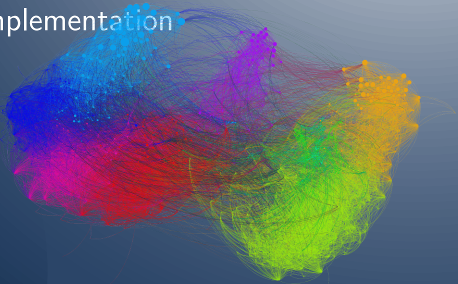
Graphs in Machine Learning

Google PageRank: Computation

Power Method and Large-Scale Implementation

Michal Valko

Inria & ENS Paris-Saclay, MVA



Partially based on material by: Andreas Krause, Branislav Kveton, Michael Kearns

Success story #2 Google `PageRank`

Google matrix: $\mathbf{G} = (1 - p)\mathbf{M} + p \cdot \frac{1}{N}\mathbb{1}_{N \times N}$, where $p = 0.15$

Success story #2 Google PageRank

Google matrix: $\mathbf{G} = (1-p)\mathbf{M} + p \cdot \frac{1}{N}\mathbb{1}_{N \times N}$, where $p = 0.15$

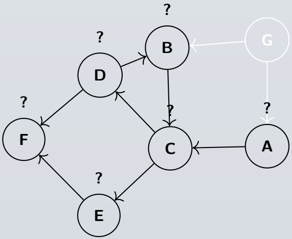
$\mathbf{G}$ is **stochastic** why? What is $\mathbf{G}\mathbf{a}$ for any $\mathbf{a}$? We look for $\mathbf{G}\mathbf{v} = 1 \times \mathbf{v}$, steady-state vector, a right eigenvector with eigenvalue 1. why?

Success story #2 Google PageRank

Google matrix: $\mathbf{G} = (1-p)\mathbf{M} + p \cdot \frac{1}{N}\mathbb{1}_{N \times N}$, where $p = 0.15$

$\mathbf{G}$ is **stochastic** why? What is $\mathbf{G}\mathbf{a}$ for any $\mathbf{a}$? We look for $\mathbf{G}\mathbf{v} = 1 \times \mathbf{v}$, steady-state vector, a right eigenvector with eigenvalue 1. why?

Perron's theorem: Such v exists and it is **unique** if the entries of $\mathbf{G}$ are positive.

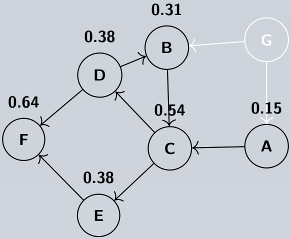


Success story #2 Google PageRank

Google matrix: $\mathbf{G} = (1-p)\mathbf{M} + p \cdot \frac{1}{N}\mathbb{1}_{N\times N}$, where $p = 0.15$

$\mathbf{G}$ is **stochastic** why? What is $\mathbf{G}\mathbf{a}$ for any $\mathbf{a}$? We look for $\mathbf{G}\mathbf{v} = 1 \times \mathbf{v}$, steady-state vector, a right eigenvector with eigenvalue 1. why?

Perron's theorem: Such v exists and it is **unique**
if the entries of $\mathbf{G}$ are positive.

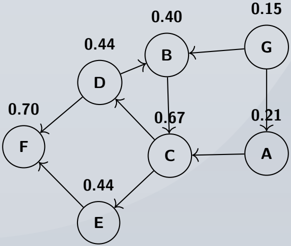


Success story #2 Google PageRank

Google matrix: $\mathbf{G} = (1-p)\mathbf{M} + p \cdot \frac{1}{N}\mathbb{1}_{N\times N}$, where $p = 0.15$

$\mathbf{G}$ is **stochastic** why? What is $\mathbf{G}\mathbf{a}$ for any $\mathbf{a}$? We look for $\mathbf{G}\mathbf{v} = 1 \times \mathbf{v}$, steady-state vector, a right eigenvector with eigenvalue 1. why?

Perron's theorem: Such v exists and it is **unique**
if the entries of $\mathbf{G}$ are positive.



Success story #2 Google PageRank

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]

Success story #2 Google PageRank

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]
- US patent for PageRank granted in 2001

Success story #2 Google PageRank

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]
- US patent for PageRank granted in 2001
- Google indexes 10's of billions of web pages (1 billion $= 10^9$)

Success story #2 Google PageRank

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]
- US patent for PageRank granted in 2001
- Google indexes 10's of billions of web pages (1 billion $= 10^9$)
- Google serves ≥ 200 million queries per day

Success story #2 Google `PageRank`

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]
- US patent for PageRank granted in 2001
- Google indexes 10's of billions of web pages (1 billion $= 10^9$)
- Google serves ≥ 200 million queries per day
- Each query processed by ≥ 1000 machines

Success story #2 Google `PageRank`

History: [Desikan, 2006]

- The anatomy of a large-scale hypertextual web search engine [Brin & Page 1998]
- US patent for PageRank granted in 2001
- Google indexes 10's of billions of web pages (1 billion $= 10^9$)
- Google serves ≥ 200 million queries per day
- Each query processed by ≥ 1000 machines
- All search engines combined process more than 500 million queries per day

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse**

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

Success story #2 Google `PageRank`

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T}$$

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T}$$
$$\mathbf{v}_1 = \mathbf{G}\mathbf{v}_0$$

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^{\mathsf{T}}$$
$$\mathbf{v}_1 = \mathbf{G}\mathbf{v}_0$$
$$\mathbf{v}_{t+1} = \mathbf{G}\mathbf{v}_t$$

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$
\begin{aligned}
\mathbf{v}_0 &= (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T} \\
\mathbf{v}_1 &= \mathbf{G}\mathbf{v}_0 \\
\mathbf{v}_{t+1} &= \mathbf{G}\mathbf{v}_t
\end{aligned}
$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t \implies \mathbf{G}\mathbf{v}_t = \mathbf{v}_t$$

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T}$$
$$\mathbf{v}_1 = \mathbf{G}\mathbf{v}_0$$
$$\mathbf{v}_{t+1} = \mathbf{G}\mathbf{v}_t$$

$\mathbf{v}_{t+1} = \mathbf{v}_t \implies \mathbf{G}\mathbf{v}_t = \mathbf{v}_t$ and we found the steady vector

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T}$$

$$\mathbf{v}_1 = \mathbf{G}\mathbf{v}_0$$

$$\mathbf{v}_{t+1} = \mathbf{G}\mathbf{v}_t$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t \implies \mathbf{G}\mathbf{v}_t = \mathbf{v}_t$$ and we found the steady vector

But wait, $\mathbf{M}$ is sparse, but $\mathbf{G}$ is dense! What to do?

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

- $n = 10^9$!!!
- luckily: **sparse** (average outdegree: 7)
- power method

$$\mathbf{v}_0 = (1_A \quad 0_B \quad 0_C \quad 0_D \quad 0_E \quad 0_F)^\mathsf{T}$$
$$\mathbf{v}_1 = \mathbf{G}\mathbf{v}_0$$
$$\mathbf{v}_{t+1} = \mathbf{G}\mathbf{v}_t$$

$\mathbf{v}_{t+1} = \mathbf{v}_t \implies \mathbf{G}\mathbf{v}_t = \mathbf{v}_t$ and we found the steady vector

But wait, $\mathbf{M}$ is sparse, but $\mathbf{G}$ is dense! What to do?

we store only $\mathbf{M}$ but do computations as with $\mathbf{G}$

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html