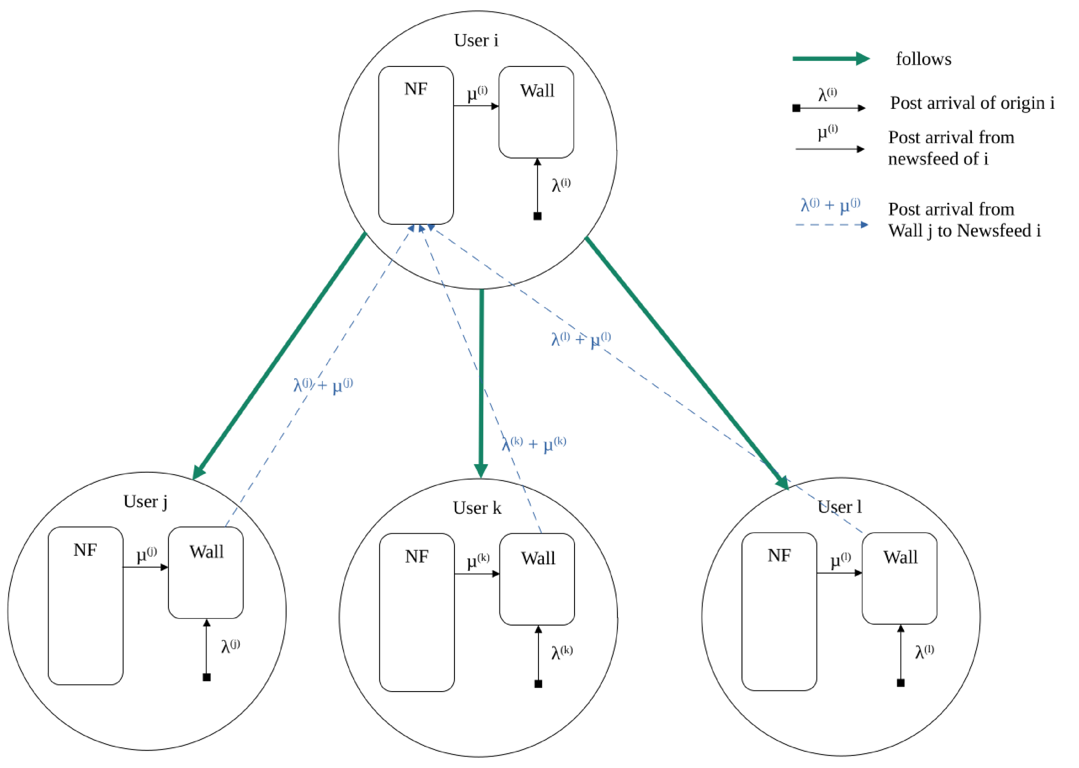
A Fast Algorithm for Ranking Users by their Influence in Online Social Platforms

Nouamane Arhachoui, Esteban Bautista, Lionel Tabourier, Anastasios Giovanidis

Sorbonne Université, CNRS – LIP6, Paris, France

Social Platform Model



Social Graph:

- $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where $(i, j) \in \mathcal{E}$ iff user i follows user j. $|\mathcal{N}| = N$, $|\mathcal{E}| = M$.
- Each node has a **Wall** and a **Newsfeed**

Activity rates for a user n:

- $\lambda^{(n)}$: **posting rate** for user n, number of posts per unit of time created by n
- $\mu^{(n)}$: **re-posting rate** for user n, number of re-posts that n does per unit of time.

Output of the model for each i in $\mathcal{N}$:

- $\mathbf{p}_i = (p_i^{(1)}\ p_i^{(2)}\ \cdots\ p_i^{(N)})^T$ where, $\forall n \in \mathcal{N}$, $p_i^{(n)}$ is the expected proportion of posts from user i on the Newsfeed of user n.
- $\mathbf{q}_i = (q_i^{(1)}\ q_i^{(2)}\ \cdots\ q_i^{(N)})^T$ where, $\forall n \in \mathcal{N}$, $q_i^{(n)}$ is the expected proportion of posts from user i on the Wall of user n.

$q_i^{(n)}$ also represents the **influence** of a user i on user n.

The ψ-score

Influence of a user i over the entire network:

$$\psi_i = \frac{1}{N}\sum_{n\in\mathcal{N}} q_i^{(n)}. \qquad (1)$$

Process to determine ψ_i:

- solve the following system of N equations:

$$\mathbf{p}_i = \mathbf{A}.\mathbf{p}_i + \mathbf{b}_i \qquad (2)$$

where $\mathbf{A}$ is a sub-stochastic matrix and $\mathbf{b}_i$ is a column vector; both defined with λ and μ

- then, map the $\mathbf{p}_i$ vector in:

$$\mathbf{q}_i = \mathbf{C}.\mathbf{p}_i + \mathbf{d}_i \qquad (3)$$

where $\mathbf{C}$ is a diagonal matrix and $\mathbf{d}_i$ is a column vector with a non-zero entry only at row i

- finally, compute the average influence with (1)

Problem

Solving N systems of N equations is required to get the whole ψ-score vector.

This is time consuming and challenging when dealing with large-scale real-world datasets.

Goal

Given a social graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where the nodes have a posting and sharing activity, we aim for an algorithm that computes the ψ-score for all nodes as fast as PageRank.

Power-ψ algorithm

What we have accomplished is a reduction of the aforementioned N systems into a single one:

$$\boldsymbol{\psi}^T = \frac{1}{N}\left[\left(\sum_{t=0}^{\infty}\mathbf{c}^T\mathbf{A}^t\right)\mathbf{B} + \mathbf{d}^T\right] \qquad (4)$$

where,

- $\boldsymbol{\psi}$ is the ψ-score vector
- $\mathbf{B}$ is the square matrix in which, for all i, the column i is the vector $\mathbf{b}_i$
- $\mathbf{c}$ is the vector representing the diagonal of $\mathbf{C}$
- $\mathbf{d}$ is the sum of all $\mathbf{d}_i$ vectors

Algorithm 1: `Power-ψ`: Power iteration based algorithm for the ψ-score vector.

input : N number of users, $N \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$, two vectors $\mathbf{c}$ and $\mathbf{d}$, s-tolerance ε
output: vector $\boldsymbol{\psi}$ with the ψ-score of all users
$\mathbf{s} \leftarrow \mathbf{c}$;
$B\_norm \leftarrow \|\mathbf{B}\|$;
$t \leftarrow 0$;
$gap \leftarrow 1$;
while $(gap > \varepsilon)$ **do**
  $\mathbf{s}_{old} \leftarrow \mathbf{s}$;
  $\mathbf{s}^T \leftarrow \mathbf{s}_{old}^T\mathbf{A} + \mathbf{c}$;
  $gap \leftarrow B\_norm\,\|\mathbf{s}_{old} - \mathbf{s}\|$;
  $t \leftarrow t + 1$;
end
$\boldsymbol{\psi}^T \leftarrow \frac{1}{N}\left(\mathbf{s}^T\mathbf{B} + \mathbf{d}^T\right)$;
return $\boldsymbol{\psi}$;

Experiments

Two types of experiments:

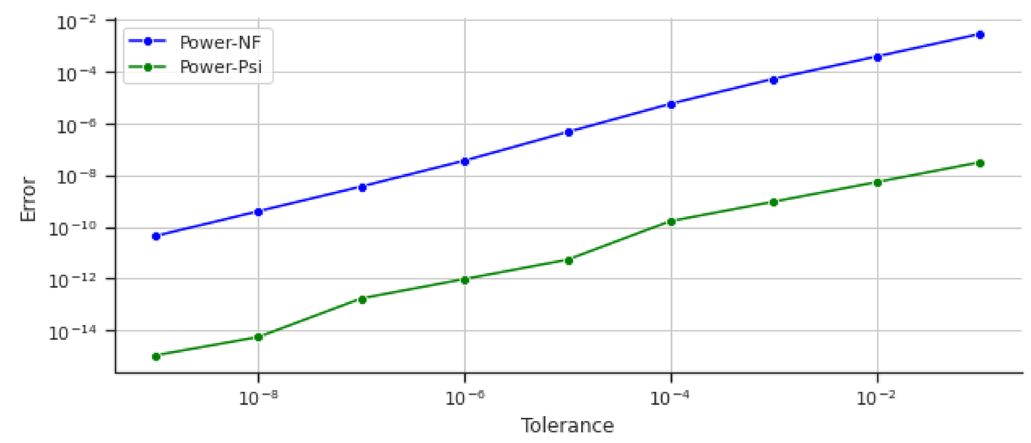
- Precision assessment for a given tolerance criterion
- Performance assessment for a measured error

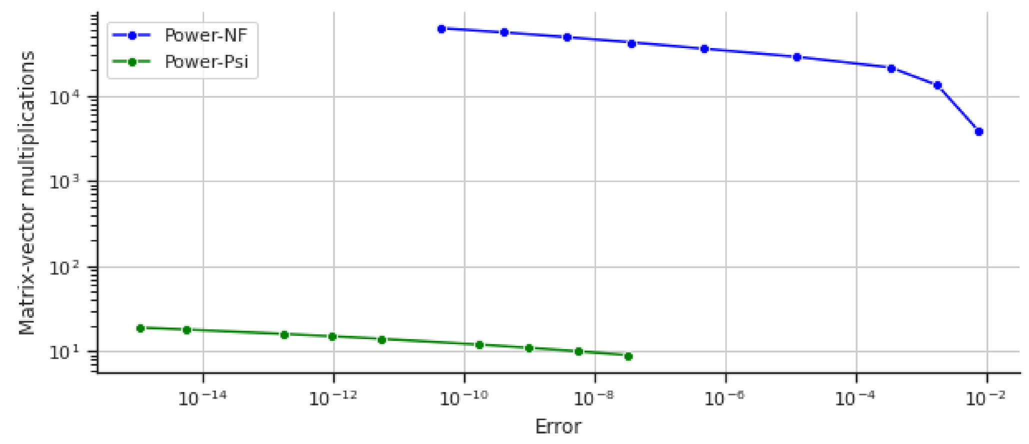
Two scenarios:

(i) heterogeneous activity scenario: users do not necessarily have the same posting and re-posting activity

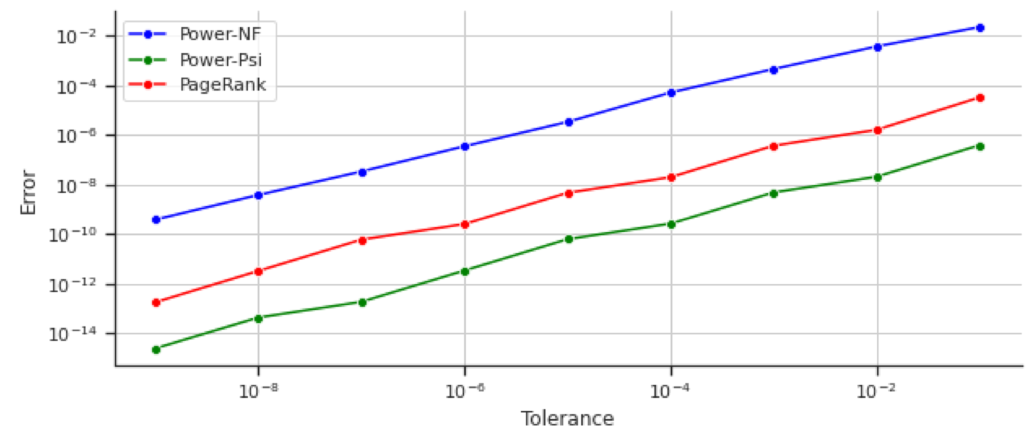
(ii) homogeneous activity scenario: all users have the same activity (i.e. the same λ and μ); in this case the ψ-score is exactly PageRank with $\alpha = \frac{\mu}{\lambda+\mu}$

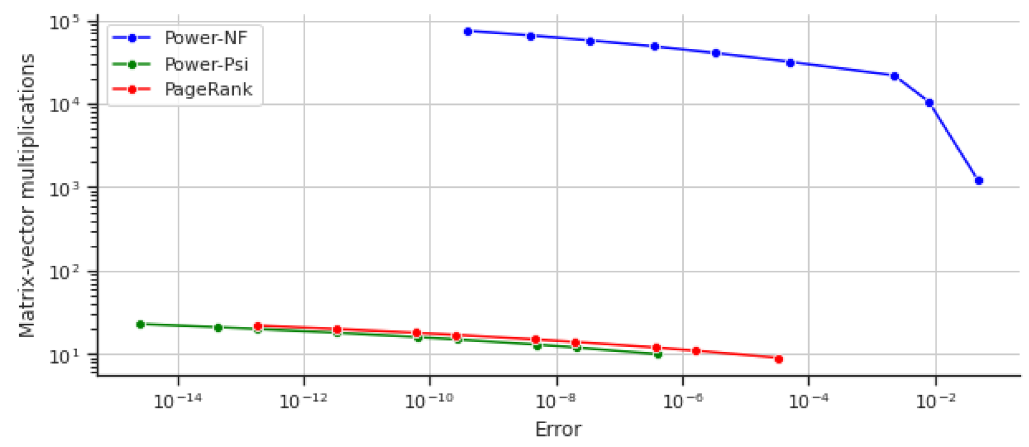
Heterogeneous activity:





Homogeneous activity:





Conclusion

The proposed `Power-ψ` algorithm outperforms the previous one and is equivalent to PageRank in terms of speed.

In the future, we plan to investigate other approaches such as the Monte Carlo methods which are widely employed in statistical prediction.

References

[1] N. Arhachoui, E. Bautista, M. Danisch, and A. Giovanidis, "A Fast Algorithm for Ranking Users by their Influence in Online Social Platforms," in *The 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Istanbul, Turkey, Nov. 2022. [Online]. Available: `https://hal.archives-ouvertes.fr/hal-03700079`.

[2] A. Giovanidis, B. Baynat, C. Magnien, and A. Vendeville, "Ranking online social users by their influence," *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 2198–2214, 2021. DOI: `10.1109/TNET.2021.3085201`. [Online]. Available: `https://doi.org/10.1109/TNET.2021.3085201`.

Contact: {nouamane.arhachoui, esteban.bautista-ruiz, lionel.tabourier, anastasios.giovanidis}@lip6.fr