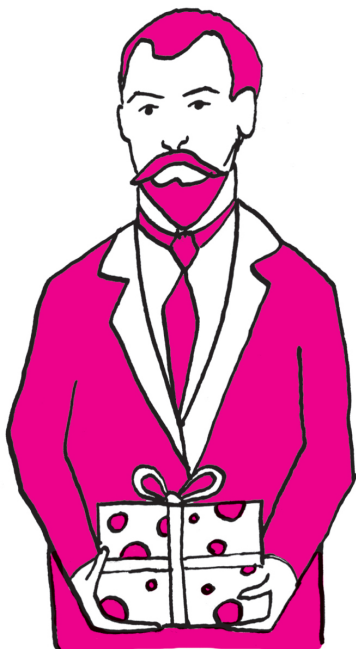


# Pudełko Smoluchowskiego, losowe ruchy w grafie i egzopeptydazy

Wojciech NIEMIRO\*

Marian Ritter von Smolan Smoluchowski – światowej sławy polski fizyk, pionier fizyki statystycznej, alpinista i taternik, żył w latach 1872–1917. Egzopeptydazy to specjalny typ enzymów. Co może mieć wspólnego Smoluchowski z egzopeptydazami, skoro ani on, ani nikt z jego współczesnych nie wiedzieli o ich istnieniu? I co do tego mają grafy? Postaram się wszystko wyjaśnić w odpowiedniej kolejności. Zacznę od uproszczonego modelu fizycznego. Podobno zbliżony model rozpatrywał Smoluchowski, ale, niestety, nie udało mi się dotrzeć do źródeł i sprawdzić tej informacji.



Wyobraźmy sobie „pudełko” (określony obszar przestrzeni), do którego losowo wpadają cząstki i z którego losowo wypadają. Mówiąc nieco dokładniej, zakładamy, że w „krótkim” odcinku czasu, powiedzmy  $[t, t + h]$ ,

- do pudełka wpada 1 cząstka z prawdopodobieństwem w przybliżeniu równym  $a_*h$ ; prawdopodobieństwo wpadnięcia 2 lub więcej cząstek jest tak małe, że możemy je zaniedbać,
- każda spośród cząstek znajdujących się w pudełku wypada z niego z prawdopodobieństwem w przybliżeniu równym  $a_+h$ , niezależnie od pozostałych cząstek.

Żeby to sformułować jeszcze dokładniej, wprowadzimy wygodną terminologię i kilka oznaczeń. Zamiast mówić „w pudełku jest  $n$  cząstek”, powiemy, że „układ znajduje się w stanie  $n$ ”. Niech  $P^h(n, n')$  oznacza prawdopodobieństwo tego, że układ w chwili  $t + h$  będzie w stanie  $n'$ , jeśli w chwili  $t$  jest w stanie  $n$ . Rozumie się, że  $n, n' \in \{0, 1, 2, \dots\}$ . Założmy istnienie granic

$$(1) \quad \lim_{h \searrow 0} \frac{1}{h} P^h(n, n') = Q(n, n') \quad (n \neq n').$$

Nazwiemy  $Q(n, n')$  intensywnością przejścia z  $n$  do  $n'$ . Ścisłe sformułowanie założeń modelu Smoluchowskiego jest takie:

$$(2) \quad \begin{cases} Q(n, n+1) = a_*, \\ Q(n, n-1) = na_+, \\ Q(n, n') = 0, \end{cases} \quad (n \neq n', n+1 \neq n', n-1 \neq n').$$

Układ znajduje się w równowadze, jeśli prawdopodobieństwa poszczególnych stanów nie zmieniają się w czasie (mówimy tu o równowadze probabilistycznej, która nie oznacza bezruchu, tylko zrównoważone losowe fluktuacje). Niech  $\pi(n)$  będzie prawdopodobieństwem stanu  $n$ . *Ogólny warunek równowagi* jest następujący:

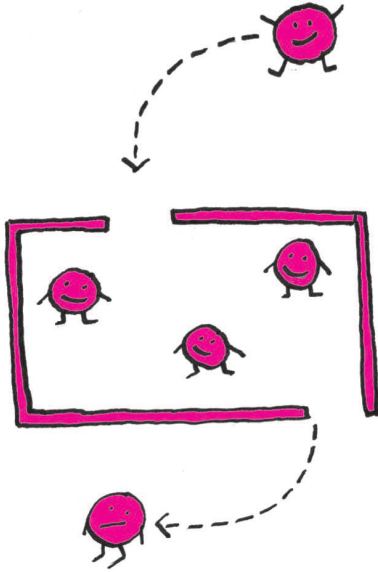
$$(3) \quad \pi(n) \sum_{n': n' \neq n} Q(n, n') = \sum_{n': n' \neq n} \pi(n') Q(n', n) \quad \text{dla każdego } n.$$

Wzór (3) mówi tyle, że prawdopodobieństwo wyjścia ze stanu  $n$  jest równe prawdopodobieństwu wejścia do stanu  $n$ . Jeśli rozkład prawdopodobieństwa  $\pi$  spełnia (3), to mówimy, że jest *rozkładem równowagi* lub inaczej stacjonarnym. Chwila zastanowienia wystarczy, żeby zrozumieć, że to dobra definicja. Silniejszy od ogólnego warunku równowagi jest następujący *warunek równowagi szczegółowej*:

$$(4) \quad \pi(n) Q(n, n') = \pi(n') Q(n', n) \quad \text{dla wszystkich } n \neq n'.$$

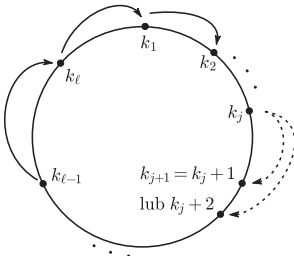
Pozostawiamy Czytelnikowi łatwe sprawdzenie, że (4) implikuje (3). Warunek (4) można interpretować jako *odwracalność w czasie*. Film, pokazujący stan układu, powiedzmy  $n(t)$ , w pewnym odcinku czasu,  $t \in [t_0, t_1]$ , wygląda tak samo, jak tenże film „puszczony wspak”:  $n(t_1 + t_0 - t)$ . To dlatego, że przejścia  $n \rightarrow n'$  zdarzają się jednakowo często jak przejścia  $n' \rightarrow n$  (przy równowagowym rozkładzie prawdopodobieństwa  $\pi$ ).

\*Instytut Matematyki Stosowanej  
i Mechaniki, Uniwersytet Warszawski;  
Wydział Matematyki i Informatyki,  
Uniwersytet Mikołaja Kopernika, Toruń



**Rozwiązanie zadania M 1454.**

Niech  $x_{k_1}$  oznacza największą z liczb  $x_1, \dots, x_n$ . Popatrzmy na wyraz  $k_1$  naszej sumy:  $\frac{x_{k_1}}{x_{k_1+1} + x_{k_1+2}}$  i oznaczmy większy z składników z mianownika,  $x_{k_1+1}$  lub  $x_{k_1+2}$ , przez  $x_{k_2}$ . Popatrzmy teraz na wyraz  $k_2$  naszej sumy:  $\frac{x_{k_2}}{x_{k_2+1} + x_{k_2+2}}$  i oznaczmy większy z składników z mianownika,  $x_{k_2+1}$  lub  $x_{k_2+2}$ , przez  $x_{k_3}$ . Kontynuując to postępowanie, otrzymujemy ciąg indeksów  $k_1, k_2, \dots$ .



Ponieważ nierówność, którą chcemy udowodnić, jest cykliczna, możemy bez utraty ogólności przyjąć, że  $k_1 = 1$ . Z definicji mamy  $k_j = k_{j-1} + 1$  lub  $k_j = k_{j-1} + 2$ . Zatem po pewnej liczbie kroków po raz pierwszy otrzymamy  $k_\ell = n - 1$  lub  $k_\ell = n$ . Ponadto  $\ell \geq n/2$ , bo w każdym kroku przesuwamy się o co najwyżej dwa. Zauważmy też, że  $k_{\ell+1} = 1 = k_1$ .

Stosując teraz nierówność między średnimi, otrzymujemy

$$\begin{aligned} \sum_{k=1}^n \frac{x_k}{x_{k+1} + x_{k+2}} &\geq \\ &\geq \sum_{j=1}^{\ell} \frac{x_{k_j}}{x_{k_j+1} + x_{k_j+2}} \geq \sum_{j=1}^{\ell} \frac{x_{k_j}}{2x_{k_j+1}} \geq \\ &\geq \ell \cdot \sqrt[\ell]{\prod_{j=1}^{\ell} \frac{x_{k_j}}{2x_{k_j+1}}} = \ell \sqrt[\ell]{\frac{1}{2^\ell}} = \frac{\ell}{2} \geq \frac{n}{4}. \end{aligned}$$

Zastanówmy się, jaki jest rozkład równowagi w modelu Smoluchowskiego. Odpowiedź jest łatwa, jeśli wpadniemy na pomysł, żeby szukać rozkładu  $\pi$  spełniającego warunek równowagi szczegółowej. Istotnie, jeśli zachodzi (4), to

$$\pi(n-1)a_* = \pi(n)na_{\dagger} \quad \text{dla } n = 1, 2, \dots$$

Stąd  $\pi(n) = \pi(n-1)\lambda/n$ , gdzie  $\lambda = a_*/a_{\dagger}$ , a więc

$$\pi(n) = \frac{\pi(n)}{\pi(n-1)} \frac{\pi(n-1)}{\pi(n-2)} \dots \frac{\pi(1)}{\pi(0)} \pi(0) = \frac{\lambda^n}{n!} \pi(0).$$

Aby wyznaczyć stałą  $\pi(0)$ , wystarczy skorzystać z oczywistego równania  $\sum_{n=0}^{\infty} \pi(n) = 1$ . Ponieważ  $\sum_{n=0}^{\infty} \lambda^n/n! = e^\lambda$ , więc  $\pi(0) = e^{-\lambda}$ . Otrzymaliśmy następujący wynik.

**Twierdzenie 1.** W modelu Smoluchowskiego rozkład prawdopodobieństwa dany wzorem

$$\pi(n) = e^{-\lambda} \frac{\lambda^n}{n!} \quad (n = 0, 1, \dots)$$

jest rozkładem równowagi.

Można pokazać, że wychodząc z dowolnego stanu początkowego, w długim okresie czasu zbliżamy się do rozkładu równowagi:  $\lim_{t \rightarrow \infty} P^t(n_0, n) = \pi(x)$  dla wszystkich  $n_0$  i  $n$ . Sławny rozkład występujący w Twierdzeniu 1 związany jest (jak wiele innych wspaniałych obiektów matematycznych) z nazwiskiem francuskiego fizyka i matematyka Siméona Denisa Poissona (1781–1840). Zastanówmy się, jaka jest *średnia* liczba cząstek w pudełku Smoluchowskiego (znajdującym się w równowadze). Jeśli tę średnią liczbę oznaczmy chwilowo przez  $m$ , to możemy przeprowadzić następujące rozumowanie. W krótkim okresie długości  $h$ , w przybliżeniu,

- do pudełka wpada *średnio*  $a_*h$  cząstek,
- z pudełka wypada *średnio*  $ma_{\dagger}h$  cząstek.

Ponieważ średnio tyle samo cząstek wpada, co wypada (równowaga!) to  $a_*h = ma_{\dagger}h$ , skąd  $m = a_*/a_{\dagger} = \lambda$ . Tak więc zmienna losowa o rozkładzie Poissona ma wartość średnią (nieszczęśliwie nazywaną też „wartością oczekiwaną”) równą  $\lambda$ . Koledzy matematycy nie uznają tego rozumowania za dowód, ale moim zdaniem to ciekawe wyprowadzenie.

Zwróćmy uwagę na zaskakujący fakt. Ponieważ model (2) spełnia warunek (4), to jest on odwracalny w czasie. Oglądając nasze pudełko znajdujące się w równowadze probabilistycznej, nie mamy możliwości rozpoznania, w którym kierunku płynie czas.

Przejdźmy do grafów. Wyobraźmy sobie wiele pudełek Smoluchowskiego, oznaczonych numerami  $1, \dots, i, \dots, k$ . Cząstki wpadają do pudełek z zewnątrz układu, wypadają na zewnątrz oraz przemieszczają się z jednego pudełka do drugiego. Zakładamy, że w „krótkim” odcinku czasu  $[t, t+h]$ ,

- do pudełka  $i$  wpada z zewnątrz 1 cząstka z prawdopodobieństwem w przybliżeniu równym  $a_{*i}h$ ,
- każda spośród cząstek znajdujących się w pudełku  $i$  wypada z niego na zewnątrz z prawdopodobieństwem w przybliżeniu równym  $a_{i\dagger}h$  lub przemieszcza się do pudełka  $j$  z prawdopodobieństwem w przybliżeniu równym  $a_{ij}h$ , niezależnie od pozostałych cząstek.

Możemy interpretować pudełka jako wierzchołki grafu. Z wierzchołka  $i$  do  $j$  prowadzi strzałka opatrzona liczbą  $a_{ij}$ . Dodatkowo można sobie wyobrazić dwa fikcyjne pudełka oznaczone symbolami  $*$  i  $\dagger$  (źródło cząstek przybywających z zewnątrz i miejsce gromadzące cząstki opuszczające układ). Jeśli  $a_{i\dagger} = 0$ , to nie ma strzałki między  $i$  i  $\dagger$ . Nasz model jest zatem opisany przez skierowany, ważony graf. Dla uproszczenia wzorów przyjmijmy jeszcze umownie, że  $a_{ii} = 0$  dla wszystkich  $i$ .

Stan układu opisuje wektor (lub „konfiguracja”)  $n = (n_1, \dots, n_i, \dots, n_k)$ , gdzie  $n_i$  oznacza liczbę cząstek w pudełku  $i$ . Definicja intensywności przejścia w wektorowym przypadku pozostaje taka sama, wzór (1) pozostaje w mocy.

Potrzebujemy jeszcze kilku oznaczeń. Niech  $n_{-i}$  oznacza wektor  $n$  z pominiętym elementem  $n_i$ , czyli  $(n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k)$ . Co oznacza  $n_{-i-j}$ , można się łatwo domyślić. Możemy teraz przetłumaczyć nasze założenia na ścisły język intensywności przejścia:

$$(5) \quad Q(n, n') = \begin{cases} a_{*i} & \text{jeśli } n'_{-i} = n_{-i}, n'_i = n_i + 1 \text{ dla pewnego } i, \\ n_i a_{ij} & \text{jeśli } n'_{-i-j} = n_{-i-j}, n'_i = n_i - 1, n'_j = n_j + 1 \text{ dla pewnych } i \neq j, \\ n_i a_{i\dagger} & \text{jeśli } n'_{-i} = n_{-i}, n'_i = n_i - 1 \text{ dla pewnego } i, \\ 0 & \text{we wszystkich pozostałych przypadkach.} \end{cases}$$

**Twierdzenie 2.** Rozpatrzmy proces zdefiniowany powyżej, w którym intensywności przejścia są dane wzorem (5). Załóżmy, że układ liczb  $\lambda_i$  spełnia następujące „równania równowagi średniej”:

$$(6) \quad \sum_j \lambda_j a_{ji} + a_{*i} = \lambda_i \left( \sum_j a_{ij} + a_{i\dagger} \right) \quad \text{dla każdego } i.$$

Wtedy rozkład prawdopodobieństwa  $\pi$  dany wzorem

$$\pi(n) = \prod_i e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}$$

jest rozkładem równowagi.

Zanim podam dowód, skomentuję założenia i tezę. Później opowiem o swojej przygodzie związanej z tym twierdzeniem.

Założenie (6) dotyczy w istocie średnich przepływów w grafie. Jeśli w pudełku  $i$  średnio znajduje się  $\lambda_i$  cząstek, to średnio  $\lambda_i a_{ij} h$  cząstek przemieszcza się z  $i$  do  $j$  w krótkim odcinku czasu  $h$ . Podobna interpretacja stosuje się do  $a_{*i}$  i  $a_{i\dagger}$ . Warunek (6) mówi zatem tyle, że *średnio* tyle samo cząstek wpływa do  $i$  co wypływa z  $i$ . Stabilizowanie się wartości średnich jest warunkiem koniecznym dla stabilizowania się rozkładów prawdopodobieństwa.

W tezie Twierdzenia 2 jest zawarty fakt, że poszczególne komponenty  $n_i$  wektora  $n$  są, w rozkładzie równowagi, statystycznie niezależne (prawdopodobieństwo stanu  $n$  jest iloczynem czynników zależnych od pojedynczych  $n_i$ ). Otóż ten fakt jest zadziwiający i przeczy naiwnej intuicji! Rozważmy bardzo prosty graf naszkicowany poniżej.

$$* \quad \xrightarrow{a_{*1}} \mathbf{1} \quad \xrightarrow{a_{12}} \mathbf{2} \quad \xrightarrow{a_{2\dagger}} \dagger$$

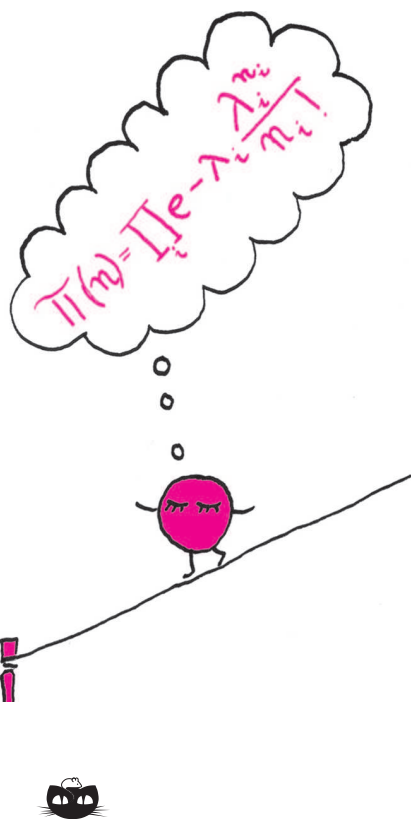
Do pudełka  $\mathbf{2}$  cząstki mogą wpaść tylko z pudełka  $\mathbf{1}$ . Trudno uwierzyć, że zawartość pudełka  $\mathbf{2}$  nie zależy od liczby cząstek w  $\mathbf{1}$ . Żeby ten pozorny paradoks wyjaśnić (przynajmniej w pewnym stopniu), zauważmy, że twierdzenie mówi o niezależności  $n_1$  i  $n_2$  w tej samej chwili. Nie wyklucza wpływu  $n_1$  w momencie  $t$  na wartość  $n_2$  w późniejszym czasie  $t + h$ . Przy okazji zauważmy, że model z dwoma pudełkami Smoluchowskiego (połączonymi tak, jak pokazano powyżej) w oczywisty sposób *nie jest* odwracalny w czasie.

**Dowód Twierdzenia 2.** Mamy wykazać, że dla dowolnego stanu  $n$  zachodzi warunek (3). Napiszmy

$$\pi(n) = \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!} = \pi(n_{-i-j}) e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!} e^{-\lambda_j} \frac{\lambda_j^{n_j}}{n_j!}.$$

Możemy uznać, że powyższy wzór czysto algebraicznie definiuje  $\pi(n_{-i})$  i  $\pi(n_{-i-j})$ . Przypomnijmy jeszcze stałe obowiązującą umowę, że  $a_{ii} = 0$ . We wszystkich poniższych wzorach indeksy  $i, j$  przebiegają zbiór  $\{1, \dots, k\}$  i nie będziemy tego powtarzać. Biorąc pod uwagę postać intensywności  $Q(n, n')$ , wnioskujemy, że lewa strona wzoru (3) (odpowiadająca wyjściu ze stanu  $n$ ) jest równa

$$\begin{aligned} \text{WYJŚCIE} &= \sum_i \sum_j \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!} n_i a_{ij} + \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!} n_i a_{i\dagger} + \\ &+ \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!} a_{*i}. \end{aligned}$$



**Rozwiązanie zadania F 878.** Energia fotonu odpowiadającego promieniowaniu o częstotliwości  $f$  wynosi  $hf$ . Naturalna szerokość linii widmowej  $\Delta E$  odpowiada więc przedziałowi częstotliwości  $\Delta f = (2\pi\tau)^{-1}$ . Fali światła o długości  $\lambda$  odpowiada częstotliwość  $f = c/\lambda$ , gdzie  $c$  oznacza prędkość światła. Jeśli źródło emitujące światło o częstotliwości  $f$  porusza się z prędkością  $v$  w kierunku odbiornika, to ze względu na zjawisko Dopplera, odbiornik zarejestruje częstotliwość  $f'$  różną od  $f$ , z dokładnością do wyrazów liniowych w  $v/c$ ,  $f' = f(1 + v/c)$ . Składowa prędkości termicznego ruchu atomów w kierunku odbiornika podlega rozkładowi Maxwella o gęstości:

$$P(v) = \sqrt{\left(\frac{M}{2\pi kT}\right)} \exp\left(-\frac{Mv^2}{2kT}\right).$$

Różnica częstotliwości emitowanej i rejestrowanej  $(f' - f)/f = v/c$ , podlega więc rozkładowi  $P(f')$ , który jest proporcjonalny do:

$$\exp\left(-\frac{Mc^2(f' - f)^2}{2kTf^2}\right),$$

co odpowiada

$$\text{FWHM} = \sqrt{\left(\frac{8kT \ln 2}{Mc^2}\right)} f.$$

Tak obliczona wartość FWHM jest równa naturalnej szerokości  $\Delta f = (2\pi\tau)^{-1}$ , dla temperatury

$$T = \frac{M\lambda^2}{4\pi^2 \cdot 8k\tau^2 c^2 \ln 2}.$$

Dla podanej linii magnezu  $T = 0,18$  K.

Pierwszy wyraz odpowiada takim przejściom z  $n$  do  $n'$ , że  $n'_i = n_i - 1$  i  $n'_j = n_j + 1$  (cząstka przemieszcza się z pudełka  $i$  do  $j$ ). Drugi wyraz – to przejścia  $n'_i = n_i - 1$  (cząstka ucieka z pudełka  $i$  na zewnątrz). Trzeci – to przejścia  $n'_i = n_i + 1$  (cząstka przychodzi z zewnątrz do pudełka  $i$ ). Łącząc pierwszy i drugi wyraz, otrzymujemy



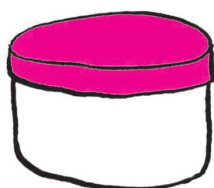
$$\text{WYJŚCIE} = \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i-1}}{(n_i-1)!} \lambda_i \left( \sum_j a_{ij} + a_{i\dagger} \right) + \pi(n) \sum_i a_{*i}.$$

Podobnie, prawa strona wzoru (3) (odpowiadająca wejściu do stanu  $n$ ) jest równa



$$\begin{aligned} \text{WEJŚCIE} &= \sum_j \sum_i \pi(n_{-j-i}) e^{-\lambda_j} \frac{\lambda_j^{n_j+1}}{(n_j+1)!} e^{-\lambda_i} \frac{\lambda_i^{n_i-1}}{(n_i-1)!} (n_j+1) a_{ji} + \\ &+ \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i-1}}{(n_i-1)!} a_{*i} + \\ &+ \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i+1}}{(n_i+1)!} (n_i+1) a_{i\dagger}. \end{aligned}$$

Pierwszy wyraz odpowiada takim przejściom z  $n'$  do  $n$ , że  $n_j = n'_j - 1$  i  $n_i = n'_i + 1$  (cząstka przemieszcza się z pudełka  $j$  do  $i$ ). Drugi wyraz – to przejścia  $n_i = n'_i + 1$  (cząstka przychodzi z zewnątrz do pudełka  $i$ ). Trzeci – to przejścia  $n_i = n'_i - 1$  (cząstka ucieka z pudełka  $i$  na zewnątrz). Analogicznie jak poprzednio,



$$\text{WEJŚCIE} = \sum_i \pi(n_{-i}) e^{-\lambda_i} \frac{\lambda_i^{n_i-1}}{(n_i-1)!} \left( \sum_j \lambda_j a_{ji} + a_{*i} \right) + \pi(n) \sum_i \lambda_i a_{i\dagger}.$$

Równania (6) implikują

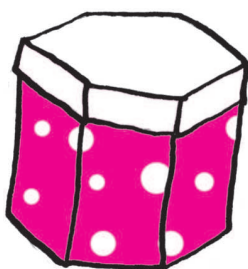
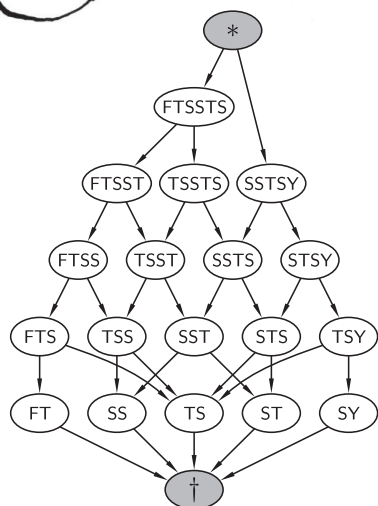
$$\text{WEJŚCIE} = \text{WYJŚCIE},$$

co kończy dowód.  $\square$

Proces zdefiniowany równaniem (5) pojawia się przy modelowaniu pewnych zjawisk biologicznych. Popatrzmy na graf na rysunku. Wierzchołki odpowiadają *peptydom*, które w uproszczeniu identyfikujemy z ciągami liter z 20-elementowego alfabetu. *Egzopeptydazy* to enzymy, które „odrywają jedną literkę z prawej lub lewej strony ciągu”. Każda krawędź grafu odpowiada działaniu specyficznej egzopeptydazy. Stan układu opisuje wektor, zawierający liczby cząstek peptydów,

$$n = (n_1, \dots, n_k).$$

Oderwanie literki jest równoważne przejściu cząstki z jednego wierzchołka do drugiego wzdłuż krawędzi grafu. „Długie ciągi” napływają z zewnątrz, ze źródła oznaczonego symbolem  $*$ . Ciągi złożone z pojedynczych liter są nieobserwowane (pochlania je stan umownie nazwany  $\dagger$ ). Jesteśmy w sytuacji przedstawionej poprzednio.



Parę lat temu pracowałem nad tym modelem wraz z dwójmą młodych, wspaniałych bioinformatyków, Anną Gambin i Bogusławem Kluge. W pewnej chwili uświadomiliśmy sobie, że symulacje komputerowe sugerują fakt sformułowany w Twierdzeniu 2. Długo nie mogliśmy w prawdziwość tego faktu uwierzyć, tak się wydawał nieintuicyjny i... podejrzenie prosty. Jak tylko przełamaliśmy barierę psychologiczną, to szybko i łatwo znaleźliśmy dowód (przytoczony powyżej). Kilka miesięcy po opublikowaniu naszej pracy dowiedzieliśmy się, że Twierdzenie 2 było już wcześniej znane. Nasz główny wynik zawarty w publikacji polegał jednak na czymś innym.