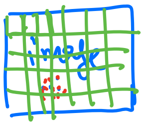
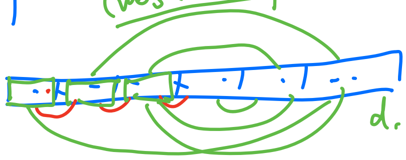
CNN RNN

(weight) dot product.





d.

$\binom{d}{2}$

input

"NLP"

"Computer Vision"

Transformer

Northeastern University, Department of Mathematics

MATH 4570 Matrix Methods for DA and ML

• Instructor: **He Wang** Email: **he.wang@northeastern.edu**

§17 Dynamical Systems

"NLP"
"Reinforcement learning"

$A\vec{x} = \lambda\vec{x}$

CONTENTS

1. Dynamical Systems and Eigenvectors.

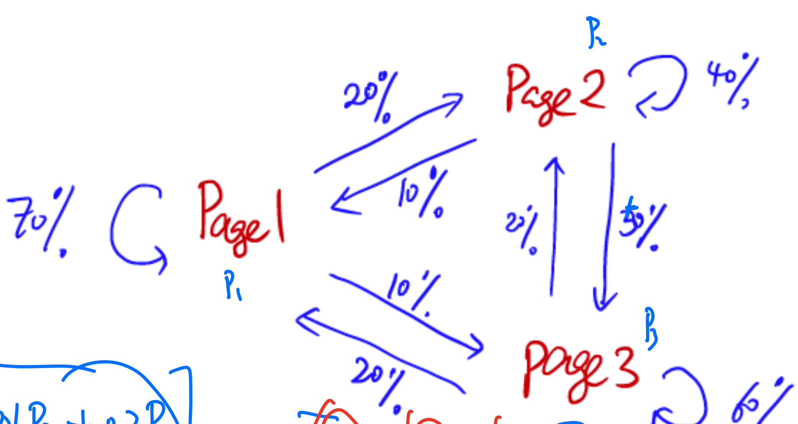
Google's PageRank Algorithm (Larry Page and Sergey Brin, 1996)

Consider a mini-web with only three pages: Page1, Page2, Page3. Initially, there is an equal number of surfers on each page. The initial probability distribution vector is

100
100
100

$$\vec{x}_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

After t_1 1 minute, some people will move onto different pages with a probability distribution vector $\vec{x}_1$, as in the following diagram



$$\vec{x}_1 = \begin{bmatrix} 0.7P_1 + 0.1P_2 + 0.2P_3 \\ 0.2P_1 + 0.4P_2 + 0.2P_3 \\ 0.1P_1 + 0.5P_2 + 0.6P_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.6 \end{bmatrix}}_{A} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

$$\vec{x}_1 = A\vec{x}_0$$

A .

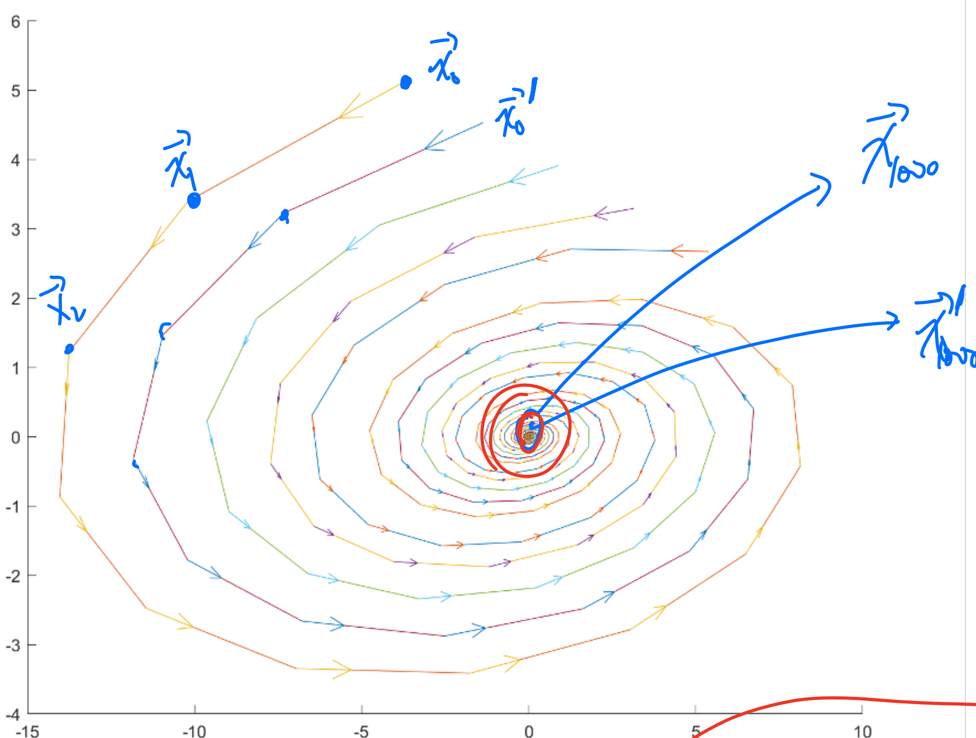
- $\vec{x}_2 = A\vec{x}_1 = A^2\vec{x}_0$
- $\vec{x}_3 = A\vec{x}_2 = A^3\vec{x}_0$

$\vdots$

- $\vec{x}_t = A^t\vec{x}_0$

Ex: $x_i \in \mathbb{R}^2$

Remark: Let A be a 2×2 matrix The endpoints of state vectors $\vec{x}(0)$, $\vec{x}(1)$, $\cdots$, $\vec{x}(t)$, ..., form the discrete **trajectory** of the system. A **phase portrait** of the dynamical system shows trajectories for various initial states.



$A^t = AAA \cdots A$ Never use it

$A = PDP^{-1}$

$A^t = PD^tP^{-1}$

PageRank Example:

$$A = \begin{bmatrix} 0.7 & 0.1 & 0.2 \\ 0.2 & 0.4 & 0.2 \\ 0.1 & 0.5 & 0.6 \end{bmatrix} \quad \text{and} \quad \vec{x}_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

Example 1. Find explicit formulas for A^t.

Example 2. Find explicit formulas for $A^t\vec{x}_0$

Example 3. Find $\lim_{t\to\infty} A^t$

Example 4. Find $\lim_{t\to\infty} A^t\vec{x}_0$

1. Find all eigenvalues by $\det(A-\lambda I)=0$

$\lambda_1=1, \quad \lambda_2=0.5, \quad \lambda_3=0.2$

2. Find an eigenbasis for A.

A basis for eigenspace E_{λ_1} is $\left\{\begin{bmatrix}7\\5\\8\end{bmatrix}\right\}$

A basis for eigenspace E_{λ_2} is $\left\{\begin{bmatrix}1\\0\\-1\end{bmatrix}\right\}$

A basis for eigenspace E_{λ_3} is $\left\{\begin{bmatrix}-1\\-3\\4\end{bmatrix}\right\}$

3. $A=PDP^{-1}$ where $P=\begin{bmatrix}7&1&-1\\5&0&-3\\8&-1&4\end{bmatrix}$ $D=\begin{bmatrix}1&&\\&0.5&\\&&0.2\end{bmatrix}$

4. $$A^t=PD^tP^{-1}=\begin{bmatrix}7&1&-1\\5&0&-3\\8&-1&4\end{bmatrix}\begin{bmatrix}1&&\\&0.5^t&\\&&0.2^t\end{bmatrix}\frac{1}{60}\begin{bmatrix}3&3&3\\44&-36&-16\\5&-15&5\end{bmatrix}$$

$$=\left[\vec{b}_1 \quad 0.5^t\vec{b}_2 \quad 0.2^t\vec{b}_3\right]P^{-1}$$

$[\vec{x}_0]_B=\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}=\begin{bmatrix}1/20\\-2/45\\-1/36\end{bmatrix}$ $\quad \vec{x}_0=c_1\vec{b}_1+c_2\vec{b}_2+c_3\vec{b}_3$ $\quad \vec{x}_0=P\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}$

$A^t\vec{X}_0=\begin{bmatrix}&&\\&&\end{bmatrix}\begin{bmatrix}1&&\\&0.5^t&\\&&0.2^t\end{bmatrix}\begin{bmatrix}p_1\\p_2\\p_3\end{bmatrix}$

$$A^t\vec{x}_0=c_1\lambda_1^t\vec{b}_1+c_2\lambda_2^t\vec{b}_2+c_3\lambda_3^t\vec{b}_3$$

$$=\frac{1}{20}\cdot\vec{b}_1-\frac{2}{45}(0.5^t)\vec{b}_2-\frac{1}{36}(0.2^t)\vec{b}_3$$

$$\lim_{t\to\infty}A^t=\left[\vec{b}_1 \quad \vec{0} \quad \vec{0}\right]P^{-1}=\begin{bmatrix}7&0&0\\5&0&0\\8&0&0\end{bmatrix}\frac{1}{60}\begin{bmatrix}3&3&3\\44&-36&-16\\5&-15&5\end{bmatrix}$$

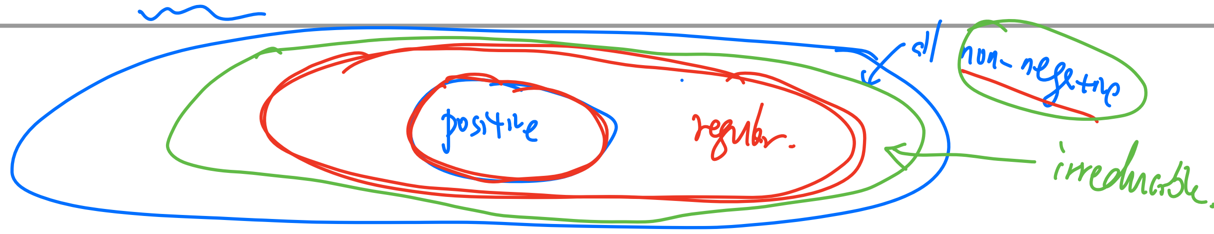
$$=\frac{1}{20}\begin{bmatrix}7&7&7\\5&5&5\\8&8&8\end{bmatrix}$$

$$\lim_{t\to\infty}A^t\vec{x}_0=\frac{1}{20}\vec{b}_1=\frac{1}{20}\begin{bmatrix}7\\5\\8\end{bmatrix}=\begin{bmatrix}7/20\\5/20\\8/20\end{bmatrix}$$

Equilibria for regular transition matrices:

Let us start with some terminologies: $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

Definition 5. $x_i \geq 0$
- A vector $\vec{x} \in \mathbb{R}^n$ is said to be a **distribution** vector if its entries are non-negative and the sum is 1. $x_1 + x_2 + \cdots + x_n = 1$
- A square matrix A is said to be a **transition matrix** (or column stochastic matrix) if all its columns are distributions vectors.



Definition 6.
- A matrix A is said to be $\boxed{\textbf{non-negative}}$ if each entry of matrix A is not negative. $a_{ij} \geq 0$
- A matrix A is said to be **positive** if each entry of matrix A is positive. $a_{ij} > 0$
- A non-negative matrix A is said to be "**regular**" (or **primitive**, or eventually positive) if the matrix A^m is positive for some integer $m > 0$.

Ex1 $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

Ex: $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ $\quad A^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ regular.

Ex: $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ $\quad A^n = \begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix}$

Ex: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\quad A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\quad A^3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\quad A^4 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ... ..

Not regular. (irreducible)

A $d \times d$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ \vdots & & & \\ a_{d1} & & \cdots & a_{dd} \end{bmatrix}$$

transition matrix :

$A^n = PD^nP^{-1}$

or $A^n = PJ^nP^{-1}$

Suppose $\vec{v_0}$ is a distribution vector.

$t_0 \quad t_1 \quad t_2 \quad \cdots \quad t_n \quad \cdots$

$\vec{V_0} \xrightarrow{A} \vec{V_1} \xrightarrow{A} \vec{V_2} \rightarrow \cdots \rightarrow \vec{V_n}$

$V_1 = A\vec{V_0}$

$\vec{V_n} = A^n \vec{V_0}$

Goal: $\lim_{n\to\infty} A^n \vec{V_0}$

If A^m is positive for some m,

then $\lambda = 1$ is the largest eigenvalue of A.

Solve. $(A - I)\vec{x} = \vec{0}$

and get $\vec{x} = \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}$ $\quad v_i > 0$

then define $\boxed{\vec{V}_{equ.} = \frac{1}{v_1 + \cdots + v_d} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix}}$

then $\lim_{n\to\infty} A^n \vec{v_0} = \vec{V}_{equ.}$

→ (cont'd) ...

Theorem 7 (Perron-Frobenius Theorem (special case for transition matrix))**.** *If A is a positive, column stochastic matrix, then:*

- *$\lambda = 1$ is an eigenvalue of multiplicity one.*
- *1 is the largest eigenvalue: all the other eigenvalues have absolute value smaller than 1.*
- *the eigenvectors corresponding to the eigenvalue 1 have either only positive entries or only negative entries. In particular, for the eigenvalue 1 there exists a unique eigenvector with the sum of its entries equal to 1.*

Theorem 8. *Let A be a "regular" transition $n \times n$ matrix.*
1. There exists exactly one distribution vector $\vec{x} \in \mathbb{R}^n$ such that $\ker(A - I)$

$$A\vec{x} = \vec{x}$$

which is called ***equilibrium*** *distribution for A denoted as $\vec{x}_{equ}$.*
2. If $\vec{x}_0$ is any distribution vector in $\mathbb{R}^n$, then

$$\lim_{m\to\infty} (A^m \vec{x}_0) = \vec{x}_{equ}$$

3. The columns of $\lim_{n\to\infty}(A^n)$ are all $\vec{x}_{equ}$, that is

$$\lim_{m\to\infty} (A^m) = [\vec{x}_{equ}\ \vec{x}_{equ} \ldots \vec{x}_{equ}]$$ rank $= 1$

Ex: Step1: Solve $(A - I)\vec{x} = \vec{0} \Rightarrow \vec{x} = \begin{bmatrix} 7 \\ 5 \\ 8 \end{bmatrix}$

Step2: $\vec{x}_{equ} = \frac{1}{7+5+8}\begin{bmatrix} 7 \\ 5 \\ 8 \end{bmatrix}$

Step3: $\lim_{n\to\infty} A^n \vec{x}_0 = \frac{1}{20}\begin{bmatrix} 7 \\ 5 \\ 8 \end{bmatrix}$

SS

$A^{1000}\vec{x}_0$

Markov Chains (1906) can be used to study real word questions like PageRank of a webpage as used by Google, automatic speech recognition systems, probabilistic forecasting, cruise control systems in motor vehicles, queues or lines of customers arriving at an airport/train station/..., currency exchange rates, animal population dynamics, music, etc.

Convention in Probability: all vectors are transposed if you read some probability books about Markov chains. A stochastic matrix P comes from a stochastic process $\{X_0, ..., X_n\}$ with values in $\{1, ..., n\}$.

$$p_{ij} = P(X_{t+1} = i \mid X_t = j) \qquad P = \begin{bmatrix} \quad \end{bmatrix}$$

Ex: $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

3. Perron-Frobenius Theorem

$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$\lambda = 1$ $\lambda = -1$

Theorem 9 (Perron-Frobenius Theorem). *Let A be an "irreducible" non-negative matrix.*

- *A has a positive (real) eigenvalue $\lambda_{\max}$ such that all other eigenvalues of A satisfy $|\lambda| \leq \lambda_{\max}$.*
- *$\lambda_{\max}$ has algebraic multiplicity 1 with a positive eigenvector $\vec{x}$.*
- ~~*Any non-negative eigenvector is a multiple of $\vec{x}$.*~~
- *If A is primitive, then all other eigenvalues of A satisfy $|\lambda| < \lambda_{\max}$.*

regular

This theorem was first proved for positive matrices by Oskar Perron in 1907 and extended by Ferdinand Georg Frobenius to non-negative irreducible matrices in 1912. .

Example 10. (Ranking of Players) The results of a round tournament be represented by the following matrix.

$$A = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \end{array} \begin{array}{c} \begin{array}{cccccc} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{array} \\ \begin{bmatrix} 0.5 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0.5 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0.5 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0.5 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0.5 \end{bmatrix} \end{array} \qquad a_{12} \qquad \vec{r}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad A\vec{r}_0$$

Here $a_{i,j} = 1$ represents player i win v.s. player j; and $a_{i,j} = 0$ represents player i loss v.s. player j.

Question: How to rank those 6 players from the results?

Suppose before the game, all ranked 1, represented by ranking vector $\vec{r_0} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ After the tournament, the ranking is $\vec{r_1} = A\vec{r_0} = \begin{bmatrix} 4.5000 \\ 2.5000 \\ 3.5000 \\ 1.5000 \\ 3.5000 \\ 2.5000 \end{bmatrix}$.

The rank is $P_1 > P_5 = P_3 > P_2 = P_6 > P_4$.

Consider the strength of the opponents, we calculate $\vec{r_2} = A\vec{r_1} = \begin{bmatrix} 14.2500 \\ 6.2500 \\ 8.2500 \\ 5.2500 \\ 9.2500 \\ 5.2500 \end{bmatrix}$, and $\vec{r_3} = A\vec{r_2} = \begin{array}{c} \begin{bmatrix} 36.1250 \\ 17.6250 \\ 20.8750 \\ 16.8750 \\ 23.3750 \\ 14.1250 \end{bmatrix} \end{array} \begin{array}{c} 1 \\ 4 \\ 3 \\ 5 \\ 2 \\ 6 \end{array}$. Now we can see the rank: $P_1 > P_5 > P_3 > P_2 > P_4 > P_6$.

The eigenvalues of A are $2.7261; 0.0028; 0.1303 \pm 1.3750i; 0.0052 \pm 1.0451i$.

λ_{max}

$\lambda = 2.7261$ is the largest eigenvalue with eigenvector $\begin{bmatrix} 0.2721 \\ 0.1372 \\ 0.1689 \\ 0.1222 \\ 0.1831 \\ 0.1165 \end{bmatrix}$. This vector is almost the same as $\vec{r}_{\geq 10}$ divided by the sum of the entries.

if A is transition matrix

$\lim_{k\to\infty} A^k = [\vec{V}_{equ}\ \vec{V}_{equ}\ \cdots\ \vec{V}_{equ}]$

A^m is positive,

rank 1

4. **More Applications**

e.g. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

4.1. Powers of a primitive matrix.

Let A be a primitive matrix. By the Perron-Frobenius theorem, let $\lambda_{\max}$ be its maximal eigenvalue.

Let $\vec{u}$ be a (right-handed) positive eigenvector of A with eigenvalue $\lambda_{\max}$, so $A\vec{u} = \lambda_{\max}\vec{u}$.

Let $\vec{v}$ be the left-handed eigenvector vector such that $\vec{v}^T A = \lambda_{\max}\vec{v}^T$ and $\vec{v}\cdot\vec{u} = 1$.

1xd dxd

?

Theorem 11. *Suppose A is primitive, with maximal eigenvalue $\lambda_{\max}$, left eigenvector $\vec{u}$ and right eigenvector $\vec{v}$ such that $\vec{v}\cdot\vec{u} = 1$, then*

$$\lim_{k\to\infty}\left(\frac{1}{\lambda_{\max}}A\right)^k = \vec{u}\vec{v}^T$$

$A^T\vec{v} = \lambda_{max}\vec{v}$

Step1: Find λ_{max}

Step2: Find $\vec{u}$ by solving $(A-\lambda_{max}I)\vec{x}=\vec{0}$

Step3: Find $\vec{v}$ by solving $(A^T-\lambda_{max}I)\vec{x}=\vec{0}$

Step4: $\lim_{k\to\infty}\left(\frac{1}{\lambda_{max}}A\right)^k = \frac{1}{\vec{u}\cdot\vec{v}}\vec{u}\vec{v}^T$

SS

Step5: Estimate $\left(\frac{1}{\lambda_{max}}A\right)^{1000}$

Ex: $A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$

Step1: $\lambda = 4$ $\lambda = -1$

2. Solve $(A-4I)\vec{x}=\vec{0} \Rightarrow \vec{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

3. Solve $(A^T-4I)\vec{x}=\vec{0} \Rightarrow \vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

4. $\lim_{k\to\infty}\left(\frac{1}{4}A\right)^k = \frac{1}{5}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \end{bmatrix}$

$= \frac{1}{5}\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix}$

$A = PDP^{-1}$

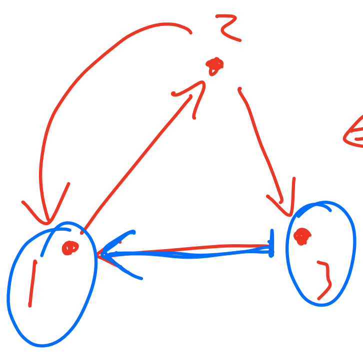
$\frac{1}{4^k}A^k = PD^kP^{-1} = \frac{4^k}{4^k}\frac{1}{5}\begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix} + \frac{(-1)^k}{4^k}\begin{bmatrix} 3/5 & -3/5 \\ -2/5 & 2/5 \end{bmatrix}$

$D = \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$ $P = \begin{bmatrix} 1 & : \\ 1 & : \end{bmatrix}$

4.2. Graphs and Non-negative matrices.

4.3. Population model (The Leslie Model).

4.4. Economic growth.



$$\Longleftrightarrow \quad V = \{1, 2, 3\}, \quad E = \{12, 23, 31, 21\}$$

$$A = \begin{array}{c} \begin{array}{ccc} & a_{12} & a_{13} \end{array} \\ \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array} \qquad a_{ij} = \begin{cases} 1 & \text{if } v_j v_i \in E \\ 0 & \end{cases}$$

$$A^k = \begin{bmatrix} \ddots & \ddots & \bullet \\ \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots \end{bmatrix}$$

$$(A^k)_{ij} = \#(\text{paths of length } k \text{ from } v_j \text{ to } v_i)$$

Further reading about the PageRank:

Other lectures:

http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

A little more professional:

https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf

https://www.math.purdue.edu/~ttm/google.pdf

http://www.ams.org/publicoutreach/feature-column/fcarc-pagerank

Original paper:

Sergey Brin and Lawrence Page http://infolab.stanford.edu/~backrub/google.html