

PIOTR H. PAWŁOWSKI

*Instytut Biochemii i Biofizyki PAN
Pawińskiego 5a, 02-106 Warszawa
E-mail: piotrp@ibb.waw.pl*

CZY JESTEŚMY SCALE-FREE?

Kiedy mamy tzw. węzły i linie oraz idee jak połączyć te węzły przy pomocy linii w jedną całość – mamy sieć. Opanowana przez *Homo reticulus*¹ zdolność dostrzegania różnego rodzaju relacji sieciowych w otaczającej go rzeczywistości jest przez niego ostatnio intensywnie wykorzystywana przy poznawaniu świata, do czego w znacznym stopniu przyczynia się rozwój i eksploatacja internetu. Współcześni uczeni, zwłaszcza eksperci z dziedziny nauk o życiu, owocnie wykorzystują ten nowy okołosieciowy styl badawczy w ich codziennej profesjonalnej praktyce. Dzisiejsi bio-badacze, a szczególnie bioinformatycy, spotykają się z sieciami w wielu obszarach swoich dociekań, zróżnicowanych zarówno pod względem specyfiki badanych organizmów, jak i poziomu uniwersalności. Sieci regulacji ekspresji genów, oddziaływań białko-białko, sygnałowe, metaboliczne, ekologiczne, ewolucyjne oraz ich kombinacyjne sprzężenia – to tylko kilka przykładów wybranych z wielu. Węzły sieci biologicznych reprezentują najczęściej biomolekuły, geny, białka, komórki, organy, osobniki lub populacje. Liniami zaś są przeważnie ukazywane w badaniach różnego rodzaju fizyczne, chemiczne bądź też socjologiczne relacje między węzłami, tak osobliwe i zróżnicowane, jak na przykład: oddziaływania, wiązania, regulacje, reakcje, transformacje, aktywacje, inhibicje, asocjacje, dysocjacje, przepływy, strumienie i tym podobne. Biologiczna sieć może zawierać tysiące węzłów i dziesiątki tysięcy linii tego typu. Informacje o nich przechowywane

są zazwyczaj w dużych wyspecjalizowanych bazach danych, na przykład, takiej jak baza DIP (ang. Database of Interacting Proteins), dostępna pod internetowym adresem <http://dip.doe-mbi.ucla.edu>.

Wszystkie domeny biologii: molekularna, komórkowa, organizmalna, populacyjna oraz biosfery są obecnie polem intensywnych poszukiwań sieciowych, myślenia o sieciach, adaptowania sieciowej terminologii i metodologii. Cokolwiek to by miało oznaczać – jest trendy.

Coraz częściej zdajemy sobie sprawę, że nie tylko naukowcy mają na co dzień do czynienia z sieciami. Żyjemy przecież w świecie różnego rodzaju „netów”, niejednokrotnie sami będąc mniej lub bardziej świadomymi ich elementami lub przynajmniej od nich w jakiś sposób zależni. Nie chodzi tu jedynie o sieć bankomatów i sieć komórkową. Ogólnoświatowa sieć komputerowa, internet oraz funkcjonalnie związana z nią sieć hipertekstów połączonych hiperłączami, World Wide Web, dziś to już nie akademickie, lecz wręcz przedszkolne ilustracje omawianego tematu (Ryc. 1). Dla czytelników preferujących przykłady mniej informatyczne, mogą równie dobrze nimi być np. sieć współpracy artystycznej w Hollywood lub też sieci lingwistyczne, a nawet sieci terrorystyczne.

Niezależnie od tego, czy rozważamy sieć fizyczną (np. sieć energetyczną), sieć relacji społecznych (np. sieć znajomych), sieć organizacyjną (np. sieć powiązań formalnych wewnątrz firmy), sieć prezentacji intelektu-

¹*Homo reticulus* — człowiek sieciowy (przyp. aut.)

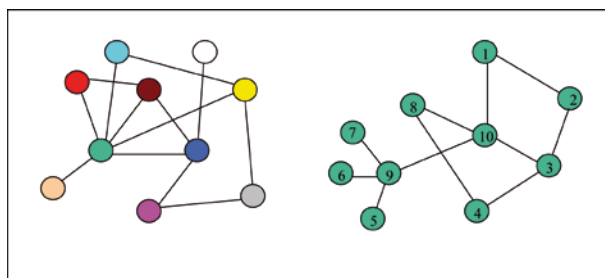


Ryc. 1. Spektakularna wizja World Wide Web z zaznaczonymi aktualnymi domenami i połączeniami na tle kosmosu.

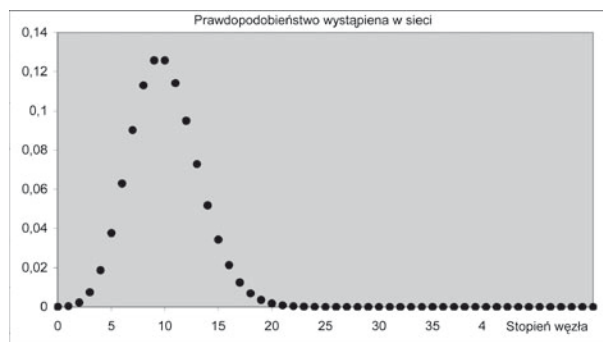
Kolorem zielonym oznaczono domeny europejskie, takie jak np. pl, blisko wschodnie, centralnej Azji i afrykańskie. Pozostałe kolory reprezentują domeny: niebieski – Ameryka Płn., żółty – Ameryka Płd. i Karaiby, czerwony – Azja i Pacyfik, turkusowy – sieci prywatne, biały – niezaklasyfikowane. (Wykorzystano i zmodyfikowano materiały należące do The Opte Project, <http://www.opte.org> oraz Astrohobby.pl, <http://www.astrohobby.pl>).

alnych dokonań (np. sieć cytowań publikacji naukowych), czy też którąś ze wspomnianych wcześniej sieci biologicznych, w każdym z tych przypadków może być ona w prosty sposób przedstawiona w formie diagramu zwanego fachowo grafem, w istocie rzeczy będącego dwuwymiarową aranżacją zestawu rozróżnialnych punktów połączonych prostymi łącznikami (Ryc. 2)

Zgodnie z matematyczną teorią grafów, podstawową wielkością opisującą węzeł sieci jest tzw. stopień węzła (ang. node degree). Określa on bowiem ile końców linii przyna-



Ryc. 2. Przykłady reprezentacji sieci w formie grafu.



Ryc. 3. Sieć Erdosa-Renyiego. Dwumianowy rozkład prawdopodobieństwa występowania w sieci węzłów o różnych stopniach.

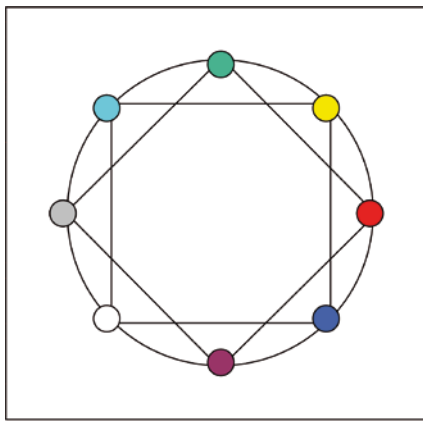
leży do danego węzła, czyli ile dany węzeł ma partnerów lub – jak kto woli – sąsiadów. Dla przykładu, w sieci na Ryc. 2 (strona prawa) węzeł nr 1 ma stopień d równy 2, węzeł nr 5 – stopień 1, a węzeł nr 10 – stopień 4.

Rozkład częstości występowania w całej sieci węzłów o różnych stopniach może, oprócz najprostszej statystycznej charakterystyki stanu powiązań, nieść także pewne wskazówki dotyczące sposobu ich powstawania. Na przykład, dwumianowy rozkład stopni węzłów, zwany czasem rozkładem Bernoulliego (ERDOS i RENYI 1959, 1960), charakteryzuje sieć o N węzłach, powstałą w taki sposób, że węzły jej łączono losowo, ze stałym prawdopodobieństwem dodając każde z możliwych $N(N-1)/2$ połączeń.

Na wykresie widać (Ryc. 3), iż rozkład ten cechuje liczebna dominacja węzłów o pewnej wyróżnionej liczbie sąsiadów. Węzły z mniejszym lub z większym node degree występują w sieci ER zdecydowanie rzadziej. Zatem można powiedzieć, że w tego typu sieci istnieje pewna charakterystyczna skala wielkości węzłów najbardziej prawdopodobnych.

Teoretycznie możliwe są również sieci, w których wszystkie węzły mają jeden i ten sam stopień (Ryc. 4), a więc rozkład prawdopodobieństwa jest w ich przypadku bardzo szczególny, bo jednopunktowy. Sieci takie można by nazwać jednoskalowymi.

Klasyczne sieci ER i jednoskalowe kraty nie często jednak spotyka się w rzeczywistości. Jak wynika z przeprowadzonych badań (ALBERT i współaut. 1999, BARABASI i ALBERT 1999), sieci empiryczne, takie jak np. www, sieć współpracy aktorów, sieć cytowań czy też sieć energetyczna, charakteryzują się rozkładami liczebności węzłów opisanymi



Ryc. 4. Przykład sieci jednoskalowej. Wszystkie węzły mają stopień równy 4.

wzorem diametralnie odmiennym od powyższych. Zdaniem badaczy, uniwersalne prawo potęgowe

$$P(d) = \text{const} \cdot d^{-\alpha} \quad 1$$

zwane też prawem Pareto, z wykładnikiem α należącym do wąskiego przedziału wartości pomiędzy 2,1 a 4, jest właśnie tym, które poprawnie opisuje omawiane rozkłady w analizowanych przypadkach.

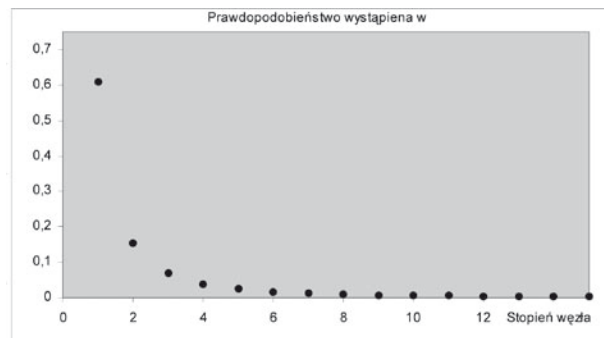
Już przy pobieżnym oglądzie (Ryc. 5) widać, iż w sieciach o rozkładzie potęgowym trudno jest wskazać charakterystyczną skalę wielkości węzłów. Nie to jednak sprawia, że sieci te nazywamy scale-free, bezskalowymi, lub też mówimy czasem prowokująco o bezskalowym skalowaniu krotności węzłów.

Bezskalowość ma tu znacznie bardziej podstawowe i dotyczy całego rozkładu. Prawo potęgowe jest, bowiem niezmiennicze względem operacji skalowania stopnia węzłów – tzn., że dla transformacji skalowania o skali s zachodzi związek:

$$P(sd) = s^{-\alpha} P(d) \quad 2$$

Mówi on, iż po przeskalowaniu stopni węzłów o rozkładzie potęgowym ich rozkład pozostaje nadal opisany prawem potęgowym, będącym przeskalowaną wersją prawa wyjściowego. Innymi słowy, analizując węzły sieci bezskalowej w różnych skalach wielkości uzyskujemy zawsze, niezależnie od skali, rozkłady podobne (Ryc. 6).

Losowo wybrana podsieć sieci bezskalowej na ogół również zachowuje potęgowe prawo rozkładu stopni, czyli jest też bezskalowa. Część sieci bezskalowych okazuje się być także bezskalowymi po zastosowaniu różnego rodzaju procedur renormalizacyjnych, polegających na zastępowaniu grup bliskich

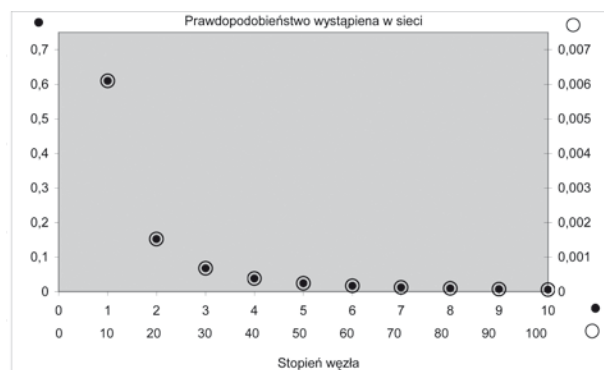


Ryc. 5. Rozkład potęgowy w typowej sieci rzeczywistej.

sobie węzłów pojedynczymi reprezentantami (Ryc. 7). Wskazuje to na samopodobieństwo tychże sieci, rozpatrywanych w różnych skalach odległości, a więc na właściwości fraktalne. Wspomniane cechy czynią sieci bezskalowe interesującymi obiektami badań.

Uwarunkowaną prawem potęgowym topologię sieci bezskalowej zwykle cechuje obecność niewielkiej liczby węzłów o wysokich stopniach, tzw. hubów oraz stosunkowo duża liczba węzłów o stopniach niskich (Ryc. 8). Sprawia to, iż sieć taka z reguły jest odporna na przypadkowe „awarie”, a podatna na precyzyjnie zaplanowane „ataki”.

Naturę ustalonego potęgowego rozkładu stopni węzłów w sieciach rzeczywistych tłumaczy model Barabasięgo-Alberta (MATSUBARA 2006), jako przejaw ciągłego wzrostu sieci i preferencyjnego dołączania nowych węzłów do węzłów już istniejących, z prawdopodobieństwem proporcjonalnym do liczby istniejących już sąsiadów. Istnieją też wyjaśnienia konkurencyjne.



Ryc. 6. Sieć bezskalowa.

Rozkład prawdopodobieństwa występowania węzłów o różnych stopniach analizowany w dwóch różnych skalach wielkości, oznaczonych odpowiednio przez punkty i kółka.

Tabela 1. Przykłady praw allosterycznych.

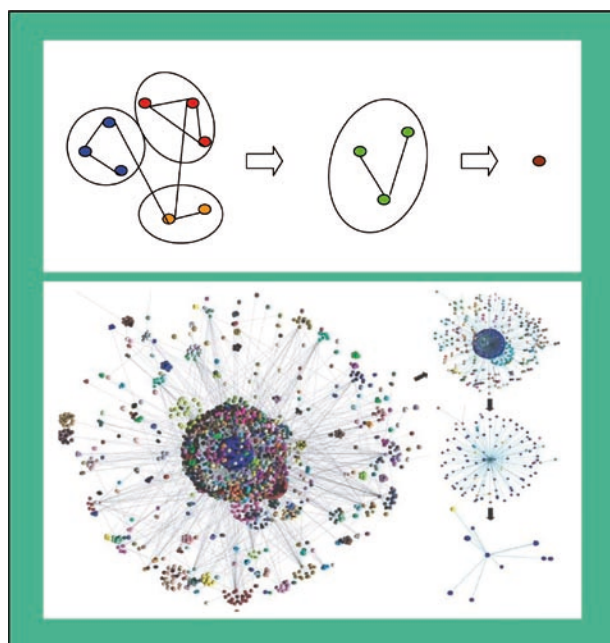
$q_0 \sim M^{\frac{3}{4}}$	q_0 — szybkość metabolizmu M — masa ciała
$t \sim M^{\frac{1}{4}}$	t — czas trwania oddechu, czas bicia serca M — masa ciała
$A \sim M^{\frac{7}{8}}$	A — powierzchnia kontaktowa transferu masy M — masa ciała
$V_{opt} \sim M^{\frac{1}{6}}$	V_{opt} — optymalna szybkość latania owadów i ptaków M — masa ciała

Prawo potęgowe, dzięki swej ujmującej prostocie, posiada dla każdego badacza przyrody niezaprzeczalny urok i pewną magię intelektualnego oddziaływania. Nie można się temu dziwić, gdyż jedne z najważniejszych praw fizyki, tzn. prawo Newtona (oddziaływania mas) i prawo Coulomba (oddziaływania ładunków elektrycznych) — to właśnie prawa potęgowe.

Prawa potęgowe wydają się być uniwersalne. Opisują zarówno fluktuacje korelacji w rozkładzie galaktyk (MATSUBARA 2006), jak i

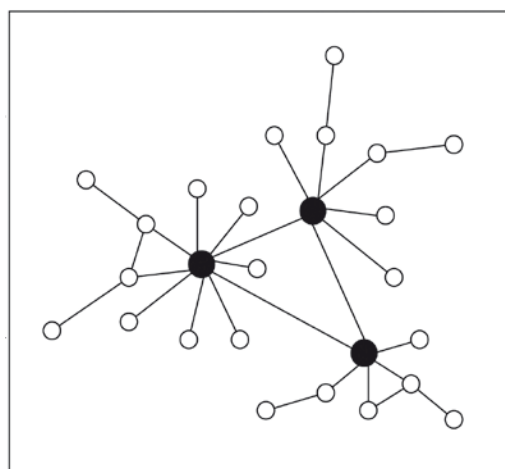
zachowanie się systemów fizycznych podczas krytycznych przejść fazowych (LI i współaut. 2005). Od ich zastosowań nie jest wolna również biologia. Dominowały tu do niedawna w formie tzw. praw allosterycznych, wiążących ze sobą różne cechy organizmów żywych. (Tabela 1).

W 2000 r. Hawkoong Jeong wraz kolegami (JEONG i współaut. 2000) w liście do Nature napisał: „... sieci metaboliczne (badanych 43 organizmów — przyp. aut.) mają te same topologiczne właściwości skalowania oraz wykazują uderzające podobieństwa do organizacji złożonych systemów niebiologicznych. Może to wskazywać, że organizacja metabolizmu jest nie tylko identyczna dla wszystkich żywych organizmów, lecz także podlega prawom projektowania odpornych



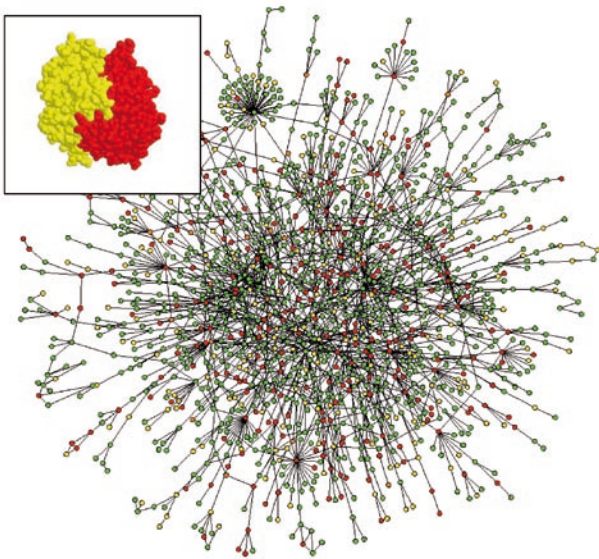
Ryc. 7. Przykład procedury renormalizacyjnej (góra) zastosowanej do analizy podzbioru sieci www (dół).

(Wykorzystano i zmodyfikowano materiały z pracy SONGA i współaut. 2005).



Ryc. 8. Sieć bezskalowa.

Na ogół posiada ona kilka tzw. hubów, czyli węzłów z dużą ilością połączeń (kolor czarny) oraz dużo węzłów z kilkoma połączeniami (kolor biały).

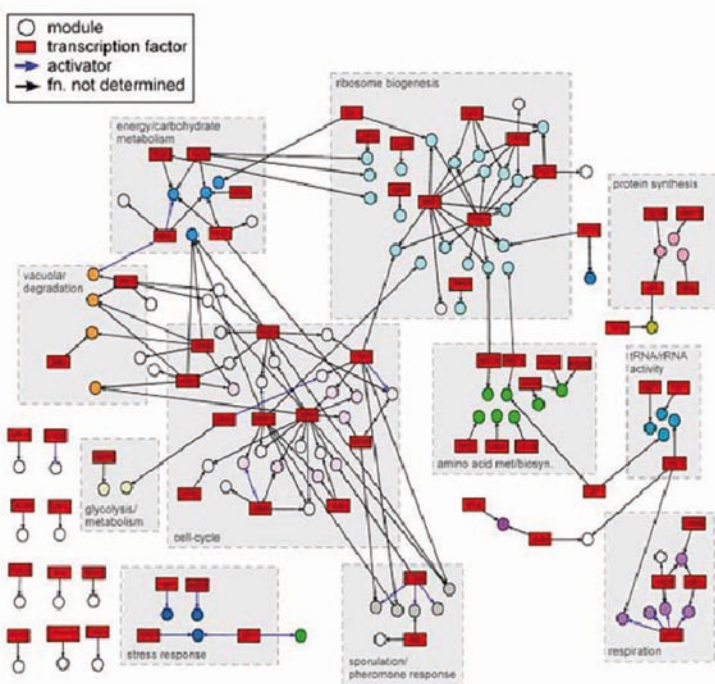


Ryc. 9. Sieć oddziaływań białko-białko u *Saccharomyces cerevisiae*.

Kółka oznaczają białka, linie zaś – interakcje. Czerwonym kolorem wyróżniono tzw. białka letalne tzn. białka, których usunięcie skutkuje śmiercią komórki. Ponadto, kolor pomarańczowy wskazuje białka, których nieobecność spowalnia wzrost komórek, zielony – białka nieletalne, zaś żółty – białka o nieustalonym znaczeniu. W insercie zamieszczono poglądowy model oddziaływania. (Wykorzystano i zmodyfikowano materiały należące do Bordalier Institute, <http://www.bordalierinstitute.com> oraz Online Digital Education Connection, <http://www.odec.ca>).

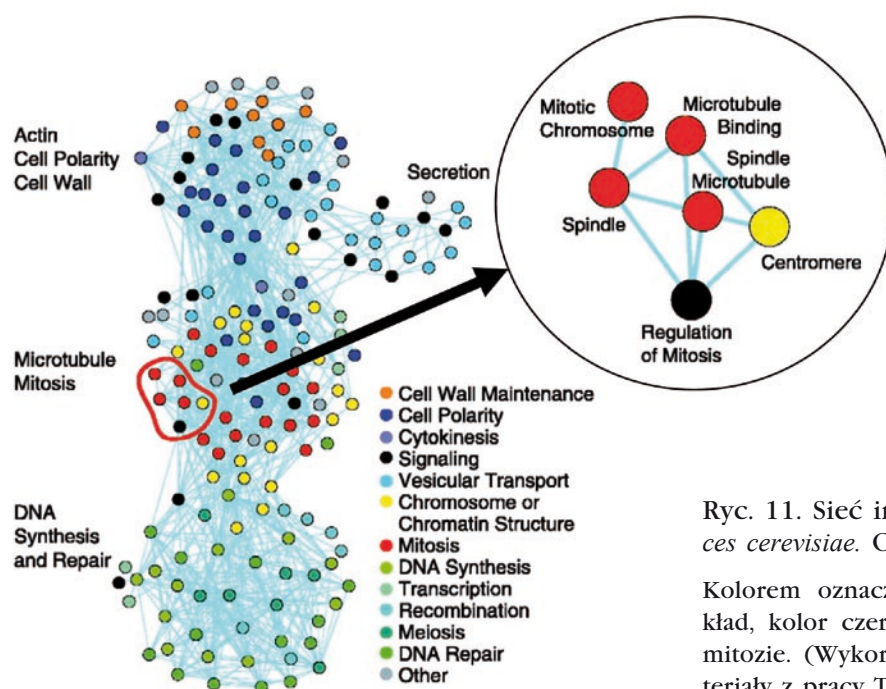
oraz tolerujących błędy sieci bezskalnych i może reprezentować powszechny plan wielkoskalowej organizacji oddziaływań pomiędzy wszystkimi składnikami komórki.” Jako, że autorzy wspomnianego listu pod pojęciem złożonych systemów niebiologicznych rozumieli głównie sieci będące dziełem celowej działalności ludzkiej, „podprogowo” niejako zostało wysłane w świat metafizyczne przesłanie, iż nieświadomie dotąd tworzyliśmy i nadal tworzymy dzieła według tych samych reguł, według których powstaliśmy. Stąd już tylko jeden krok pozostał by odżył na nowo spór współczesnej nauki z kreacjonistami lub zwolennikami koncepcji Inteligentnego Projektu, skupionymi teraz wokół idei Powszechnego Planu. Na szczęście, jak dotąd, tak się nie stało, a cały wysiłek badaczy został skupiony na uzyskaniu odpowiedzi na pytanie - czy rzeczywiście sieci biologiczne, jak np. sieć oddziaływań białkowych (Ryc. 9) lub też sieć regulacji genów (Ryc. 10), są scale-free?

Dotychczasowe próby rozstrzygnięcia powyższej kwestii nie przyniosły jednak oczekiwanego rezultatu. Dla przykładu, za będącą scale-free uważa się sieć interakcji genów *Saccharomyces cerevisiae* (TONG i współaut. 2004) (Ryc.11). W zależności zaś od przyjętej metodologii jest nią (NOORT i współaut. 2004), bądź nie jest, sieć ich koekspresji (MAGWENE i KIM 2004) (Ryc. 12). Wiele też sieci uznaje się jako scale-free jedynie w ograniczonym zakresie zmienności node degree (KHANIN i WIT 2006). Tego typu rozbieżności nie



Ryc. 10. Sieć regulacji genów u *Saccharomyces cerevisiae*.

Kółka obrazują moduły, czyli kolekcje genów posiadających wspólne regulatory. Kolorami oznaczono wyróżnione kategorie modułów. Dla przykładu, kolor zielony wskazuje biosyntezę i metabolizm aminokwasów. (Wykorzystano i zmodyfikowano materiały należące do Computational Genomics, <http://www.psrg.lcs.mit.edu>).

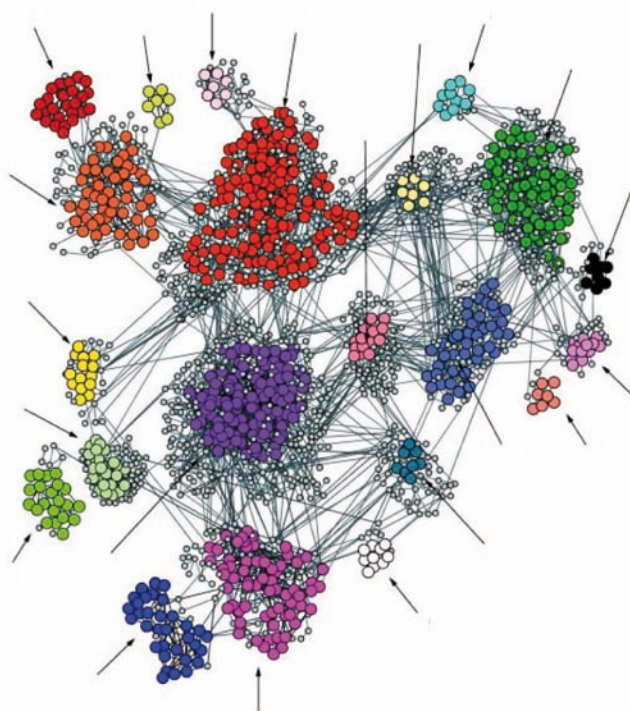


Ryc. 11. Sieć interakcji genów *Saccharomyces cerevisiae*. Oznaczono funkcję genu.

Kolorem oznaczono funkcję genu. Na przykład, kolor czerwony oznacza udział genu w mitozie. (Wykorzystano i zmodyfikowano materiały z pracy TONGA i współaut. 2004).

koniecznie muszą oddawać sedno sprawy. Mogą bowiem pojawiać się w tej stosunkowo młodej dziedzinie badań, na przykład w wyniku zastosowania różnych, często nie uzgodnionych jeszcze, procedur badawczych, różniących się między sobą zarówno ilością, jak i jakością, uzyskiwanych danych, a także rozumieniem przyjmowanych założeń, co może nie zawsze prowadzić do porównywalnych

wyników końcowych. Mogą być też skutkiem błędów popełnionych podczas analizy. Funkcja potęgowa, bowiem, jest w swym przebiegu bardzo podobna do funkcji wykładniczej i z tego powodu, przy małej liczbie niekorzystnie układających się lub tendencyjnie pogrupowanych danych, obie funkcje mogą być ze sobą mylone. W celu zmniejszenia ryzyka tego typu pomyłki stosuje się ostatnio



Ryc. 12. Sieć koekspresji genów *Saccharomyces cerevisiae*.

Kolory odróżniają geny o różnej funkcji, strzałki zaś wskazują wyróżnione podsieci genów połączonych jedynie w ramach swojej grupy. (Wykorzystano i zmodyfikowano materiały należące do Genome Biology, <http://genomebiology.com>).

różnego rodzaju metody statystyczne, takie jak np. MLE (ang. *Maximum Likelihood Estimation*) lub AICc (ang. *corrected Akaike's Information Criterion*), oceniające jakość dopasowania rozkładu teoretycznego do rozkładu obserwowanego oraz testujące propozycje alternatywne. Wśród tych ostatnich wymienić można: uogólnione prawo potęgowe, prawo potęgowe z eksponencjalnym obcięciem, rozciągnięty rozkład eksponencjalny lub rozkład geometryczny. Również i autor tego artykułu wraz z kolegami jest współtwórcą pracy pokazującej, iż rozkład częstości występowania węzłów danego stopnia w sieci oddziaływań białko-białko jest lepiej opisany przez superpozycję dwóch eksponentów, niż przez rozkład potęgowy (PAWŁOWSKI i współaut. 2008) – a więc podważającej w rezultacie ideę uniwersalnej bezskalowalności. Każda ze wspomnianych alternatyw preferuje określone mechanizmy ewolucyjne, zatem, rozstrzygnięcie tu kwestii *stricte* matematyczno-statystycznej może mieć w konsekwencji doniosłe znaczenie dla nauki o ewolucji życia lub też, generalnie, o ewolucji systemów złożonych.

Biologia jest tą dziedziną wiedzy, w której stopień uszczegółowienia znacznie przekracza stopień uogólnień, dlatego też każda jed-

nolita idea obejmująca jak najszerszy obszar badawczy jest przyjmowana przez biologów z wielkim entuzjazmem. Jednak, choć bezskalową architekturę niektórych sieci już teraz łatwo dałoby się wytłumaczyć teoretycznie przy pomocy prostego, a zarazem atrakcyjnego, modelu ewolucyjnego, zakładającego duplikację, powstawanie nowych funkcji, lub utratę genów (ALOY i RUSSELL 2004) – na klarowne z punktu widzenia nauki rozstrzygnięcie w tej kwestii musimy jeszcze trochę poczekać.

Pomijając kwestie praktyczne, jak np. możliwość ekstrapolacji własności sieci bezskalowalnej poza obszar poznany, wydaje się, że odpowiedź „tak” na pytanie, czy jesteśmy scale-free, zbliżyłaby nas znacznie do osiągnięcia jednego z najważniejszych celów biologii, jakim jest odkrycie fundamentalnych reguł tworzenia się struktury i funkcjonowania wszystkich komórek – reguł być może będących przejawem uniwersalnych praw organizowania się materii w ogóle. Tym samym wielka unifikacja zapoczątkowana przez fizyków mogłaby być niespodziewanie kończona przez... biologów. Czy tak będzie? – Zobaczmy. Przedstawiciele innych dyscyplin nauki nie zechcą przecież być gorsi.

ARE WE SCALE-FREE?

Summary

This article describes the scale-free property of biological networks. In practice, a scale-free network can be constructed by progressively adding nodes to an existing network by introducing new links to the

existing nodes using preferential attachment mechanism. Are biological networks scale free? Is it an impact of universal mysterious power law on biological networks architecture?

LITERATURA

- ALBERT R., JEONG H., BARABASI A. L., 1999. *Diameter of the world-wide web*. Science 401, 130–131.
- ALOY P., RUSSELL R. B., 2004. *Taking the mystery out of biological networks*. EMBO Rep. 5, 349–350.
- BARABASI A. L., ALBERT R., 1999. *Emergence of scaling in random networks*. Science 286, 509–512.
- ERDOS P., RENYI A., 1959. *On Random Graphs. I*. Publicationes Mathematicae 6, 290–297.
- ERDOS P., RENYI A., 1960. *The evolution of Random Graphs*. Magyar Tud. Akad. Mat. Kutató Int. Kozl. 5, 17–61.
- JEONG H., TOMBOR B., ALBERT R., OLTVAI Z. N., BARABASI A. L., 2000. *The large-scale organization of metabolic networks*. Nature 407, 651–654.
- KHANIN R., WIT E., 2006. *How scale-free are biological networks*. J. Comput. Biol. 13, 810–8.
- LI L., ALDERSON D., DOYLE J. C., WILLINGER W., 2005. *Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications*. Internet Math. 2, 431–523.
- MAGWENE P., KIM J., 2004. *Estimating genomic coexpression networks using first-order conditional independence*. Genome Biology 5, R100.
- MATSUBARA T., 2006. *Statistics and Dynamics in the Large-scale Structure of the Universe*. J. Physics: Conference Series 31, 27–34.
- NOORT V., SNEL B., HUYNEN M. A., 2004. *The yeast co-expression network has a small-world, scale-free architecture and can be explained by a simple model*. EMBO Rep. 5, 280–284.
- PAWŁOWSKI P. H., KACZANOWSKI S., ZIELENKIEWICZ P., 2008. *Protein interaction network. Double exponential model*. JPB 1, 061–067.
- SONG C., HAVLIN S., MAKSE H. A., 2005. *Self-similarity of complex networks*. Nature 433, 392–395.
- TONG A. H., LESAGE G., BADER G. D., DING H., XU H., XIN X., YOUNG J., BERRIZ G. F., BROST R. L., CHANG M. i współaut., 2004. *Global mapping of the yeast genetic interaction network*. Science 303, 808–813.