

K velkým datům přes matice a grafy

Miroslav Tůma

Katedra numerické matematiky, MFF UK

`mirektuma@karlin.mff.cuni.cz`

MFF UK, 10.4.2019

- 1 Motivace
- 2 Šíření infekční choroby
- 3 Jiné motivace
- 4 K algoritmickému řešení problému vlastních čísel
- 5 Rozděl a spojuj
- 6 Přemítání o problému
- 7 Dělení grafů na části
- 8 Rozděl a spojuj
- 9 Závěr

- Značná část viditelného makrosvěta se dá popsat
 - Spojitými zákony pro fyzikální veličiny

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

- ★ Toto je například rovnice šíření elektromagnetických vln v jedné dimenzi
- ★ Aby se takový vztah dal zpracovávat v počítači, musí se převést na diskrétní – **diskretizovat**
- ★ Tedy “převést” na matice, vektory, ...

$$A, b, v, \dots$$

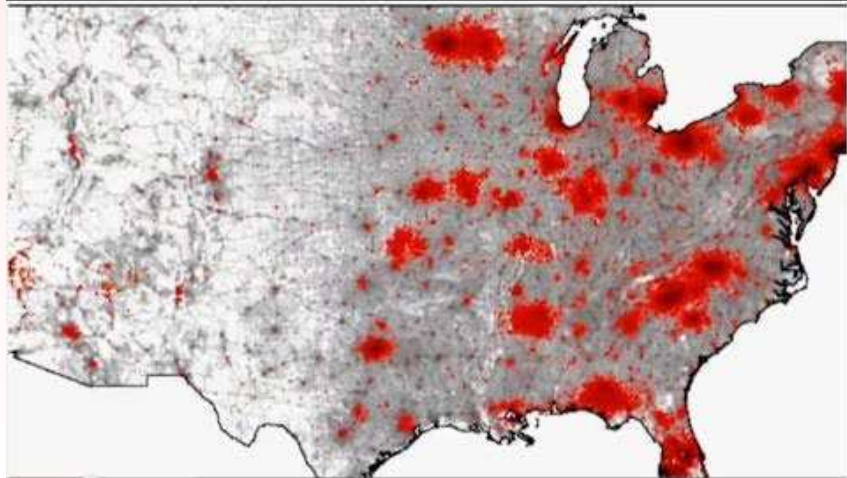
- Nebo přímo diskrétními strukturami
- Jenže co pak s těmi diskrétními strukturami dál ...

- Tato přednáška ukáže problémy, které s těmi diskrétními strukturami máme a kde hledat matematickou pomoc
 - Příklad šíření infekční choroby
 - násobení maticí → problém vlastních čísel
 - Použitelné matematické techniky
 - v pozadí řešení soustav lineárních rovnic – eliminace

Outline

- 1 Motivace
- 2 Šíření infekční choroby**
- 3 Jiné motivace
- 4 K algoritmickému řešení problému vlastních čísel
- 5 Rozděl a spojuj
- 6 Přemítání o problému
- 7 Dělení grafů na části
- 8 Rozděl a spojuj
- 9 Závěr

Šíření infekční choroby ve dnech $1, \dots, n$

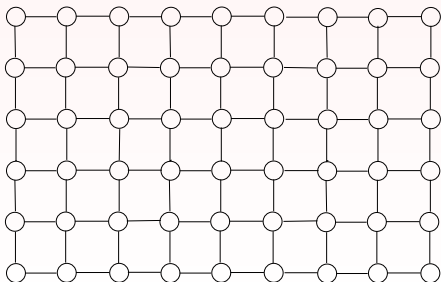


Šíření infekční choroby ve dnech $1, \dots, n$

- Den t , soubor n osob.
- $I(t)$ značí procento **infikovaných** osob
- $S(t)$ procento osob, které **nebyly nikdy infikovány**
- $R(t)$ procento **uzdravených** osob: $R(t) = 100 - I(t) - S(t)$ (po k dnech)
- **Ještě potřebujeme vědět parametry:**
 - Jaké procento τ lidí se v průměru od souseda nakazí, když se nemoc šíří například mezi lůžky nemocnice, to jest při daném rozložení.
 - Složitější: Jaká část ν lidí je třeba mobilní.
 - Model můžeme činit složitější až se dostane skoro k realitě
 - To je tak skoro vždycky a matematika nám musí říci, co to je **skoro k realitě**.

Šíření infekční choroby ve dnech $1, \dots, n$

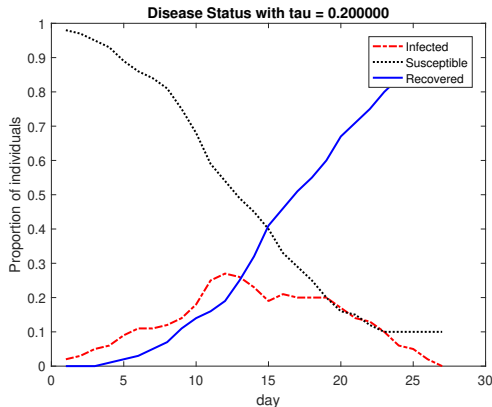
- Jak na to?
- Například simulací.
- Simulace má prvek náhodnosti:
 - ▶ Nějaké procento sousedů onemocní, ale není jasné kteří to budou.
 - ▶ Simulace se musí nechat probíhat víckrát a pak výsledky zprůměrovat. Říkáme tomu, že model je stochastický.
 - ▶ Nejjednodušší situace: předpoklad o rozmístění



Šíření infekční choroby ve dnech $1, \dots, n$

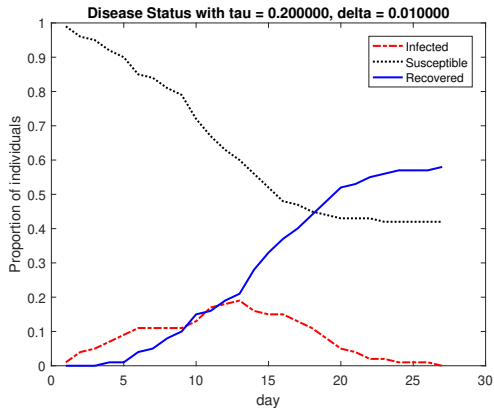
- Nejjednodušší simulace

- ▶ 10×10 postelí v nemocnici, $k = 2$, $\tau = 0.2$ (nakažlivost)
- ▶ Ten prostřední to chytne nejdřív a šíří a šíří ...



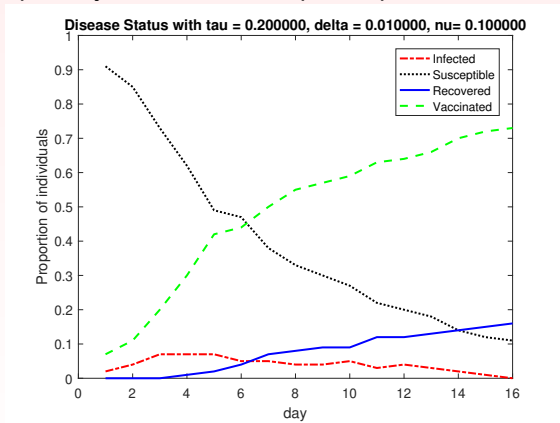
Šíření infekční choroby ve dnech $1, \dots, n$

- Simulace s mobilitou $\delta = 0.01$



Šíření infekční choroby ve dnech $1, \dots, n$

- Simulace s přidáním očkováním s pravděpodobností 10%



- Co je za simulací?
 - ▶ Model v simulaci má takzvanou **Markovskou** vlastnost:
 - ▶ Označme si **stav** každého vrcholu (postele) $1, \dots, p$ pomocí p -tice

$$(d_1, d_2, \dots, d_p)$$

- ★ očkovaný, $d_i = -2$
- ★ vyléčený, $d_i = -1$
- ★ ohrožený (ještě nemoc neprodělal), $d_i = 0$
- ★ infikovaný j -tý den, $d_i = j, j \leq k$
- ★ hrany mezi **sousedy**

Šíření infekční choroby ve dnech $1, \dots, n$

- Model v pozadí je matice pravděpodobností přechodů mezi jednotlivými stavy.
- $A = (a_{ij})$, a_{ij} je pravděpodobnost přechodu ze stavu i do stavu j , součty ve sloupcích jsou 1.
- Začneme-li v prvním dnu, pak Ae_1 jsou pravděpodobnosti pro druhý den, A^2e_1 jsou pravděpodobnosti pro třetí den atd.
- Jak najdeme cílový stav, třeba i pro komplikovanější model?
- Chceme, aby platilo pro nějaké rozložení

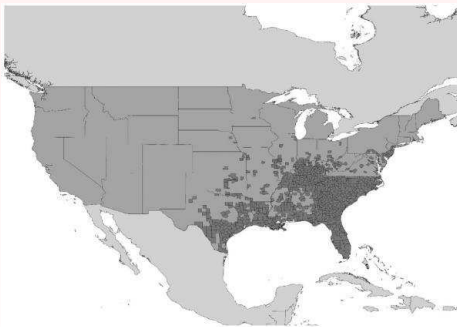
$$Az = z.$$

- To je ale řešení **problému vlastních čísel**

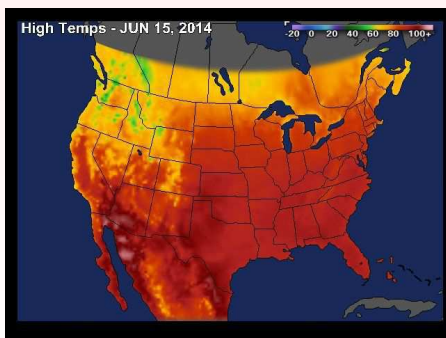
$$Ax = \lambda x, \lambda = 1.$$

- 1 Motivace
- 2 Šíření infekční choroby
- 3 Jiné motivace**
- 4 K algoritmickému řešení problému vlastních čísel
- 5 Rozděl a spojuj
- 6 Přemítání o problému
- 7 Dělení grafů na části
- 8 Rozděl a spojuj
- 9 Závěr

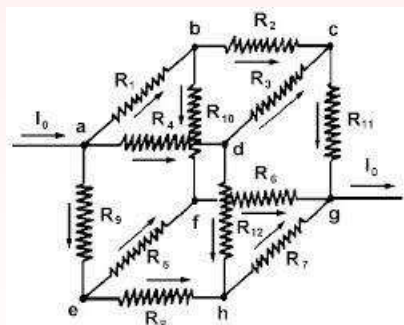
- Za šířením byla nějaká struktura, ale ta může být složitější
- Rozšíření komára



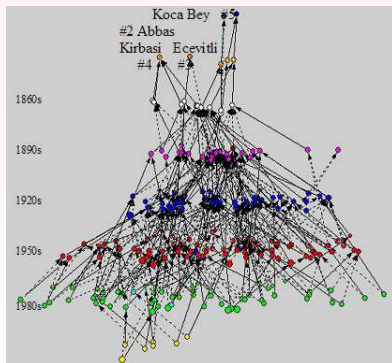
- Za šířením byla nějaká struktura, ale ta může být složitější
- Teplotní extrém



- Za šířením byla nějaká struktura, ale ta může být složitější
- Elektrická síť



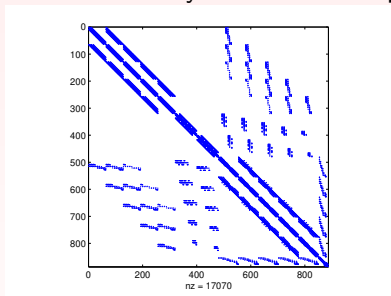
- Za šířením byla nějaká struktura, ale ta může být složitější
- Genealogická informace



- 1 Motivace
- 2 Šíření infekční choroby
- 3 Jiné motivace
- 4 K algoritmickému řešení problému vlastních čísel**
- 5 Rozděl a spojuj
- 6 Přemítání o problému
- 7 Dělení grafů na části
- 8 Rozděl a spojuj
- 9 Závěr

K řešení problému

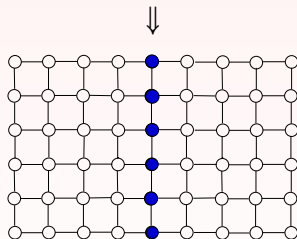
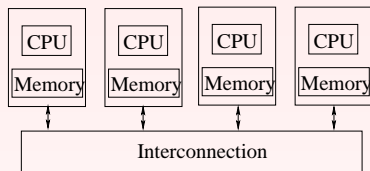
- Jak vypadají naše problémy a jak se s nimi vyrovnáme?
 - Především, struktura matice může být obrovská a nepřehledná



- Spousta nul: říkáme, že matice je **řídka**
- Dimenze klidně až milióny, miliardy
- Takovou strukturu můžeme reprezentovat **grafem**.
- Co s takovým grafem budeme dělat?

- 1. Potřebujeme matici rozložit (eliminovat nenuly)
→ uspořádání matice, aby nám vzniklo co nejméně nenulových prvků rozkladu
- 2. Potřebujeme se vyrovnat s moderními počítačovými architekturami
→ rozdělení grafu (a práce) mezi více procesorů/jader

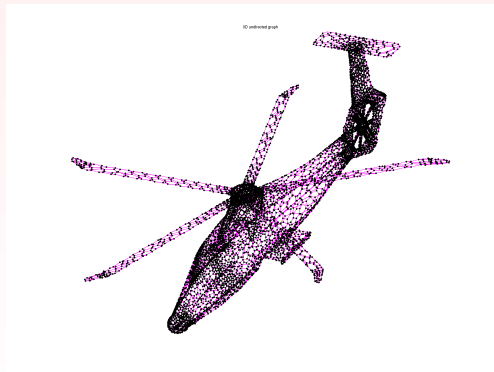
Řešíme problém: 2.



- Ale jak rozdělíme graf, kde není na první pohled vidět jak?
- Spousta technik a stále nové ve vývoji.

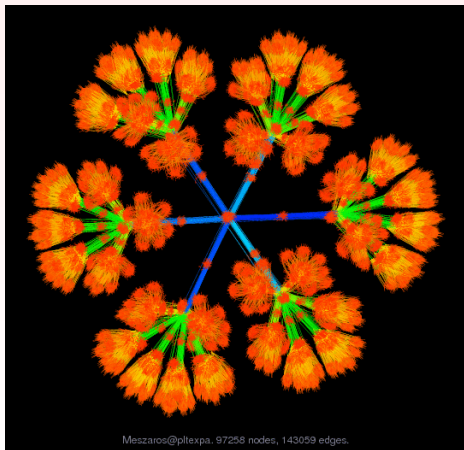
Řešíme problém: 2.

- Jak rozdělit například tenhle graf?



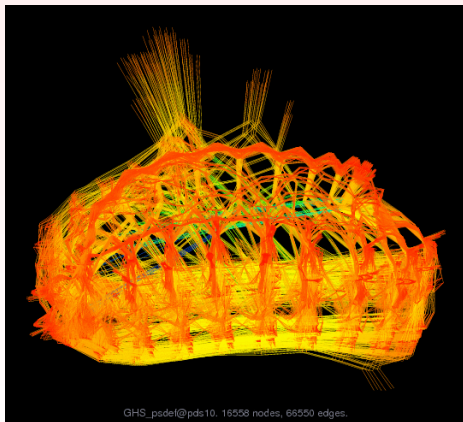
Řešíme problém: 2.

- A jak rozdělit tenhle graf?



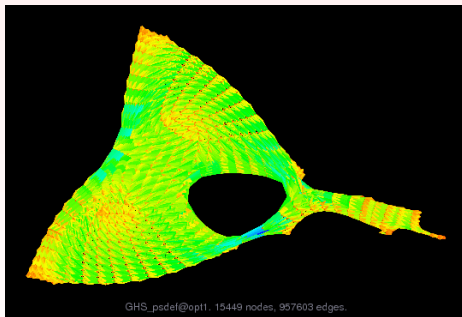
Řešíme problém: 2.

- A jak rozdělit tenhle graf?



Řešíme problém: 2.

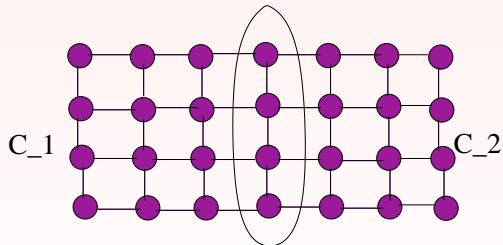
- A jak rozdělit tenhle graf?



Outline

- 1 Motivace
- 2 Šíření infekční choroby
- 3 Jiné motivace
- 4 K algoritmickému řešení problému vlastních čísel
- 5 Rozděl a spojuj**
- 6 Přemítání o problému
- 7 Dělení grafů na části
- 8 Rozděl a spojuj
- 9 Závěr

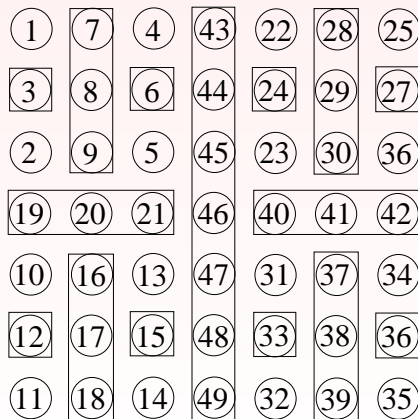
- Druhý, ale i první problém můžeme řešit následovně.



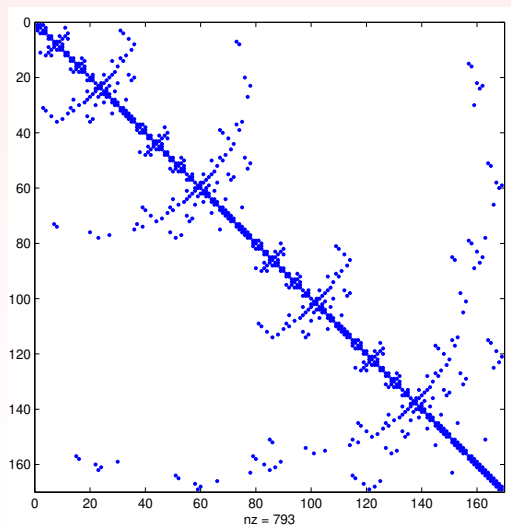
Matice po jednom dělení

	C_1	C_2	S
C_1			
C_2			
S			

Rozdělit můžeme vícrát - rekurzivně



Dělíme vícekrát: krásnoučká pravidelná matice



- Řada problémů praxe vede k diskrétním úlohám, které mají pod sebou strukturu vztahů / strukturu grafu
- Například naše úloha šíření infekční nemoci
- Řešení výsledných problémů vede k maticím, vektorům řadě zajímavých úloh mezi trojlístkem
 - Řídká matice
 - Problém vlastních čísel a řešení soustav lineárních rovnic
 - Vlastnosti počítačové architektury.