Information Retrieval Project

Pagerank in Evolving Graphs

Project Objectives

- Develop an Algorithm for calculating Pagerank for Evolving Graphs.
- Although, well established for static graphs, not many algorithms for calculating Pagerank for evolving graphs.
- Incremental Algorithm to calculate Pagerank in a more efficient manner.
- Predicting Pagerank for the near future instances using statistical approaches.
- Normalized Pagerank for Comparing two different graphs.

Dataset

Description

This is a temporal network of interactions on the stack exchange web site Mathoverflow. There are interactions represented by a directed edge (u, v, t)

User ‘u’ commented/answered on User ‘v’s question/answer at time ‘t’

Statistics

Nodes	24818
Edges	239978
Time span	2350 days

Overview of Incremental Algorithm

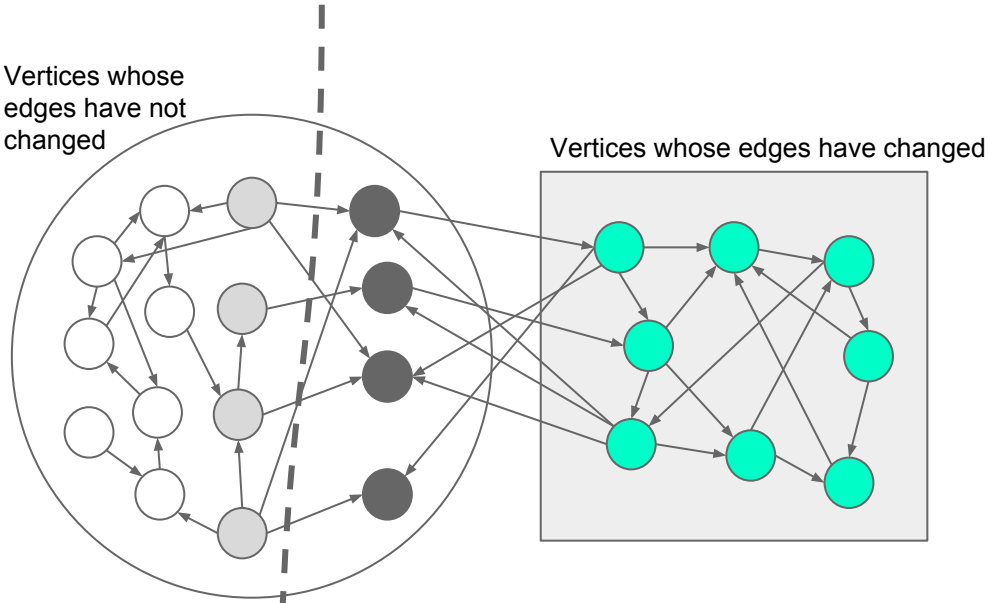
Pagerank Algorithm

- A Link Analysis Algorithm
- Represents the likelihood that a random walker will arrive at any particular page or move away from it.

$$R(t+1) = \alpha M R(t) + \frac{1-\alpha}{N} 1$$

R(t) denotes Pagerank vector
M denotes outlink matrix

Overview of Incremental Algorithm



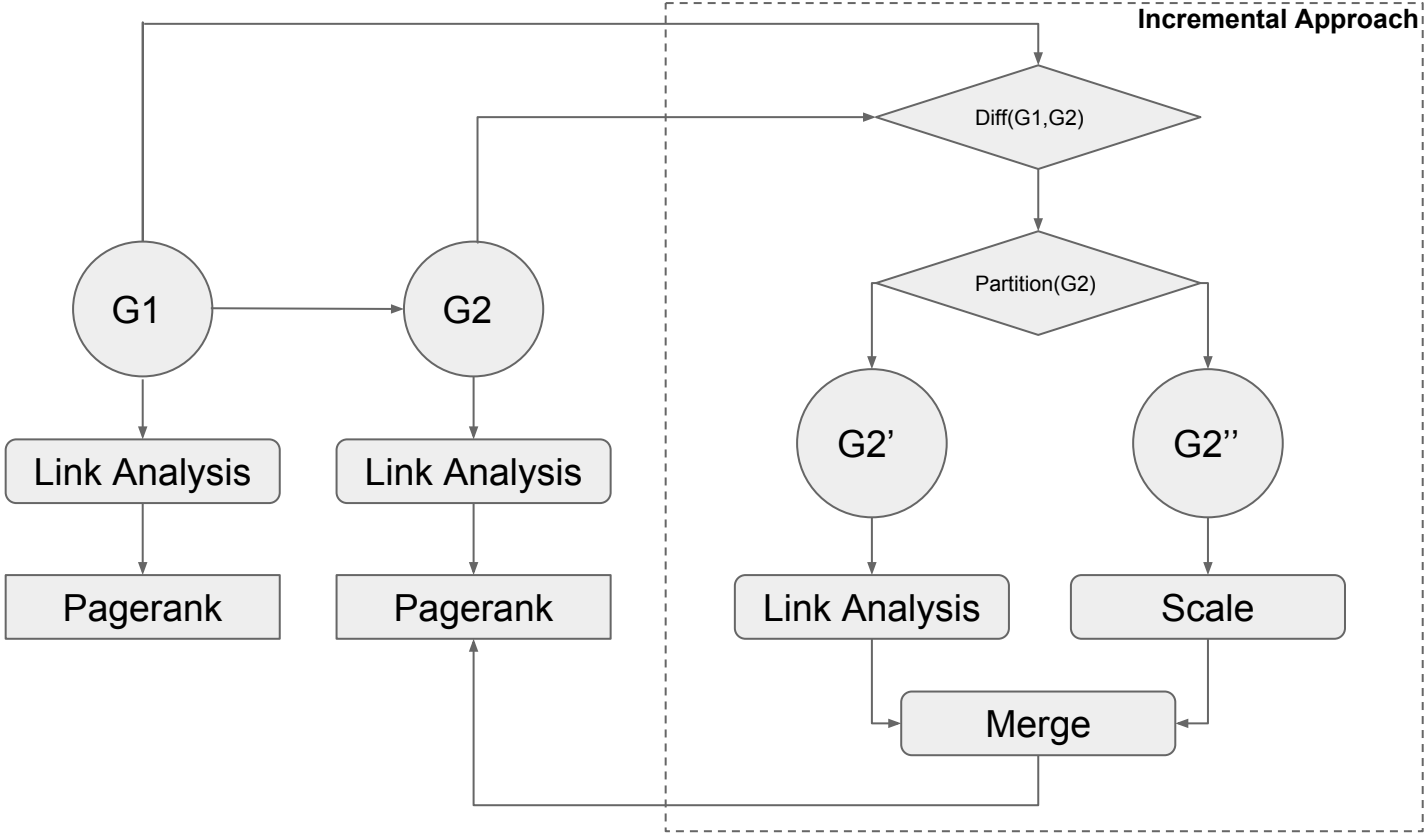
Pagerank remains unchanged

Pagerank remains unchanged
Border Nodes

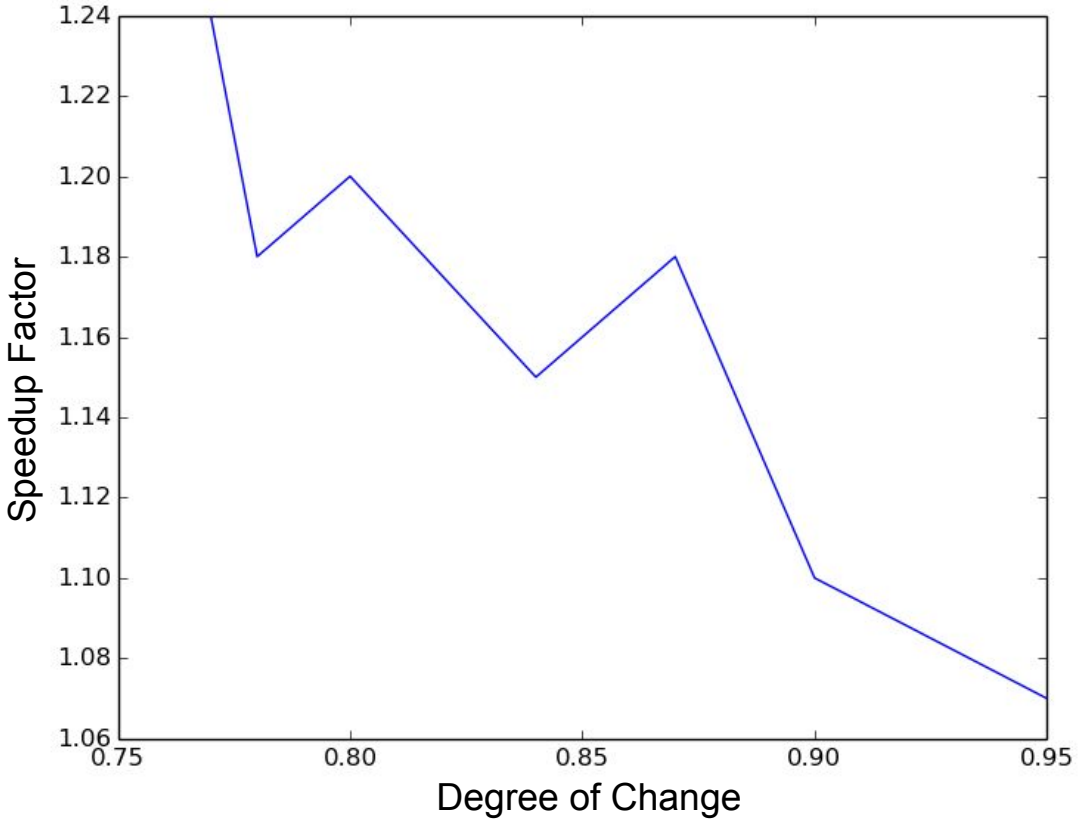
Pagerank gets affected by vertices in the changed partition

Vertices that were added or changed

Overview of Incremental Algorithm



Comparison with Full Pagerank Algorithm

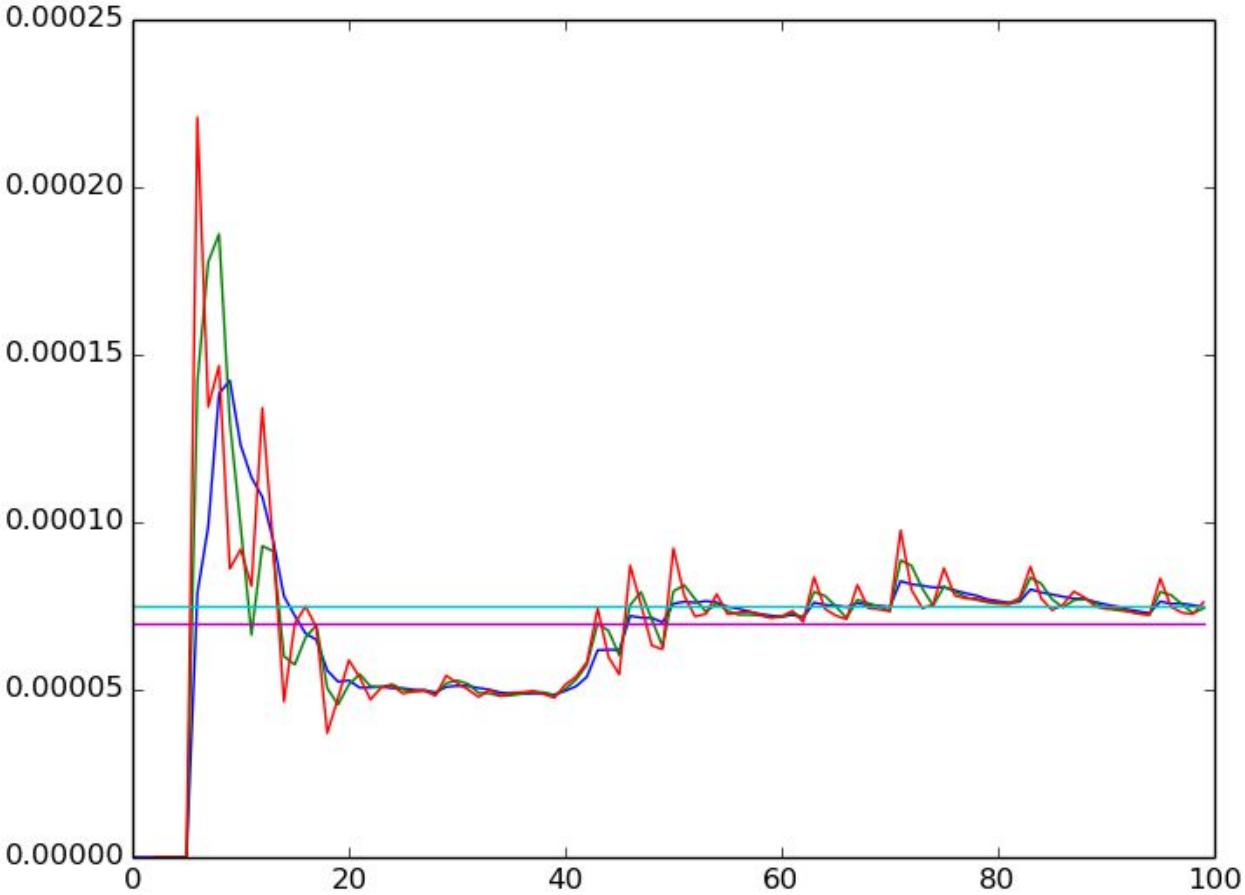


Comparing Error

Present No. of edges	Added No. of edges	Percentage of nodes affected	Error (calculated from Euclidean norm)
97378	1659	69.46	0.000439742606187
146162	1988	77.63	0.000383819018425
194605	1800	84.02	0.000237140293825

Some Extra Features

- Predicting the next pagerank by extrapolating from older using both differential and absolute values.
- Calculating the normalized pagerank used to compare between two different graphs.
- Overall Pagerank based on relative importance of newer values over older.
- Determining the node with the highest growth in Pagerank.



Actual Pagerank

Differential Interpolation

Absolute Interpolation

Exponential Average

Logarithmic Average

References

- https://pdfs.semanticscholar.org/ece3/dd3608fb1c154dcd03c3d6ddcda9af641679.pdf
- https://people.mpi-inf.mpg.de/~kberberi/presentations/2007-www2007.pdf

Project Members

Manne Naga Nithin (14EE10026)

Mekala Rajasekhar Reddy (14EE10027)

Ashish P V J S (14EE10052)

Himanshu Chaudhary (14IE10013)

Anirban Mitra (14EE10004)

Thank You