Another realistic application of linear algebra arises in search engines

- Markov matrices and Markov chains
- PageRank algorithm, Brin-Page algorithm

- Motivational problems:
  1. Urban-suburban populations
     - Let $p_{k,1}$, $p_{k,2}$ denote the average urban, suburban populations in year k
     - People move from city to suburb with probability $r_{2,1}$
     - People move from suburb to city with probability $r_{1,2}$
     - Next year average population is expected to be

$$\begin{aligned} p_{k+1,1} &= (1-r_{2,1})p_{k,1} + r_{1,2}p_{k,2} \\ p_{k+1,2} &= r_{2,1}\, p_{k,1} + (1-r_{1,2})p_{k,2} \end{aligned}$$

     - In matrix form, with $\boldsymbol{p}_k, \boldsymbol{p}_{k+1} \in \mathbb{R}^2$, $\boldsymbol{R} \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{p}_{k+1} = \begin{pmatrix} p_{k+1,1} \\ p_{k+1,2} \end{pmatrix} = \boldsymbol{R}\boldsymbol{p}_k = \begin{pmatrix} 1-r_{2,1} & r_{1,2} \\ r_{2,1} & 1-r_{1,2} \end{pmatrix} \begin{pmatrix} p_{k,1} \\ p_{k,2} \end{pmatrix}$$

2. Wolves eat sheep and grass, sheep eat grass, grass uses nutrients from corpses, manure

- Let $p_{k,1}$, $p_{k,2}$, $p_{k,3}$ denote the average wolf, sheep, grass populations in year k
- Wolves catch sheep with probability r_{12}, consume fraction r_{13} of grass
- Sheep consume fraction r_{23} of grass, benefit from wolves eating competitors at rate r_{21}
- Grass growth enhanced with probability r_{31} of wolf population, r_{32} of sheep population
- Next year average population is expected to be

$$\begin{aligned} p_{k+1,1} &= (1-r_{21}-r_{31})p_{k,1}+r_{12}p_{k,2}+r_{13}p_{k,3} \\ p_{k+1,2} &= r_{21}\,p_{k,1}+(1-r_{32}-r_{12})p_{k,2}+r_{23}p_{k,3} \\ p_{k+1,3} &= r_{31}\,p_{k,1}+r_{32}p_{k,2}+(1-r_{13}-r_{23})p_{k,3} \end{aligned}$$

- In matrix form, with $\boldsymbol{p}_k, \boldsymbol{p}_{k+1} \in \mathbb{R}^3$, $\boldsymbol{R} \in \mathbb{R}_+^{3\times 3}$

$$\boldsymbol{p}_{k+1} = \begin{pmatrix} p_{k+1,1} \\ p_{k+1,2} \\ p_{k+1,3} \end{pmatrix} = \boldsymbol{R}\boldsymbol{p}_k = \begin{pmatrix} 1-r_{21}-r_{31} & r_{12} & r_{13} \\ r_{21} & 1-r_{32}-r_{12} & r_{23} \\ r_{31} & r_{32} & 1-r_{13}-r_{23} \end{pmatrix} \begin{pmatrix} p_{k,1} \\ p_{k,2} \\ p_{k,3} \end{pmatrix}$$

Definition. *A Markov matrix $\boldsymbol{R} \in \mathbb{R}_+^{m \times m}$ (a.k.a. stochastic matrix) is a matrix with positive components and column vectors of unit 1-norm, $\|\boldsymbol{r}_j\|_1 = \sum_{i=1}^m r_{ij} = 1$. The matrix entry r_{ij} is the probability of transition from state j to state i.*

Definition. *A Markov chain $\{\boldsymbol{p}_k \in \mathbb{R}^m, k = 0, 1, \ldots\}$ (a.k.a. probability vectors) is a sequence of vectors generated by repeated application of a Markov matrix from the initial vector $\boldsymbol{p}_0$*

$$\boldsymbol{p}_{k+1} = \boldsymbol{R}\boldsymbol{p}_k$$

Definition. *The steady state of a Markov chain (if it exists) is the probability vector $\boldsymbol{x}$ for which*

$$\boldsymbol{x} = \boldsymbol{R}\boldsymbol{x}$$

- Note: The steady state of a Markov chain is simply the eigenvector associated with eigenvalue $\lambda = 1$

Theorem. *If $\boldsymbol{R} \in \mathbb{R}_+^{m \times m}$ is a Markov matrix there exists $\boldsymbol{x} \neq \boldsymbol{0}$ such that $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{x}$. If $r_{i,j} > 0$, $\boldsymbol{x}$ is unique.*

Google PageRank algorithm

- Given m interlinked webpages, rank them in order of "importance"
- Let x_i denote the importance score of page i, form vector $\boldsymbol{x} \in \mathbb{R}^m$
- Use the link structure of the web to determine "importance"
- "Importance" (PageRank 1993):
  1. x_i number of links to page i (does not take into account importance of source)
  2. x_i sum of importance scores of all pages linking to i
  3. x_i sum of importance scores of pages that link to i divided by the number of outgoing links on each page

$$x_i = \sum_{j=1}^{m} a_{i,j} \frac{x_j}{n_j} = \sum_{j=1}^{m} r_{i,j} x_j \Rightarrow \boldsymbol{x} = \boldsymbol{R}\boldsymbol{x}$$

with $a_{i,j} = 1$ if j links to i, $a_{i,j} = 0$ otherwise (*adjacency matrix*)

$$r_{k,j} = \frac{1}{n_j}$$

- Brin-Page $\boldsymbol{P} = 0.85\,\boldsymbol{R} + \frac{0.15}{m}\,\mathrm{ones}(m)$ (Stochastic matrices form a vector space!)

- Analyze importance of node within a foodchain