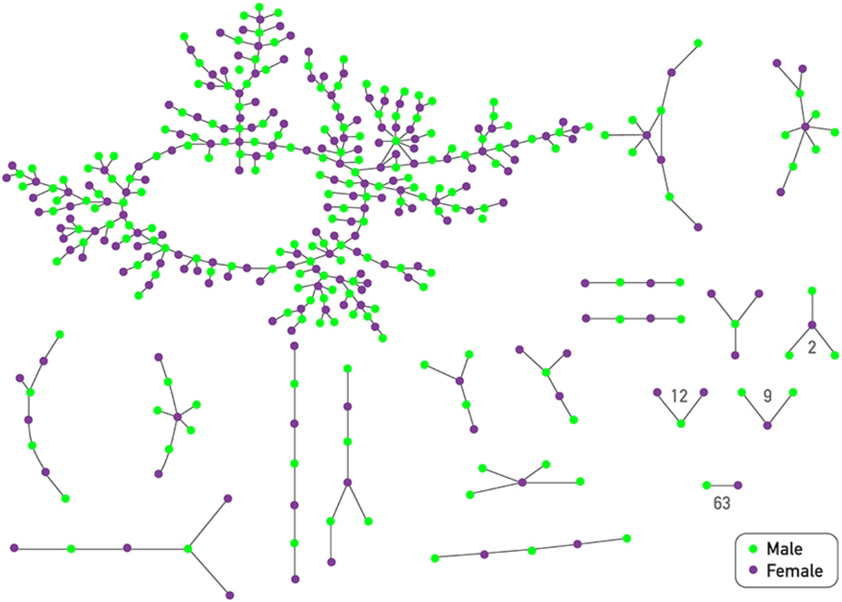
Teorie epidemií

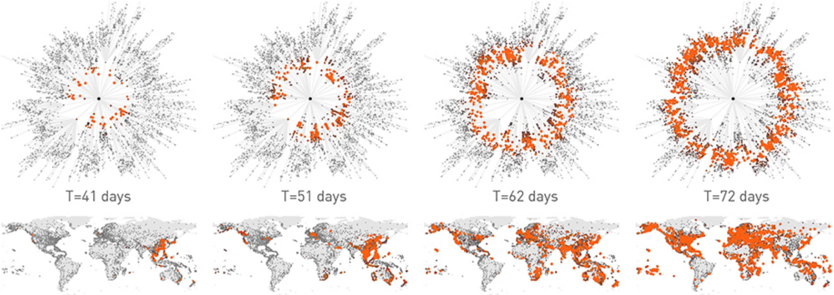
Modelování a teorie sítí

Petr Liška

http://networksciencebook.com/

26.04.2022


2
12
9
63
Male
Female


T=41 days
T=51 days
T=62 days
T=72 days

Co je to síť?

Síť má dva základní parametry:

- *Počet uzlů*, nebo-li N, reprezentující počet bodů v systému. K rozeznání jednotlivých uzlů si je značíme pomocí $i = 1, 2, ..., N$. Tento parametr často nazýváme *velikost sítě*.
- *Počet vazeb*, nebo-li L, reprezentující celkový počet interakcí mezi uzly. Vazby zpravidla nebývají označené.

Síť se nazývá *orientovaná*, jestliže všechny její vazby jsou orientované, tzn. že mají mezi uzly pevně daný směr. V opačném případě mluvíme o síti *neorientované*, čili nerozlišujeme směr jejích vazeb.

Graf je uspořádána dvojice $G = (V, E)$, kde

- V je množina vrcholů,
- $E \subseteq \{[x, y] \mid x, y \in V, x \neq y\}$.

Reálné sítě

Síť	*Uzly*	*Vazby*	*Ori*	N	L
Herecká	herci	společný film	N	702 388	29 397 908
Vědecká	vědci	společný článek	N	23 133	93 437
WWW (část)	stránky	URL adresy	A	325 729	1 497 134
Citační	články	citace	A	449 673	4 689 479

Základní charakteristiky

- *stupeň* k_i – vyjadřuje počet vazeb, které daný uzel i má k ostatním uzlům; v případě orientované sítě rozlišujeme *vstupní* k_i^{in} a *výstupní stupeň* k_i^{out}.
- *celkový počet vazeb*

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i,$$

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$

- *průměrný stupeň*

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

$$\langle k^{in} \rangle = \langle k^{out} \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{in} = \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \frac{L}{N}$$

Základní charakteristiky

- *rozložení stupňů* p_k – označuje pravděpodobnost, že náhodně vybraný uzel v síti má stupeň k, tedy platí

$$\sum_{k=0}^{\infty} p_k = 1.$$

Označíme-li N_k počet uzlů mající stupeň k v sítí o N uzlech, obdržíme:

$$p_k = \frac{N_k}{N}.$$

Potom máme jiný vztah pro p_k

$$\langle k \rangle = \sum_{k=0}^{\infty} \frac{kN_k}{N} = \sum_{k=0}^{\infty} kp_k \, .$$

Náhodná síť

Náhodná síť - definice podle Gilberta

Každý pár z N uzlů je spojen s pravděpodobností p.

Náhodnou síť sestrojíme pomocí následujících kroků:

- začneme s N izolovanými uzly
- vybereme náhodnou dvojici uzlů a vygenerujeme náhodné číslo mezi 0 a 1, pak v případě, že toto číslo překročí hodnotu p, spojíme dané uzly vazbou, v opačném případě je necháme nespojené
- předchozí krok opakujeme pro každou z $\frac{N(N-1)}{2}$ dvojic uzlů

Pravděpodobnost, že má náhodná síť přesně L vazeb, je součinem tří výrazů:

- p^L, což vyjadřuje pravděpodobnost, že L pokusů o připojení $N\,(N-1)/2$ dvojic uzlů skončí úspěšným připojením
- $(1-p)^{\frac{N\,(N-1)}{2}-L}$, což je pravděpodobnost, že zbývajících $\frac{N\,(N-1)}{2}-L$ pokusů neskončí úspěšným připojením
- $\binom{\frac{N\,(N-1)}{2}}{L}$, což je kombinační číslo, které nám vyjadřuje, kolika způsoby můžeme L vazeb rozmístit mezi $N\,(N-1)/2$ dvojic uzlů

$$p_L = \binom{\frac{N(N-1)}{2}}{L} p^L (1-p)^{\frac{N(N-1)}{2}-L}$$

Očekávaný počet hran

$$\langle L \rangle = \sum_{L=0}^{\frac{N(N-1)}{2}} L p_L = p \frac{N(N-1)}{2}$$

je tedy součin pravděpodobnosti p (vyjadřující, že dva uzly jsou spojeny) a $L_{max} = \frac{N(N-1)}{2}$ (vyjadřující počet párů, které chceme spojit).

Odsud průměrný stupeň uzlu je

$$\langle k \rangle = \frac{2\langle L \rangle}{N} = p(N-1)$$

Distribuce uzlů

Pravděpodobnost, že uzel i v náhodné síti má přesně k vazeb je součinem tří výrazů:

p^k, což je pravděpodobnost, že k vazeb daného uzlu existuje

$(1-p)^{N-1-k}$, což je pravděpodobnost, že $(N-1-k)$ zbývajících vazeb chybí

$\binom{N-1}{k}$ nám říká, kolika způsoby můžeme z potenciálních $N-1$ vazeb (které uzel může mít) vybrat k z nich

Distribuce uzlů tedy je

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \approx \mathrm{e}^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

Svět je malý

$$N(d) \approx 1 + \langle k \rangle + \langle k \rangle^2 + \cdots + \langle k \rangle^d = \frac{\langle k \rangle^{d+1} - 1}{\langle k \rangle - 1} \approx \langle k \rangle^d$$

$$N(d_{max}) \approx N \approx \langle k \rangle^{d_{max}}$$

$$d_{max} \approx \frac{\ln N}{\ln \langle k \rangle}$$

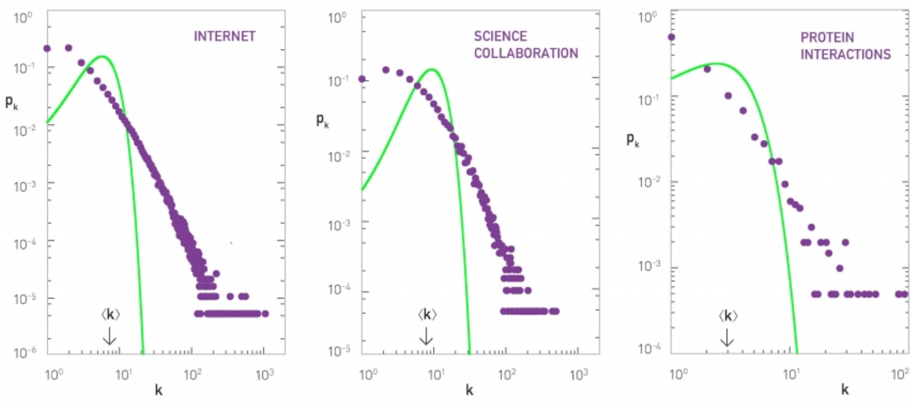
$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$$

$$\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle} = \frac{\ln(8 \cdot 10^9)}{\ln 10^3} = 3{,}3$$

Reálná data

Síť	$\langle k \rangle$	$\langle d \rangle$	d_{max}	$\ln N / \ln \langle k \rangle$
Herecká	83,71	3,91	14	3,04
Vědecká	8,08	5,35	15	4,81
WWW (část)	4,60	11,27	93	8,31
Citační	10,43	11,21	42	5,55

Kde je problém?



Bezškálová síť

Definice

Bezškálová síť je taková síť, jejíž rozložení stupňů se řídí mocninným zákonem, tj.

$$p_k \sim k^{-\gamma}\,.$$

$$p(k) = (\gamma - 1) k_{\mathrm{min}}^{\gamma-1} k^{-\gamma}$$

$$k_{max} = k_{min}\, N^{\frac{1}{\gamma-1}}$$

Momenty

$$\langle k^n \rangle = \sum_{k_{min}}^{\infty} k^n p_k \approx \int_{k_{min}}^{\infty} k^n p(k)\, \mathrm{d}k$$

Nižší momenty mají důležitou interpretaci:

- $n = 1$: První moment odpovídá průměrnému stupni, čili $\langle k \rangle$.
- $n = 2$: Druhý moment , $\langle k^2 \rangle$, je důležitý při výpočtu *rozptylu* $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$. Jeho druhá odmocnina, σ, se nazývá *směrodatná odchylka*.
- $n = 3$: Třetí moment, $\langle k^3 \rangle$, určuje *šikmost* rozložení a říká nám, jak symetrické je p_k kolem průměru $\langle k \rangle$.

Význam bezškálovosti

$$\langle k^n \rangle = C \, \frac{k_{max}^{n-\gamma+1} - k_{min}^{n-\gamma+1}}{n - \gamma + 1}$$

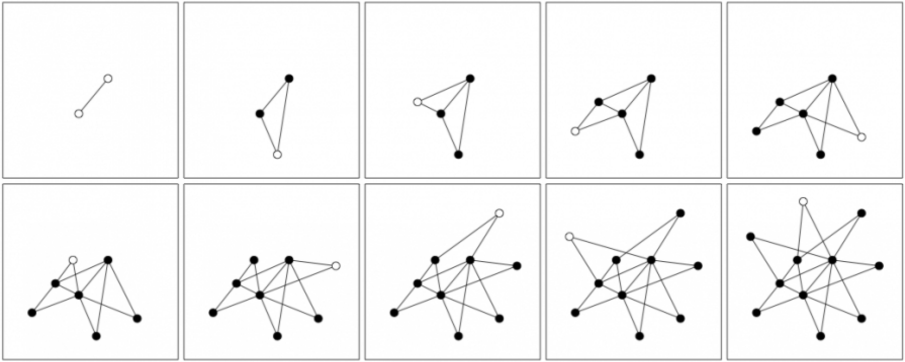
Pro $k_{max} \to \infty$ dostaneme

- Pokud $n - \gamma + 1 \leq 0$, potom s rostoucím k_{max} jde člen $k_{max}^{n-\gamma+1}$ do nuly. Proto jsou všechny momenty, které splňují $n \leq \gamma - 1$, konečné.
- Pokud $n - \gamma + 1 > 0$, potom s $k_{max} \to \infty$ jde i $\langle k^n \rangle$ do nekonečna. Proto všechny momenty větší než $\gamma - 1$ divergují.

Bezškálová síť, Barabási-Albert model

Růst - v každém kroku přidáme uzel s m novými spojeními
Preferential attachment - pravděpodobnost $\prod(k)$, že spojení nového uzlu bude navázáno na starý uzel i je dána

$$\prod(k_i) = \frac{k_i}{\sum_j k_j}$$



Jaký je stupeň uzlu?

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = m \prod(k_i) = m \frac{k_i}{\sum\limits_{j=1}^{N-1} k_j} = \frac{mk_i}{2mt - m} = \frac{k_i}{2t - 1} \approx \frac{k_i}{2t}$$

$$\frac{\mathrm{d}k_i}{k_i} = \frac{1}{2}\frac{\mathrm{d}t}{t}, \qquad k_i(t_i) = m$$

$$k_i(t) = m \left(\frac{t}{t_i}\right)^{\frac{1}{2}}$$

Jaká je distribuce stupňů uzlů?

Kolik uzlů má stupeň menší než k?

$$m\left(\frac{t}{t_i}\right)^{\frac{1}{2}} < k \Longrightarrow t_i < t\left(\frac{m}{k}\right)^{\frac{1}{2}}$$

Dohromady máme $N = m_0 + t \approx t$ uzlů. Pravděpodobnost, že vybereme uzel stupně menšího než k

$$P(k) = 1 - \left(\frac{m}{k}\right)^{\frac{1}{2}}$$

Distribuce pak je

$$p_k = \frac{\partial P(k)}{\partial k} = \frac{2m^2}{k^3}$$

Friendship paradox

$$k_n(k_i) = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Pro náhodné sítě platí

$$k_n(k_i) = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle (1 + \langle k \rangle)}{\langle k \rangle} = 1 + \langle k \rangle$$

Pro bezškálovou síť platí

$$\langle k^2 \rangle \to \infty \quad \text{pro} \quad N \to \infty$$

Síť	N	L	$\langle k \rangle$	$\langle k^2 \rangle$
Internet	192 244	609 066	6,34	240,1
Vědci	23 133	93 439	8,08	178,2
Herci	702 388	29 397 908	83,71	47 353