- The generalized voltage v_k
  - $\Pr[\text{A random walk starting at } k \text{ visits } j \text{ before } i]$
- Calculating $\boldsymbol{v} \triangleq (v_1\ v_2\ \cdots\ v_n)^\top$
  - $v_i = 0$ and $v_j = 1$
  - $\forall k \neq i, j,\ v_k = \sum_l p_{kl} \cdot v_l$
  - Split $\boldsymbol{P} = \begin{pmatrix} \hat{\boldsymbol{P}} & \boldsymbol{c}_i & \boldsymbol{c}_j \\ \boldsymbol{r}_i^\top & 0 & p(i,j) \\ \boldsymbol{r}_j^\top & p(j,i) & 0 \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} \hat{\boldsymbol{v}} & 0 & 1 \end{pmatrix}^\top$
  - Then $\hat{\boldsymbol{v}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{v}} + \boldsymbol{r}_j \;\Rightarrow\; \hat{\boldsymbol{v}} = (\boldsymbol{I} - \hat{\boldsymbol{P}})^{-1}\boldsymbol{r}_j$
- $\mathrm{ep}(i \to j) = \sum_k p_{ik} \cdot v_k = \boldsymbol{r}_i^\top (\boldsymbol{I} - \hat{\boldsymbol{P}})^{-1}\boldsymbol{c}_j + p(i,j)$
- Computing all $\mathrm{ep}(i \to j)$ requires $\Theta(n^2)$ matrix inversions

Theorem. *Let*

$$\boldsymbol{Q} = [q(i,j)] \triangleq (\boldsymbol{I} - c\boldsymbol{P})^{-1}.$$

$\forall i \neq j$, *there is*

$$\mathrm{ep}(i \to j) = \frac{q(i,j)}{q(i,i)q(j,j) - q(i,j)q(j,i)}.$$

- Fast solution to all-pair proximities
  - Compute $\boldsymbol{Q} = (\boldsymbol{I} - c\boldsymbol{P})^{-1}$
  - For all pair of nodes i and j, compute $\mathrm{Prox}(i,j) = \frac{q(i,j)}{q(i,i)q(j,j)-q(i,j)q(j,i)}$
- Time complexity $\Theta(1 \text{ matrix inversion}) + \Theta(n^2)$
- When only the proximity of one pair of nodes
  - Only need two columns of $\boldsymbol{Q}$
  - Taylor expansion

$$(\boldsymbol{I} - c\boldsymbol{P})^{-1} = \boldsymbol{I} + c\boldsymbol{P} + (c\boldsymbol{P})^2 + \cdots$$

  - Computing i-th column of $\boldsymbol{Q}$

$$\boldsymbol{Q}\boldsymbol{e}_i = (\boldsymbol{I} - c\boldsymbol{P})^{-1}\boldsymbol{e}_i = \boldsymbol{e}_i + c\boldsymbol{P}\boldsymbol{e}_i + (c\boldsymbol{P})^2\boldsymbol{e}_i + \cdots$$

- Time complexity $\Theta\left(t(n+m)\right)$, where t is the number of iterations

Lemma. *The expected time r_i for a random walk starting at node i to return to i is the reciprocal of the stationary probability of i. That is*

$$r_i = \frac{1}{\pi_i}.$$

- Intuitively
  - A long walk always ends up in stationary distribution $\boldsymbol{\pi}$
  - Suppose the walk length is T, then the expected number of times it visits i is $\pi_i \cdot T$
  - The average length between two visits is $\frac{T}{\pi_i \cdot T} = \frac{1}{\pi_i}$
  - A rigorous proof requires the Strong Law of Large Numbers

Theorem. *The probability that a random walk starting at node i visits j before returning to i, which equals* $\mathrm{ep}(i \to j)$*, satisfies*

$$\mathrm{ep}(i \to j) c(i,j) = \frac{1}{\pi_i},$$

where $c(i,j)$ is the commute time between i and j.

- Proof
  - Consider a random walk w starting at i, and random variables
    - $X =$ the first time w returns to i
    - $Y =$ the first time w returns to i after visiting j
  - By definition $E(X) = \frac{1}{\pi_i}$ and $E(Y) = c(i,j)$
  - Clearly $X \leq Y$, and $\Pr[X = Y] = p \triangleq \mathrm{ep}(i \to j)$
    - $E(Y - X) = p \cdot 0 + (1-p) \cdot E(Y) = (1-p)c(i,j)$
  - Also $E(Y - X) = E(Y) - E(X) = c(i,j) - \frac{1}{\pi_i}$
- $\mathrm{ep}(i \to j) + \mathrm{ep}(j \to i) = \frac{1}{c(i,j)} \left(\frac{1}{\pi_i} + \frac{1}{\pi_j} \right)$
- Recall that $h(i,j)$ is small whenever π_j is large, which is bad for personalization
  - Sarkar et al. (2008) alleviate this by restricting the length of random walk
  - Tong et al. (2007) alleviate this by reducing the dependence on stationary distribution