

IV117: Úvod do systémové biologie

David Šafránek

1.10.2008

Obsah

Pojem modelu a simulace in silico

Statická analýza modelu

Dynamická analýza modelu

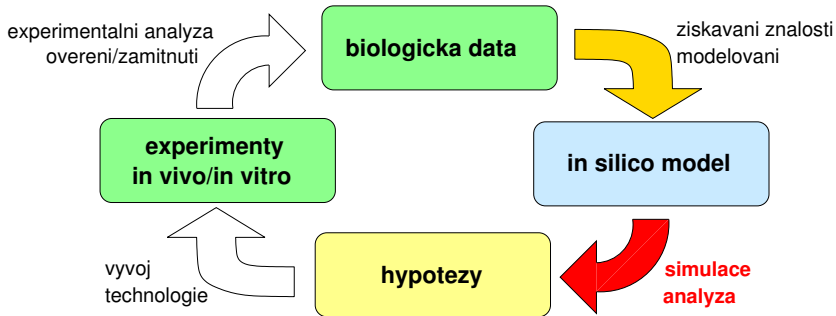
Obsah

Pojem modelu a simulace in silico

Statická analýza modelu

Dynamická analýza modelu

Workflow systémové biologie



Proč dělat model in silico?

- omezené možnosti in vivo/in vitro experimentů
- náročnost experimentů na laboratorní prostředí
- nelze nastavit libovolné iniciální podmínky
- experimenty náročné na čas
- vysoké náklady na biologický materiál, zařízení a zabezpečení experimentů

Co očekávat od modelu?

- nestačí jen znalost o chemických látkách a genech
- důležité je pochopení jejich významu – **funkce**
 - chemické reakce a jejich interakce
- statická analýza modelu – konstrukce sítí
- dynamická analýza modelu – simulace chování při daných iniciálních podmínkách a prostředí
- predikce chování – inspirace pro experimenty in vivo/in vitro
 - tvorba hypotéz na základě in silico modelu
- detailní analýza důvodů (ne)platnosti hypotézy

In silico model

- abstraktní model
- teoretický (idealizovaný) obraz skutečného organismu
- sestává z množiny **proměnných** a množiny **relací** mezi proměnnými
- relace popisují logické i funkční vztahy (interakce proměnných)
- relace mohou obsahovat kvalitativní i kvantitativní informace
- umožňuje simulovat funkční stavy organismu za určitých podmínek

Reprezentace modelu

- biologická síť – komplexní systémový popis organismu
 - neexistuje jednoznačná definice
 - orientovaný nebo neorientovaný graf
 - uzly představují typicky proměnné
 - hrany představují typicky (funkční) relace mezi proměnnými
- k uzlům a relacím jsou přiřazeny kvalitativní i kvantitativní informace potřebné k simulaci (dynamická analýza)
- biologické sítě lze přímo zkoumat – statická analýza
 - vyhledávání alternativních cest
 - srovnávání sítí různých organismů
 - zkoumání měřitelných vlastností sítí

Simulace modelu

- simulace představuje “spuštění” modelu pro dané výchozí nastavení proměnných v daném prostředí
- simulace imituje skutečné chování modelovaného objektu na určité úrovni abstrakce
- simulace umožňuje predikci hypotéz
- simulace zobecňuje a doplňuje in vivo/in vitro experimenty
 - predikce má smysl pouze pro validovaný model!

Obsah

Pojem modelu a simulace in silico

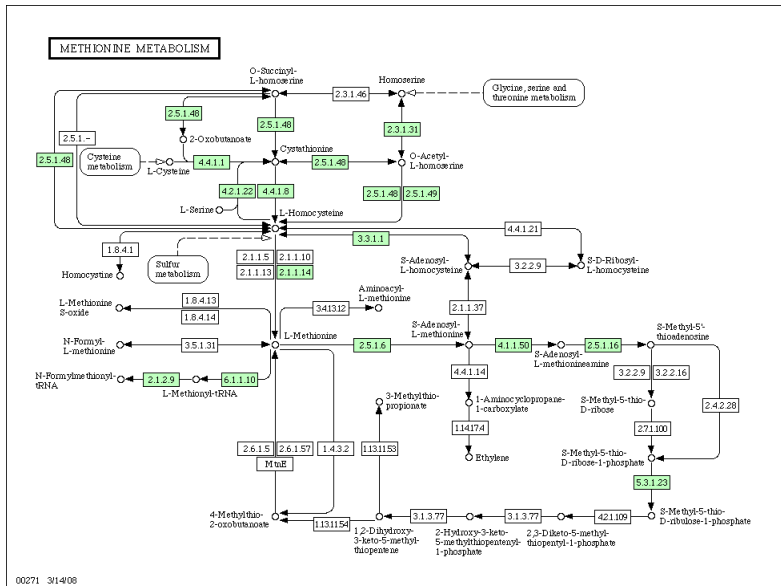
Statická analýza modelu

Dynamická analýza modelu

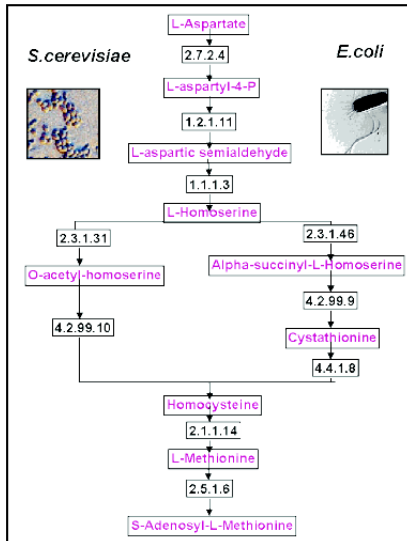
Dráhy vs. síť

- dráhy jsou podsítě lineárního tvaru
 - sekvence metabolických reakcí
 - specifické zaměření na určité proměnné
 - analyzované problémy: délka dráhy, existence alternativních drah
- síť reprezentují komplexní data (zohledňují širokou množinu proměnných a všech relevantních interakcí)
 - komplexy určitého druhu reakcí (transkripce, metabolismus, protein-protein, ...)
 - analyzované problémy: stupeň větvení, délka nejkratší dráhy, modularita, motivy, ...

Vyhledávání alternativních drah – *S. cerevisiae*



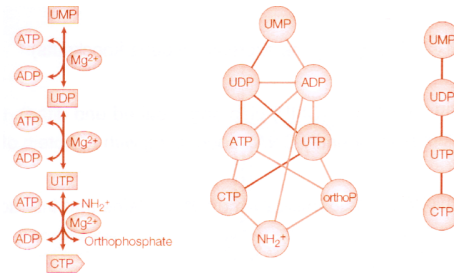
Alternativní dráhy



Alternativní dráhy

- význam v genetice a genomice
 - modifikace drah souvisí přímo s vývojem genomu (evoluce)
 - identifikace neznámých genů
- význam v biotechnologii
 - identifikace a implementace alternativních variant enzymů, substrátů, produktů i celých drah
- význam ve farmakologii
 - laterální náhrada genu
 - metabolicky-specifické medikamenty

Biologické sítě



- různé typy sítí:
 - regulatorní síť (popis transkripční regulace)
 - proteinové síť (popis interakce proteinů)
 - metabolické síť (popis metabolismu)
 - signální síť (popis aktivačních/deaktivačních kaskád proteinů a neurotransmiterů)
 - další typy (např. síť genových interakcí)

Biologická síť jako graf

Definition

Nechť V je konečná množina uzlů a $E \subseteq V \times V$ relace.

Biologickou sítí nazveme graf G reprezentovaný uspořádanou dvojicí $G \equiv (V, E)$.

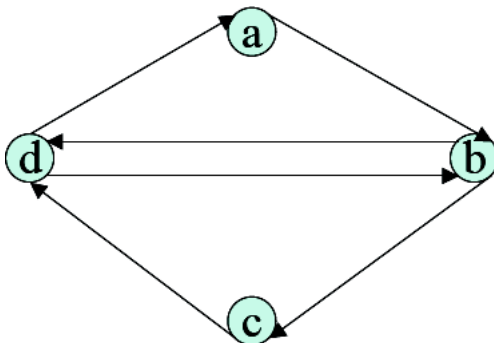
- Pokud $\forall \langle a, b \rangle \in E. \langle a, b \rangle \in E \rightarrow \langle b, a \rangle \in E$, G nazýváme *neorientovaný*.
- V ostatních případech hovoříme o *orientovaném* grafu.

typ sítě	V	E	G
genová regulační	geny (resp. proteiny)	regulace exprese	or.
proteinová	proteiny	proteinové interakce	neor.
metabolická	metabolity	enzymové reakce	or.
signální	molekuly	aktivace/deaktivace	or.

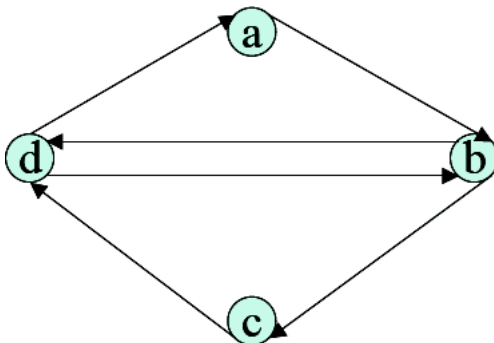
Cesty a kružnice

- *cesta* v grafu je libovolná sekvence uzlů $[a_1, a_2, \dots, a_n]$ t.ž. $\forall i \in \{1, \dots, n-1\}. \langle a_i, a_{i+1} \rangle \in E$, číslo $n-1$ nazýváme *délkou cesty* (počet hran)
- cestu nazveme *primitivní* pokud se na ní každá hrana vyskytuje právě jednou
- cestu nazveme *elementární* pokud se na ní každý vrchol vyskytuje právě jednou
- *kružnice* v grafu je libovolná cesta $[a_1, a_2, \dots, a_n]$ t.ž. $a_1 = a_n$
- *smyčkou* nazýváme libovolnou kružnici délky 1

Cesty a kružnice

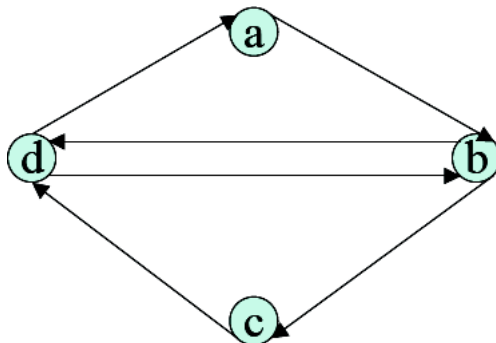


Cesty a kružnice



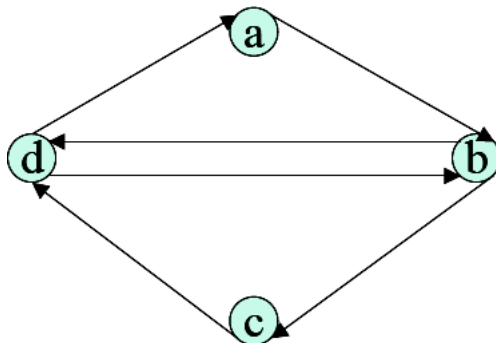
- kolik kružnic ...

Cesty a kružnice



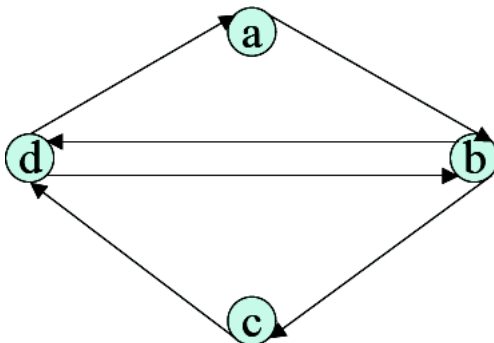
- kolik kružnic ... 4
- kolik cest z a do d ...

Cesty a kružnice



- kolik kružnic ... 4
- kolik cest z a do d ... 2
- délka nejkratší cesty z d do c ...

Cesty a kružnice



- kolik kružnic ... 4
- kolik cest z a do d ... 2
- délka nejkratší cesty z d do c ... $d(d, c) = 2$

Vlastnosti grafu

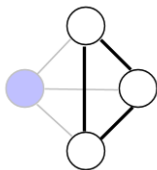
- délku nejkratší cesty z a do b značíme $d(a, b)$
- *charakteristickou délkou cesty* grafu $G \equiv (V, E)$ značíme L_G a definujeme:

$$L_G = \frac{2}{|V|(|V| - 1)} \sum_{a, b \in V} d(a, b)$$

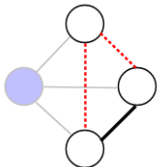
- množinu *sousedních uzlů* uzlu a značíme N_a a definujeme $N_a = \{b \in V \mid \langle a, b \rangle \in E \vee \langle b, a \rangle \in E\}$
- *stupeň* uzlu a značíme k_a a definujeme jako počet všech sousedních uzlů uzlu a , tedy $k_a = |N_a|$
- *koeficient seskupení uzlu* (clustering coefficient [Watts, Strogatz]) a značíme C_a a definujeme:

$$C_a = \frac{|\{\langle c, d \rangle \in E \mid c \in N_a \wedge d \in N_a\}|}{k_a(k_a - 1)}$$

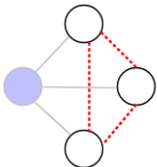
Vlastnosti grafu



$$c = 1$$



$$c = 1/3$$

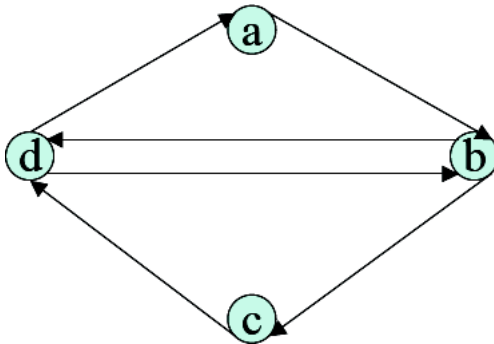


$$c = 0$$

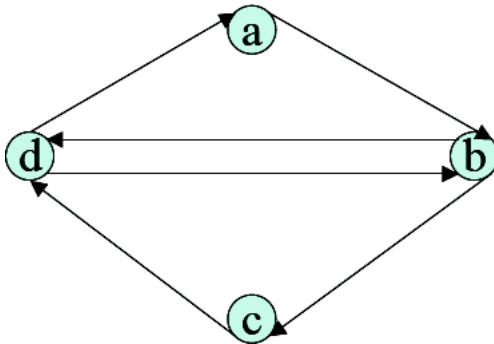
- *koeficient seskupení grafu* G je značen C_G a definován jako průměr klastrovacích koeficientů všech uzlů:

$$C_G = \frac{1}{|V|} \sum_{a \in V} C_a$$

Vlastnosti grafu

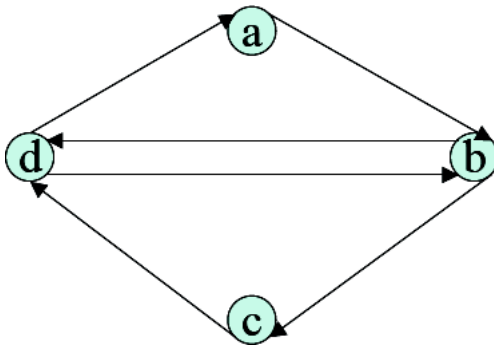


Vlastnosti grafu



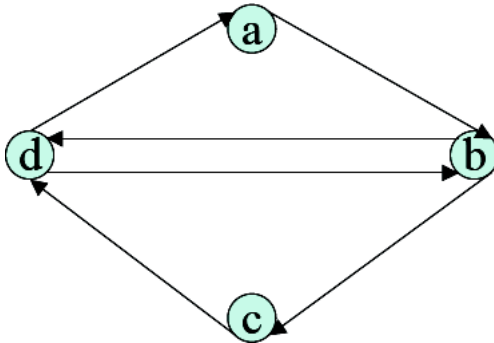
$$C_a =$$

Vlastnosti grafu



$$C_a = \frac{2}{2} = 1$$

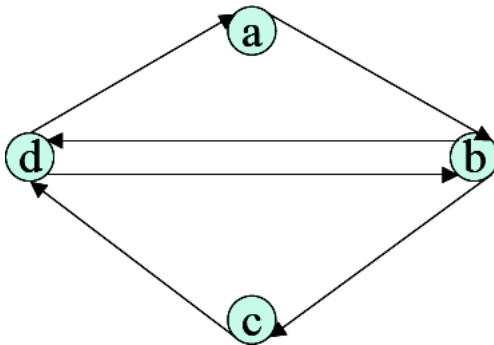
Vlastnosti grafu



$$C_a = \frac{2}{2} = 1$$

$$C_b =$$

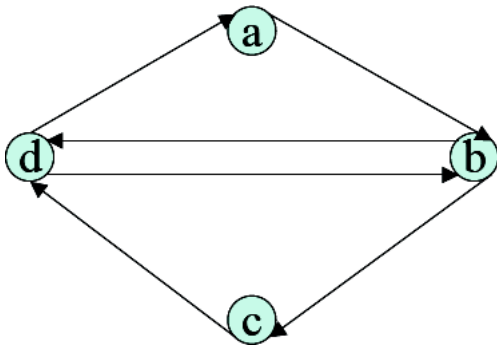
Vlastnosti grafu



$$C_a = \frac{2}{2} = 1$$

$$C_b = \frac{2}{6} = \frac{1}{3}$$

Vlastnosti grafu

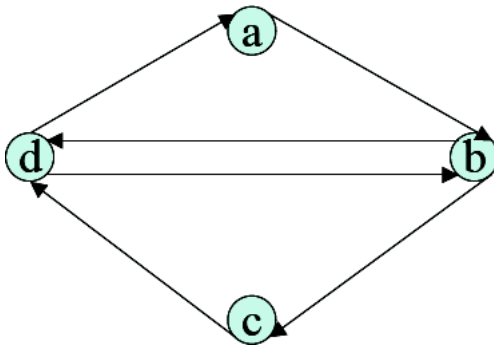


$$C_a = \frac{2}{2} = 1$$

$$C_b = \frac{2}{6} = \frac{1}{3}$$

$$C_G =$$

Vlastnosti grafu



$$C_a = \frac{2}{2} = 1$$

$$C_b = \frac{2}{6} = \frac{1}{3}$$

$$C_G = \frac{1}{4} \cdot \left(2 + \frac{2}{3}\right) = 0.65$$

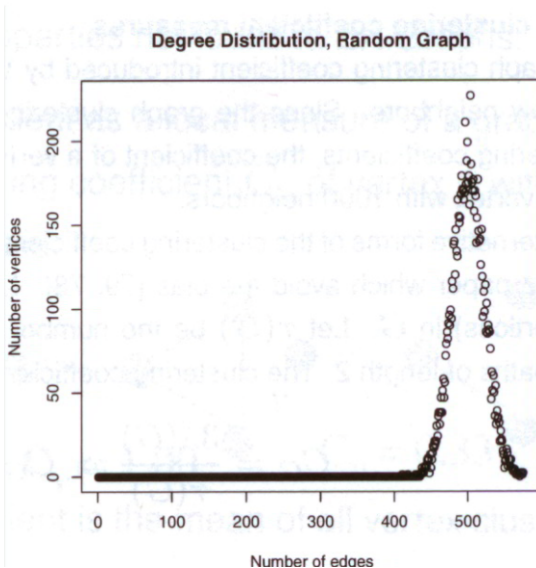
Náhodný graf

- *náhodný graf* je definován pevným počtem uzlů a pravděpodobností p existence hrany mezi libovolnými dvěma uzly
- pravděpodobnost, že v náhodném grafu má daný uzel stupeň k , je charakterizována Poissonovým rozložením (s konst. λ):

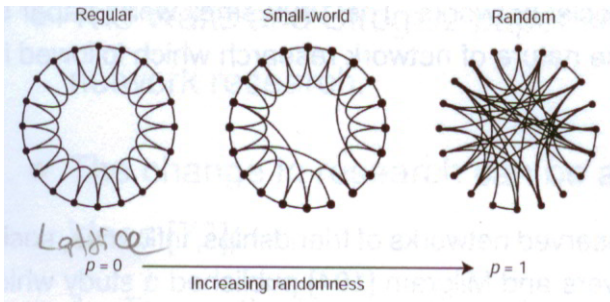
$$f(k|\lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

[Erdős, Rényi, “On the evolution of random graphs”]

Náhodný graf



Vlastnosti sítí

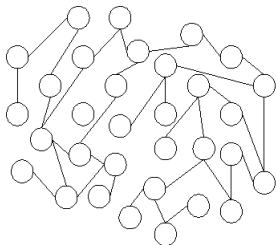


typ grafu	C_G	L_G
svaz	vysoké	dlouhé
náhodný graf	nízké	krátké
small-world	vysoké	krátké

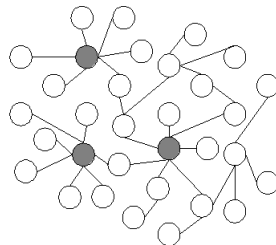
Small-world síť

- zavedeny Wattsem a Strogatzem, “Collective dynamics of ‘small-world’ networks”, Nature 393, 1998
- klíčem jsou lokální a globální metriky seskupení uzlů a metrika charakteristické délky cesty
- identifikovány jako grafy s vysokým koeficientem seskupení ale krátkou charakteristickou délkou cesty
- bylo prokázáno, že mnoho reálných sítí má tento charakter
 - např. graf filmových herců propojených dle společného účinkování
 - neuronové síť v *C. elegans*
- výrazný posun v porozumění chování rozsáhlých dynamických systémů
- zavedení pojmu “real-world graphs”

Scale-free síť



(a) Random network



(b) Scale-free network

- zavedl Barabási a Albert, “Emergence of Scaling in Random Networks”, Science 286, 1999

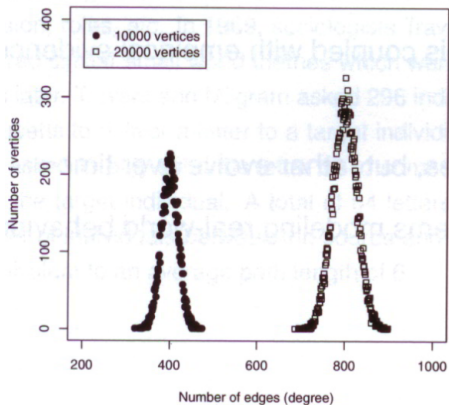
Scale-free sítě

- reálné sítě nejsou statické (nemají pevný počet uzlů), ale vyvíjejí se dynamicky v čase, tzv. “rostou”
- nové uzly se napojují nejvíce k těm uzlům, které jsou se zbytkem sítě již dobře propojeny
- např. metabolické sítě *E. coli* jsou scale-free [Wagner, Fell, 2001]
- označíme-li $P(k)$ pravděpodobnost, že libovolný uzel má stupeň k , pak pro scale-free sítě platí následující úměra (pro konst. λ):

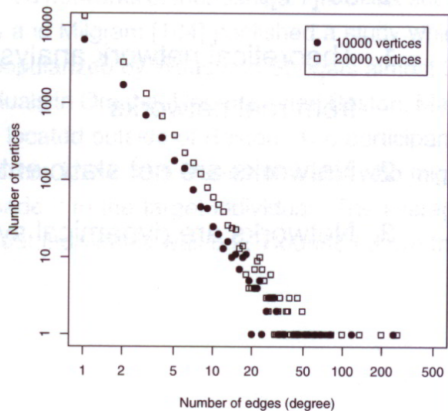
$$P(k) \sim k^{-\lambda}$$

Scale-free síť

Erdos-Renyi random graph



Barabasi-Albert scale-free graph



Motivy ve scale-free sítích

- ve scale-free sítích se vyskytují specifické uzly, tzv. huby – uzly s vysokým stupněm propojení na kostru síťové struktury
- ostatní uzly jsou lokálně napojeny k hubům
- objeveno např. při studiu proteionvé sítě kvasinky pivovarské [Jeong, Mason, 2001]
- díky hubům jsou sítě robustní proti náhodnému vyjmutí uzlu, ale naopak vyjmutí hubu znamená výrazné porušení sítě
- tato struktura vede k hierarchičnosti a modulárnímu charakteru
- jako moduly jsou identifikovány často opakující se výrazné podsítě (motivy) [Alon et.al., “Network Motifs: Simple Building Blocks of Complex Networks”, 2002]

http:

[//www.weizmann.ac.il/mcb/UriAlon/coliData.html](http://www.weizmann.ac.il/mcb/UriAlon/coliData.html)

Motivy

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)				Feed-forward loop			Bi-fun				
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	800 ± 40	41			
Neurons				Feed-forward loop			Bi-fun			Bi-parallel	
<i>C. elegans</i> *	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs				Three chain			Bi-parallel				
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)				Feed-forward loop			Bi-fun			Bi-parallel	
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)				Three-node feedback loop			Bi-fun			Four-node feedback loop	
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s8384	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web				Feedback with two mutual dyads			Fully connected triad			Uplinked mutual dyad	
nd.edu§	325,729	1,466f	1.1e5	2e3 ± 1e2	800	6.8e6	5e4 ± 4e2	15,000	1.2e6	1e4 ± 2e2	5000

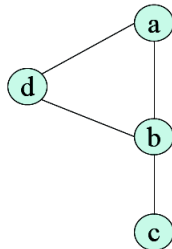
Reprezentace grafů

- maticí sousednosti – matice rozměru $|V| \times |V|$
 $M[a, b] = 1$, pokud $\langle a, b \rangle \in E$;
 $M[a, b] = 0$, jinak.
- maticí incidence – matice rozměru $|V| \times |E|$
 $M[a, e] = 1$, pokud $\exists b \in V. e \equiv \langle a, b \rangle \in E$;
 $M[a, e] = -1$, pokud $\exists b \in V. e \equiv \langle b, a \rangle \in E$;
 $M[a, e] = 0$, jinak.
- nároky na paměť – $\mathcal{O}(|V|^2)$
- reprezentace efektivní pro malé grafy
- pro větší nutno použít seznam sousedů

Reprezentace grafů – matice sousednosti

	8	4	2	1
	a	b	c	d
a	0	1	0	1
b				
c				
d				

Val
5



Metody statické analýzy

- cílem je zkoumat vlastnosti individuální biologické sítě nebo vzájemné porovnání biologických sítí
- vyhledávání v grafu
 - do šířky (breadth-first search – BFS)
 - do hloubky (depth-first search – DFS)
 - iterativní zanořování (kombinace DFS a BFS)
- vyhledávání kružnic
- vyhledávání nejkratších cest
- vyhledávání cest incidujících s danou množinou uzlů
- vyhledávání a identifikace motivů

Nástroje pro statickou analýzu

- uplatnění tradičních grafových algoritmů
- Pathway Hunter Tool
 - analýza nejkratších cest v metabolických drahách
 - identifikace uzlů – enzymy v EC notaci
 - možnost porovnání metabolických drah v různých organismech
 - vyhledávání metabolických cest v databázi KEGG
 - <http://pht.tu-bs.de/PHT/>
- KEGG – <http://www.genome.ad.jp/kegg/>
 - databáze metabolických drah
 - vyhledávání dle enzymových čísel
 - umožňuje nalézt cesty incidující s danou množinou enzymů
 - podporuje grafické vyobrazení metabolických drah (staticky předdefinovaná schémata)

Obsah

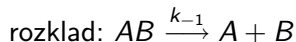
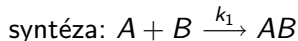
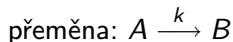
Pojem modelu a simulace in silico

Statická analýza modelu

Dynamická analýza modelu

Modely dynamiky – základní vlastnosti

- chemické reakce — syntéza a rozklad látek v čase
- **stav** — situace pozorovaná v určitém okamžiku
- kinetika — změna stavu za limitní jednotku času
 - rychlost reakce určuje konstanta (reaction rate, *mol/s*)
- ekvilibrium — rovnovážný stav
 - vyčerpání energie (zdrojů)
 - reakce vyrovnávají vzájemný účinek



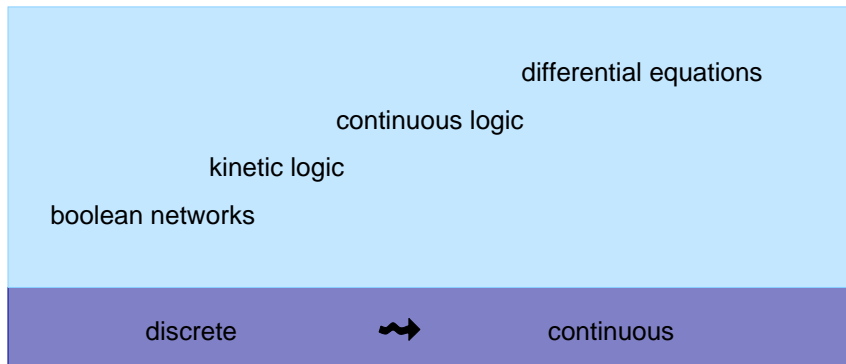
Deterministické modely dynamiky

- makropohled — předpoklad vysoké molární koncentrace látek
 - stav \equiv vektor aktuálních koncentrací látek
 - vyžadují nejvíce kvantitativních znalostí
- diskrétní modely abstrahující čas a hodnoty koncentrací
 - boolovské sítě a kinetická logika
 - stav zachycuje určitou diskrétní úroveň koncentrací
- diskrétní modely abstrahující čas a průběh interakcí
 - stav definován jako u spojitých modelů
 - dynamika řídících interakcí diskretizována

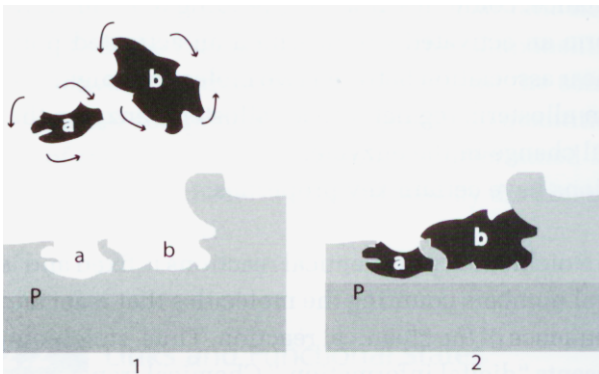
Stochastické modely dynamiky

- mikropohled — interakce individuálních molekul [Gillespie]
- stav zachycuje počty molekul jednotlivých látek
 - může být přímo diskrétní konfigurace
 - nebo distribuční funkce
- vzhledem k čase diskrétní i spojité modely
 - markovovy řetězce
 - stochastické diferenciální rovnice

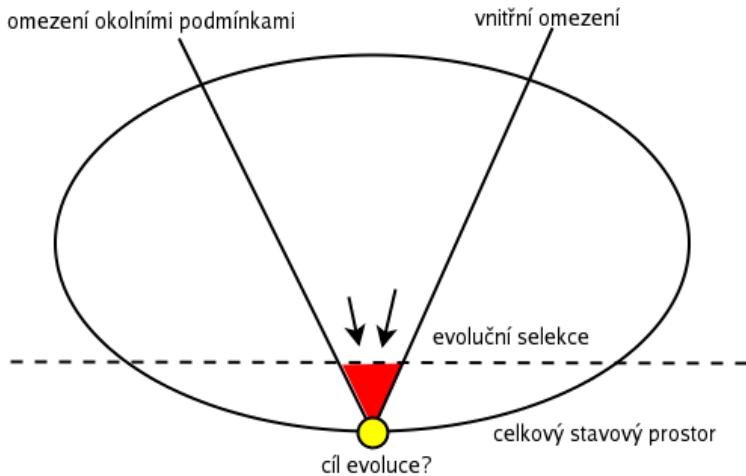
Deterministické modely dynamiky



Koncept vymezení relevantního podprostoru



Koncept vymezení relevantního podprostoru



Rizika systémové biologie

- dogmata nesmí být přeceňována
 - “le programme génétique” [Monod, Jacob]
 - genetický determinismus
 - “the selfish gene” [Dawkins]
- hypotézy musí být řádně testovány
- etická rizika

Shrnutí

- biologický systém definován interakcemi mezi jeho komponentami
- interakce jsou omezeny základními zákony chemie ale i evolučním vývojem
- syntaxí organismu-systému je síť komponent
- sémantikou organismu-systému je jeho funkce (dynamika)
- základní koncepty systémové biologie
 - modely a simulace in silico
 - důraz na interakci, součinnost
 - hierarchie (např. separace časových škál)
 - vymezení relevantního podprostoru možností

Shrnutí

