

# Modelování komplexních sítí

Radek Pelánek



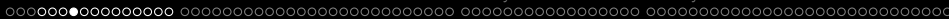






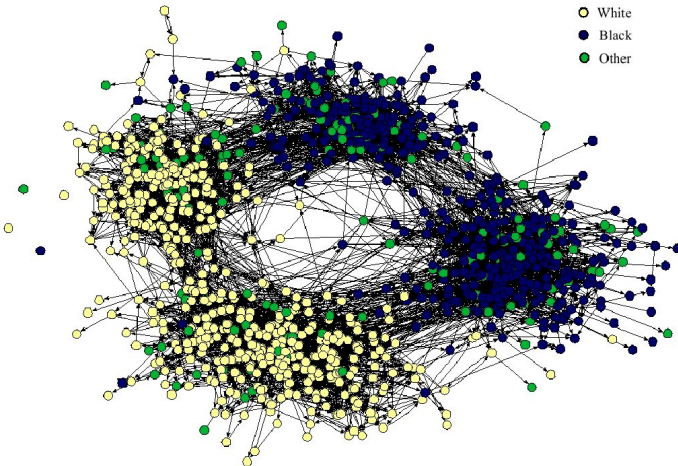






# Sociální sítě

uzly: lidé, hrany: známost

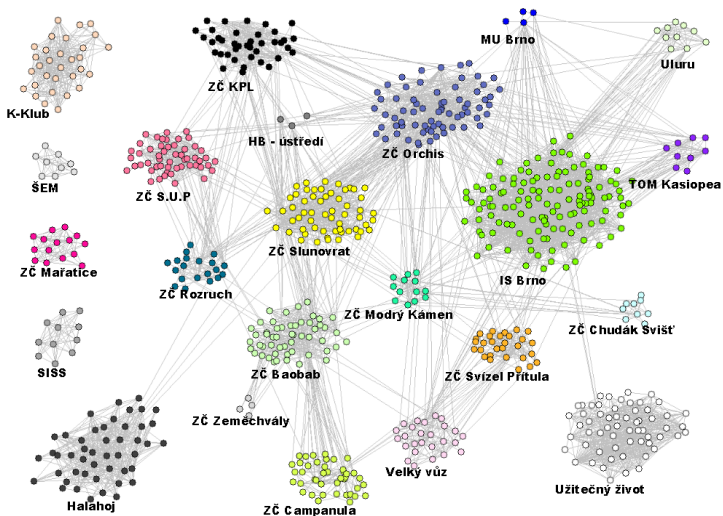








# Spolupráce organizátorů volnočasových akcí



















# Grafy: základní pojmy

- cesty, **vzdálenosti**
- **stupeň** vrcholu  $d_v =$  počet hran z vrcholu  $v$  vycházejících (u orientovaných grafů rozlišujeme výstupní stupeň a vstupní stupeň)
- **distribuce stupňů**  $P(k)$  – pravděpodobnost, že náhodně vybraný uzel má stupeň  $k$



# Malý svět

„malý svět“  $\sim$  potkáte cizího člověka a po chvíli zjistíte, že máte společné známé

průměrné **nejkratší vzdálenosti** mezi uzly v komplexních sítích jsou „malé“

# Milgramův experiment

- Stanley Milgram, 1967
- 60 balíčků, z Kansasu do Massachusetts
- balíčky povoleno posílat jen známým osobám (známost na úrovni křestního jména  $\sim$  tykání)
- identifikace adresáta: jméno, zaměstnání, přibližné místo bydliště
- cíl: co nejrychleji k adresátovi









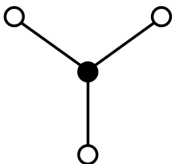
# Shlukování formálněji

- uzel  $v$  má  $k_v$  sousedů
- $e_v$  = počet vzájemně propojených sousedů
- **koeficient shlukování:**

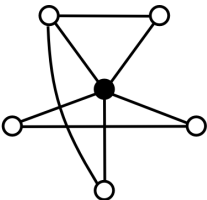
$$C_v = \frac{e_v}{\binom{k_v}{2}} = \frac{2e_v}{k_v(k_v - 1)}$$

# Koeficient shlukování: příklad

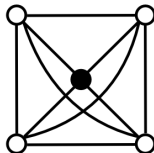
koeficient  
shlukování  
uzlu ●



$$0/3 = 0$$



$$3/10 = 0,3$$

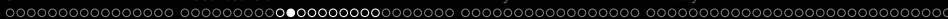


$$6/6 = 1$$

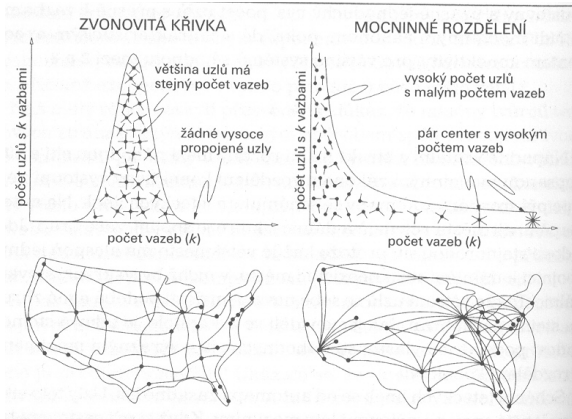
# Distribuce stupňů

Existuje typická hodnota, kolem které se stupeň uzlů pohybuje („škála grafu“)?

- náhodné grafy  $\Rightarrow$  ano
- komplexní síť  $\Rightarrow$  ne



# Poisson a mocninný zákon: příklad

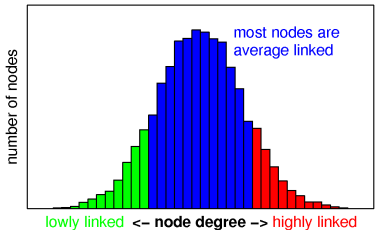


A.-L. Barabási: V pavučině sítí

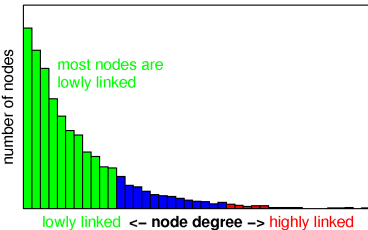


# „Typické“ uzly

random networks



real networks (power-law, scale-free)



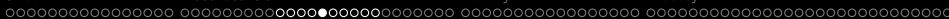
# Distribuce stupňů formálněji

- Poissonova distribuce (pro velké  $\lambda \sim$  normální distribuce)

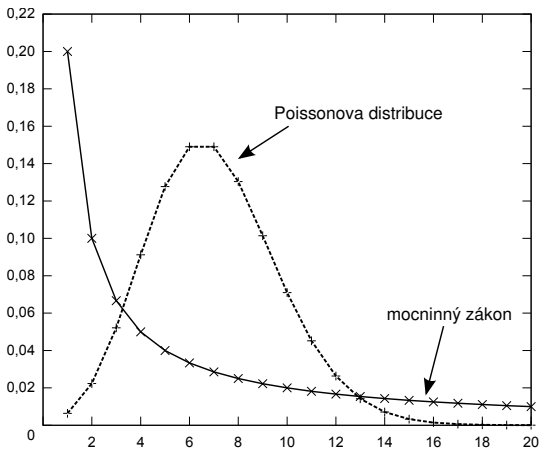
$$P(n) = \lambda^n e^{-\lambda} / n!$$

- Mocninný zákon

$$P(n) \sim n^{-\gamma}$$



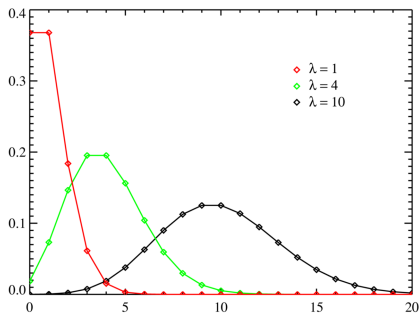
# Poissonova distribuce, mocninný zákon



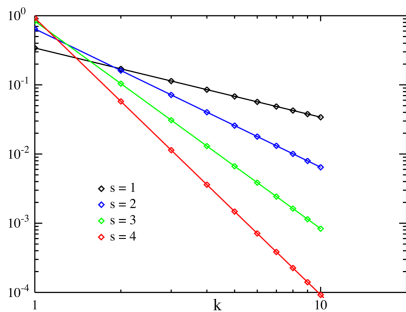


# Poissonova distribuce, mocninový zákon

## Poissonova distribuce



## mocninový zákon



Wikipedia



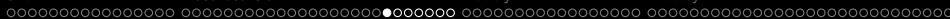






# Vlastnosti sítí – výzkum

	network	type	$n$	$m$	$z$	$\ell$	$\alpha$	$C^{(1)}$	$C^{(2)}$	$r$	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
sexual contacts	undirected	2 810				3.2				265, 266	
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421



# Assortativity

korelace  $r$  mezi stupni vrcholů

	network	$n$	$r$
real-world networks	physics coauthorship <sup>a</sup>	52 909	0.363
	biology coauthorship <sup>a</sup>	1 520 251	0.127
	mathematics coauthorship <sup>b</sup>	253 339	0.120
	film actor collaborations <sup>c</sup>	449 913	0.208
	company directors <sup>d</sup>	7 673	0.276
	Internet <sup>e</sup>	10 697	-0.189
	World-Wide Web <sup>f</sup>	269 504	-0.065
	protein interactions <sup>g</sup>	2 115	-0.156
	neural network <sup>h</sup>	307	-0.163
	food web <sup>i</sup>	92	-0.276
models	random graph <sup>u</sup>		0
	Callaway <i>et al.</i> <sup>v</sup>		$\delta/(1 + 2\delta)$
	Barabási and Albert <sup>w</sup>		0

# Motivy

- podgrafy vyskytující se daleko častěji než v náhodném grafu
- různé motivy pro různé typy sítí
- většina komplexních sítí má nějaké motivy

## Další vlastnosti

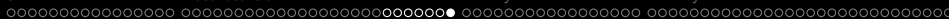
Network	Nodes	Edges	$N_{real}$	$N_{rand} \pm SD$	Z score	$N_{real}$	$N_{rand} \pm SD$	Z score	$N_{real}$	$N_{rand} \pm SD$	Z score
<b>Gene regulation (transcription)</b>				<b>Feed-forward loop</b>			<b>Bi-fan</b>				
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> <sup>9</sup>	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
<b>Neurons</b>				<b>Feed-forward loop</b>			<b>Bi-fan</b>			<b>Bi-parallel</b>	
<i>C. elegans</i> <sup>†</sup>	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
<b>Food webs</b>				<b>Three chain</b>			<b>Bi-parallel</b>				
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			
Couchella	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			
<b>Electronic circuits (forward logic chips)</b>				<b>Feed-forward loop</b>			<b>Bi-fan</b>			<b>Bi-parallel</b>	
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
<b>Electronic circuits (digital fractional multipliers)</b>				<b>Three-node feedback loop</b>			<b>Bi-fan</b>			<b>Four-node feedback loop</b>	
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
s838 <sup>‡</sup>	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
<b>World Wide Web</b>				<b>Feedback with two mutual dyads</b>			<b>Fully connected triad</b>			<b>Uplinked mutual dyad</b>	
nc.edu <sup>§</sup>	325,729	1,466 <sup>6</sup>	1.1e5	2e3 ± 1e2	800	6.8e6	5e44e2	15,000	1.2e6	1e4 ± 2e2	5000











# Analýza sociálních sítí

## *social network analysis*

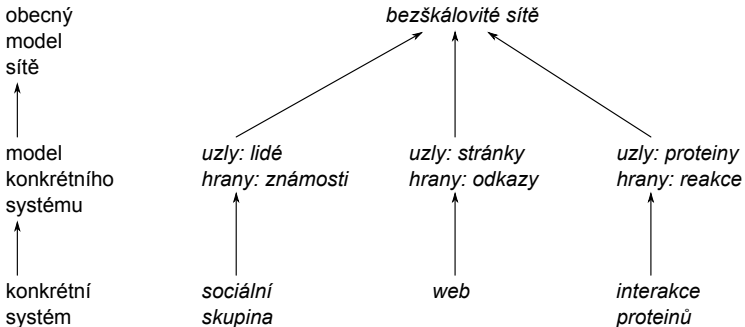
- studováno dlouho – donedávna však data pouze v malém (dotazníky)
- nyní data ve velkém (mobily, e-mail, Facebook, ...)
- analýzy: metriky centrality, detekce shluků
- komerční využití:
  - mobilní operátoři a nabídky zákazníkům
  - doporučující algoritmy



# Modely komplexních sítí

- komplexní sítě mají **typické vlastnosti**
- dokážeme tyto vlastnosti **modelovat** na abstraktní úrovni?

# Úrovně abstrakce



# Základní modely komplexních sítí

- náhodné grafy (Erdős-Renyi model)
- grafy malého světa (small-world graphs, Watts-Strogatz model)
- bezškálovité sítě (scale-free networks, Barabási-Albert model)

modely jednoduché, umožňují simulaci i částečné analytické řešení





# Zajímavé výsledky

- „téměř všechny grafy mají vlastnost  $Q$ “ =  
pravděpodobnost se blíží k 1 (v limitě pro  $n \rightarrow \infty$ )
- pro hodně vlastností **fázový přechod**:
  - grafy s pravděpodobností hran  $p \Rightarrow$  téměř všechny mají danou vlastnost nebo ji nemají (např. souvislost)
  - skokový přechod
- NetLogo: Networks / Giant Component: náhodný graf, velikost největší komponenty

# Vlastnosti

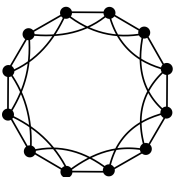
- průměrná délka cesty:  $\sim \log n$
- distribuce stupňů: **Poissonova distribuce**
- shlukování: průměrný shlukovací koeficient  $C \sim p \sim \frac{m}{n}$ , výrazně **méně** než u reálných komplexních sítí

vlastnosti reálných sítí: splňují malý svět, nesplňují ostatní

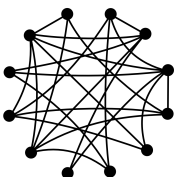
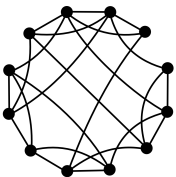
# Grafy malého světa

(Small-world graphs, Watts-Strogatz model)

- 1 **pravidelná inicializace**:  $N$  vrcholů, uspořádáme do kruhu, každý spojíme s  $K$  sousedy ( $K/2$  na každé straně)
- 2 **náhodné předrátování**:  $\forall$  hrany – s pravděpodobností  $p$  nahradíme náhodnou hranou

 $p = 0$ 

rostoucí míra náhody

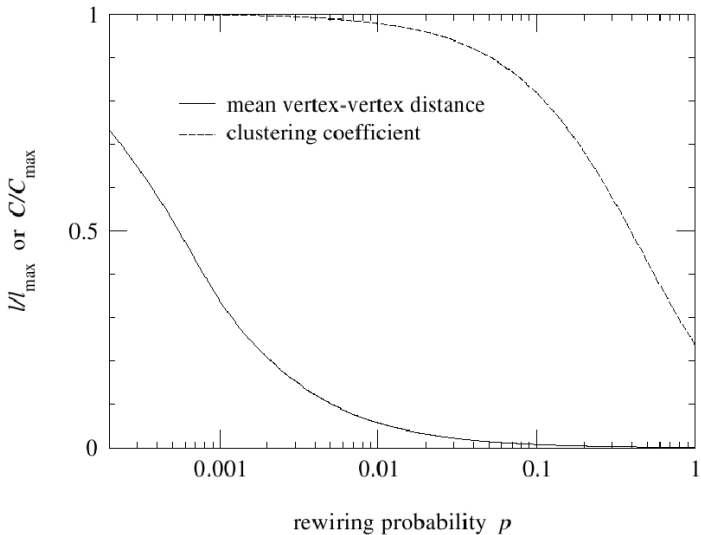
 $p = 1$ 

NetLogo: Networks / Small Worlds

# Vlastnosti

- průměrná délka cesty  $\sim \log N$  (pro vhodné  $p$ , přesná charakteristika složitá)
- shlukování: pro  $p = 0$  máme  $C = \frac{3(K-2)}{4(K-1)}$ , pro větší  $p$  trochu menší, ale stále **dosti velké** (reálné)
- distribuce stupňů  $\sim$  Poissonova distribuce (přesná charakterizace složitá), **jiná** než u reálných komplexních sítí

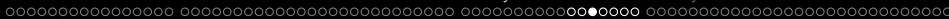
# Mezi řádem a náhodou



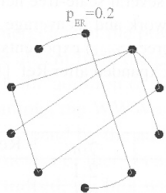




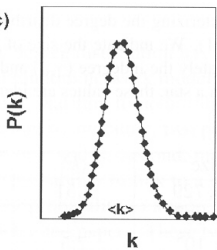




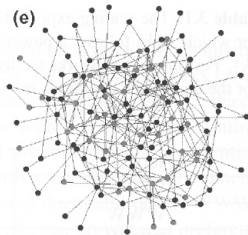
(a)



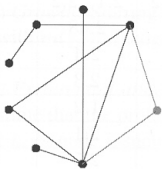
(c)



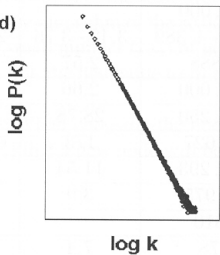
(e)



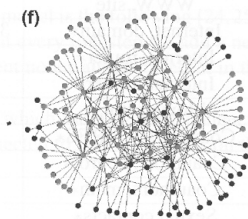
(b)



(d)



(f)

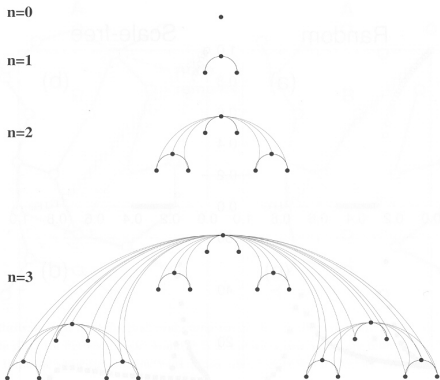




# Další modely rostoucích sítí

- nelineární  $\Pi(k)$  (upřednostněné připojení)
- počáteční atraktivnost
- stárnutí, způsobilost (fitness)
- rušení, přesměrování hran

# Deterministický model bezškálovité sítě I



**Figure 3.6:** Construction of the deterministic scale-free network, showing the first four steps of the iterative process. After [42].



# K čemu to je?

- komplexní sítě mají zajímavé vlastnosti
- máme modely, které tyto vlastnosti (do určité míry) reprodukují

# K čemu to je?

- komplexní sítě mají zajímavé vlastnosti
- máme modely, které tyto vlastnosti (do určité míry) reprodukují
- no a co?
- **můžeme to nějak využít?**

# Procesy na sítích

- chyby, útoky, stabilita
- šíření epidemií, informací
- hledání v sítích
- spolupráce

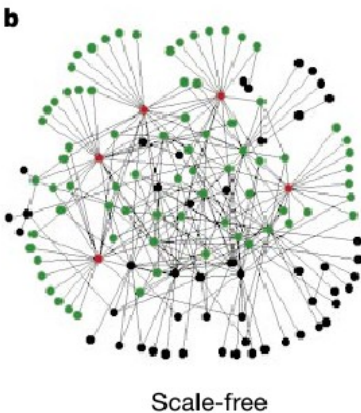
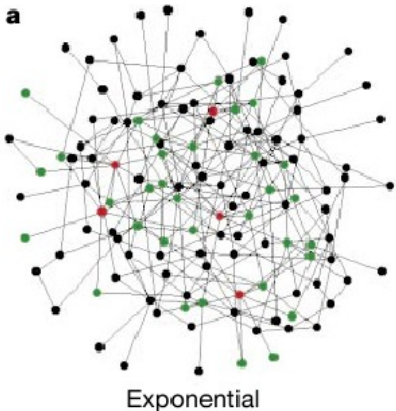
analýzu procesů provádíme zejména pomocí **simulace**

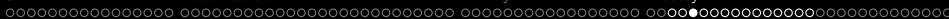




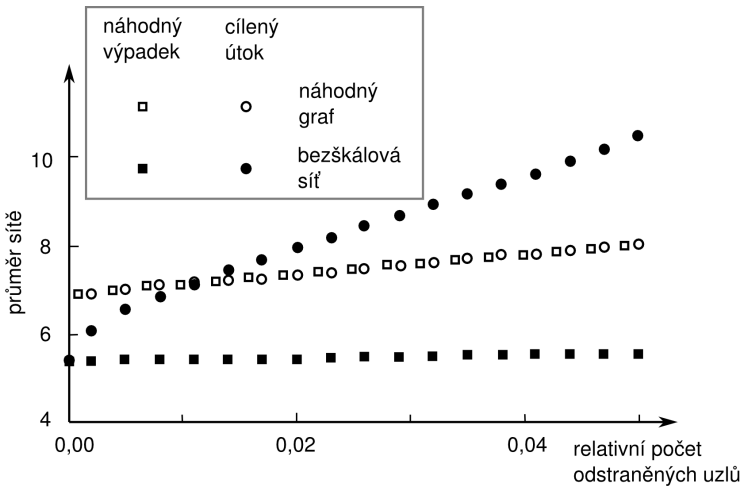


# Role klíčových uzlů





# Odolnost proti chybám a útokům



# Robustnost a topologie

Bezškálovité sítě (oproti náhodným):

- **vyšší odolnost proti chybám**
- **náchylnější proti útokům** – „Achilova pata komplexních sítí“



# Příklady

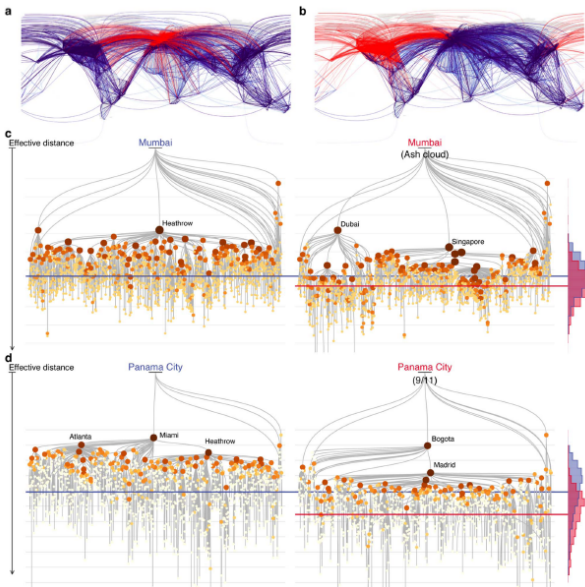
Bezškálovité sítě s uvedenými vlastnostmi (odolnost proti chybám, náchylnost k útokům):

- komunikační sítě (Internet, www)
- ekonomické sítě (a např. teroristický útok na NY)
- proteiny – funkčnost vysoce propojených proteinů je životně důležitá



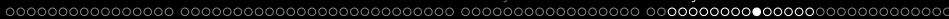


## Robustnost sítí

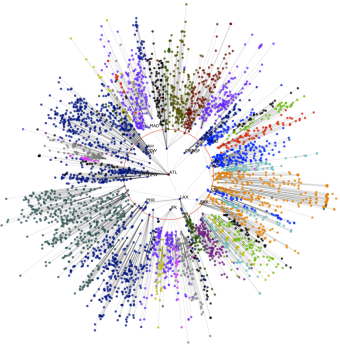
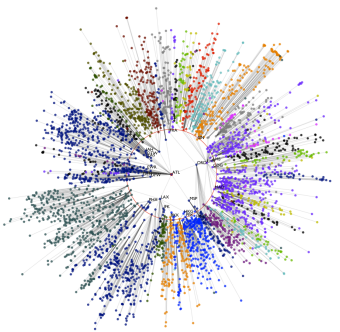








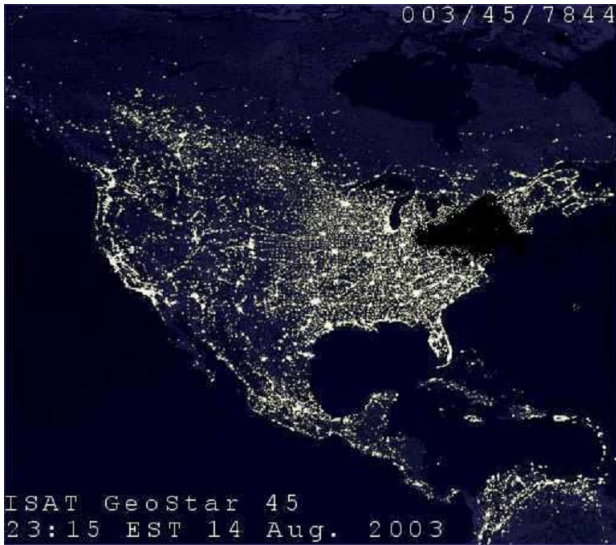
Robustnost sítí





# Dynamické efekty

- změna v síti může vyvolávat další změny
- např. elektrická energie v elektrické síti, energie v potravních řetězcích
- tok ovlivňuje funkčnost sítě
- jaká je robustnost sítě?
- jak souvisí s topologií sítě?





# Dynamické efekty

- narušení sítě může vést k neočekávaným důsledkům (vlivem zpětných vazeb)
- elektřina: výpadek celé sítě vlivem kumulace zátěže
- potravní řetězce: vyhubení predátora vedoucí k poklesu kořisti

# Robustnost: příklady studovaných otázek

- Jaký je vliv stability/robustnosti a složitosti sítě?
- Jaký vliv na stabilitu ekosystému má složitost potravního řetězce (a potažmo biodiverzita)?
- Proč jsou potravní řetězce krátké?
- různé výsledky pro sítě modelované náhodně a sítě modelované realističtěji

# Šíření epidemií

- šíření počítačových virů po Internetu
- HIV po sexuální síti
- šíření náboženství, informací v sociální síti
- šíření nových technologií (na základě sociální sítě)





# Dopad použití sítí na simulace

epidemii můžeme simulovat za využití:

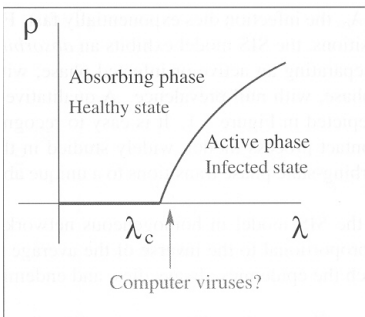
- homogenního prostředí / náhodném pohybu  
NetLogo / epiDEM / epiDEM Basic
- modelů sítí  
NetLogo / Networks / Virus on Network

Je to důležitý rozdíl? V čem se mohou lišit výsledky?

# Epidemie v homogenním prostředí

V „homogenním“ prostředí (de facto náhodný graf):

- **kritická hranice**
  - infekčnost menší  $\Rightarrow$  epidemie se nešíří
  - infekčnost větší  $\Rightarrow$  epidemie se výrazně šíří
- **uniformní imunizace**

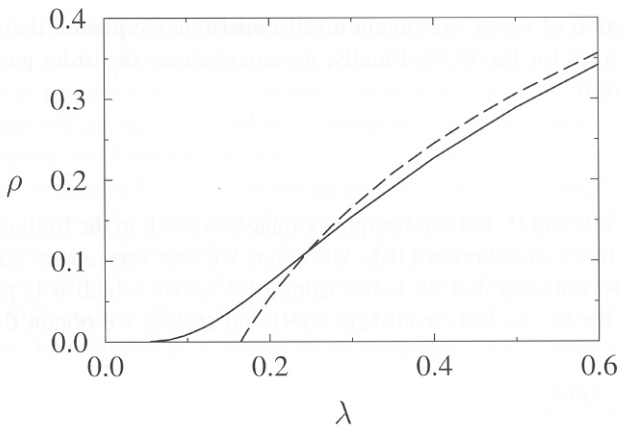


# Epidemie v bezškálovitých sítích

- **neexistuje kritická hranice**
- nemoc s velmi malou infekčností se může rozšířit – díky uzlům s vysokým stupněm
- uniformní imunizace je poměrně neúčinná
- **cílená imunizace** zasahující hlavně uzly s vysokým stupněm však může být velmi účinná

Praktické poučení: např. pro boj s AIDS

# Epidemie v bezškálovitých sítích



# Hledání v grafu

- vlastnost „malého světa“ – mezi většinou uzlů **existují** krátké cesty
- Jak tyto cesty najít?
- Milnerův experiment:
  - nejen, že existuje krátký řetězec známostí
  - účastníci experimentu jej byli schopni najít **bez znalosti celého grafu**, tj. jen za použití **lokálních informací** (srovnej ABM)

# Kleinbergův model

- graf je založen na dvou rozměrné mřížce, uzly spojeny se sousedy
- náhodně přidány dlouhé vazby
- informace o poloze v této mřížce je využívána pro navigaci směrem k cíli

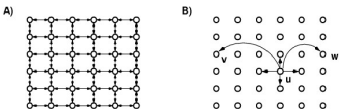
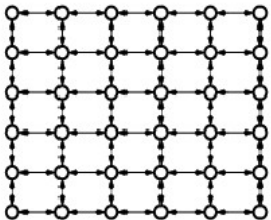


Figure 1: (A) A two-dimensional grid network with  $n = 6$ ,  $p = 1$ , and  $q = 0$ . (B) The contacts of a node  $u$  with  $p = 1$  and  $q = 2$ .  $v$  and  $w$  are the two long-range contacts.

A)



B)

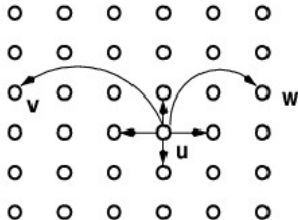


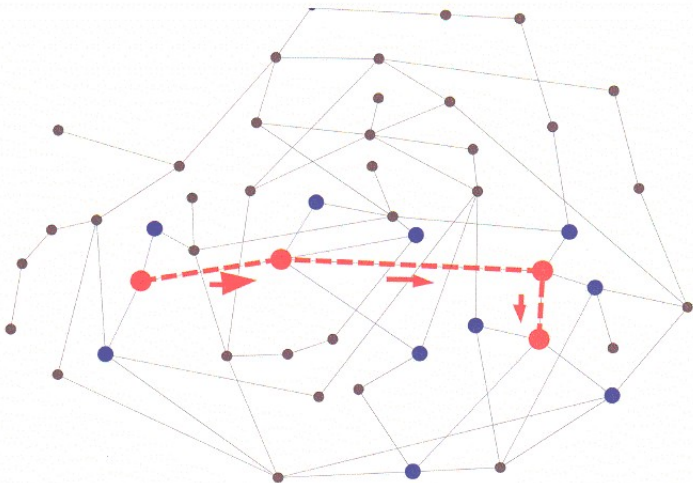
Figure 1: (A) A two-dimensional grid network with  $n = 6$ ,  $p = 1$ , and  $q = 0$ . (B) The contacts of a node  $u$  with  $p = 1$  and  $q = 2$ .  $v$  and  $w$  are the two long-range contacts.

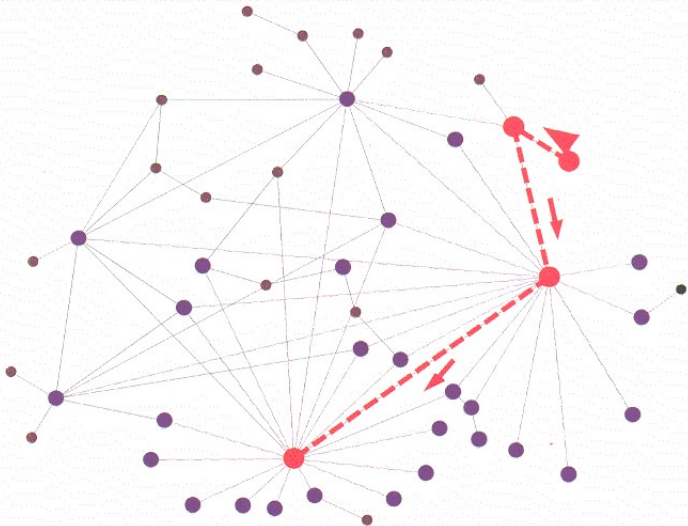


# Hledání s využitím náhodné procházky

- velký a neznámý graf  $\Rightarrow$  náhodná procházka
- **náhodná procházka** (random walk) = vždy vybírá další uzel pro navštívení čistě náhodně
- bezškálovité grafy: cíleně preferujeme uzly s vyšším stupněm
- výsledky simulace: lepší pokrytí než čistá náhodná procházka
- aplikace: Gnutella (peer-to-peer filesharing system)







# Dynamika spolupráce v síti

- **dilema vězně** - dříve studováno v homogenním prostředí (každý s každým nebo pravidelná mřížka)
- jaký je **vliv topologie** sítě na vývoj spolupráce?
- pozorování:
  - **shlukování podporuje rozvoj spolupráce**
  - na grafech malého světa se spolupráci daří lépe než na náhodných grafech

# Další procesy na sítích

- buněčné automaty na sítích
- synchronizace
- formování názorů
- iterované hry

# NetLogo poznámky

- Models Library / Networks
- entita link, ask link-neighbors
- vykreslení grafu: layout-circle, layout-spring
- cyklení: repeat, loop, while
- Preferential attachment model – elegantní realizace náhodného výběru podle distribuce stupňů:  
report [one-of both-ends] of one-of links

# Procesy na sítích

- Diffusion on a Directed Network – toky na sítích (např. voda, energie, peníze), základ modelu
- Virus on a Network – variace na SIR model epidemie, „geografická“ topologie sítě
- Team Assembly – formace týmů, vznik sítě spolupráce

# Shrnutí

- **příklady** komplexních sítí
- **společné vlastnosti**: krátké cesty, shlukování, bezškálovitost
- **abstraktní modely**: náhodné grafy, malý svět, bezškálovité sítě
- **procesy na sítích**: útoky, výpadky, epidemie, šíření informací, spolupráce, hledání
- chování modelů studováno pomocí simulace