



Graphs in Machine Learning

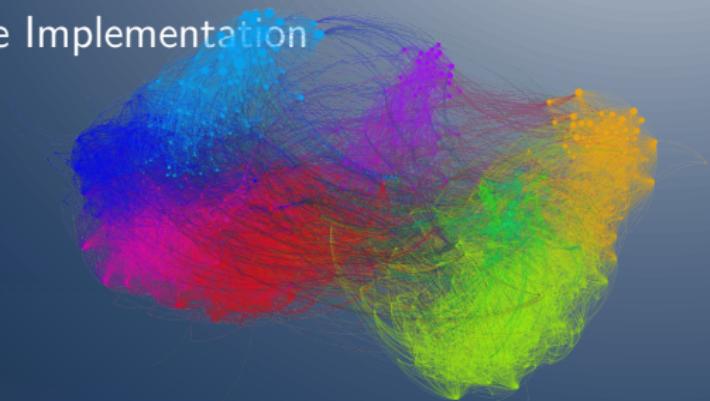
Google PageRank: Computation

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

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 ($1 \text{ billion} = 10^9$)
- 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 ($1 \text{ billion} = 10^9$)
- 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 ($1 \text{ billion} = 10^9$)
- 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

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

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

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

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

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

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

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

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

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

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

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

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

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

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

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

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

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

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

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

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

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

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

Success story #2 Google PageRank

Problem: Find an eigenvector of a stochastic matrix.

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

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

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

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

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

But wait, **M** is sparse, but **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

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

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

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

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

But wait, **M** is sparse, but **G** is dense! What to do?

we store only **M** but do computations as with **G**

Michal Valko

`michal.valko@inria.fr`

Inria & ENS Paris-Saclay, MVA



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