Modelowanie i analiza sieci złożonych

VI. Sieci ewoluujące.

Grzegorz Siudem

Politechnika Warszawska

Zadanie 10 pn.
„Przygotowanie i uruchomienie nowego kierunku studiów na studiach II stopnia - Inżynieria i Analiza Danych (IAD)”

realizowane jest w ramach projektu
„NERW PW. Nauka – Edukacja – Rozwój – Współpraca”
współfinansowanego ze środków Unii Europejskiej
w ramach Europejskiego Funduszu Społecznego

Przed zajęciami

Do przypomnienia przed zajęciami

Z MASZ_2:

- kontekst historyczny pracy Barábasiego i Alberty.

Z innych przedmiotów:

- rozkład Bernoulliego $\mathbb{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$,
- równanie Chapmana-Kołmogorowa,
- metody rozwiązywania równań rekurencyjnych.

Wykład

Przypomnienie kontekstu historycznego

Znane modele sieci nie miały cech sieci rzeczywistych

- problemem była nieobecność rozkładów potęgowych, ale także
- wysoki współczynnik gronowania,
- bezskalowość.

A.-L. Barabási, R. Albert, Emergence on scaling in random networks, Science, 286:509-512, 1999.

Przypomnienie kontekstu historycznego

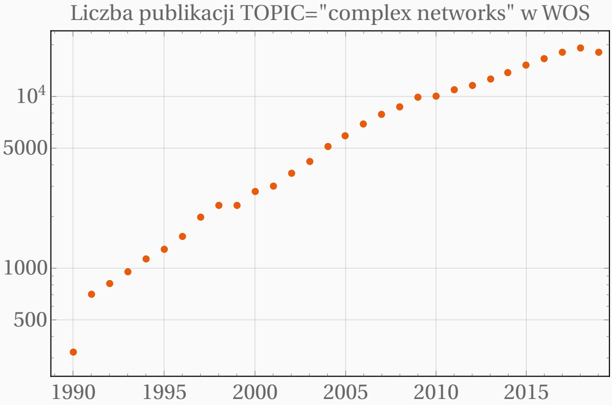
Znane modele sieci nie miały cech sieci rzeczywistych

- problemem była nieobecność rozkładów potęgowych, ale także
- wysoki współczynnik gronowania,
- bezskalowość.

A.-L. Barabási, R. Albert, Emergence on scaling in random networks, Science, 286:509-512, 1999.

Od tej pracy datuje się sieciologię!
Czy słusznie?

Rozwój sieciologii



Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.

Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.
- przy wyborze do których wierzchołków w grafie dołączamy krawędzie nowego wierzchołka kierujemy się regułą preferencyjnego dołączania.

Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.
- przy wyborze do których wierzchołków w grafie dołączamy krawędzie nowego wierzchołka kierujemy się regułą preferencyjnego dołączania.

Obie te reguły są zdroworozsądkowe:

Motywacją autorów był rozrost sieci www.

Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.
- przy wyborze do których wierzchołków w grafie dołączamy krawędzie nowego wierzchołka kierujemy się regułą preferencyjnego dołączania.

Obie te reguły są zdroworozsądkowe:

Motywacją autorów był rozrost sieci www.

- ta sieć zdecydowanie rośnie.

Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.
- przy wyborze do których wierzchołków w grafie dołączamy krawędzie nowego wierzchołka kierujemy się regułą preferencyjnego dołączania.

Obie te reguły są zdroworozsądkowe:

Motywacją autorów był rozrost sieci www.

- ta sieć zdecydowanie rośnie.
- działa w niej reguła św. Mateusza.

Motywacja empiryczna

Procedura konstrukcyjna sieci BA ma dwa założenia:

- są to sieci rosnące (ewoluujace), w każdym kroku czasowym dodajemy do grafu kolejny wierzchołek.
- przy wyborze do których wierzchołków w grafie dołączamy krawędzie nowego wierzchołka kierujemy się regułą preferencyjnego dołączania.

Obie te reguły są zdroworozsądkowe:

Motywacją autorów był rozrost sieci www.

- ta sieć zdecydowanie rośnie.
- działa w niej reguła św. Mateusza.

Dotyczy to zreszą większej liczby sieci...

Jakich?

Procedura konstrukcyjna BA

- W chwili $t = 0$ zaczynamy z grafem pełnym o $m_0 \geqslant 1$ wierzchołkach.

Procedura konstrukcyjna BA

- W chwili $t = 0$ zaczynamy z grafem pełnym o $m_0 \geqslant 1$ wierzchołkach.
- W kolejnych krokach czasowych dołączamy do sieci nowe węzły, a każdy z nich wnosi $m \leqslant m_0$ nowych połączeń, dołączanych zgodnie z regułą preferencyjnego dołączania

$$\Pi(k_i) \propto k_i.$$

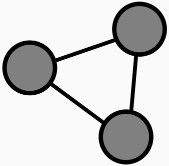
Procedura konstrukcyjna BA

- W chwili $t = 0$ zaczynamy z grafem pełnym o $m_0 \geqslant 1$ wierzchołkach.
- W kolejnych krokach czasowych dołączamy do sieci nowe węzły, a każdy z nich wnosi $m \leqslant m_0$ nowych połączeń, dołączanych zgodnie z regułą preferencyjnego dołączania

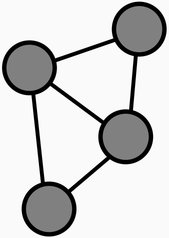
$$\Pi(k_i) \propto k_i.$$

- Wzrost sieci kończymy w dowolnej chwili t.

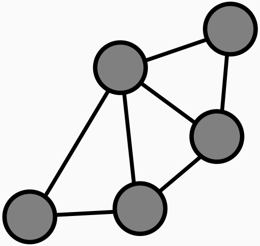
Wizualizacja procedury konstrukcyjnej sieci BA



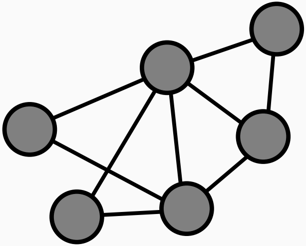
Wizualizacja procedury konstrukcyjnej sieci BA



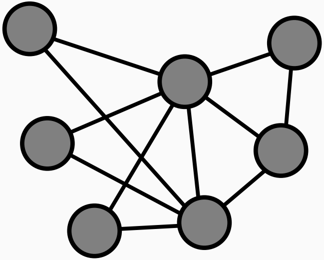
Wizualizacja procedury konstrukcyjnej sieci BA



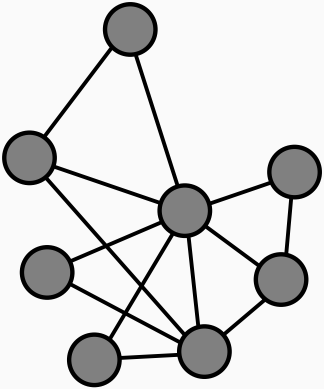
Wizualizacja procedury konstrukcyjnej sieci BA



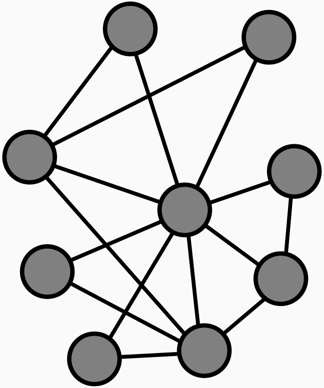
Wizualizacja procedury konstrukcyjnej sieci BA



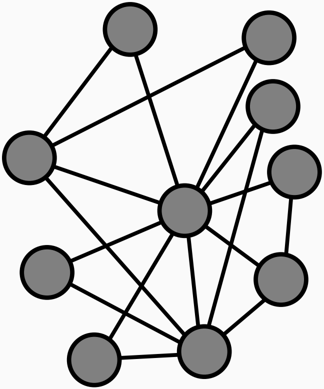
Wizualizacja procedury konstrukcyjnej sieci BA



Wizualizacja procedury konstrukcyjnej sieci BA



Wizualizacja procedury konstrukcyjnej sieci BA



po chwili...

Wizualizacja procedury konstrukcyjnej sieci BA

Opis ilościowy modelu

Liczby krawędzi i węzłów

$$N = t + m_0 \approx t,$$
$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Opis ilościowy modelu

Liczby krawędzi i węzłów

$$N = t + m_0 \approx t,$$
$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Reguła preferencyjnego dołaczania

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

Dlaczego?

Opis ilościowy modelu

Liczby krawędzi i węzłów

$$N = t + m_0 \approx t,$$
$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Reguła preferencyjnego dołaczania

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

Dlaczego?

Typowe podejście do problemu to jeden z dwóch sposobów:

- metoda czasu ciągłego (w przybliżeniu średniego pola),
- równanie master.

Opis ilościowy modelu

Liczby krawędzi i węzłów

$$N = t + m_0 \approx t,$$

$$E = mt + \frac{m_0(m_0 - 1)}{2} \approx mt.$$

Reguła preferencyjnego dołaczania

$$\Pi(k) = \frac{k}{\sum_{i=1}^{t} k_i} = \frac{k}{2mt}.$$

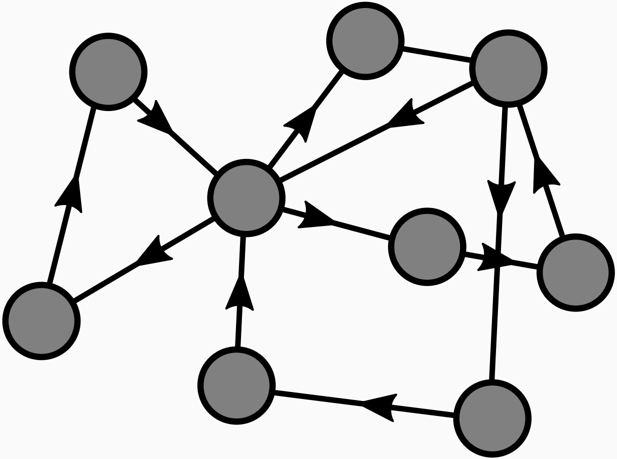
Dlaczego?

Typowe podejście do problemu to jeden z dwóch sposobów:

- metoda czasu ciągłego (w przybliżeniu średniego pola),
- równanie master.

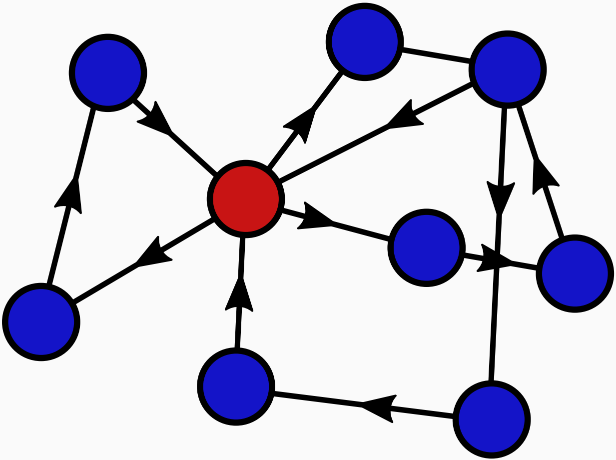
A jak Państwo podeszliby do rozwiązania tego problemu?

Średnie pole – tak robią to fizycy



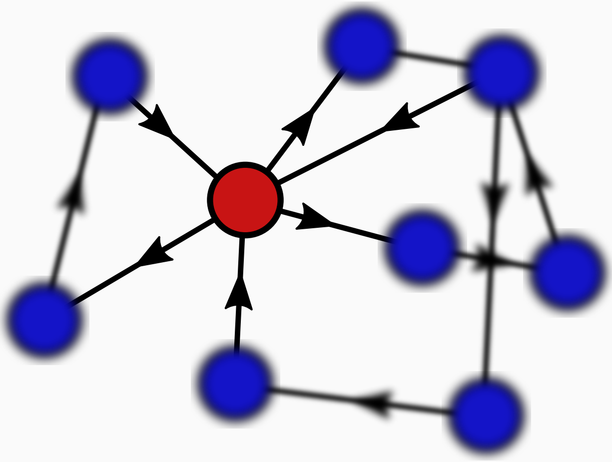
Problemy fizyki statystycznej często są trudne bo *poplątane*...

Średnie pole – tak robią to fizycy



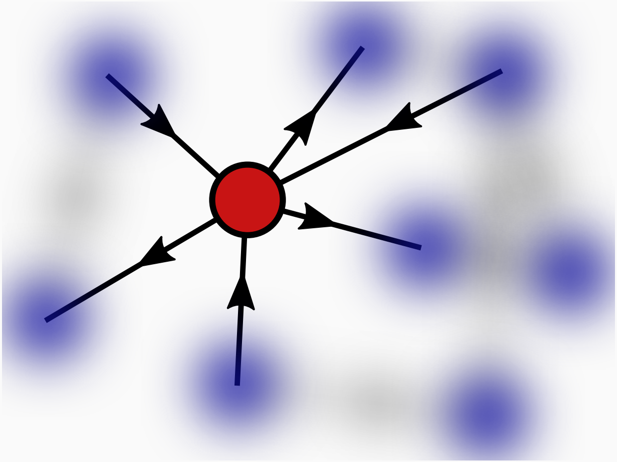
Jak je rozplątać?

Średnie pole – tak robią to fizycy



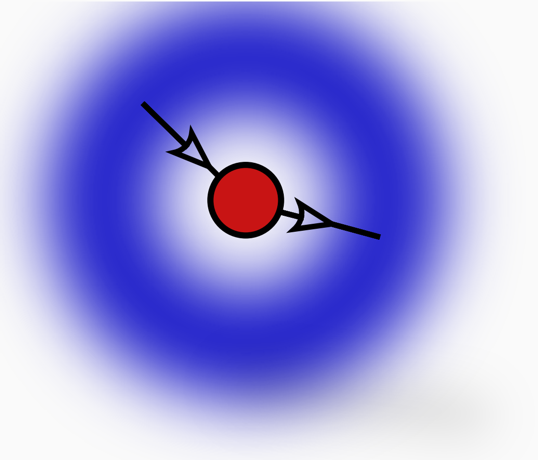
Uprośćmy nieco wyjsciowy problem...

Średnie pole – tak robią to fizycy



... zastępujemy resztę układu *średnim polem*.

Średnie pole – tak robią to fizycy



Otrzymujemy układ podobny, ale łatwiejszy do rozwiązania!

Jak to zrobić dla sieci BA?

Wartość oczekiwana zastępuje zmienną losową:

- Niech k_i oznacza *oczekiwany* (średni) stopień i-tego węzła.
- Czyli dopuszczamy jego niecałkowite wartości!

Jak to zrobić dla sieci BA?

Wartość oczekiwana zastępuje zmienną losową:

- Niech k_i oznacza *oczekiwany* (średni) stopień i-tego węzła.
- Czyli dopuszczamy jego niecałkowite wartości!

Ciągły czas:

- Przyjmujemy, że nowe krawędzie dołączamy w sposób ciągły, a nie dyskretny.
- Choć nowe krawędzie pojawiają się w dyskretnych chwilach t_i.

Jak to zrobić dla sieci BA?

Wartość oczekiwana zastępuje zmienną losową:

- Niech k_i oznacza *oczekiwany* (średni) stopień i-tego węzła.
- Czyli dopuszczamy jego niecałkowite wartości!

Ciągły czas:

- Przyjmujemy, że nowe krawędzie dołączamy w sposób ciągły, a nie dyskretny.
- Choć nowe krawędzie pojawiają się w dyskretnych chwilach t_i.

Średnie pole:

- Nowe krawędzie rozdzielane są niezależnie od siebie.

Przybliżenie ciągłego czasu

Zmiany stopni wierzchołków

Zgodnie z przyjętymi założeniami dane są rozkładem Bernoulliego:

$$\frac{dk_i}{dt} = \sum_{l=0}^{m} l \binom{m}{l} [\Pi(k_i)]^l [1 - \Pi(k_i)]^{m-l} = m\Pi(k_i) = \frac{k_i}{2t}$$

$$k_i(t_i) = m,$$

gdzie t_i to czas dołączenia i-tego węzła.

Przybliżenie ciągłego czasu

Zmiany stopni wierzchołków

Zgodnie z przyjętymi założeniami dane są rozkładem Bernoulliego:

$$\frac{dk_i}{dt} = \sum_{l=0}^{m} l \binom{m}{l} [\Pi(k_i)]^l [1 - \Pi(k_i)]^{m-l} = m\Pi(k_i) = \frac{k_i}{2t}$$

$$k_i(t_i) = m,$$

gdzie t_i to czas dołączenia i-tego węzła.

Rozwiązujemy równanie różniczkowe

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Przybliżenie ciągłego czasu cd.

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Przejdźmy do rozkładu $\mathcal{P}(k_i)$

$$\mathcal{P}(k_i) = T(t_i)\left|\frac{dk_i}{dt_i}\right|^{-1},$$

gdzie $T(t_i)$ to gęstość prawdopodobieństwa czasów t_i

$$T(t_i) = \frac{1}{t}$$

Dlaczego?

Przybliżenie ciągłego czasu cd.

$$k_i(t) = m\sqrt{\frac{t}{t_i}}.$$

Przejdźmy do rozkładu $\mathcal{P}(k_i)$

$$\mathcal{P}(k_i) = T(t_i)\left|\frac{dk_i}{dt_i}\right|^{-1},$$

gdzie $T(t_i)$ to gęstość prawdopodobieństwa czasów t_i

$$T(t_i) = \frac{1}{t}$$

Dlaczego?

Połączenie trzech powyższych równań prowadzi do:

$$\mathcal{P}(k) = \frac{2m^2}{k^3}.$$

Wyprowadźmy to!

O istotności założeń

Algorytm Barábasiego i Alberty ma tylko dwa założenia:

- sieć rośnie w każdym kroku,
- nowe krawędzie dołączane są preferencyjnie.

Czy któreś z tych założeń można pominąć?

Sprawdźmy!

Losowe dołączanie węzłów (model A)

Stosując metodę średniego pola wyznacz rozkład stopni sieci w której

- w każdym kroku dodajemy nowy wierzchołek.
- krawędzie rozdajemy przypadkowo

$$\Pi(k_i) = \frac{1}{t + m_0} \approx \frac{1}{t}.$$

Losowe dołączanie węzłów (model A)

Stosując metodę średniego pola wyznacz rozkład stopni sieci w której

- w każdym kroku dodajemy nowy wierzchołek.
- krawędzie rozdajemy przypadkowo

$$\Pi(k_i) = \frac{1}{t + m_0} \approx \frac{1}{t}.$$

Rozwiązanie:

Równanie różniczkowe postaci

$$\frac{dk_i}{dt} = \frac{m}{t},$$

ma rozwiązanie

$$k_i(t) = m \ln\left(\frac{t}{t_i}\right) + m,$$

Sieć o ustalonym rozmiarze (model B)

Stosując metodę średniego pola (na ile się to uda!) wyznacz rozkład stopni sieci w której

- Sieć od początku na N wierzchołków.
- krawędzie rozdajemy preferencyjnie.

Sieć o ustalonym rozmiarze (model B)

Stosując metodę średniego pola (na ile się to uda!) wyznacz rozkład stopni sieci w której

- Sieć od początku na N wierzchołków.
- krawędzie rozdajemy preferencyjnie.

Rozwiązanie:

Równanie różniczkowe postaci

$$\frac{dk_i}{dt} = \frac{N-1}{N}\frac{k_i}{2t} + \frac{1}{N},$$

ma rozwiązanie

$$k_i(t) = \frac{2(N-1)}{N(N-2)}t \approx \frac{2}{N}t,$$

jak z tego jednak otrzymać rozkład?

Podsumowanie

Praca domowa

Przeczytaj *Personal Introduction* do A.-L. Barabási, *Network Science*
`http://networksciencebook.com/chapter/0`

Fundusze Europejskie
Wiedza Edukacja Rozwój

Politechnika Warszawska

Unia Europejska
Europejski Fundusz Społeczny

Zadanie 10 pn.
„Przygotowanie i uruchomienie nowego kierunku studiów na studiach II stopnia - Inżynieria i Analiza Danych (IAD)”

realizowane jest w ramach projektu
„NERW PW. Nauka – Edukacja – Rozwój – Współpraca”
współfinansowanego ze środków Unii Europejskiej
w ramach Europejskiego Funduszu Społecznego

Dziękuję za uwagę!