



भारतीय सूचना प्रौद्योगिकी संस्थान धारवाड़

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

Pseudo-Cores

Name: Somyadeep Shrivastava (17BCS028)

Dheeraj Chaudhary (17BCS009)

Under the Guidance of: Dr Yayati Gupta

OBJECTIVE

- To study the virality of a meme by artificially inducing it
- View the network as K-shells
- Examine the cascading power of core node and pseudo cores. Henceforth their importance in virality of a meme
- Check whether pseudo cores are alternative to core nodes
- Examine pathfinding algos targeting core/pseudo cores
- Visualize Leakage power of shells

Importance

- Study of virality of a meme helps scientists to address problems pertaining to disciplines like epidemiology and digital marketing.
- Though there are many memes out there but only some get popular
- Core nodes play an important role in virality of a meme as they possess maximum spreading power and high reachability
- If we somehow find the path to core nodes, we can make a meme viral
- Even the pseudo cores possess the same characteristics as core and it's easier to reach pseudocore than core nodes.
- Key algorithms such as shell based hill climbing and degree based hill climbing have been used to find path to pseudocore/core.

DATASETS

- Social circles: Facebook
- This dataset consists of 'circles' (or 'friends lists') from Facebook. Facebook data was collected from survey participants using the Facebook app
- It has 4039 nodes and 88234 edges
- Link : https://snap.stanford.edu/data/facebook_combined.txt.gz

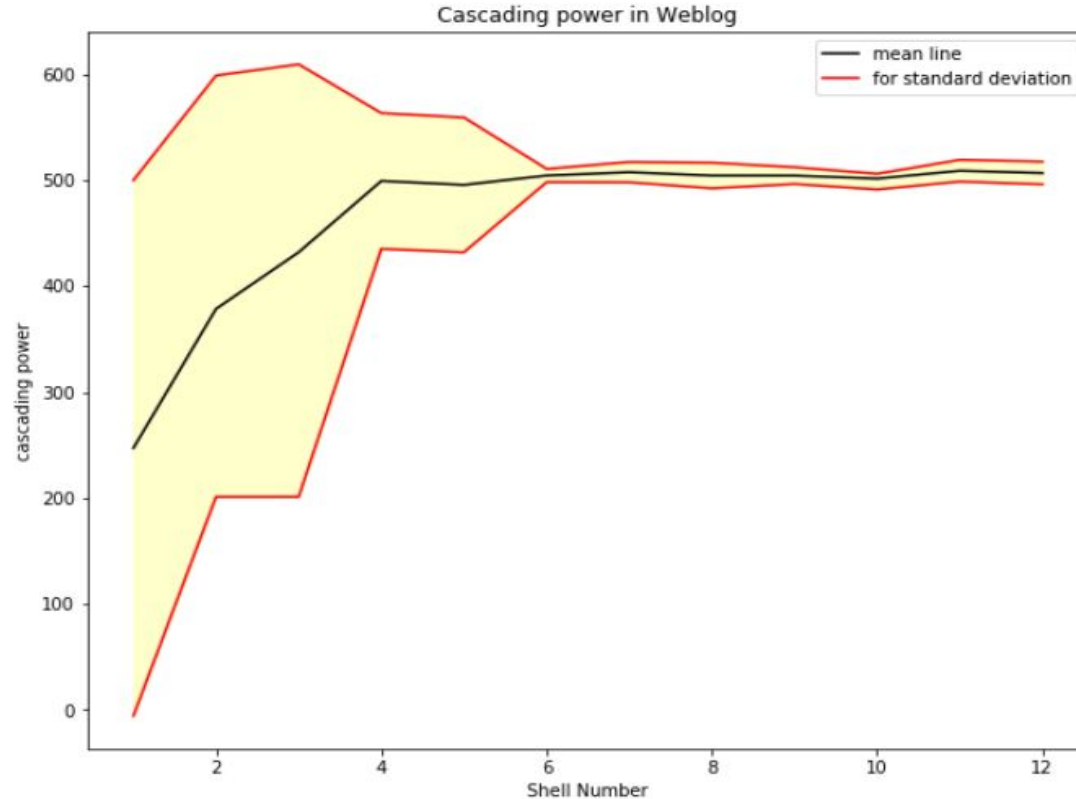
- Webgraph
- Webgraph of poll blogs
- It has 643 nodes and 2280 edges
- Link : <http://nrvis.com/download/data/web/web-polblogs.zip>

- Road Network
- Road Network of Minnesota which is a midwestern U.S. state
- It has 2642 vertices and 3303 edges
- Link : <http://nrvis.com/download/data/road/road-minnesota.zip>

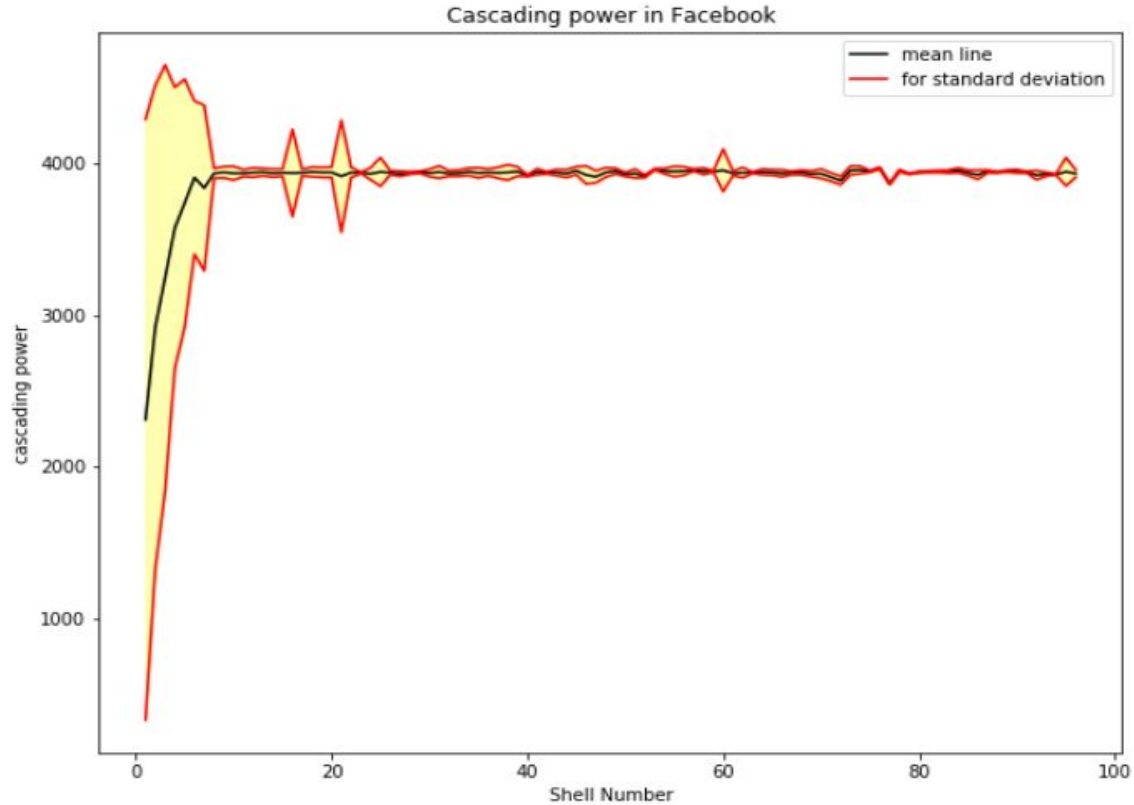
Analysis

- This section covers plots of cascading power, path traversal steps and leakage power.
- The road network has not been used to show pseudo core and leakage power , though it was a large network after applying K-shell decomposition , we obtained only two shells. Though it is used in showing path traversal for core node.
- The K- shell decomposition give 96 shells for FB data and 12 shells for webgraph.
- For each plot of hill climbing algos y axis has been taken as no. of initial nodes reached core nodes corresponding to number of steps which is the x axis.

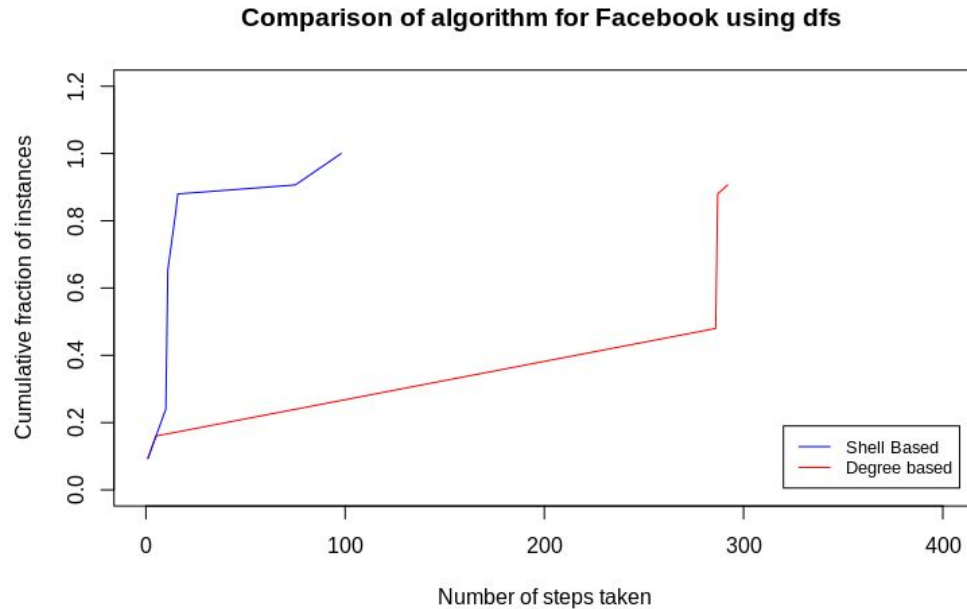
Cascading Behaviour in weblog



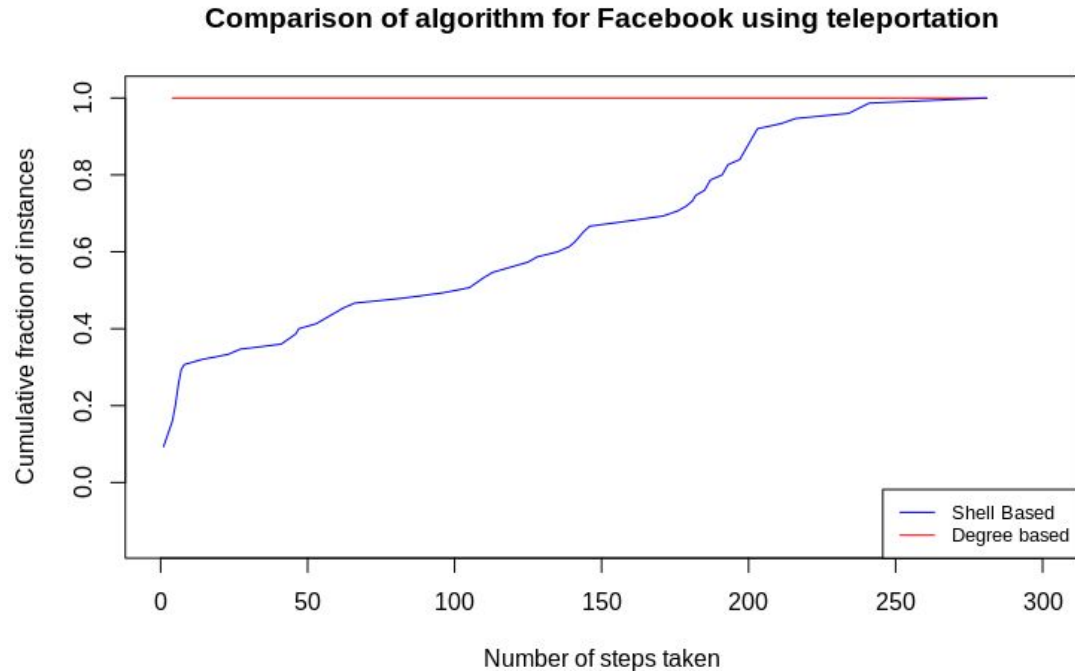
Cascading Behaviour in FB



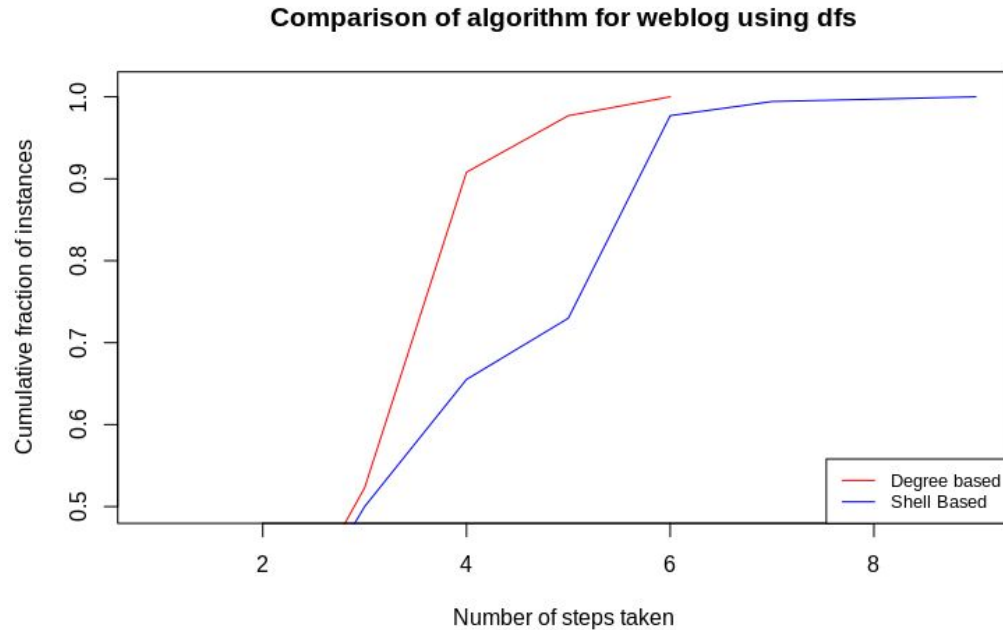
DFS based shell and shell degree based on FB data



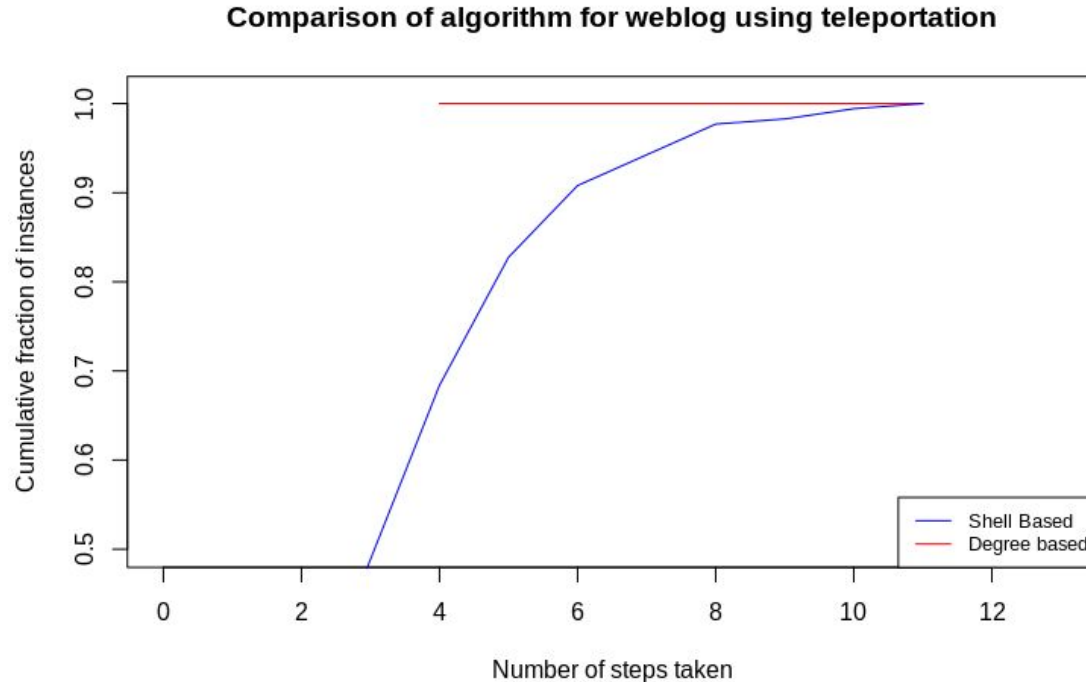
Teleportation based shell and shell degree based on FB data



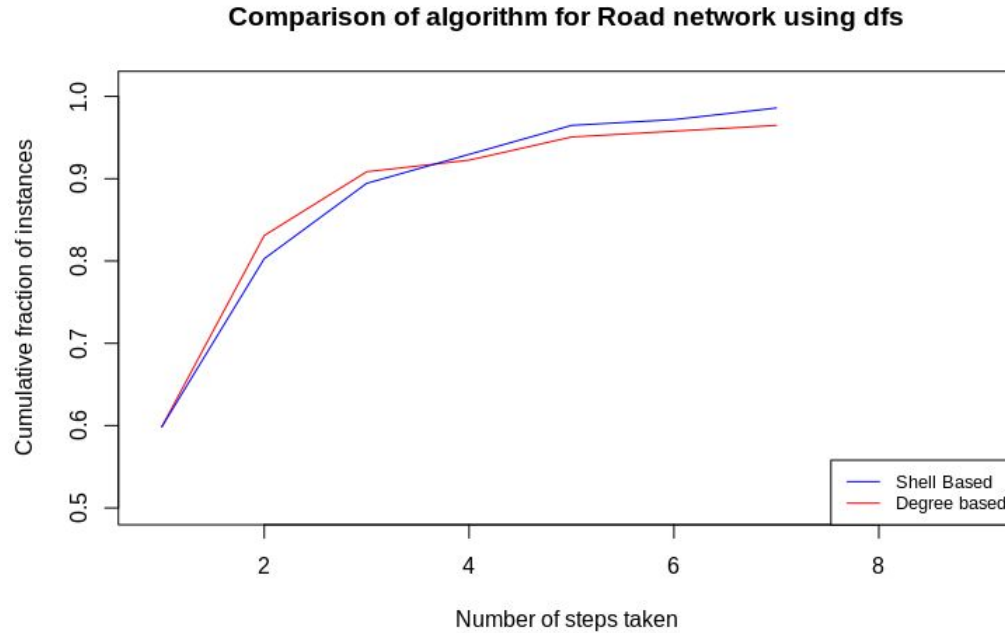
DFS based shell and shell degree based on weblog data



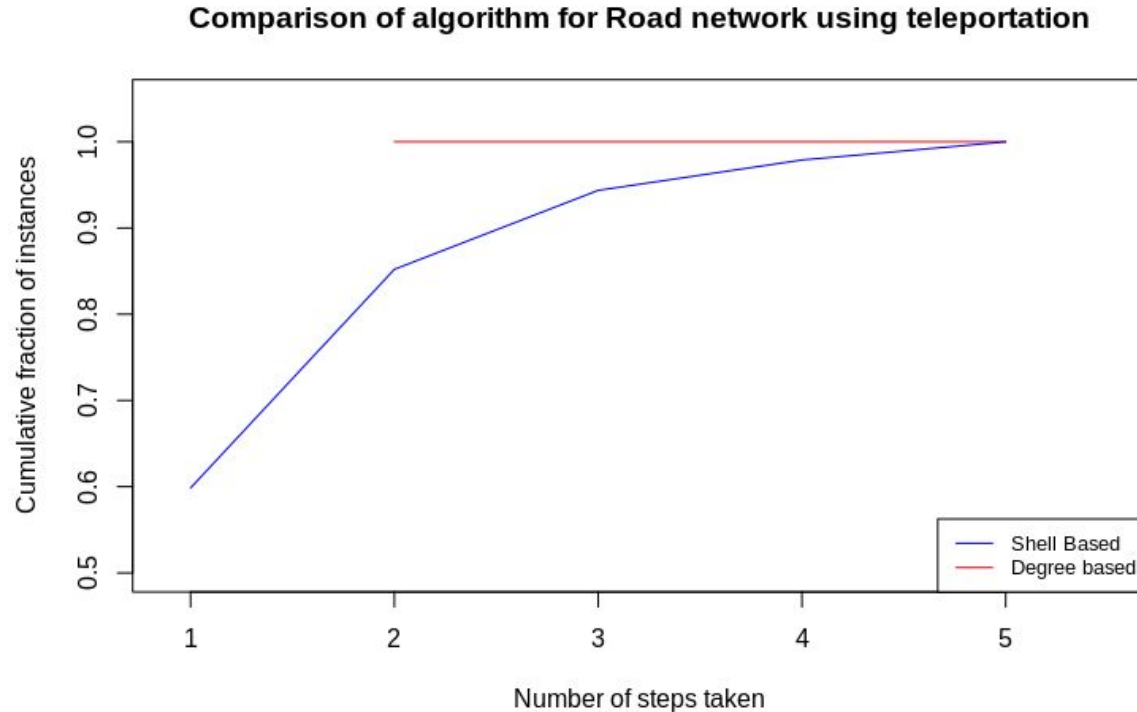
Teleportation based shell and shell degree based on weblog data



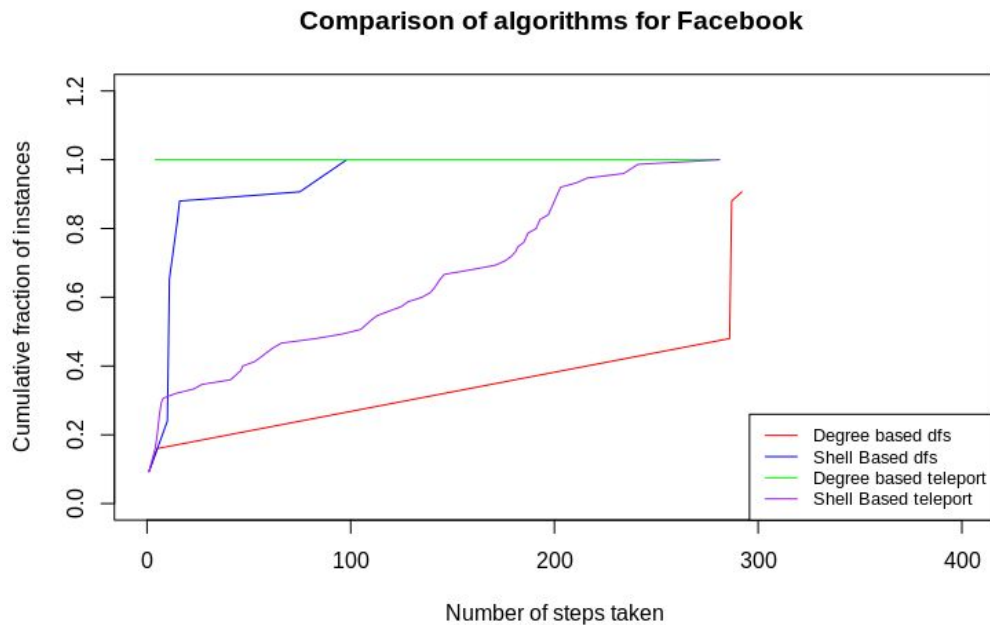
DFS based shell and shell degree based on road data



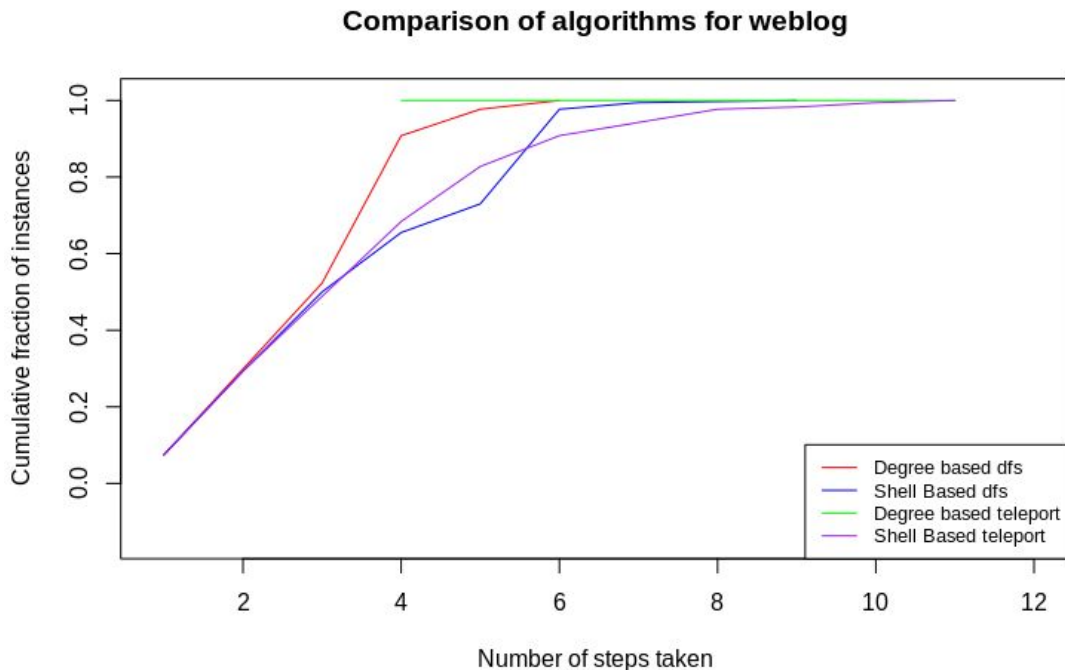
Teleportation based shell and shell degree based on weblog data



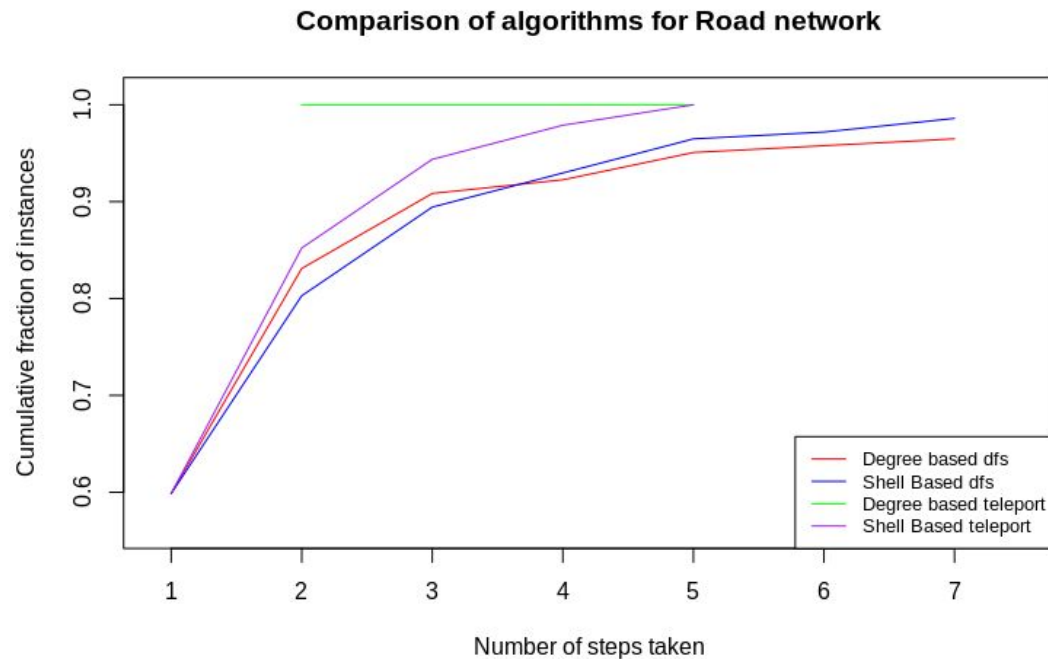
Comparison of various algo on FB DATA



Comparison of algo on weblog

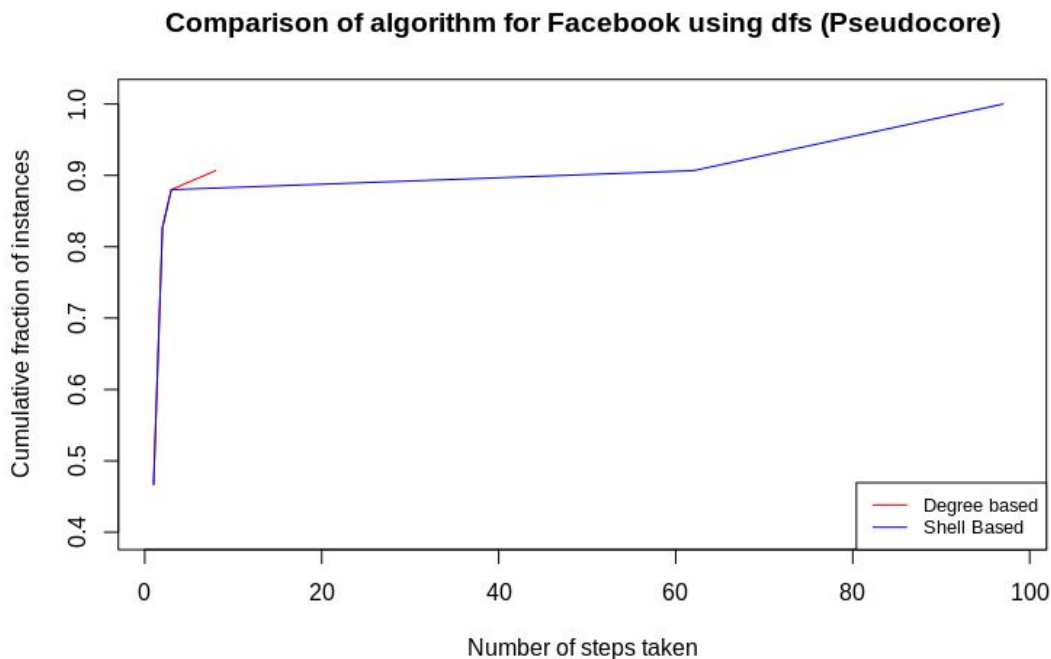


Comparison of algo on road network

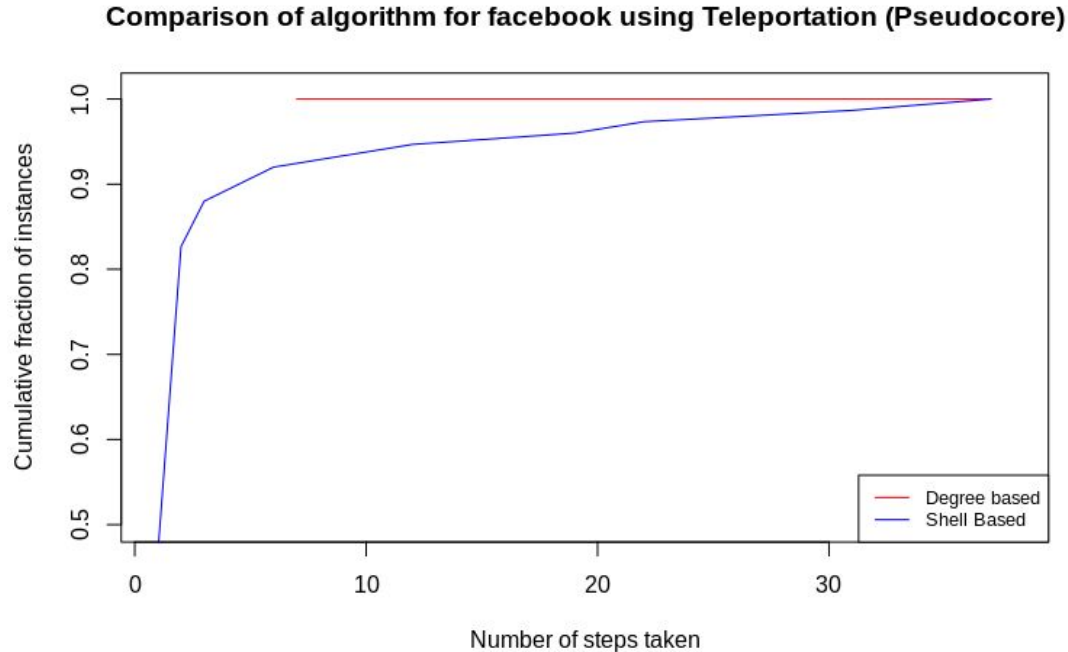


Pseudo Cores

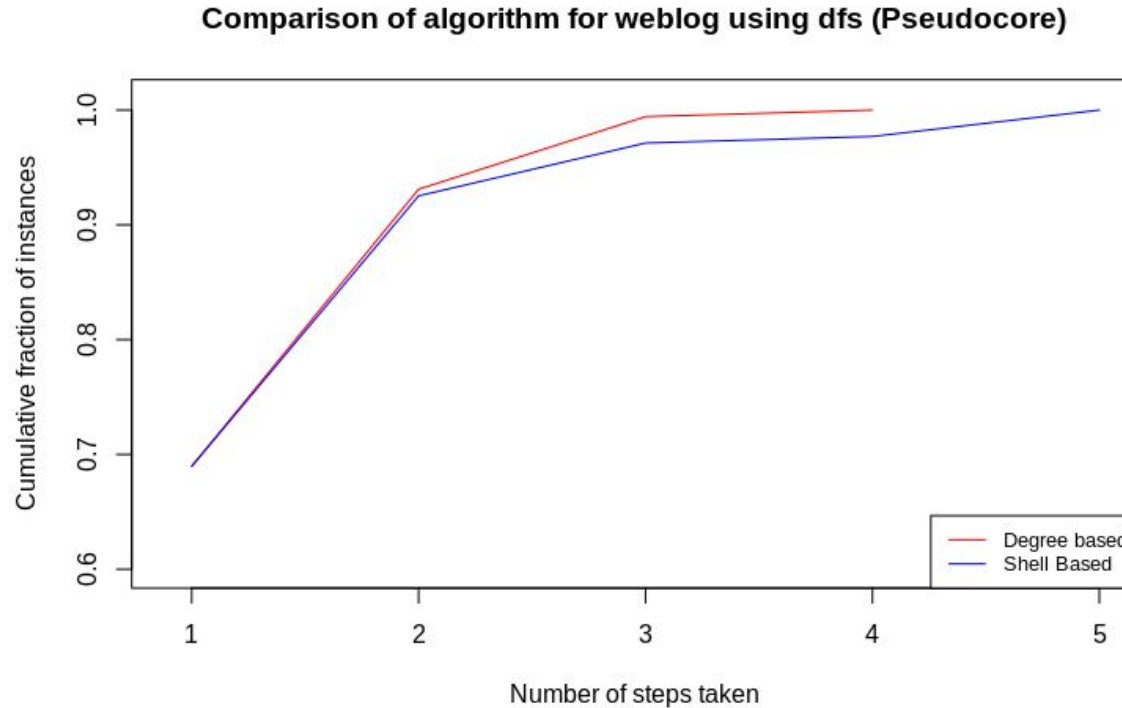
DFS based shell and shell degree based on FB data for pseudocore



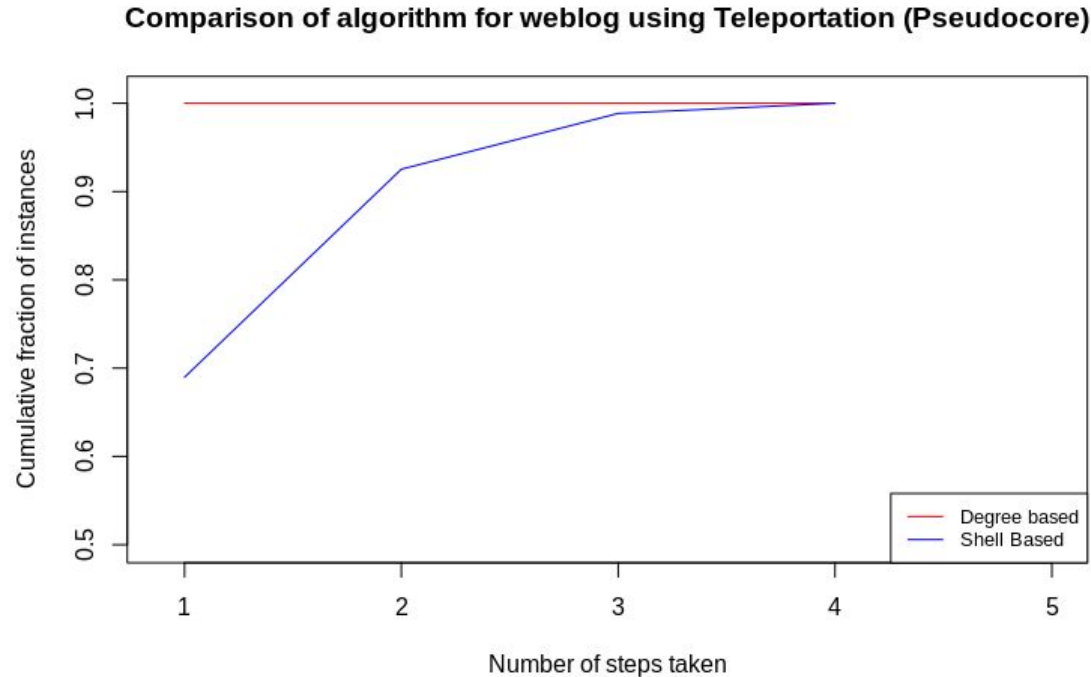
Teleportation based shell and shell degree based on FB data for Pseudocore



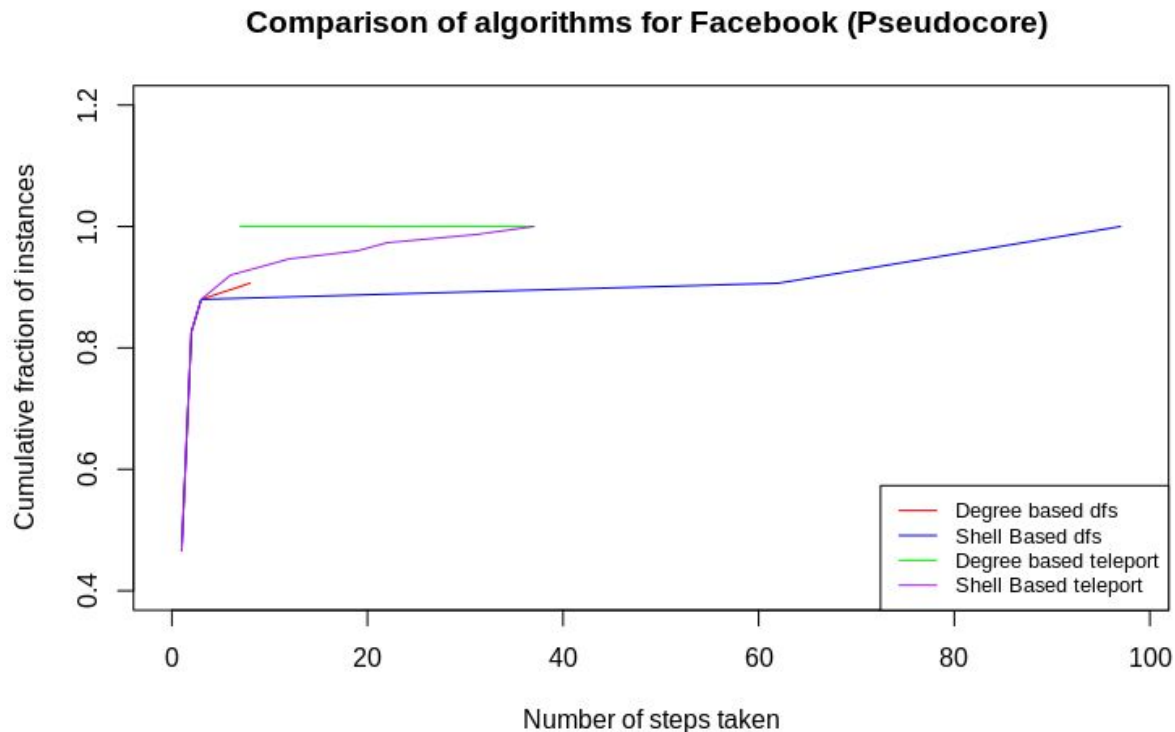
DFS based shell and shell degree based on Weblog data for pseudocore



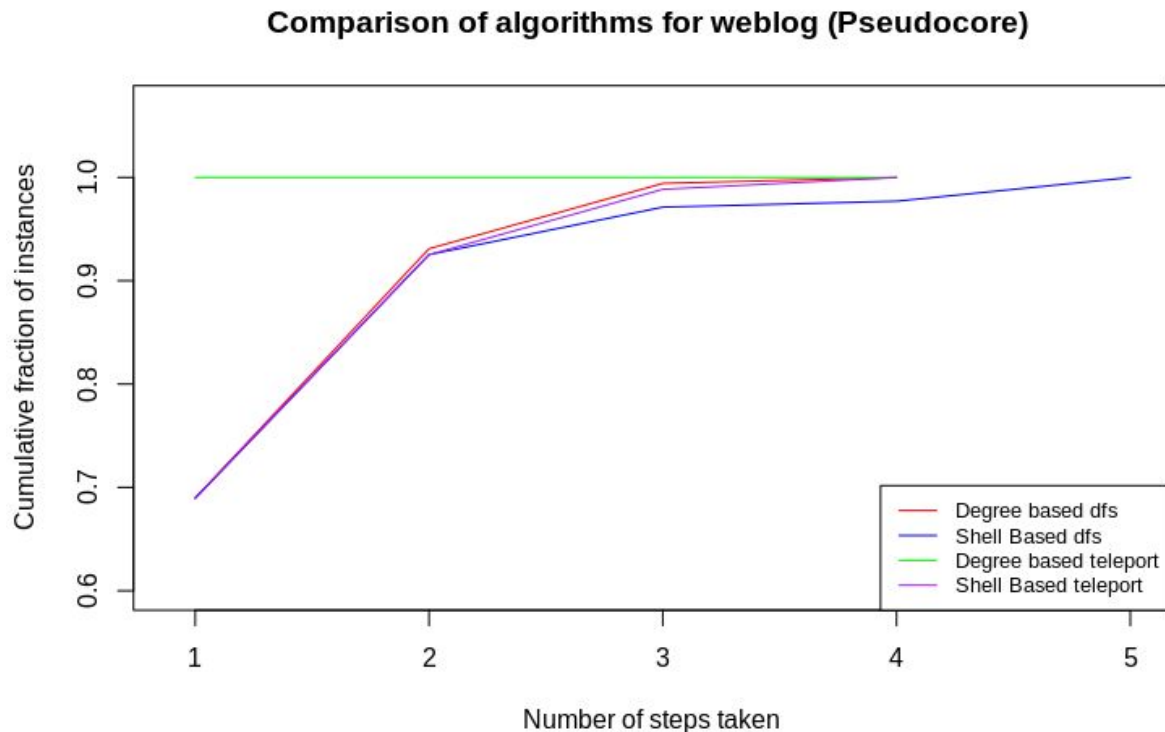
Teleportation based shell and shell degree based on FB data for Pseudocore



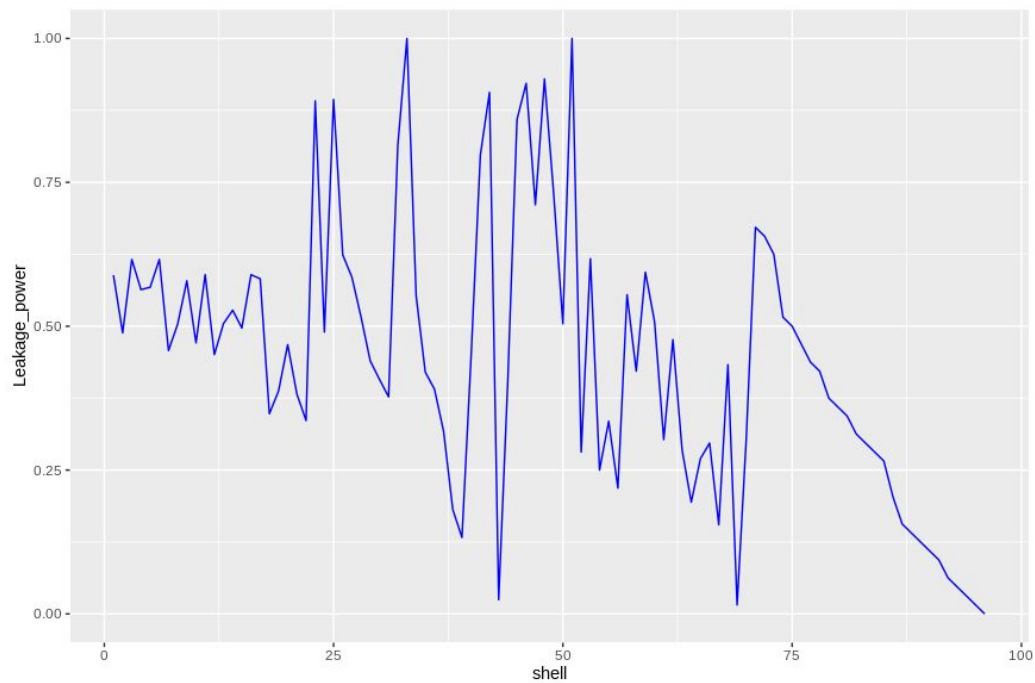
Comparison of algo on FB for pseudocore



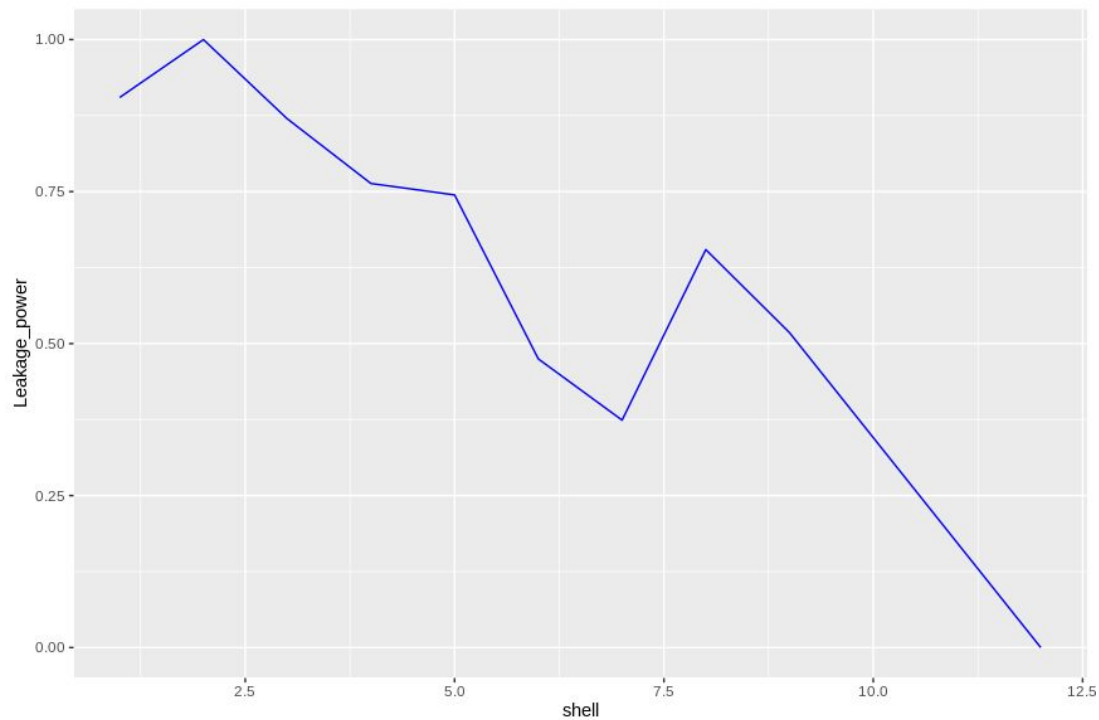
Comparison of algo on FB for pseudocore



Leakage Power of Facebook Network



Leakage Power of Webplot



Implementation and Variations

- Cascading power of core nodes has been calculated as per the paper and plotted apart from that plot showing standard deviation of power has also been plotted.
- Idea of shell based and shell degree based algorithm has been used to come up with own version of DFS and teleportation and corresponding plots have been shown
- Leakage power of shells for the network is plotted.
- Leaving the naive algos such as random walk(which was not part of core paper) and large dataset for which computation was unfeasible . Rest all part of the paper is implemented and discussed.

Conclusions and Future Work

- Reaching Pseudo cores was much faster than reaching core nodes
- There is always a trade-off between teleportation algos and DFS algos. As initial nodes always converge to target node in teleportation but can take a longer time in case of large network whereas in DFS based algos it might not take as much time as teleportation but it won't converge to target nodes in some cases.

- While running algos we also find out that there are few shells in which traversal going inside them takes a long time and can be stated as trapping shells. Which could also be seen in Leakage Power.
- We also observed that convergence of initial node to reach target nodes in DFS algos is somewhat dependent on number of nodes in last shell. If there are only two nodes in last shell, we end up not reaching the core node. While this is not a problem in pseudocores.

- Thus the need to implement hill climbing algos based on Leakage Power arise and like work in this area. So that we can get to target node much faster.
- Till now we were unable to find any suitable technique other than cascading power comparison for pseudocore identification. If in future we come across any other technique we would like to incorporate that.

Thank You