

IV124 Komplexní sítě

Jan Fousek, Eva Hladká

Fakulta informatiky, Masarykova univerzita

19. dubna 2018

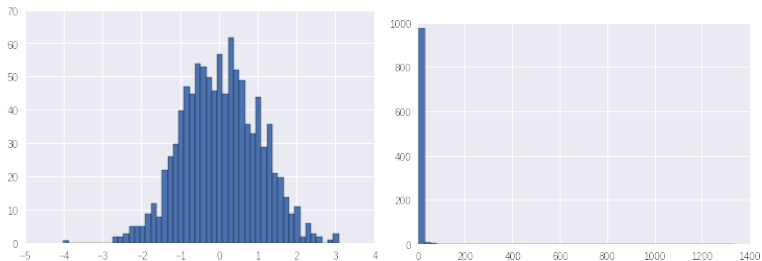
Huby a bezškálové sítě

Minule jsme se věnovali hubům:

- huby: uzly s překvapivě velkým stupněm
- dnes se systematicky podíváme na to, co to znamená „překvapivý“

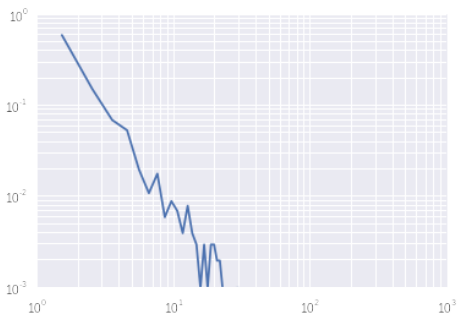
Huby a distribuce stupně uzlů

náhodná síť vs. síť s huby

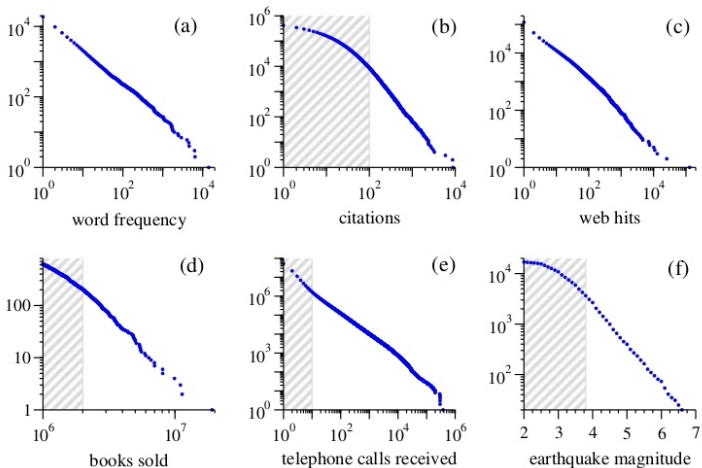


Log-log zobrazení distribuce

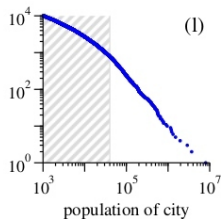
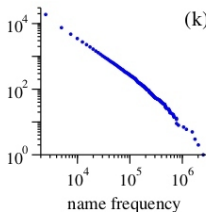
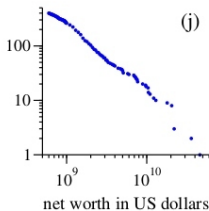
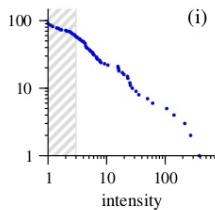
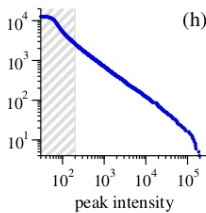
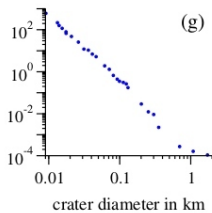
Na běžném histogramu toho moc nevidíme, log-log zobrazení je mnohem zajímavější



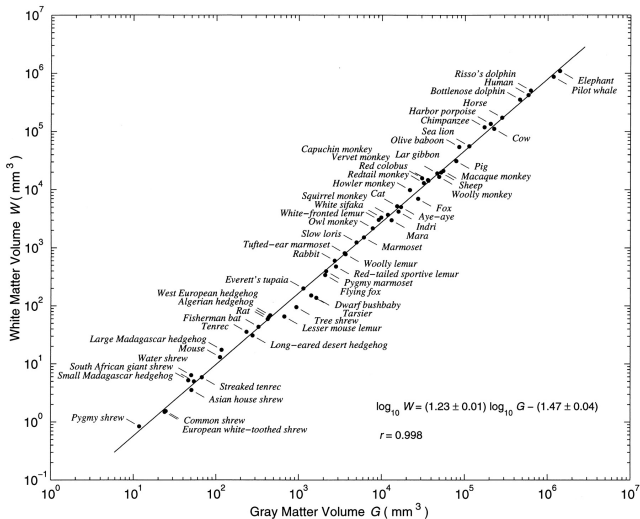
Povědomá distribuce¹



Povědomá distribuce



Povědomá distribuce



Od zobrazení ke vzorci

Na log-log osách máme přímku:

$$\log(p(x)) = c - \gamma \log(x)$$

Po umocnění:

$$p(x) = cx^{-\gamma}$$

Tzv. mocninný zákon (mocninné pravidlo, *power-law*).

Mocninné pravidlo: variance

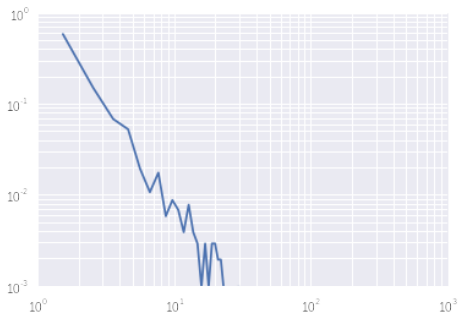
Druhý moment (variance) obecně není omezený \implies čím větší máme vzorek (sít'), tím větší bude maximální hodnota.

	\bar{k}	σ
E. Coli metab.	5.58	20.79
WWW	4.60	30.27
kvasinky prot.	2.9	4.88

Log-log zobrazení²

Problém:

- s rostoucím stupněm klesá počet vzorků, roste šum

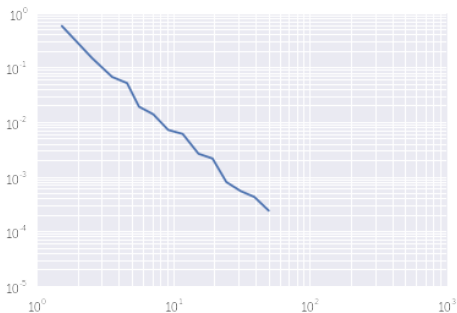


²Poznámka: vždy používat logaritmus o základu 10.

Log-log zobrazení

Řešení:

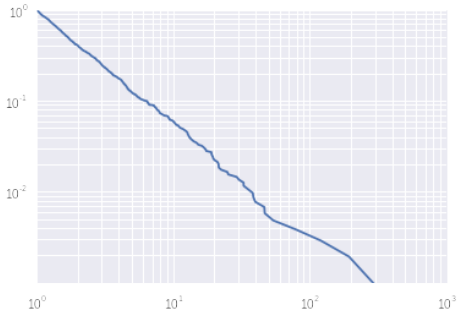
- logaritmické třídy



Log-log zobrazení

Řešení

- komplementární kumulativní distribuce
- $P_{\geq}(x) = x^{1-\gamma}$



Vlastnosti bezškálových sítí

Velikost sítě vs velikost hubů

- lze odvodit³, že $k_{max} \approx k_{min} N^{\frac{1}{\gamma-1}}$
- tedy k_{max} je polynomiálně závislá na N

Velké množství systémů se zdá být v režimu $2 < \gamma < 3$. Má to nějaký důvod?

Třídy bezškálových sítí

Anomální režim $\gamma \leq 2$

- stupeň největšího hubu roste rychleji než velikost sítě
- protože exponent $\frac{1}{\gamma-1}$ je větší než jedna
- jinými slovy, taková síť není bez smyček asymptoticky možná

Třídy bezškálových sítí

Bezškálový režim $2 < \gamma < 3$

- první moment distribuce (průměr) je konečný, ostatní momenty divergují (rozptyl, šikmost, ...)
- nejzajímavější vlastnosti:
 - specifické chování dynamických procesů (difuze)
 - robustní proti náhodnému výpadku, zranitelná na cílený útok (narozdíl od náhodné sítě)
 - ...

Třídy bezškálových sítí

Režim náhodné sítě $\gamma > 3$

- pravděpodobnost velkých hubů klesá příliš rychle
- těžce rozeznatelná od náhodné sítě
- konkrétně z předchozí rovnice: $N \gg \frac{k_{max}}{k_{min}} \gamma^{-1}$

Tedy pokud požadujeme rozpětí alespoň 3 řádů, tak pro $\gamma = 5$, $k_{min} = 1$ a $k_{max} = 10^2$ potřebujeme síť o velikosti $N \gg 10^8$.

Empirické hledání exponentu

Motivace:

- kromě předchozí klasifikace důležité pro nulové modely např. pro studium dynamických procesů

Postup:

- vycházíme z log-log transformace, fitujeme přímkou
- použijeme log. intervaly, případně kumulativní distribuci pro odstranění šumu
- aplikujeme metodu nejmenších čtverců, max. věrohodnost a pod. v oblíbeném statistickém software

Praktická ukázka

...