Data Sci Discover Project

Ryan Greenup

September 3, 2020

Contents

1 Simplifying Power Walk to be solved with Sparse Matrices

The Random Surfer model is:

$$\mathbf{S} = \alpha \mathbf{T} + \mathbf{F}$$

where:

- $\mathbf{T}$
  - is an $i \times j$ matrix that describes the probability of travelling from vertex j to i
    - This is transpose from the way that `igraph` produces an adjacency matrix.
- $\mathbf{F} = \begin{bmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \frac{1}{n} \vdots \end{bmatrix}$

Interpreting the transition probability matrix in this way is such that $\mathbf{T} = \mathbf{A}\mathbf{D}_A^{-1}$ under the following conditions:

- No column of $\mathbf{A}$ sums to zero
  - If this does happen the question arises how to deal with $\mathbf{D}_\mathbf{A}^{-1}$
    - I've been doing $D'_{\mathbf{A},i,j} := \mathtt{diag}\left(\frac{1}{\mathtt{colsums}(\mathbf{A})}\right)$ and then replacing any 0 on the diagonal with 1.
  - What is done in the paper is to make another matrix $\mathbf{Z}$ that is filled with 0, if a column sum of $\mathbf{A}$ adds to zero then that column in $\mathbf{Z}$ becomes $\frac{1}{n}$
    - This has the effect of making each row identical
    - The probability of going from an orphaned vertex to any other vertex would hence be $\frac{1}{n}$
    - The idea with this method is then to use $D^{-1}_{(\mathbf{A}+\mathbf{Z})}$ this will be consistent with the *Random Surfer* the method using $\mathbf{F}$ in]

  where each row is identical that is a 0

The way to deal with the *Power Walk* is more or less the same.

observe that:

$$\mathbf{B} = \beta^{\mathbf{A}} \wedge \mathbf{A}_{i,j} \in \mathbb{R} \implies |\mathbf{B}_{i,j}| > 0 \quad \forall i,j > n \in \mathbb{Z}^+$$

Be mindful that the use of exponentiation in] is not an element wise exponentiation and not an actual matrix exponential (which would be defined by using power series and logs but is defined)

So if I have:

- $\mathbf{O}_{i,j} := 0, \quad \forall i,j \leq n \in \mathbb{Z}^+$
- $\vec{p_i}$ as the state distribution, being a vector of length n

Then It can be shown (see (2)):

$$\mathbf{O}\mathbf{D}_\mathbf{B}^{-1}\vec{p_i} = \vec{\delta}^{\mathbf{T}}\vec{p_i}$$

where:

- $\vec{\delta_i} = \frac{1}{\mathtt{colsums}(\mathbf{B})}$

This means we can do:

$$
\begin{aligned}
\vec{p_{i+1}} &= \mathbf{B}\mathbf{D}_{\mathbf{B}}^{-1}\vec{p_i} \\
&= (\mathbf{B} - \mathbf{O} + \mathbf{O})\,\mathbf{D}_{\mathbf{B}}^{-1}\vec{p_i} \\
&= \left((\mathbf{B} - \mathbf{O})\,\mathbf{D}_{\mathbf{B}}^{-1} + \mathbf{O}\mathbf{D}_{\mathbf{B}}^{-1}\right)\vec{p_i} \\
&= (\mathbf{B} - \mathbf{O})\,\mathbf{D}_{\mathbf{B}}^{-1}\vec{p_i} + \mathbf{O}\mathbf{D}_{\mathbf{B}}^{-1}\vec{p_i} \\
&= (\mathbf{B} - \mathbf{O})\,\mathbf{D}_{\mathbf{B}}^{-1}\vec{p_i} + \vec{\delta'}\vec{p_i}\vec{1}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{O}\mathbf{D}_{\mathbf{B}}^{-1}\overrightarrow{p_i} &= \begin{pmatrix} 1 & 1 & 1 & \\ 1 & 1 & 1 & \dots \\ 1 & 1 & 1 & \\ & \vdots & & \ddots \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n}[\boldsymbol{A}_{i,1}] & 0 & 0 & \\ 0 & \sum_{i=1}^{n}[\boldsymbol{A}_{i,2}] & 0 & \dots \\ 0 & 0 & \sum_{i=1}^{n}[\boldsymbol{A}_{i,3}] & \\ 0 & \vdots & 0 & \ddots \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \end{pmatrix} \\
&= \begin{pmatrix} \frac{p_1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,1}]} & \frac{p_2}{\sum_{i=1}^{n}[\boldsymbol{A}_{2,1}]} & \frac{p_3}{\sum_{i=1}^{n}[\boldsymbol{A}_{3,1}]} & \\ \frac{p_1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,1}]} & \frac{p_2}{\sum_{i=1}^{n}[\boldsymbol{A}_{2,1}]} & \frac{p_3}{\sum_{i=1}^{n}[\boldsymbol{A}_{3,1}]} & \dots \\ \frac{p_1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,1}]} & \frac{p_2}{\sum_{i=1}^{n}[\boldsymbol{A}_{2,1}]} & \frac{p_3}{\sum_{i=1}^{n}[\boldsymbol{A}_{3,1}]} & \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,1}]} & \frac{1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,2}]} & \frac{1}{\sum_{i=1}^{n}[\boldsymbol{A}_{i,3}]} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} && (1) \\
&= \boldsymbol{\delta}^{\boldsymbol{T}}\overrightarrow{p}\,\overrightarrow{1} && (2)
\end{aligned}
$$