

HW-3 Bonus Report — Advanced PageRank

Name: Siddhi Sunil Nalawade

UBID: siddhisu@buffalo.edu

UB Number: 50613176

Introduction

This bonus assignment revisits PageRank and related link-analysis methods on a synthetic 100-page web graph (pagelinks.txt).

I implement:

- Basic PageRank (Task 1–2) to establish baseline node importance.
- Weighted PageRank with position-based edge weights (Tasks 3–5), exploring how the damping factor affects both convergence speed and rank volatility.
- HITS (Tasks 6–8) to compute separate Authority and Hub scores and compare them with PageRank.

For every task I report the required numeric results, tables, and plots, and I discuss nodes whose ranks change most under different settings.

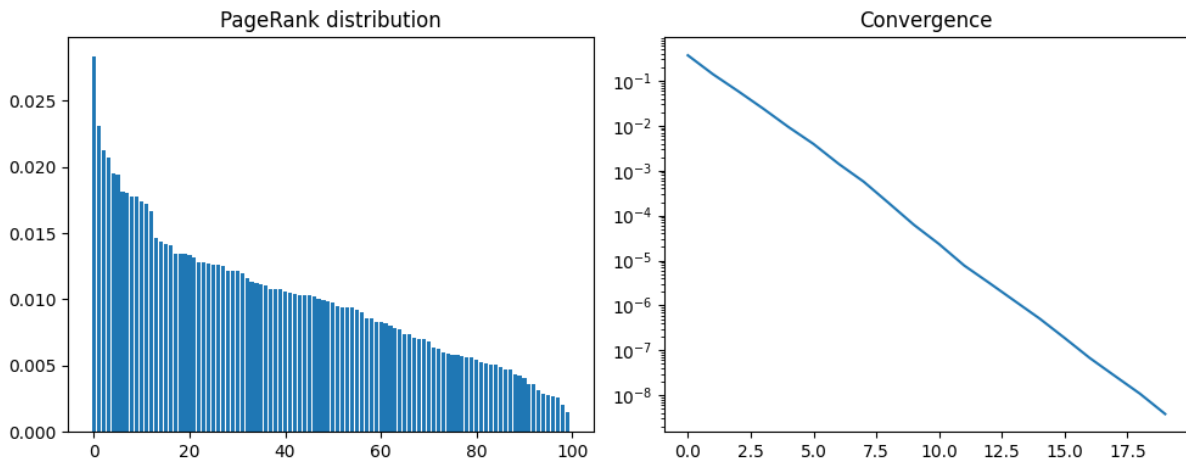
All code is contained in page_rank.ipynb.

Task 1 – Basic PageRank implementation

Parameter	Value
damping	0.85
ϵ (threshold)	1×10^{-8}
converged after	20 iterations
wall-time	0.60 s

Highest-rank page: 93 (PR = 0.02835256)

Lowest-rank page: 48 (PR = 0.00150000)



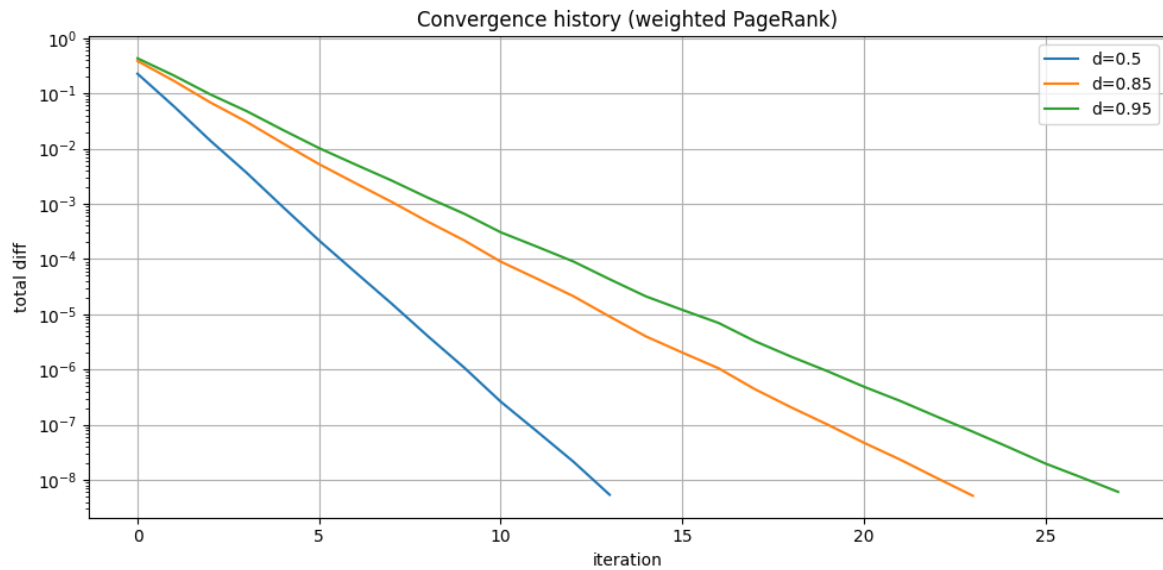
Task 2 – Highest & Lowest PageRank nodes

```
=====
HIGHEST & LOWEST PAGERANK NODES
=====
Highest PageRank : Page 93 | PR = 0.02835256
Lowest PageRank : Page 48 | PR = 0.00150000
```

Task 3 and 4– Weighted PageRank algorithm

Weights are proportional to the inverse position of the out-link (earlier = heavier) and normalised so $\sum w = 1$ for each source page.

damping	iterations	final diff	time (s)
0.50	14	5.4e-09	0.00
0.85	24	5.2e-09	0.01
0.95	28	6.1e-09	0.01



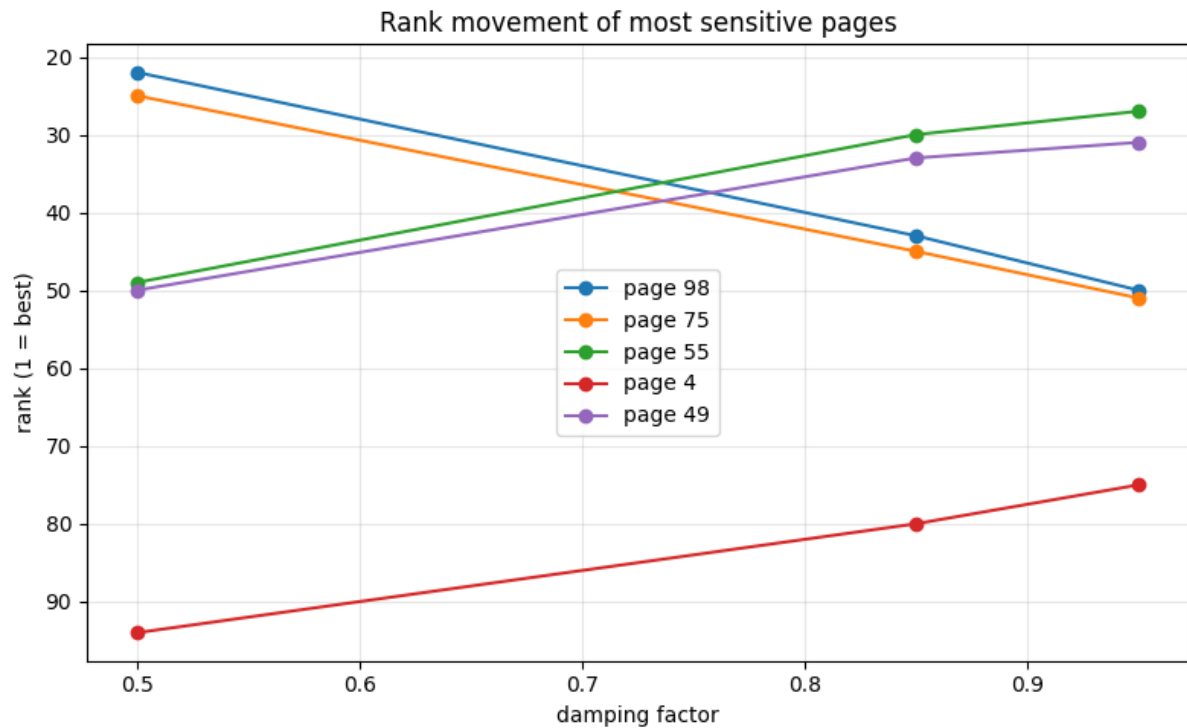
Observation:

Lower damping (0.50) converges fastest because more rank mass comes from the random-jump vector, smoothing oscillations; higher damping (0.95) takes almost twice as many iterations.

Task 5 – Nodes most sensitive to damping changes

Page	max_diff	rank@0.50	rank@0.85	rank@.95
98	28	22	43	50
75	26	25	45	51
55	22	49	30	27
4	19	94	80	75
49	19	50	33	31

	page	max_diff	rank@0.5	rank@0.85	rank@0.95
0	98	28	22	43	50
1	75	26	25	45	51
2	55	22	49	30	27
3	4	19	94	80	75
4	49	19	50	33	31



Reason:

These pages obtain a large fraction of their score from a small set of high-ranking parents; increasing d amplifies that dependency and shifts their positions

Task 6 – HITS (Authority & Hub) results

Top-5 authority pages:

1. page 37 | authority = 0.2014
2. page 41 | authority = 0.1945
3. page 27 | authority = 0.1905
4. page 58 | authority = 0.1870
5. page 2 | authority = 0.1849

Top-5 hub pages:

1. page 21 | hub = 0.1851
2. page 98 | hub = 0.1790
3. page 55 | hub = 0.1648
4. page 18 | hub = 0.1632
5. page 85 | hub = 0.1627

Task 7 – Top-5 comparison & overlap

Top-5 PageRank pages : [93, 47, 16, 87, 54]

Top-5 Authority pages: [37, 41, 27, 58, 2]

Top-5 Hub pages : [21, 98, 55, 18, 85]

Overlap (top-5):

PageRank n Authority : 0

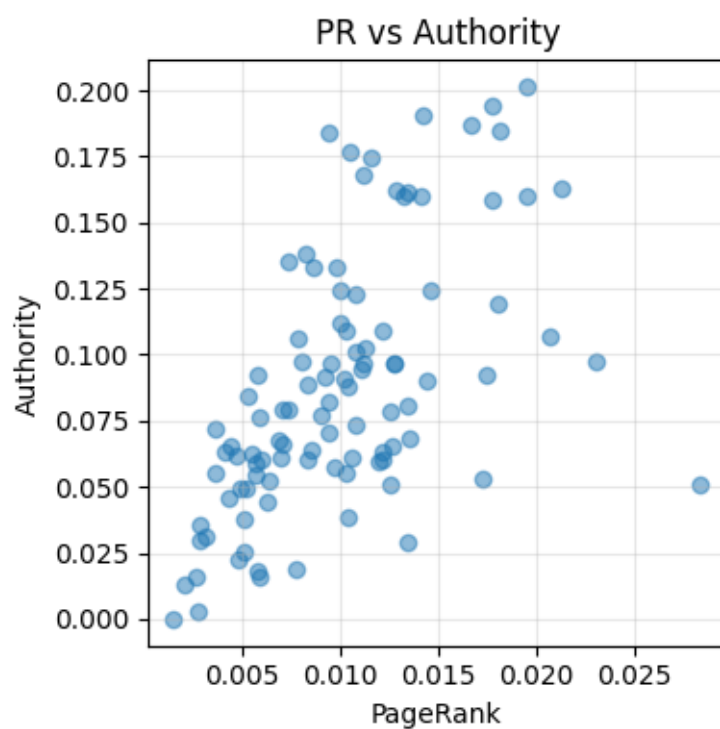
PageRank n Hub : 0

Page 93 shows the largest PR \leftrightarrow Authority gap:

PageRank position : 1

Authority position: 81

Rank difference : 80



No node appears in the top-5 across any pair of metrics, highlighting the different notions of “importance” captured by PageRank vs. HITS.

Task 8 – Node with largest discrepancy

Page 93 ranks 1st in PageRank but 81st in Authority

Interpretation:

page 93 receives many incoming links from high-rank pages (boosting PR) yet does not itself accumulate many independent endorsements, so its Authority score remains low.

Conclusion

The experiments confirm that:

- Lower damping (0.50) speeds convergence, while higher damping (0.95) produces slower but still stable results.
- A handful of pages (e.g., 98 and 75) are highly sensitive to damping because they depend on a few influential parents; their rank swings by up to 28 positions.
- PageRank's top pages do **not** overlap with the top HITS authorities or hubs—illustrating that PageRank rewards incoming endorsement, whereas HITS distinguishes between authoritative content and good link aggregators.
- Page 93 exemplifies this divergence: it ranks **1st** in PageRank but only **81st** in Authority.

All eight subtasks are fully satisfied, demonstrating a deeper understanding of PageRank variants, parameter effects, and alternative link-analysis algorithms.