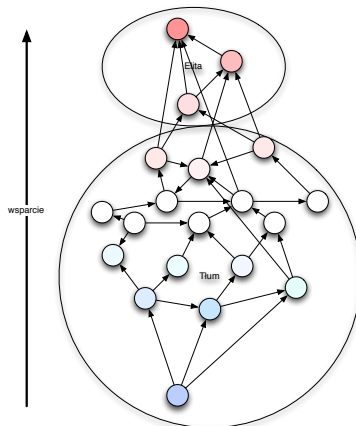
# Grafy przepływu impulsów

- To co nas najbardziej interesowało w modelu przepływowym, oprócz tego dokąd zmierza potencjał, było to którądy następują przepływy
- Przepływy te indukują pewien graf, zwany grafem przepływowym o zbiorze wierzchołków identycznym ze zbiorem neuronów, oraz krawędziami dokładnie dam gdzie nastąpił przepływ.
- Krawędzi na których przepływ nastąpił wielokrotnie traktujemy jako multikrawędzie, lub krawędzie z wagą.
- Obserwacje numeryczne wskazywały, iż grafy przepływu dla modelu przepływowego mają strukturę bezskalową z  $\gamma \approx 2$ , naszym następnym celem badań było zanalizować ten fakt i podać jego formalne wyjaśnienie.



# Grafy przepływu impulsów

Schemat grafu przepływowego. W obrębie elity, potencjał płynie praktycznie zawsze w kierunku rosnącego wsparcia (interakcje elita-elita są zaniedbywalne), w tłumie poszczególne wagi mają większe znaczenie.





# Model matematyczny i dowód bezskalowości

- Ustalmy  $K \ll N$  oznaczające rozmiar elity. Niech  $K \rightarrow \infty$  przy  $N \rightarrow \infty$
- Niech  $u_i$ ,  $i = 1 \dots K$  będą jednostkami elity, w kolejności malejącego wsparcia
- Jednostka potencjału wpada do elity w jednostce  $u_{k_0}$ , a następnie pnie się po  $u_{k_l}$ ,  $k_{l+1} < k_l$  (losowo) aż dotrze do  $u_1$  i tam już pozostaje.
- Stopnie wejściowe wierzchołków elity w oryginalnej sieci są przybliżane, przez liczby  $D_i$ , oznaczające ile jednostek potencjału przeszło przez  $u_i$  w drodze do  $u_1$ .



## Model matematyczny i dowód bezskalowości

Rozważmy ciąg zmiennych losowych  $X_0, X_1, X_2, \dots$  takich, że  $X_0$  jest jednostajna na  $(0, 1)$ ,  $X_{l+1}$  jest jednostajna na  $(0, X_l)$  dla  $l > 0$ . Wtedy łatwo widać, iż:

$$k_l = \lceil KX_l \rceil \quad (22)$$

Zdefiniujmy  $\pi_i$ ,  $i = 1 \dots$  jako prawdopodobieństwo, że jednostka potencjału odwiedza  $u_i$ . Wtedy  $\pi_i = \mathbb{P}(\exists_l k_l = i)$  i dla odpowiednio dużych  $K$  mamy:

$$\pi_i = \mathbb{E} \left| \left\{ l, X_l \in \left[ \frac{i-1}{K}, \frac{i}{K} \right] \right\} \right|, i > 1 \quad (23)$$



## Model matematyczny i dowód bezskalowości

W takim przypadku wartości  $D_i$  są rozmieszczone zgodnie z rozkładem dwumianowym  $b(\pi_i, n)$ , gdzie  $n$  jest całkowitą liczbą jednostek potencjału. Zauważmy, że ciąg  $T_I = -\log X_I$  jest procesem punktowym Poissona o intensywności 1 na  $R_+$ . Stąd mamy:

$$\pi_i = \mathbb{E} \left| \left\{ I, T_I \in \left[ -\log \left( \frac{i}{K} \right), -\log \left( \frac{i-1}{K} \right) \right] \right\} \right| \approx \frac{1}{i} \quad (24)$$

Zatem dla dużych  $n$  na mocy prawa wielkich liczb

$$D_i \approx \frac{n}{i} \quad (25)$$



# Model matematyczny i dowód bezskalowości

Z tego wynika, że dla dużych  $k$  mamy:

$$|\{i, D_i > k\}| \approx \frac{n}{k} \quad (26)$$

czyli

$$|\{i, D_i \approx k\}| \approx \frac{n}{k^2} \quad (27)$$

co dowodzi, iż powstała sieć jest siecią bezskalową z  $\gamma = 2$ .



# Model matematyczny i dowód bezskalowości

Z tego wynika, że dla dużych  $k$  mamy:

$$|\{i, D_i > k\}| \approx \frac{n}{k} \quad (26)$$

czyli

$$|\{i, D_i \approx k\}| \approx \frac{n}{k^2} \quad (27)$$

co dowodzi, iż powstała sieć jest siecią bezskalową z  $\gamma = 2$ .



## Warianty modelu przepływowego

- Przebadaliśmy także pewne warianty modelu przepływowego dla zmodyfikowanej funkcji energetycznej:

$$\mathbb{E} = \frac{1}{2} \sum_{i \neq j} w_{i,j} |\sigma_i - \sigma_j|^\alpha \quad (28)$$

dla  $\alpha \in (0, 1)$ .

- Warto zauważyć, że przypadek  $\alpha = 1$  sprowadza się do poprzedniego modelu.
- Intuicyjnie widać, iż dla  $\alpha < 1$  stan w którym jeden element zgarnia cały potencjał nie jest już optymalny energetycznie. W tym przypadku optimum jest znacznie bardziej skomplikowane i polega na odpowiednim rozmieszczeniu potencjału w wielu jednostkach.



Podobnie jak poprzednio wyróżniamy pewien podzbiór (elitę) i rozbijamy funkcję energetyczną następująco:

$$\mathbb{E}(\bar{\sigma}) = \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{elite}} w_{ij} |\sigma_i - \sigma_j|^\alpha \quad (29)$$

Podobnie jak w poprzednim wypadku możemy zignorować czynnik elita-elita (ze względu na jej zaniedbywalną wielkość) oraz tłum-tłum ze względu na niewielki wkład energetyczny wynikający z niewielkich wag.



Podobnie jak poprzednio wyróżniamy pewien podzbiór (elitę) i rozbijamy funkcję energetyczną następująco:

$$\mathbb{E}(\bar{\sigma}) = \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{elite}} w_{ij} |\sigma_i - \sigma_j|^\alpha \quad (29)$$

Podobnie jak w poprzednim wypadku możemy zignorować czynnik **elita-elita** (ze względu na jej zaniedbywalną wielkość) oraz tłum-tłum ze względu na niewielki wkład energetyczny wynikający z niewielkich wag.



Podobnie jak poprzednio wyróżniamy pewien podzbiór (elitę) i rozbijamy funkcję energetyczną następująco:

$$\mathbb{E}(\bar{\sigma}) = \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha + \frac{1}{2} \sum_{i, j \in \text{elite}} w_{ij} |\sigma_i - \sigma_j|^\alpha \quad (29)$$

Podobnie jak w poprzednim wypadku możemy zignorować czynnik elita-elita (ze względu na jej zaniedbywalną wielkość) oraz **tłum-tłum** ze względu na niewielki wkład energetyczny wynikający z niewielkich wag.



Ostatecznie funkcja energetyczna wyraża się następująco:

$$\mathbb{E}(\bar{\sigma}) \simeq \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha \simeq \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i|^\alpha = \frac{1}{2} \sum_{i \in \text{elite}} S_i |\sigma_i|^\alpha \quad (30)$$

Ze względu na fakt, iż tłum przeciętnie ma mało potencjału.  $S_i$  są niezależnymi zmiennymi Gaussowskimi o wariancji rzędu  $N$ . Aby przejść dalej trzeba oszacować wartości sum Gaussowskich  $S_i$ .



Ostatecznie funkcja energetyczna wyraża się następująco:

$$\mathbb{E}(\bar{\sigma}) \simeq \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i - \sigma_j|^\alpha \simeq \frac{1}{2} \sum_{i \in \text{elite}, j \in \text{bulk}} w_{ij} |\sigma_i|^\alpha = \frac{1}{2} \sum_{i \in \text{elite}} S_i |\sigma_i|^\alpha \quad (30)$$

Ze względu na fakt, iż tłum przeciętnie ma mało potencjału.  $S_i$  są niezależnymi zmiennymi Gaussowskimi o wariancji rzędu  $N$ . Aby przejść dalej trzeba oszacować wartości sum Gaussowskich  $S_i$ .



Niech  $N = 2^M$  dla pewnego  $M$ . Istnieje permutacja  $\delta$  taka, że

$$S_{\delta(i)} \simeq \sqrt{\frac{2N}{\log_2 N}} \left( \log 2 \log_2 N - \log \sqrt{\log_2 N} \frac{\log i}{2\sqrt{\log 2}} \right) \quad (31)$$

(czynniki  $\sqrt{\frac{N}{M/2}} = \sqrt{\frac{2N}{\log_2 N}}$  jest odpowiedzialny za skalowanie wariancji)



Przekształcając otrzymane wyrażenie mamy, że:

$$S_i = \underbrace{\sqrt{2N} \sqrt{\log_2 N} (1 + o(1))}_{a_N} - \frac{\sqrt{2N}}{\sqrt{\log_2 N} 2\sqrt{\log 2}} \log i \quad (32)$$

Zauważmy, iż warunek równowagi energetycznej w rozważanym układzie można określić jako stan w którym nie następują żadne przepływy potencjału, czyli pochodna energii ze względu na stan powinna być stała:

$$[(\sigma_i)^\alpha]' S_i = \alpha \sigma_i^{\alpha-1} S_i = \text{const} \quad (33)$$

zatem:

$$\sigma_i = c \cdot S_i^{\frac{1}{1-\alpha}} \quad (34)$$



Podsumowując mamy ostatecznie:

$$S_i^{\frac{1}{1-\alpha}} \simeq a_N^{\frac{1}{1-\alpha}} + \frac{1}{1-\alpha} a_N^{\frac{1}{1-\alpha}-1} \log i \simeq c_1 - c_2 \log i \quad (35)$$

Ile jest takich  $\sigma_i$  w elicie dla których  $\sigma_i = k$  dla ustalonego  $k$ ? To pytanie można postawić następująco: ile jest takich  $i$ , że  $c_1 - c_2 \log i \simeq k$ ? Mamy:

$$\log i \simeq \frac{c_1 - k}{c_2} \quad (36)$$

równoważnie

$$i \simeq \exp\left(\frac{c_1 - k}{c_2}\right) \quad (37)$$

Jest zatem w przybliżeniu  $\exp\left(\frac{c_1 - k}{c_2}\right)$  węzłów o stopniu  $k$ . Jest to zanik wykładniczy, ale niezmiernie powolny.

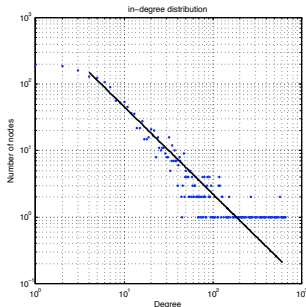


## Wyniki numeryczne

- Wynik teoretyczny jest w znakomitej zgodności z obserwacjami naumerycznymi [Piekniewski and Schreiber(2007)]
- Zależnie od rozmiaru sieci i sposobu estymacji współczynników wynik oscyluje w okolicach 2.0
- Estymowanie gęstości rozkładów Pareto jest trudne, dlatego czasem lepiej estymować współczynnik prawa potęgowego dystrybuanty, który różni się od współczynnika potęgowego gęstości o 1.



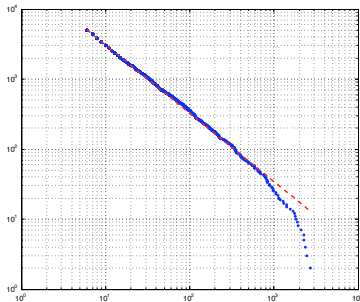
## Wyniki numeryczne



**Rysunek:** Histogram stopni wejściowych wierzchołków dla sieci zawierającej 2000 jednostek (wykres log-log). Czarna kreska jest estymacją nachylenia za pomocą metody najmniejszych kwadratów. Wyliczony współczynnik jest w okolicach 2.1.



## Wyniki numeryczne



**Rysunek:** Dystrybuanta empiryczna stopni wejściowych wierzchołków dla układu 6000 neuronów (wykres log-log). Czerwona kreska jest estymacją nachylenia za pomocą metody najmniejszych kwadratów. Wyliczony współczynnik dla gęstości to 1.956.



## Bezskalowość w sieciach dynamicznych?

- Podstawową cechą modelu przepływowego jest to, że każda jednostka posiada dokładną *pamięć* ilości potencjału którą posiada.
- Jeśli do danej jednostki wiele razy sygnał przyszedł, to zwiększa to szanse iż pewne sygnały wyjdą (o ile dany neuron nie jest w elicie)
- Aktywny neuron (dużo dostający) ma większe szanse pozostać aktywnym w dalszym ciągu, pojawia się zjawisko preferowanego doczepiania (*preferential attachment*), niezbędnego do utworzenia sieci bezskalowej.
- Dlaczego zatem zwykłe neurony nie ujawniają takich własności? Może mają zbyt krótką pamięć stanu?

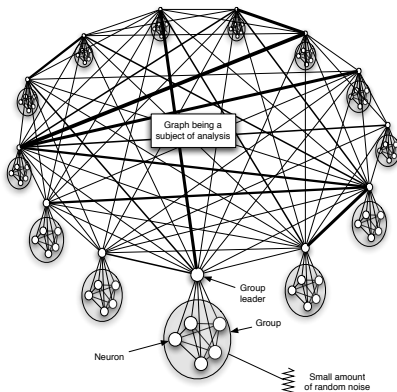


## Bezskalość w sieciach dynamicznych?

- Najprostszym sposobem zrobienia z neuronu czegoś mądrzejszego jest wzięcie kilku neuronów i połączenie ich w sieć
- Jeśli pojedynczy neuron (dynamiczny) ma za małą *pamięć* to może grupka kilkunastu będzie już więcej *pamiętać*?
- Na fali tego pomysłu, wiosną 2007 przeprowadziliśmy pewien zaskakujący eksperyment numeryczny z wykorzystaniem dynamicznych neuronów Eugene Izhikevicha.

# Bezskalość w sieciach dynamicznych?

**Rysunek:** Badany układ składał się z kilku tysięcy grup, w każdej po około 15 neuronów dynamicznych Izhikevicha, pobudzanych lekkim szumem (w celu wywołania rezonansu stochastycznego) oraz połączonych z pozostałymi za pomocą pojedynczego "neuronu bramy".





# Bezskalowość w sieciach dynamicznych?

- Każda z powstałych grup w trakcie dynamiki pobudzała się lub wygaszała, pewne z nich stale impulsowały
- W trakcie trwania dynamiki połączenia między grupami były ustalone jako losowe z rozkładu normalnego. Wewnątrz każdej grupy około 20% połączeń było inhibicyjnych, reszta była pobudzająca.
- Dynamika układu indukuje pewien nowy graf, ze zbiorem wierzchołków tożsamym ze zbiorem grup, oraz zbiorem ważonych krawędzi, na których waga zdefiniowana jest następująco:
  - Przebiegi czasowe impulsów grup leżących na krańcach krawędzi zostały lekko rozmazane
  - Następnie zostały wymnożone po współrzędnych i scałkowane
  - Waga została zdefiniowana jako wartość całki. Miara ta mierzy korelację między przebiegami aktywności

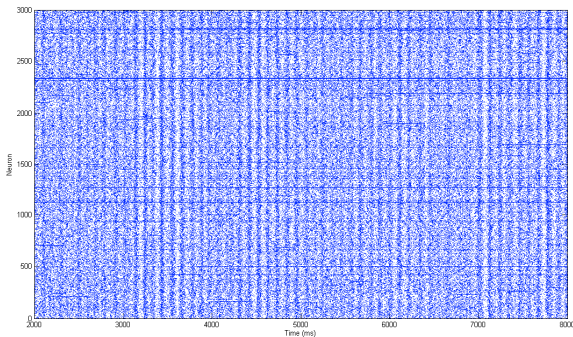


Spis treści  
Wstęp  
Sieci neuronowe  
Model przepływowy  
Podsumowanie  
Bibliografia

Grafy przepływu impulsów  
Model matematyczny i dowód bezskalości  
Warianty modelu przepływowego  
Wyniki numeryczne  
Bezskalość w sieciach dynamicznych?



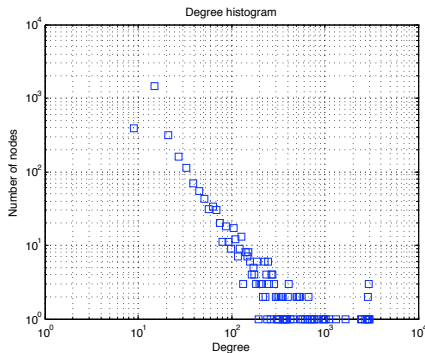
## Bezskalość w sieciach dynamicznych?



**Rysunek:** Przykładowy przebieg aktywności układu 3000 grup neuronów, po około 15 neuronów każda.



# Bezskalość w sieciach dynamicznych?



**Rysunek:** Histogram stopni węzłów w indukowanym grafie (wykres log-log). Wyraźnie widoczny jest rozkład potęgowy.



## Bezskalowość w sieciach dynamicznych?

- Graf indukowany przez grupy neuronów okazuje się grafem bezskalowym z wykładnikiem  $\gamma \approx 2$
- Dokładnie ten sam eksperyment w którym grupy neuronów zastąpiono pojedynczymi jednostkami owocuje rozkładami wykładniczymi, podobnymi do tych które obserwuje się z biologicznych sieciach neuronowych
- Jest to zatem pewne wyjaśnienie obserwacji empirycznych, pojedyncze neurony (nawet dynamiczne) nie mają dość *pamięci stanu* aby wykształcić *preferowane przyczepianie* konieczne do powstania bezskalowości
- Jak się okazuje już grupki kilkunastu neuronów mają dość skomplikowaną neurodynamikę, aby pamiętać swój stan i swoją historię dość długo [Piekniński(2007)]



## Co dalej?

- Model przepływowy ma szereg ciekawych uogólnień nad którymi obecnie pracujemy
- Grupy neuronów analizowane w ostatniej pracy są nieco sztuczne, potrzebny jest model spontanicznie wykształcających się grup, a następnie analiza połączeń pomiędzy takimi grupami  
[Izhikevich et al.(2004)Izhikevich, Gally, and Edelman]
- Być może prezentowane modele mają jakąś interpretację socjologiczną lub ekonomiczną?



## Podsumowanie

- Przedstawiony został model przepływowy stanowiący pomost pomiędzy sieciami neuronowymi a grafami bezskalowymi
- Przedstawiona została analiza teoretyczna dwóch wariantów modelu przepływowego wraz z dowodem oraz wyniki badań numerycznych
- Przedstawiony został przedsięwzięcie badań w dziedzinie dynamicznych sieci impulsujących w kontekście bezskalowych połączeń między grupami impulsujących neuronów



D. Ackley, G. Hinton, and T. Sejnowski.  
A learning algorithm for boltzmann machines.  
*Cognitive Science*, 9:147–169, 1985.



Reka Albert and Albert-László Barabási.  
Statistical mechanics of complex networks.  
*Reviews of modern physics*, (74):47–97, January 2002a.



Reka Albert and Albert-László Barabási.  
Statistical mechanics of complex networks.  
*Reviews of Modern Physics*, 74:47, 2002b.  
URL <http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0106096>.



Réka Albert, Hawoong Jeong, and Albert-László Barabási.  
Diameter of the world-wide web.  
*Science*, 401:130–131, Septmeber 1999.



L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley.  
Classes of small-world networks.  
*Proc Natl Acad Sci U S A*, 97(21):11149–11152, October 2000.  
ISSN 0027-8424.  
doi: 10.1073/pnas.200327197.  
URL <http://dx.doi.org/10.1073/pnas.200327197>.



Albert-László Barabási and Reka Albert.  
Emergence of scaling in random networks.  
*Science*, (286):509–512, October 1999.



Albert-László Barabási, Hawoong Jeong, Zoltan Neda, Erzsebet Ravasz, A. Schubert, and Tamas Vicsek.





Evolution of the social network of scientific collaborations.

*Physica A*, 311(4):590–614, 2002.



P. Erdős and A. Rényi.

On random graphs. i.

*Publicationes Mathematicae*, 6:290–297, 1959.



P. Erdős and A. Rényi.

The evolution of random graphs.

*Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.



R. FitzHugh.

Impulses and physiological states in models of nerve membrane.

*Biophysics Journal*, 1:445–466, 1961.



Wulfram Gerstner and Werner M. Kistler.

*Spiking Neuron Models: Single Neurons, Populations, Plasticity*.

Cambridge University Press, Cambridge, UK, 2002.



D.O. Hebb.

*The Organization of Behavior*.

Wiley, New York, 1949.



A. L Hodgkin and A. F. Huxley.

A quantitative description of membrane current and its application to conduction and excitation in nerve.

*Journal of Physiology*, 117:500–544, 1952.



J.J Hopfield.

Neural networks and physical systems with emergent collective computational abilities.



In *Proceedings of the National Academy of Sciences*, volume 79, pages 2554–2558. Washington : The Academy, 1982.



Ramon Ferrer i Cancho and Ricard V. Solé.

The small-world of human language.

*Proceedings of the Royal Society of London B*, 268(1482):2261–2265, November 7 2001.



Eugene M. Izhikevich.

*Dynamical systems in Neuroscience: The Geometry of Excitability and Bursting*.

MIT Press., Boston, 2006.



Eugene M. Izhikevich.

Simple model of spiking neurons.

*IEEE Transactions on Neural Networks*, (14):1569– 1572, 2003.

URL <http://www.nsi.edu/users/izhikevich/publications/spikes.pdf>.



Eugene M. Izhikevich, Joe A. Gally, and Gerald M. Edelman.

Spike-timing dynamics of neuronal groups.

*Cerebral Cortex*, (14):933– 944, 2004.

URL <http://vesicle.nsi.edu/users/izhikevich/publications/reentry.pdf>.



Hawoong Jeong, B. Tombor, Réka Albert, Zoltan N. Oltvai, and Albert-László Barabási.

The large-scale organization of metabolic networks.

*Nature*, 407(6804):651–653, October 5 2000.



Christof Koch and Gilles Laurent.

Complexity and the Nervous System.

*Science*, 284(5411):96–98, 1999.

doi: 10.1126/science.284.5411.96.



URL <http://www.sciencemag.org/cgi/content/abstract/284/5411/96>.



Jose M. Montoya and Ricard V. Solé V.

Small world patterns in food webs.

*Journal of Theoretical Biology*, 214(3):405–412, February 7 2002.



J. S. Nagumo, S. Arimoto, and S. Yoshizawa.

An active pulse transmission line simulating a nerve axon.

*Proceedings of the IRE*, 50:2061–2070, 1962.



Filip Piekiewicz.

Emergence of scale-free graphs in dynamical spiking neural networks.

In *Proc. IEEE International Joint Conference on Neural Networks*, Orlando, Florida, USA, August 2007. IEEE Press.



Filip Piekiewicz and Tomasz Schreiber.

Emergence of scale-free spike flow graphs in recurrent neural networks.

In *Proc. IEEE FOCI*, pages 357–362, 2007.



S. Redner.

How popular is your paper? an empirical study of the citation distribution.

*European Physical Journal B*, 4(2):131–134, 1998.



D. E. Rumelhart, G. E. Hinton, and R. J. Williams.

Learning internal representations by error propagation.

In D. Rumelhart and J. McClelland, editors, *Parallel Distributed Processing*, volume 1, pages 318–362. MIT Press, Cambridge, MA, 1986.



Duncan J. Watts and Steven H. Strogatz.



Collective dynamics of small-world networks.

*Nature*, (393):440–442, June 1998.



P. Werbos.

*Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.*

PhD thesis, Harvard, 1974.