Spectral Analysis of real world networks

By :-

Anshuman Tripathi (07CS3024)

Gautam Kumar (07CS1021)

Parin Chheda (07CS3023)

(under guidance of Animesh Shrivastav and Prof Niloy Ganguly)

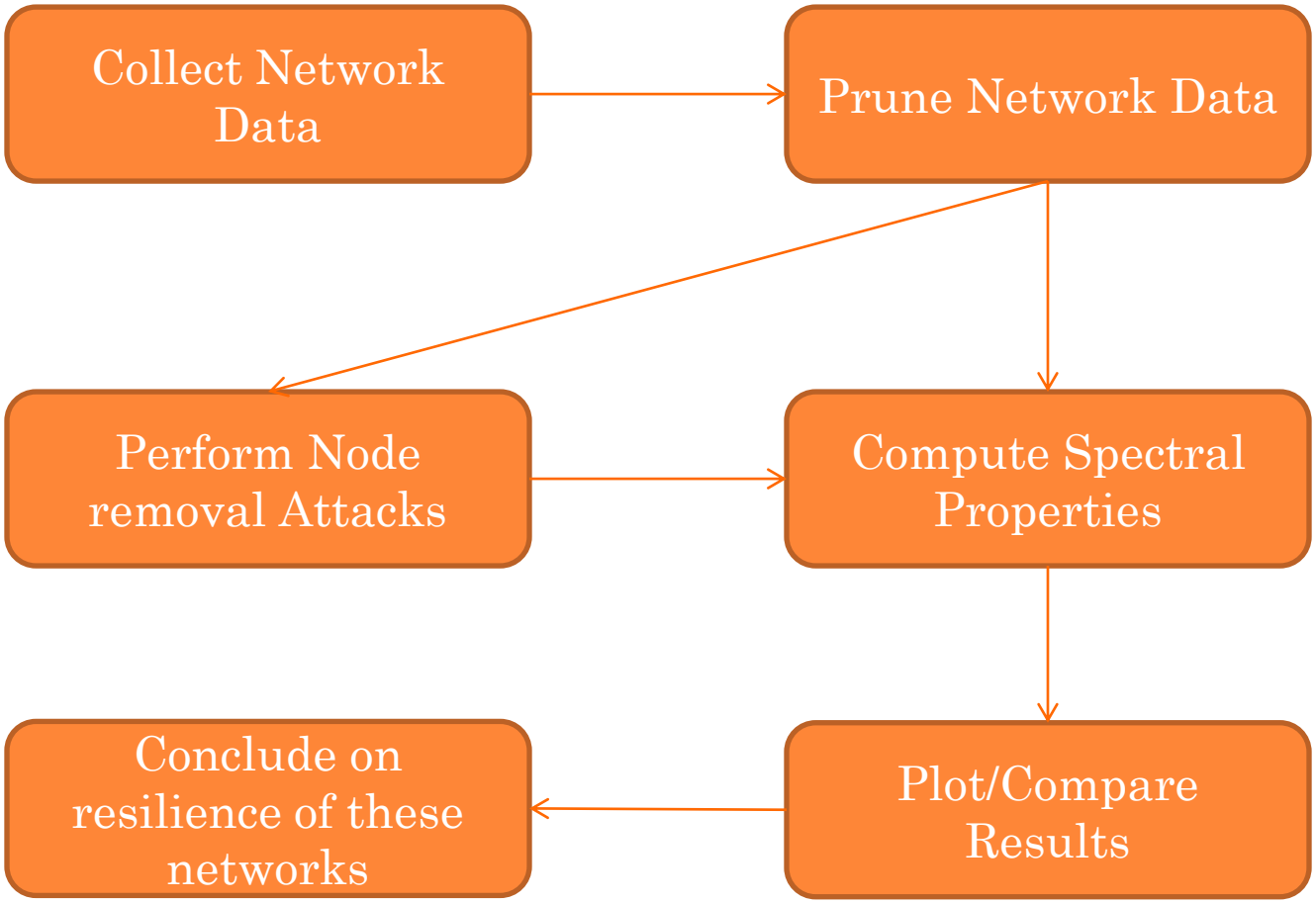
Project Goal

- Collect Network data from real world networks like World Wide Web, Facebook, Twitter … etc
- Compute spectral properties of the graphs
  - Laplace spectrum
  - Adjacency spectrum
  - Degree distribution
  - Assortativity … etc
- Study these spectral properties under certain type of network attacks to conclude resilience of these networks

Real – World networks

- Autonomous System Graph
  - Every AS router is viewed as a node in the graph
  - A trace route from a router to another router denotes an edge
- Facebook
  - Every individual is a node
  - Friendship denotes an undirected edge
- Twitter
  - Followers (who follow ‘x’) and Friends (who ‘x’ follows) define directed edges adjacent to ‘x’

Work Flow


Collect Network Data
Prune Network Data
Perform Node removal Attacks
Compute Spectral Properties
Conclude on resilience of these networks
Plot/Compare Results

Collecting Data (AS)

- The network data for AS router network was downloaded from http://snap.stanford.edu/data/as-skitter.html
- The Data organized in for of edge-list
- Undirected Graph

Statistics :-

Number of nodes (|V|)	1696415 ~ 1.7M
Number of Edges (|E|)	11095298 ~ 11.1M
Highest Degree	1008
Assortativity	0.04
Clustering Coef.	0.2963

Collecting Data (Facebook & Twitter)

- Designed python based crawlers
  - Facebook
    - Used cloudlight python module
    - The friend list dynamically fetched from Facebook server
    - Used mobile version of Facebook (http://m.facebook.com) to browse friends (10 friends per page)
    - Crawled ~2000 nodes in 3 days
  - Twitter
    - OAuth2 authentication
    - Used Twitter API for python (twython) (https://github.com/ryanmcgrath/twython)
    - Crawling limited by number of api-calls per hour from a client (350 calls/hour)
    - Crawled ~1900 nodes in 1 day

Collecting Data (Facebook & Twitter)

- Facebook data downloaded from
- Twitter data downloaded from

Statistics

	Facebook	**Twitter**
Number of nodes (|V|)	258912 ~ 2.5M	40103281 ~ 40M
Number of Edges (|E|)	60022032 ~ 60M	1468365182 ~ 1.5B
diameter	6.5	5.9

Pruning Of Networks

- Data collected too huge for performing spectral computations
- Entire data is not necessary for studying statistical properties
- Prune the data obtained w.r.t degree of node
- Selecting Threshold
  - Should conserve the degree distribution of the original network
  - Should reduce number of nodes to computationally feasible levels ~ 10K

Statistics of pruned networks

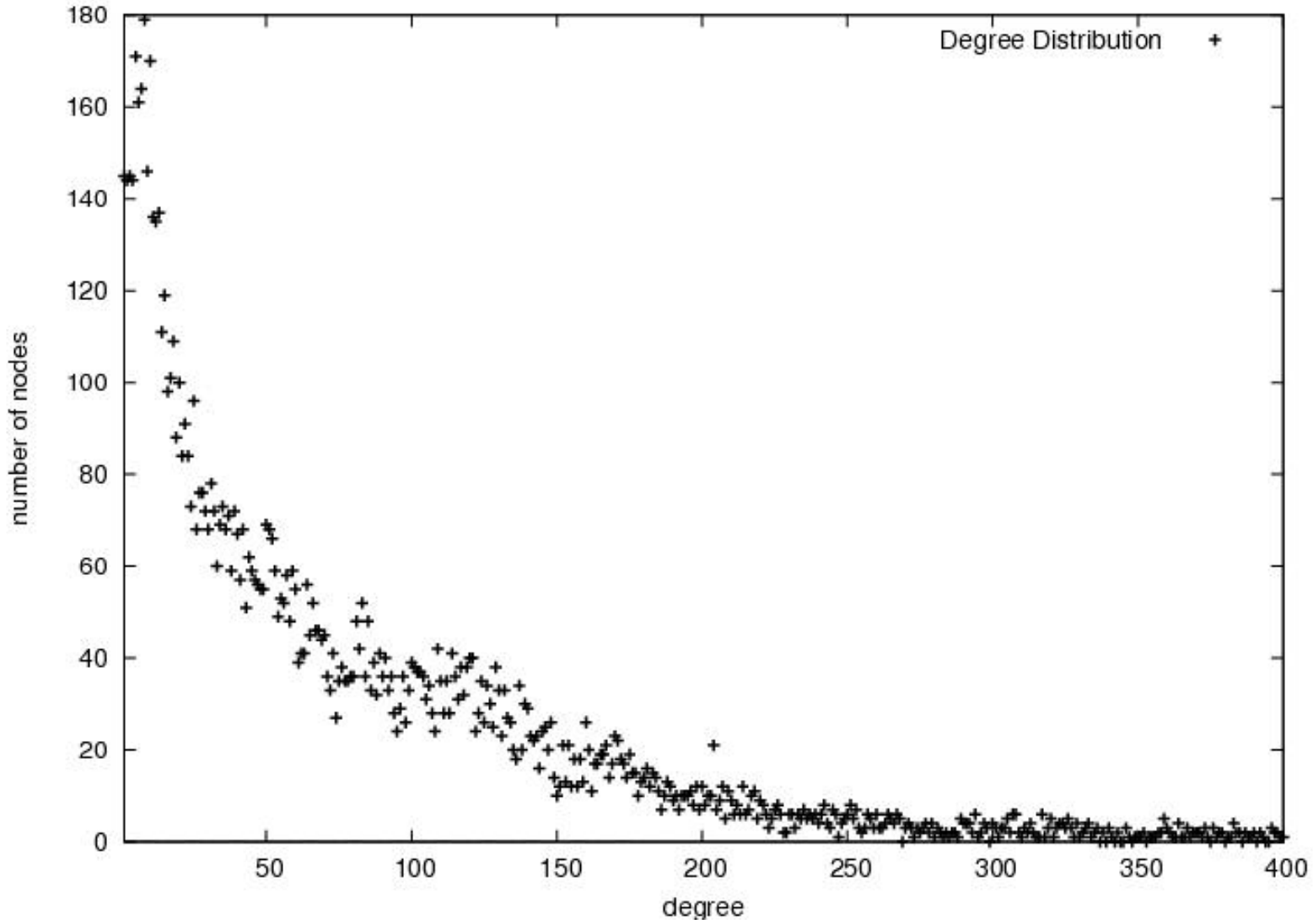
Metric	AS	Facebook	Twitter
Number of nodes	9881 ~ 10K	10707 ~ 10K	1030869 ~ 1M
Number of edges	403474 ~ 403K	328926 ~ 329K	55921630 ~ 55M
Threshold	>175	>800	>100 and < 500
Assortativity	0.0398	0.3589	N/A[2]
Clustering Coef.	0.3095	0.3143	N/A[2]
Diameter[1]	9	13	10
Size of Big Component	99.78%	99.75%	99.99%
Number of components	11	7	4

[1] Diameter of the big component

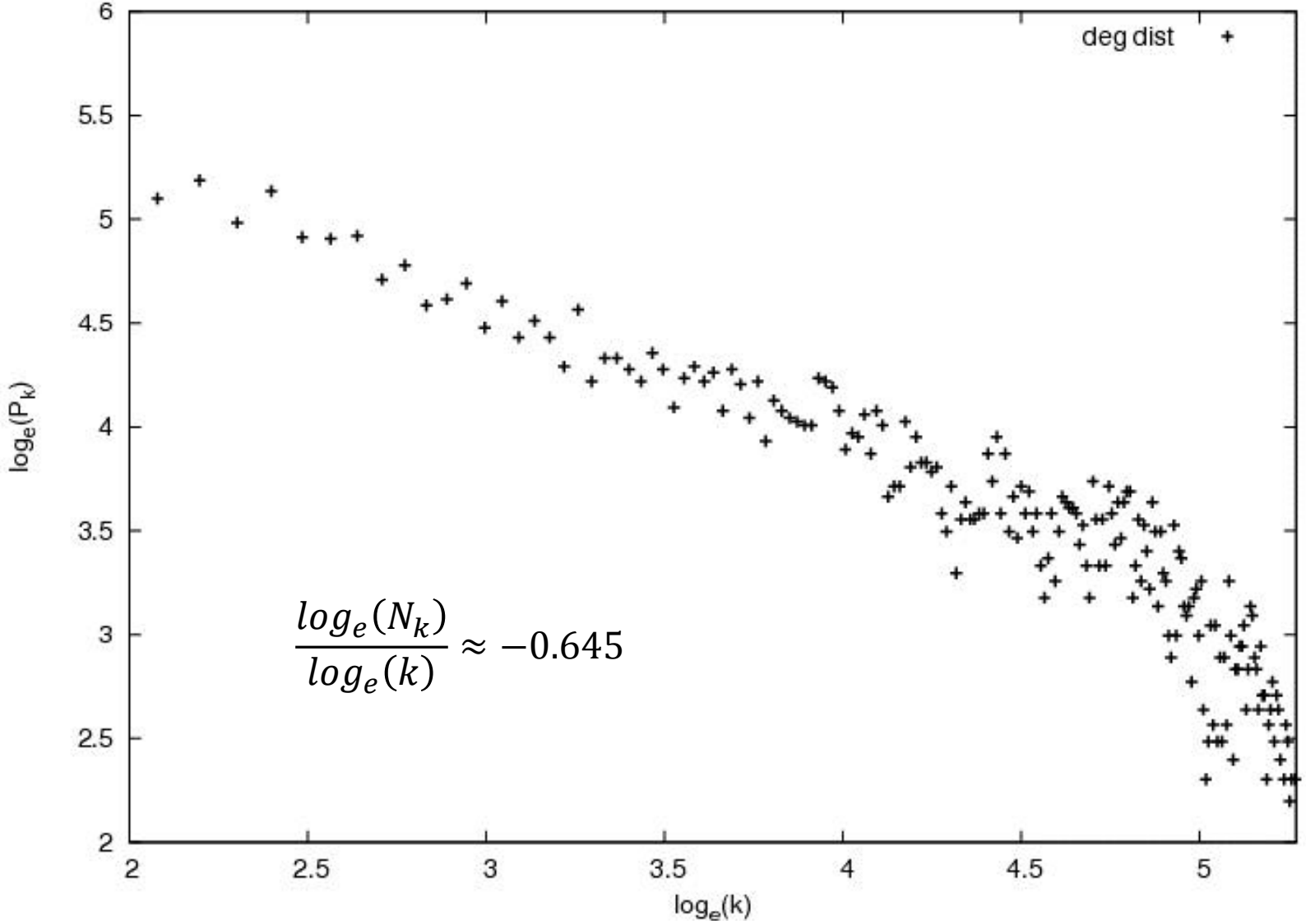
[2] unable to compute => graph too big

Pruning (AS)

- Threshold = 175

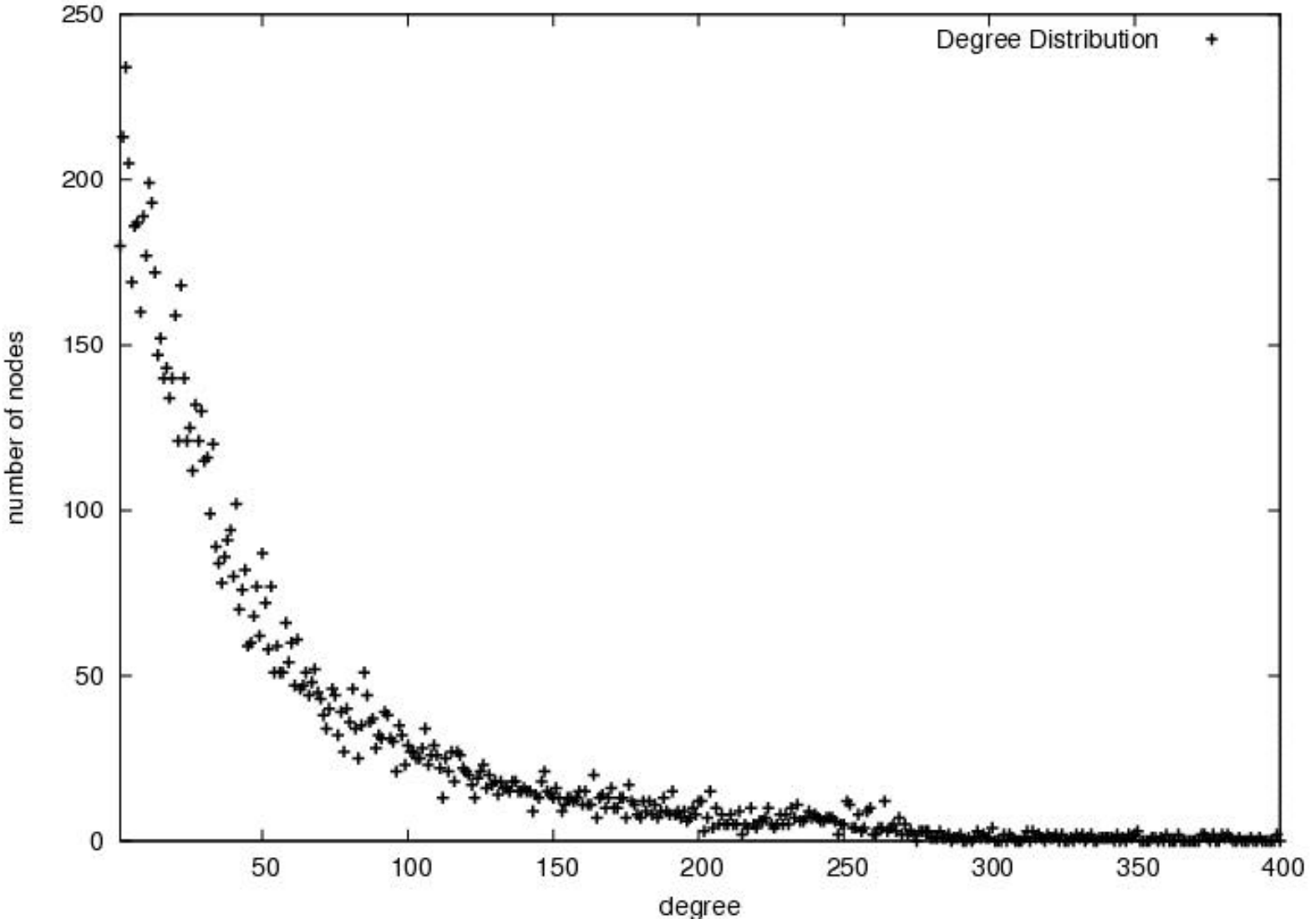


Degree Distribution: AS (log-scale)

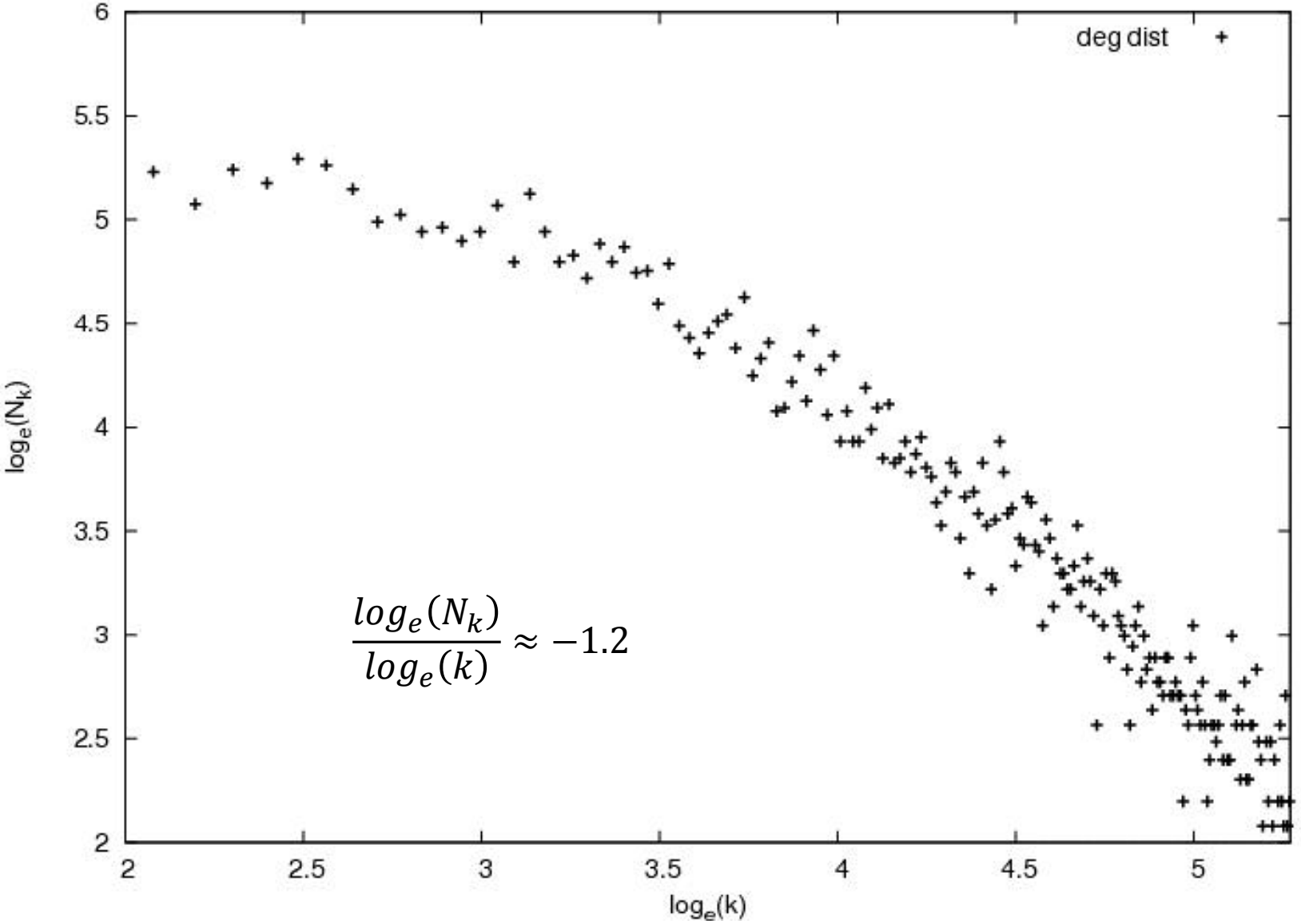


Pruning (Facebook)

- Threshold = 800

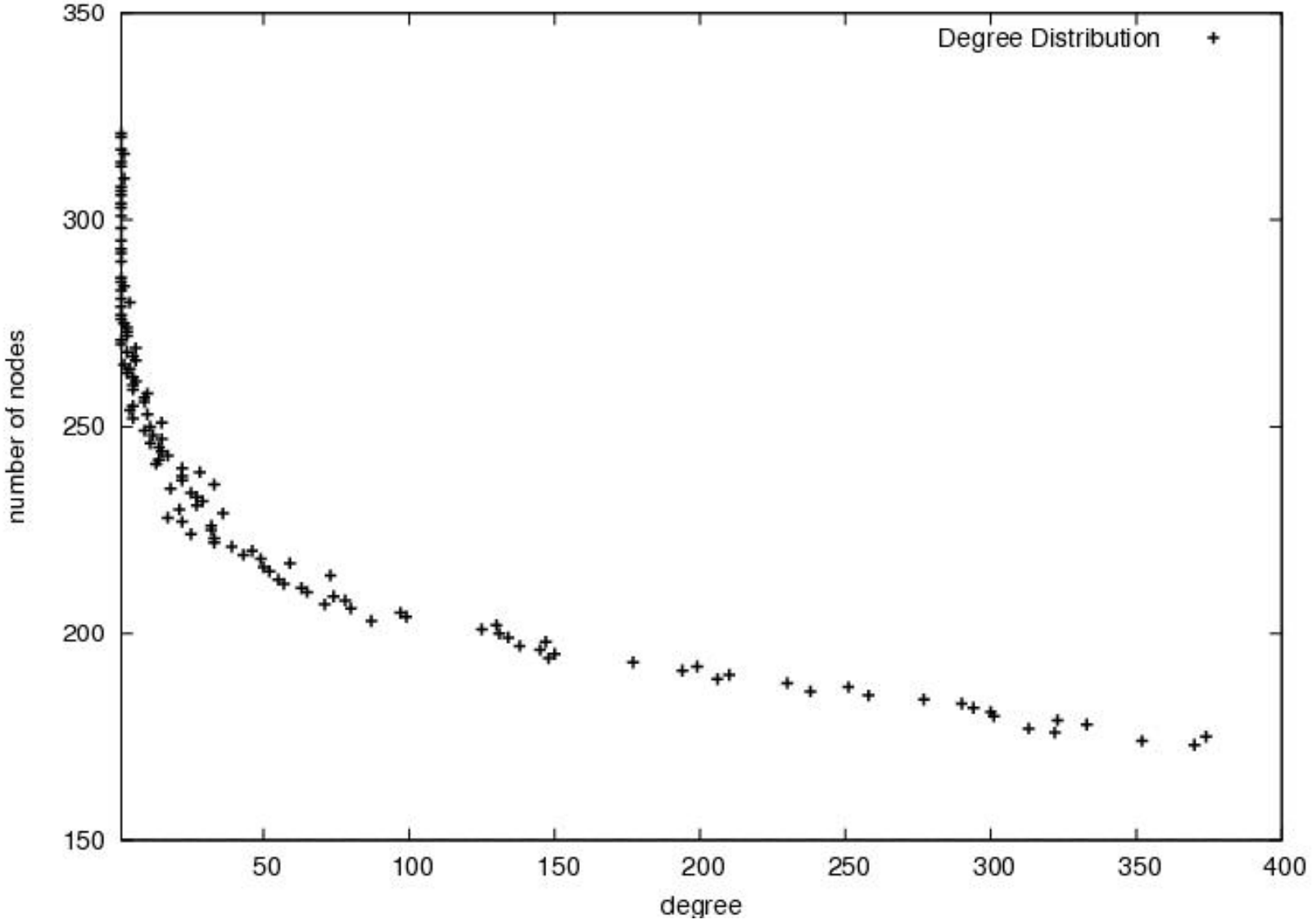


Degree Distribution: Facebook (log-scale)

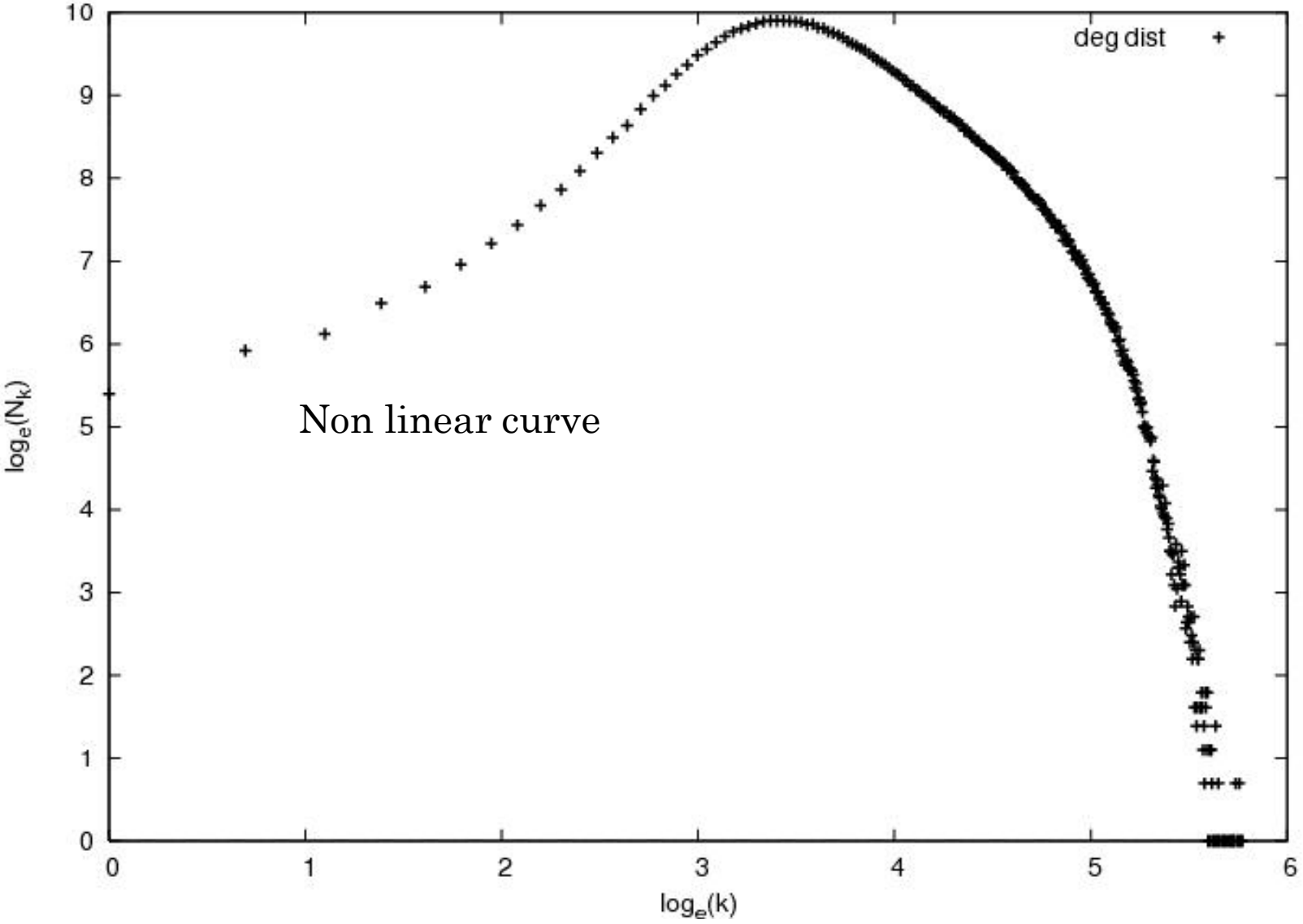


Pruning (Twitter)

- Threshold = 100 to 500 (out-degree)

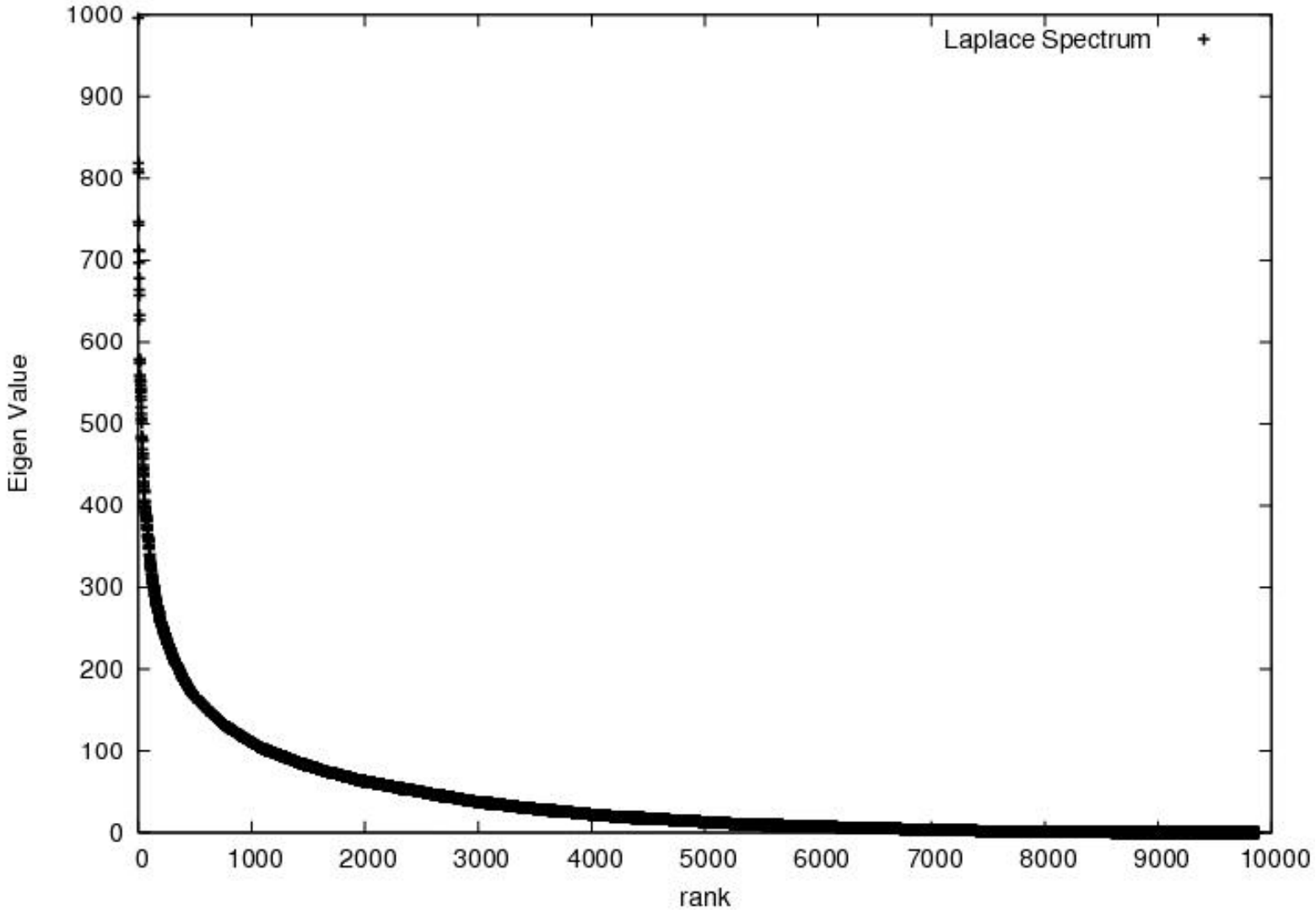


Degree Distribution: Twitter (log-scale)



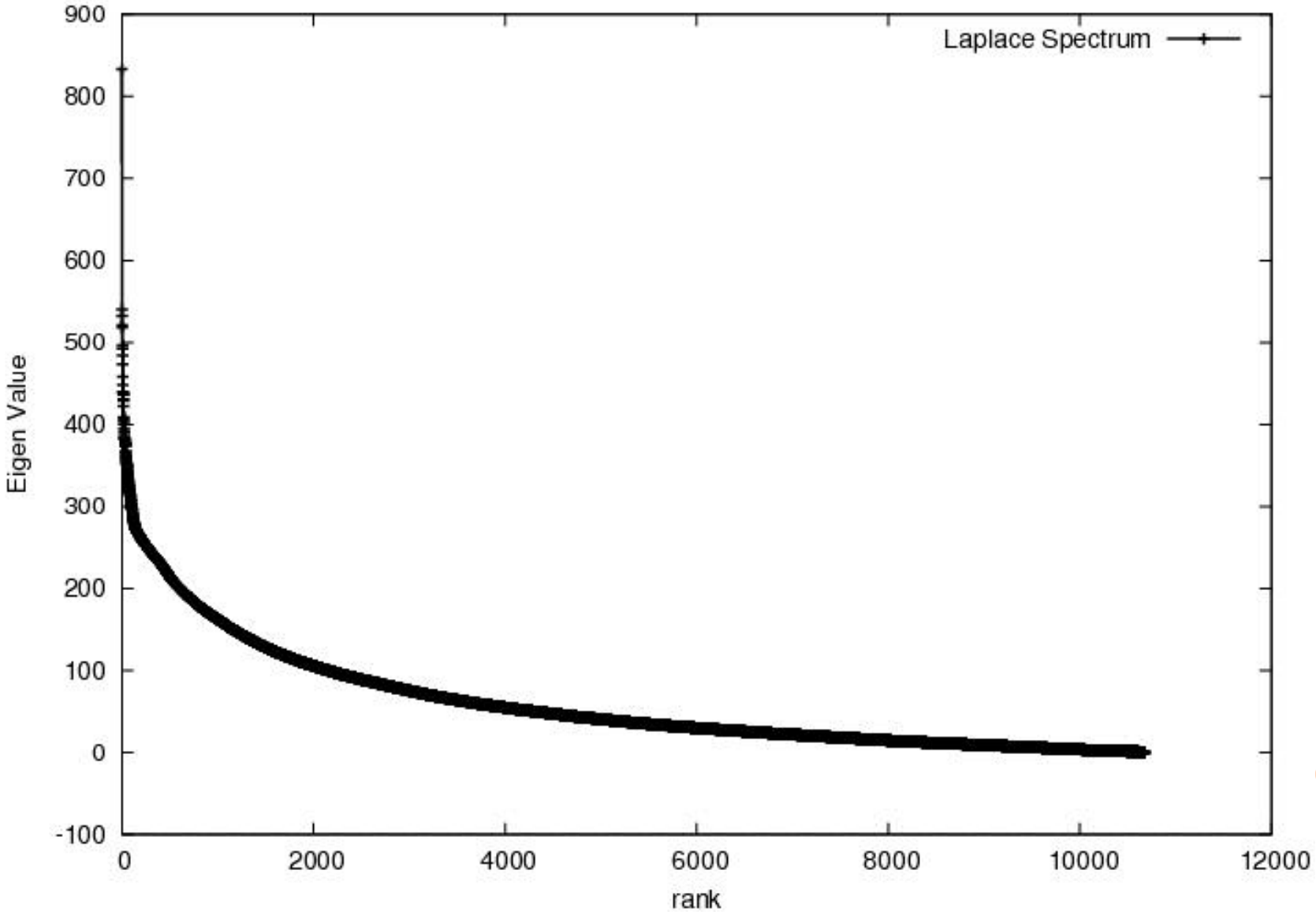
Spectral Analysis (laplace spectrum)

- AS



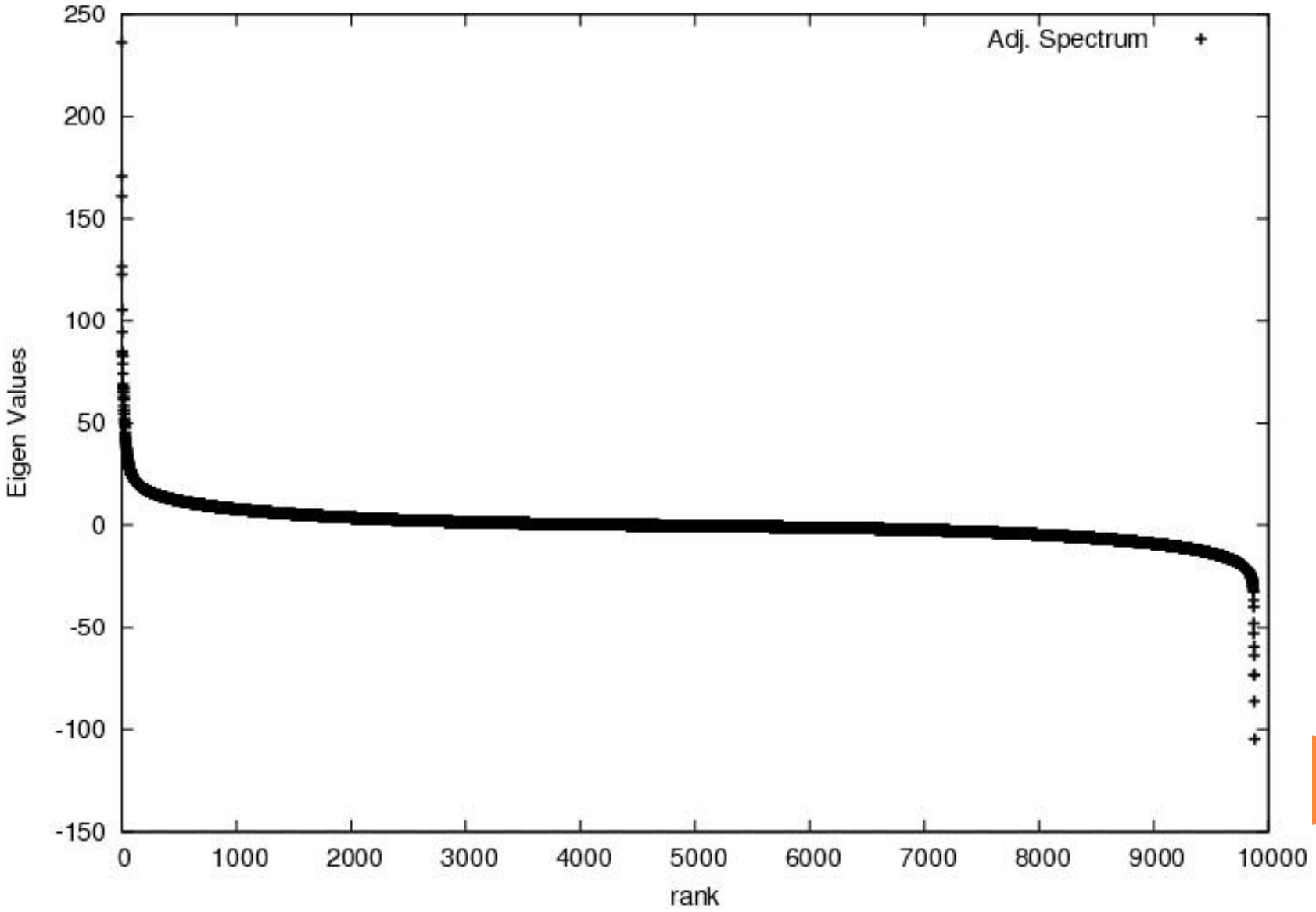
Spectral Analysis (laplace spectrum)

- Facebook



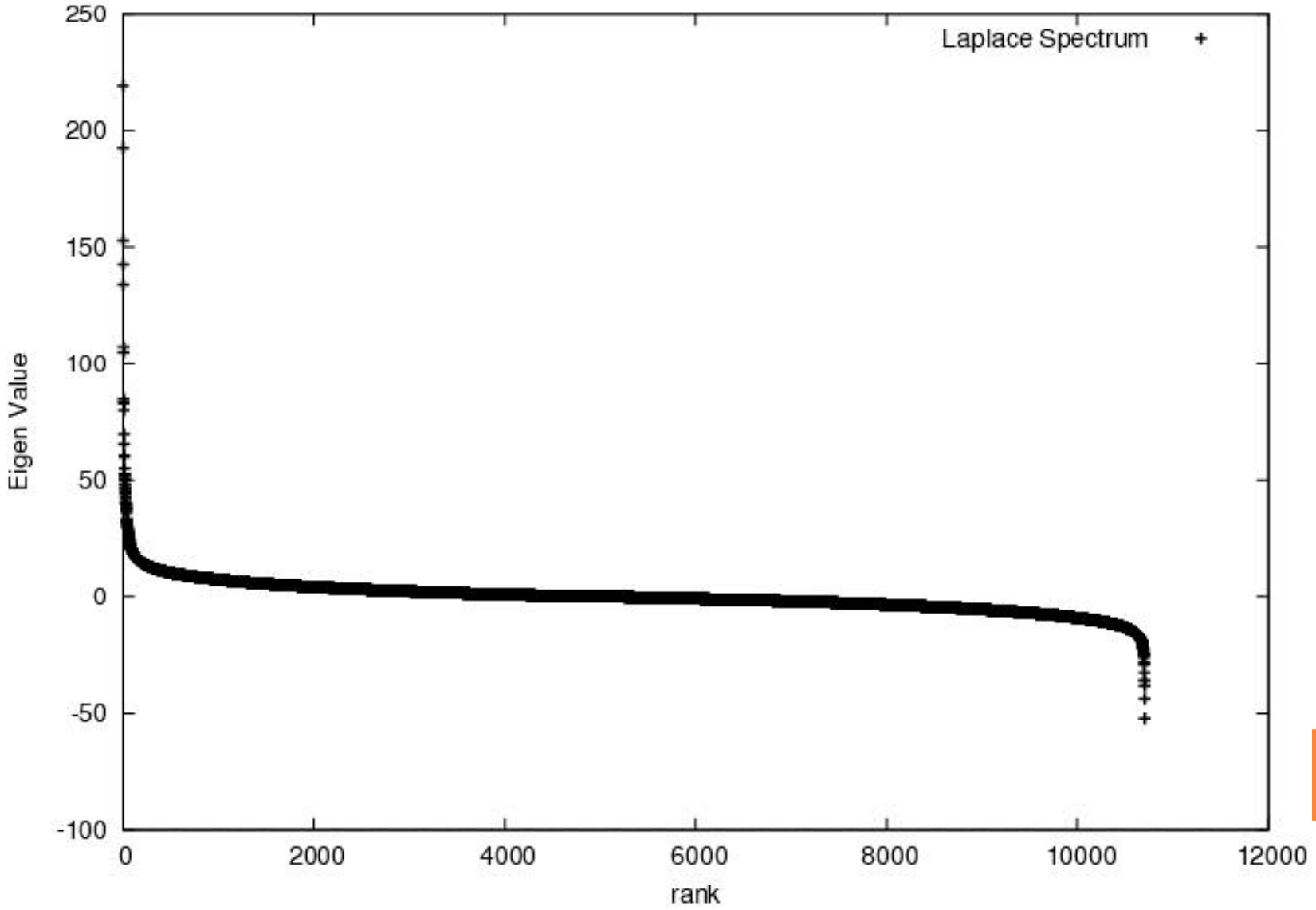
Adjacency spectrum

- AS



Adjacency Spectrum

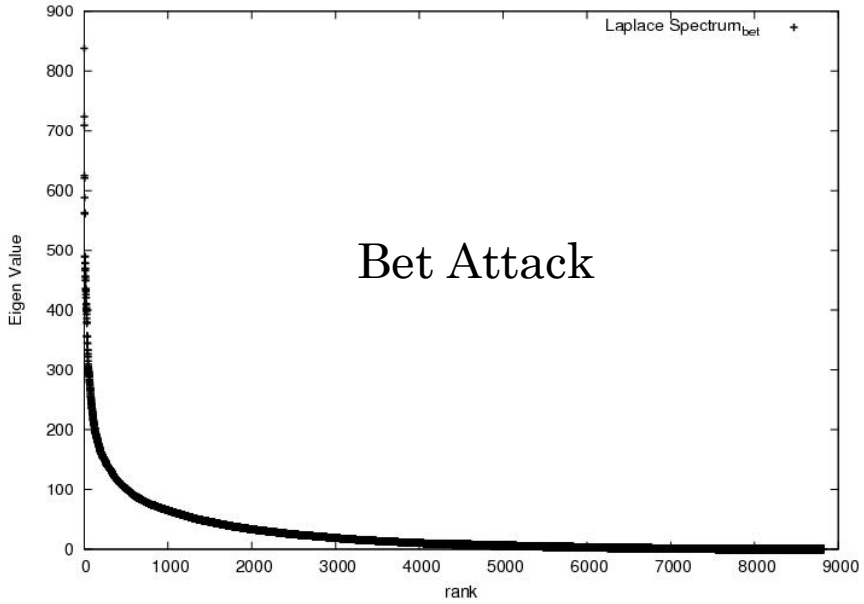
- Facebook

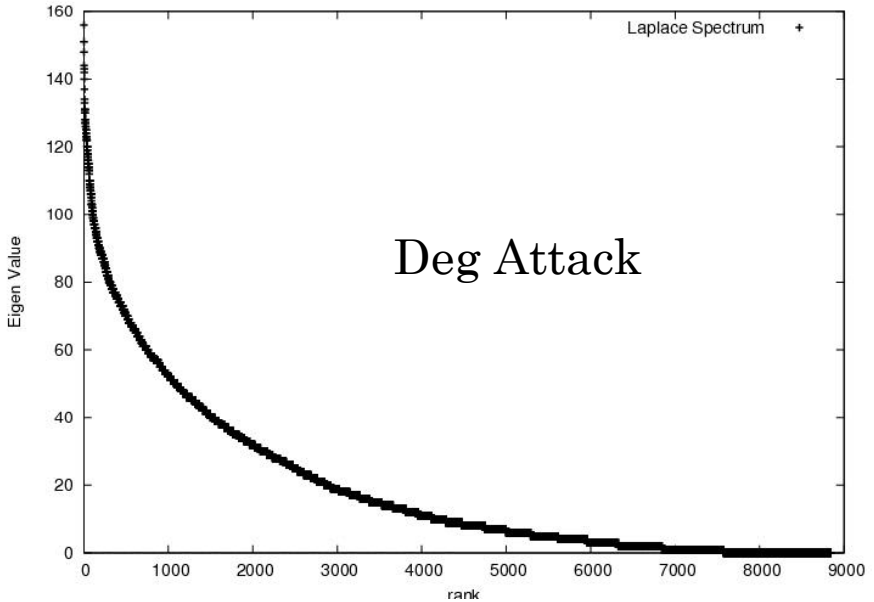


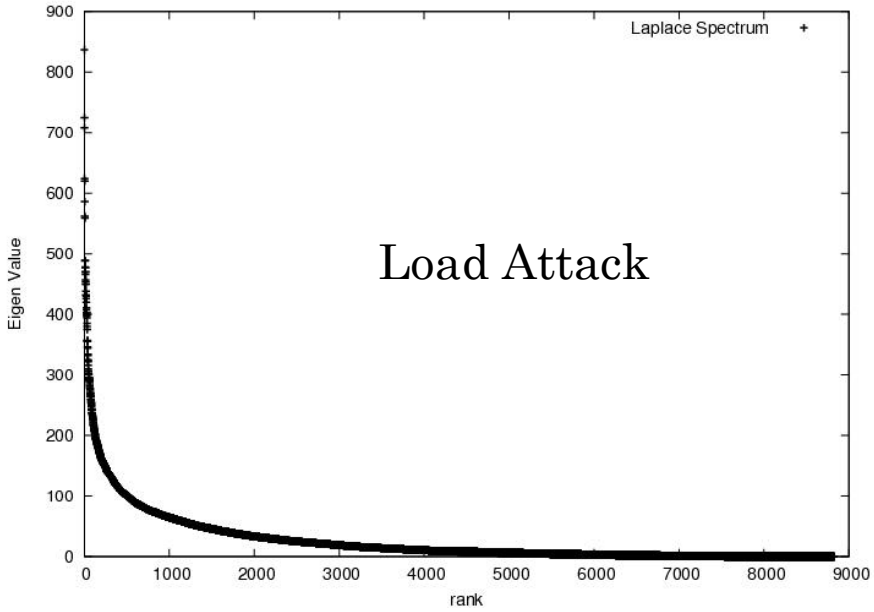
Node removal

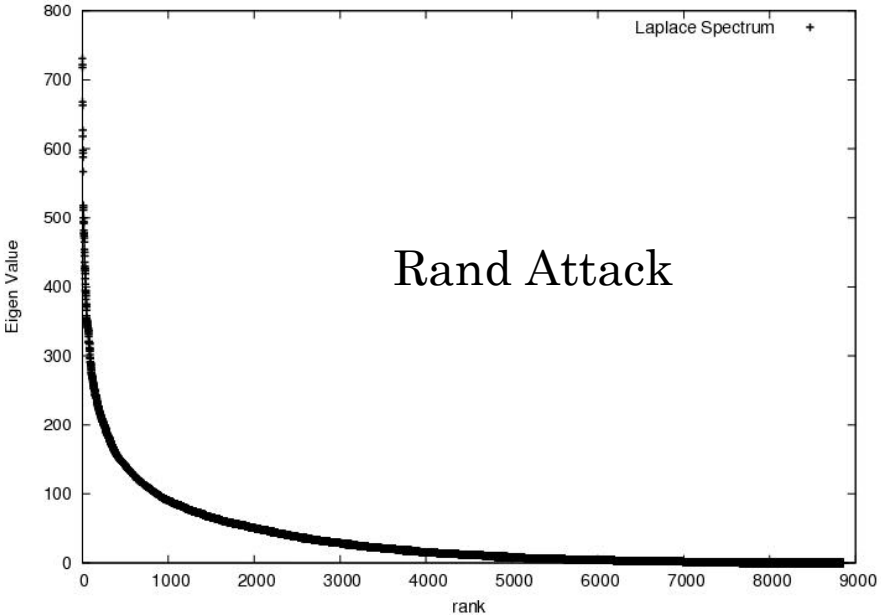
- Node removal: Top k node removed based on four metrics
  - Random node removal (‘rand’ attack)
  - Degree based (nodes with high degree centrality) (‘deg’ attack)
  - Based on betweeness centrality (‘bet’ attack)
  - Based on closeness centrality (‘load’ attack)
- Sort the nodes based on a particular centrality and remove Top ‘k’ nodes : size of attack = k

Node removal on AS (Laplace Spectrum)

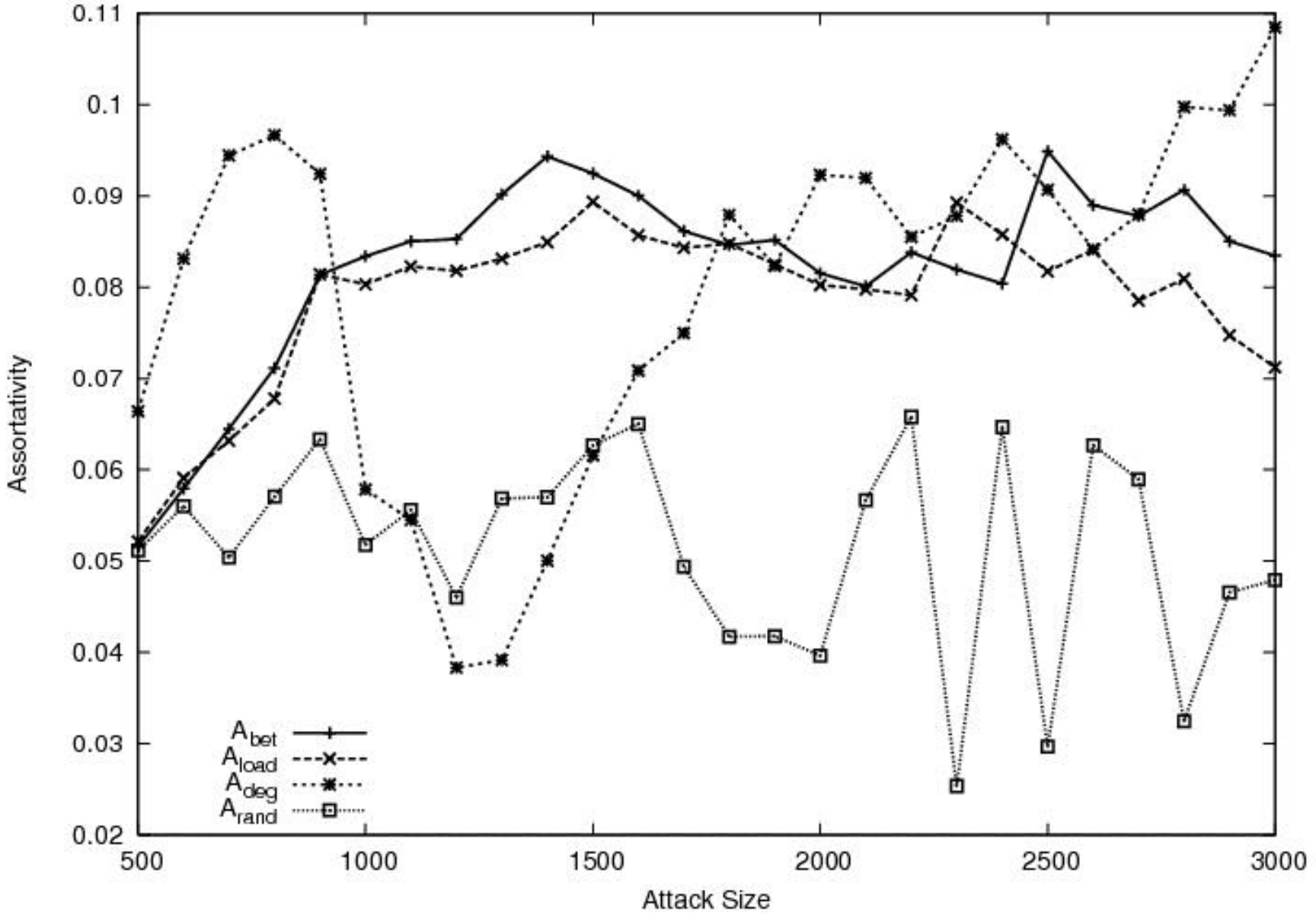




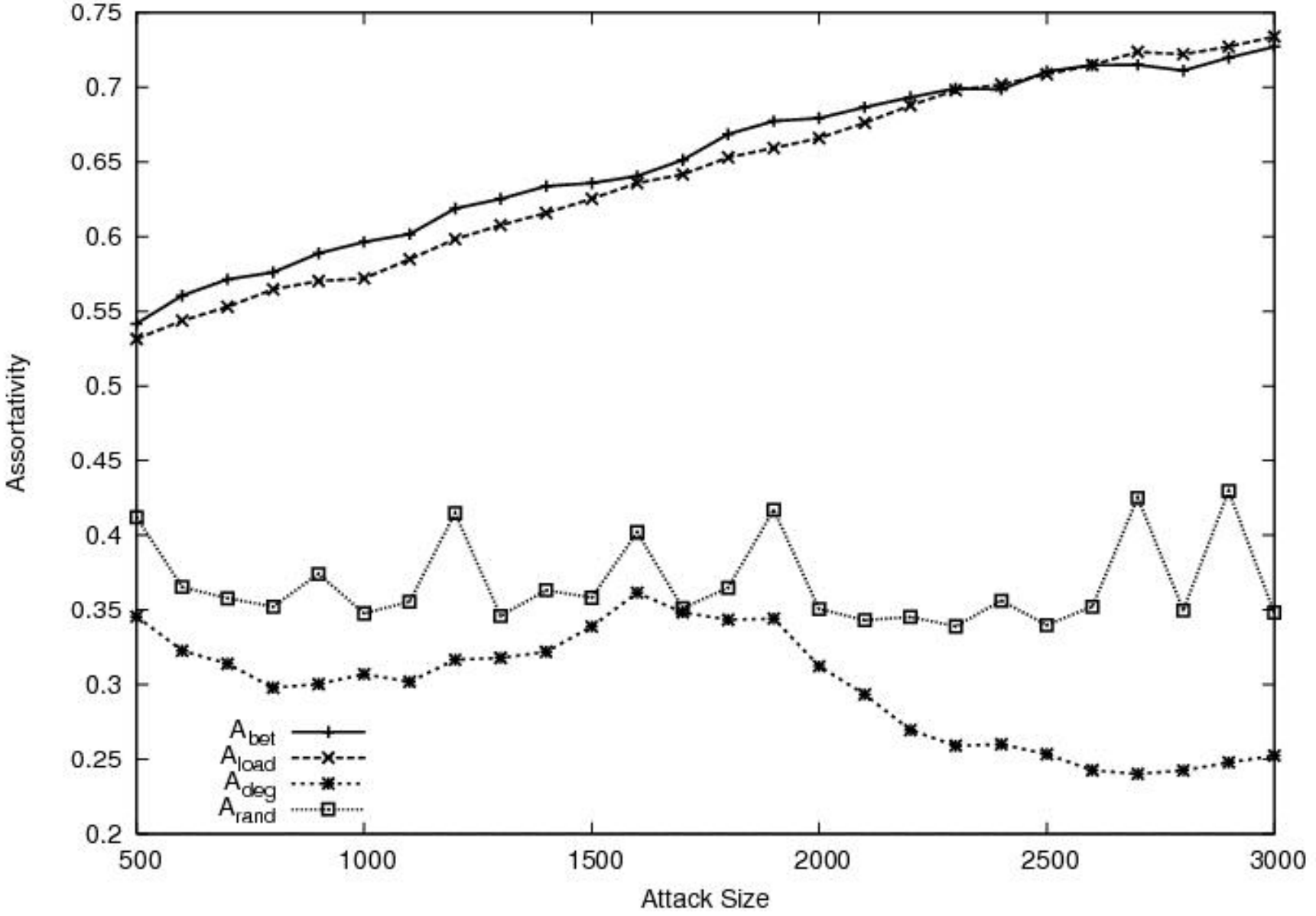




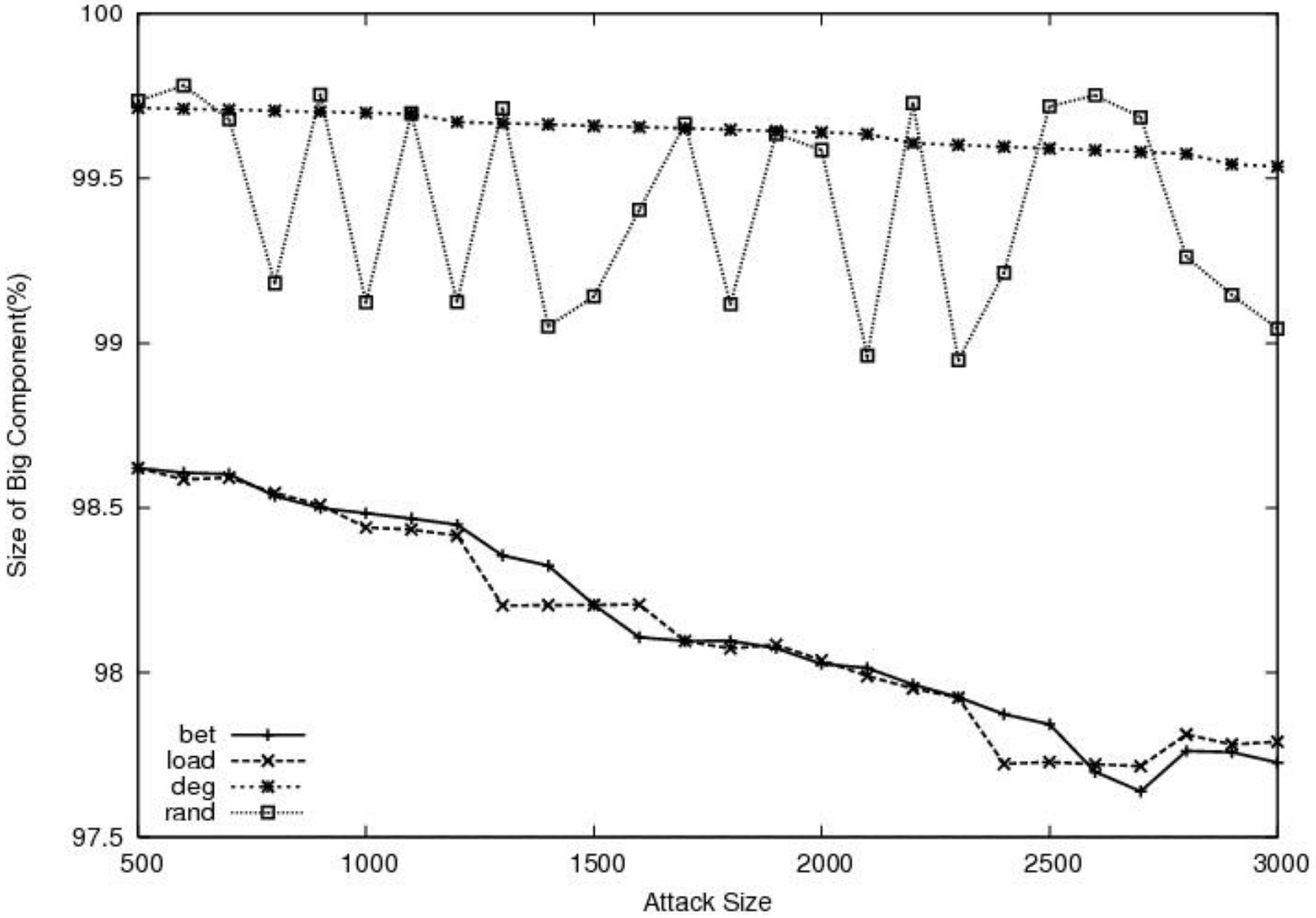
Node removal AS (Assortativity)



Node removal on Facebook



Node removal Facebook (Size of Big Components)



Bimodal network simulation

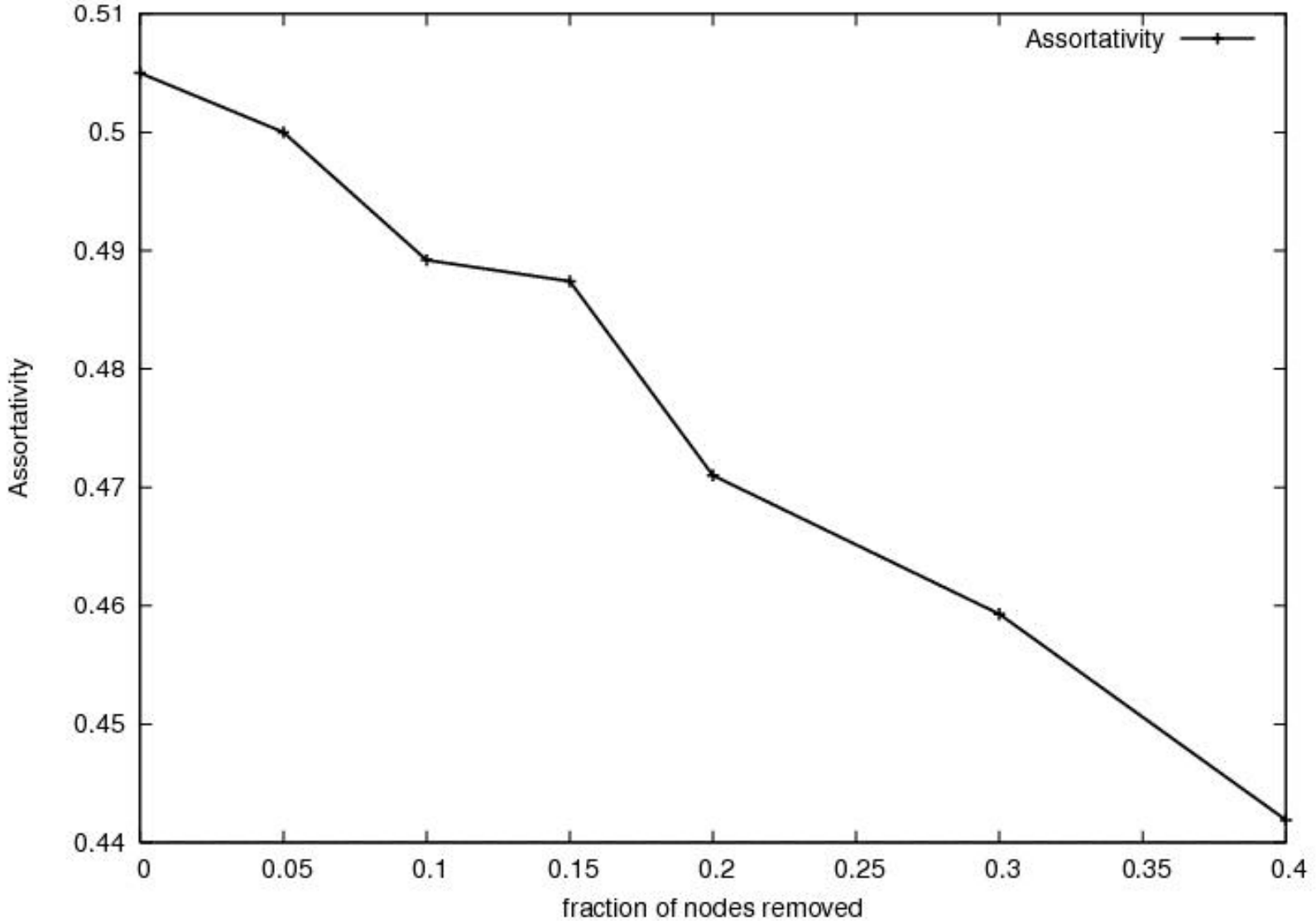
- Bimodal networks are networks in which a node can have either low degree or high degree (super nodes)
- Bimodal network simulation
  - Simulation done using a C code(courtesy Animesh Srivastav)
  - Variation of Assortativity with random node removal

Statistics of simulated Bimodal Network generated:

Low degree	5
High degree	20
Prob. Of low degree	0.8
Assortativity of network	0.5

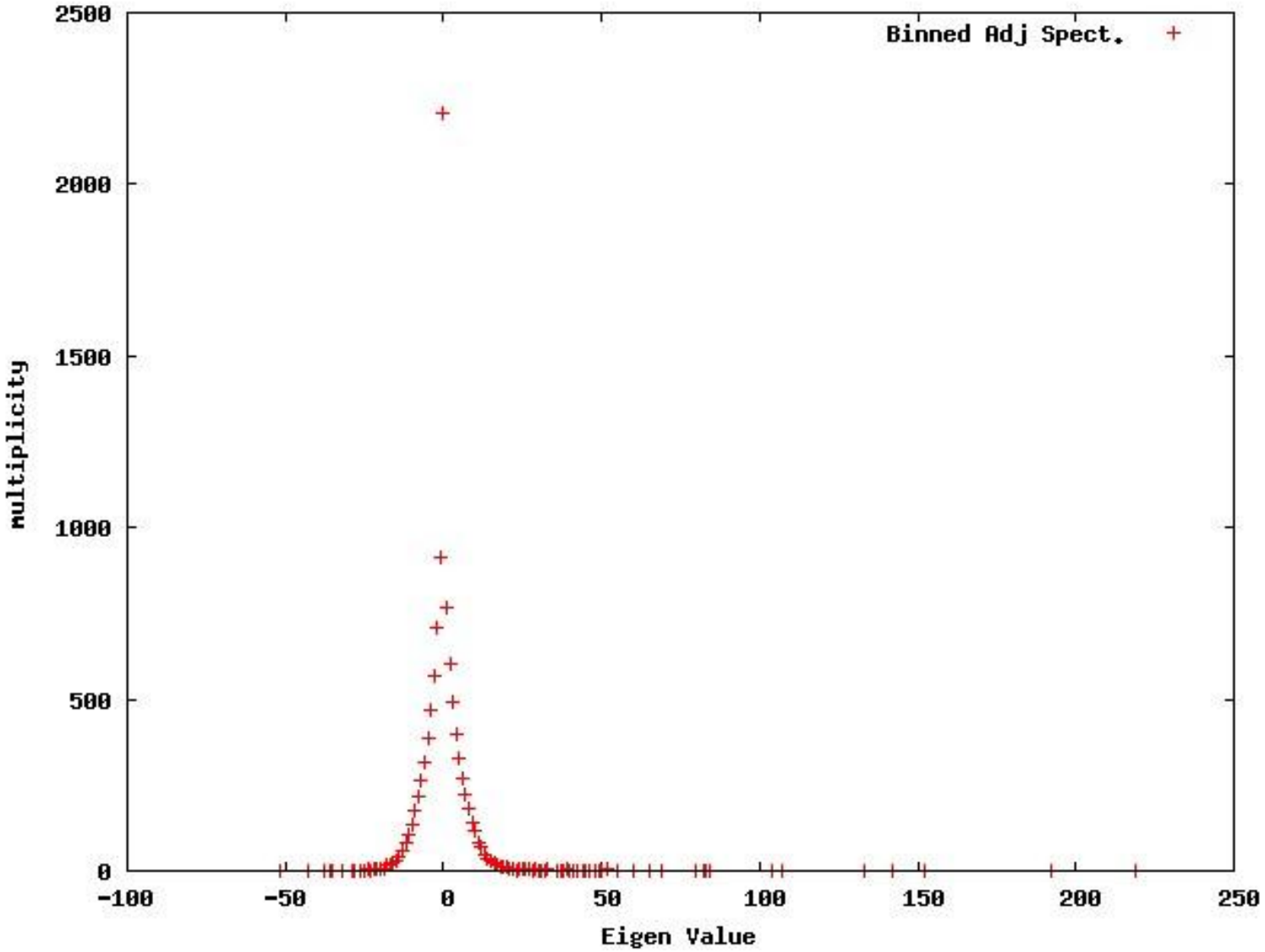
Bimodal Network Simulation (Assortativity vs. node removal)

- Number of iterations = 10

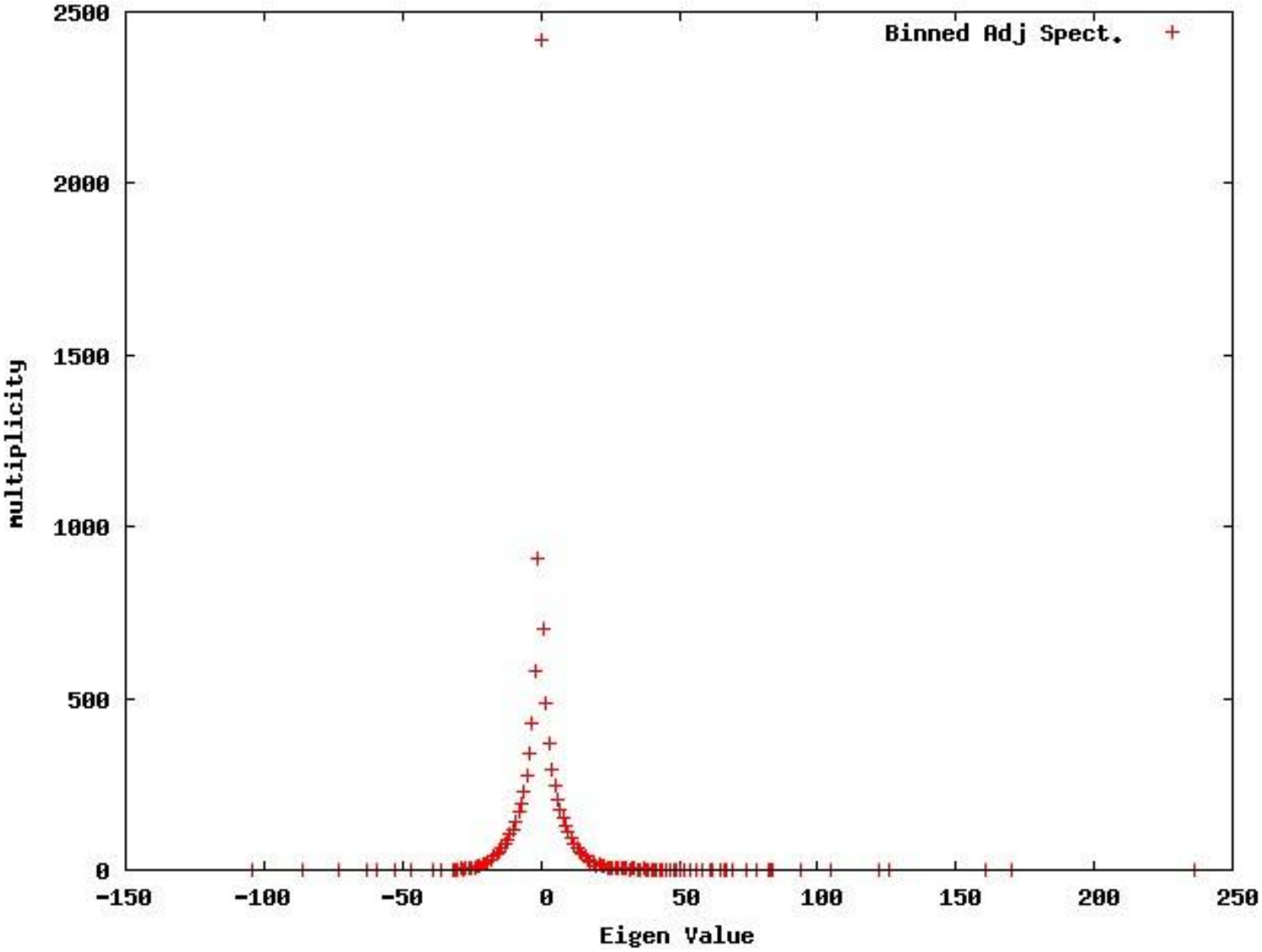


POST-MIDSEM

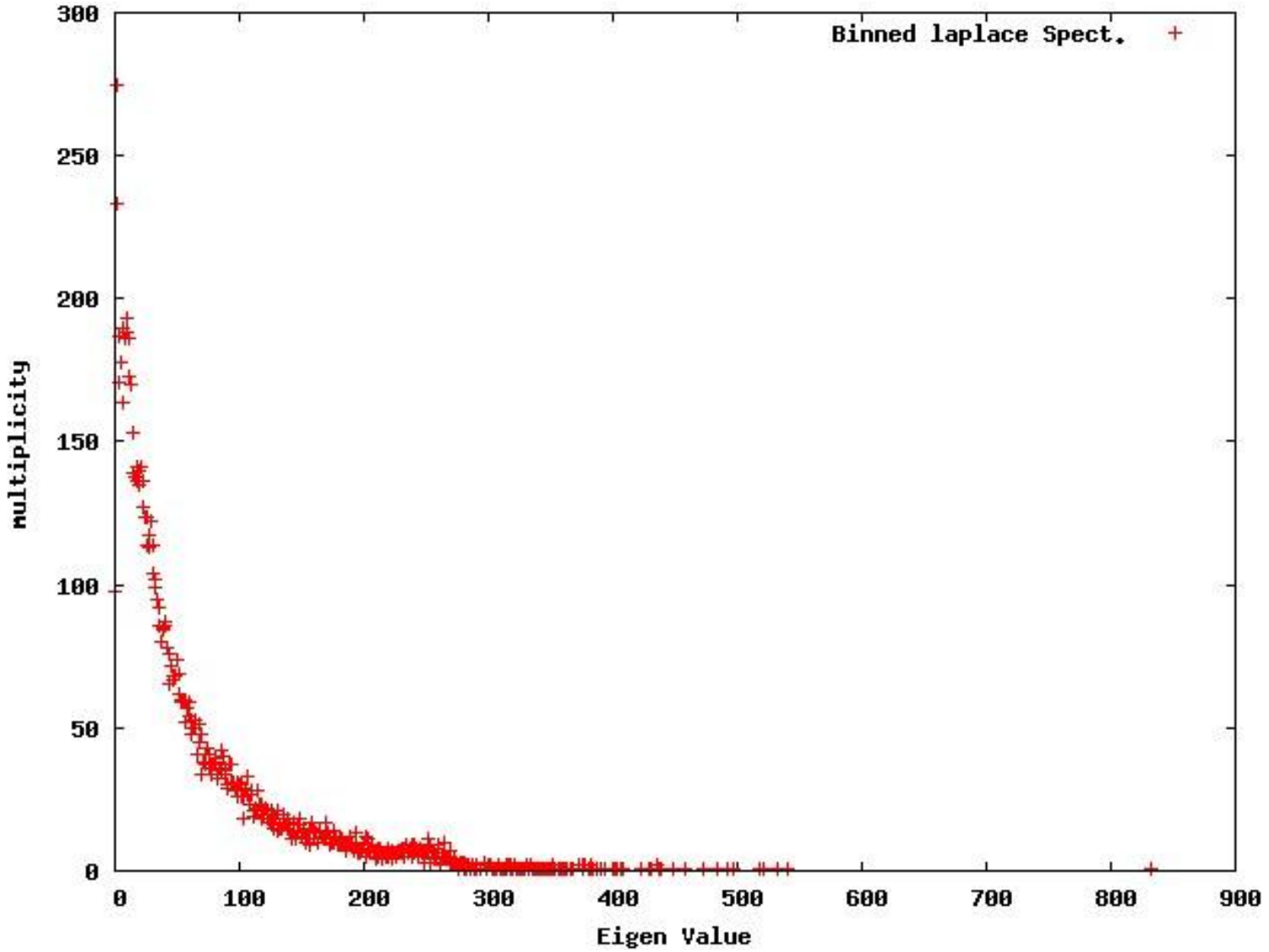
Adjacency Binned Spectrum (FACEBOOK)



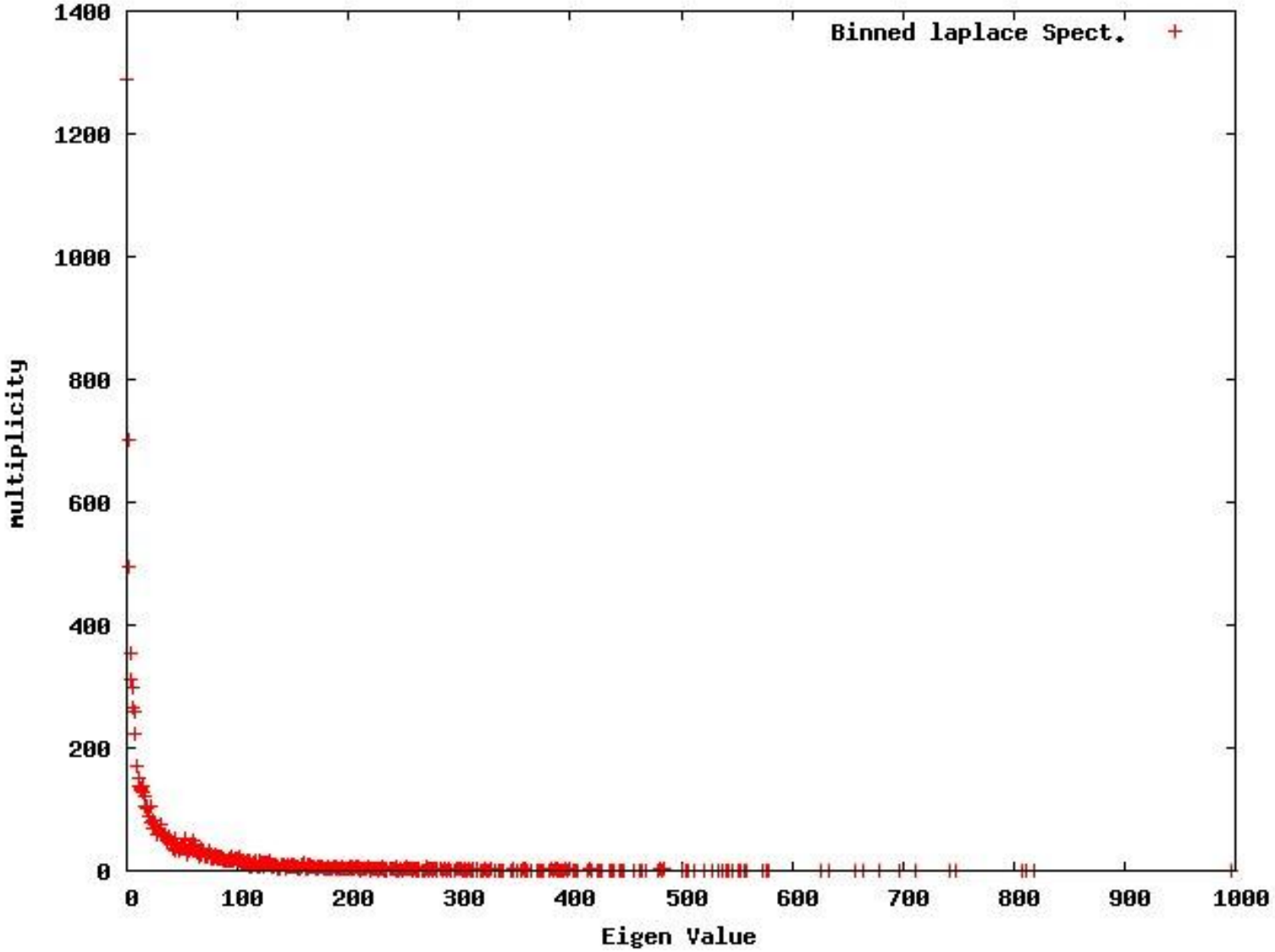
Adjacency Binned Spectrum (AS)



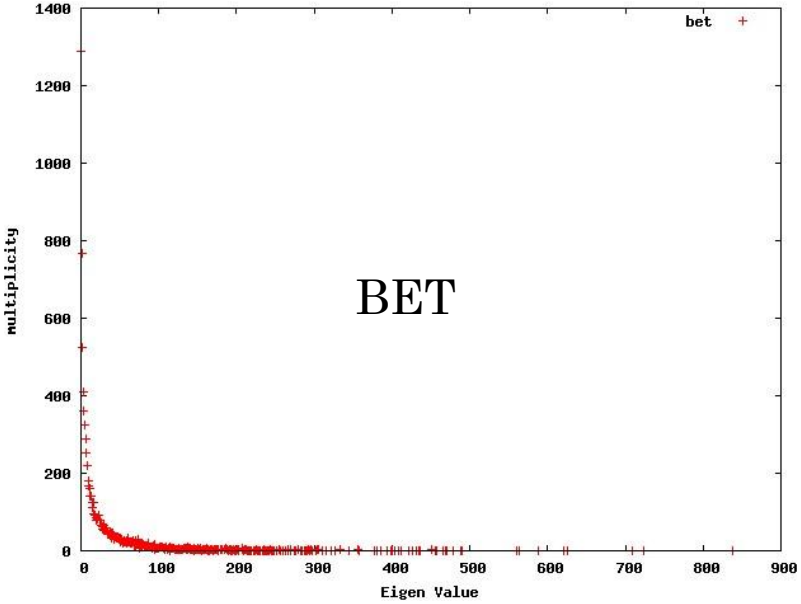
Laplace Binned Spectrum (Facebook)



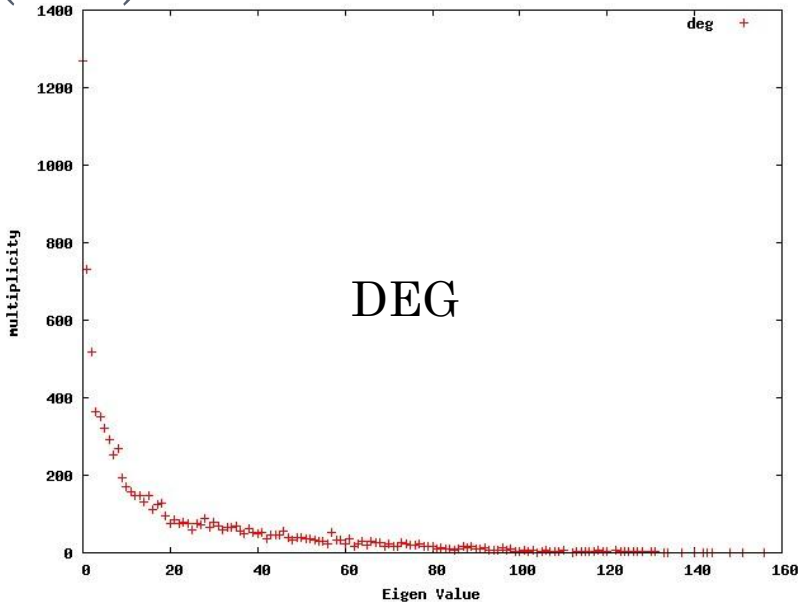
Laplace Binned Spectrum (AS)



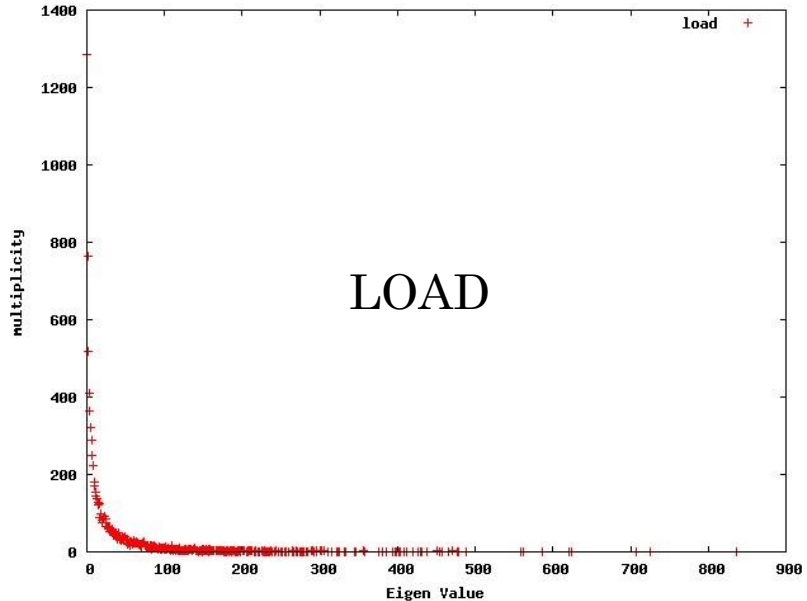
Laplace Binned Spectrum Of Attacked Network (AS)



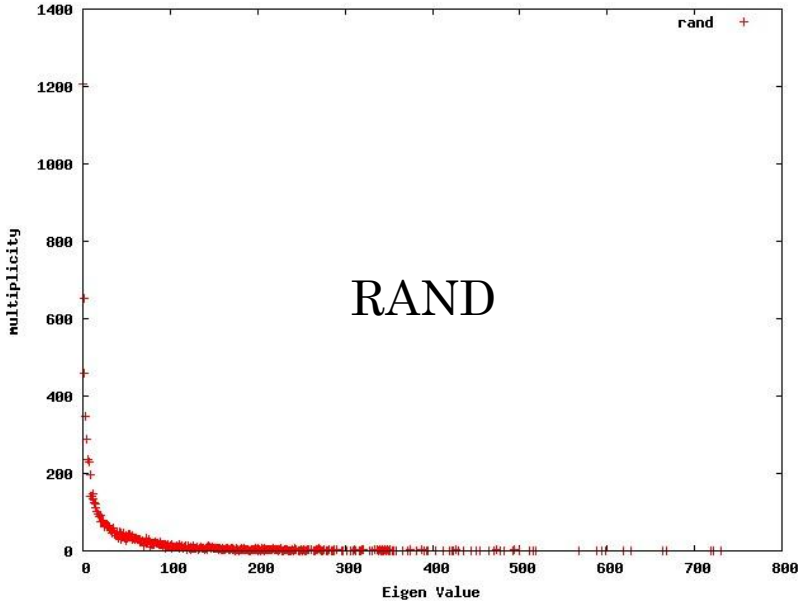
BET



DEG



LOAD

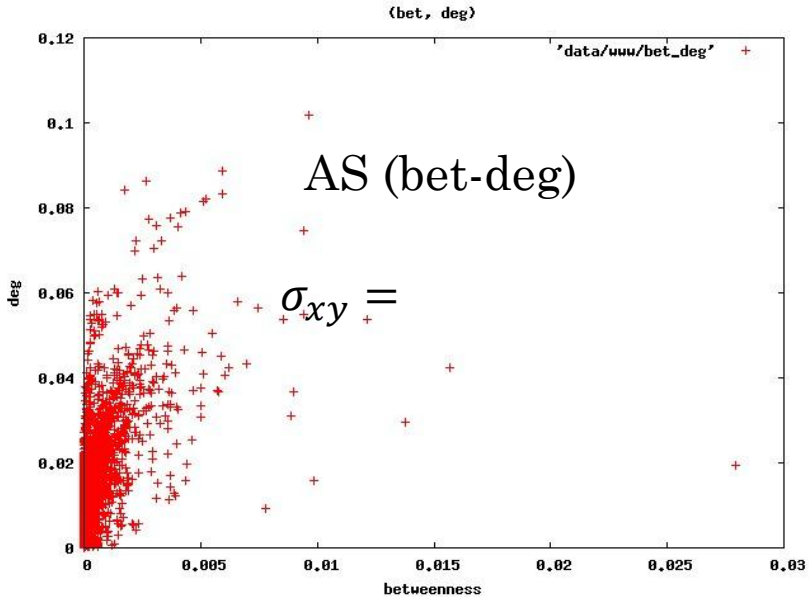


RAND

Deductions

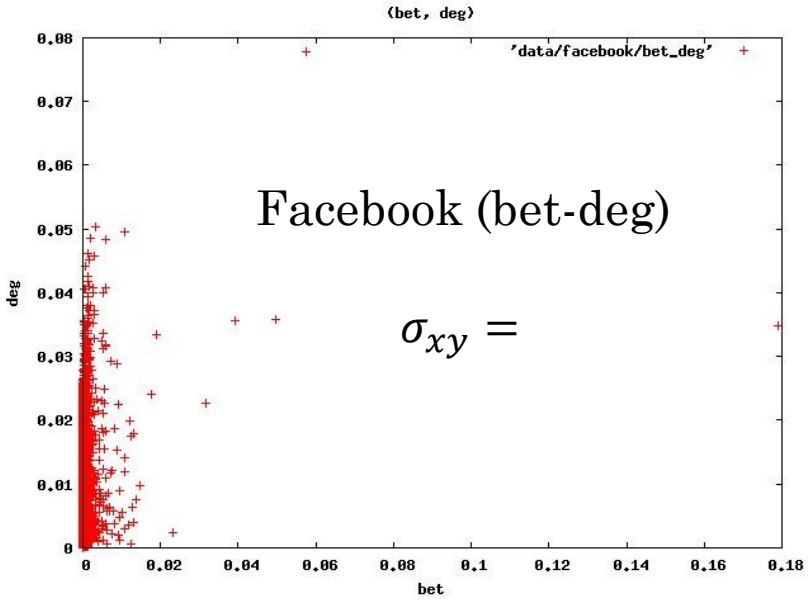
- From the attacks done on AS and Facebook Network is clearly visible that attacks based on load and betweenness centrality behave in same way
- The trend of change in assortativity is different for AS and Facebook network.
- Correlation between various centralities in AS and Facebook ?

CO-RELATION



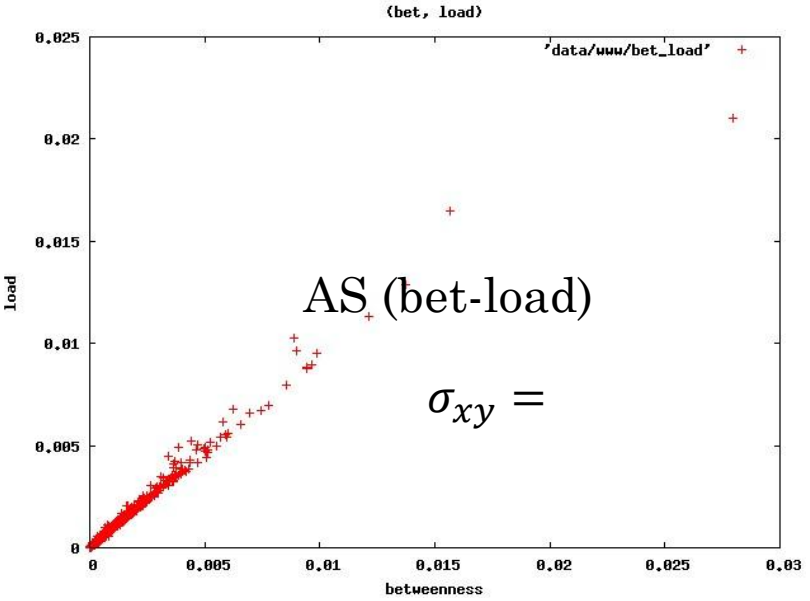
AS (bet-deg)

$\sigma_{xy} =$



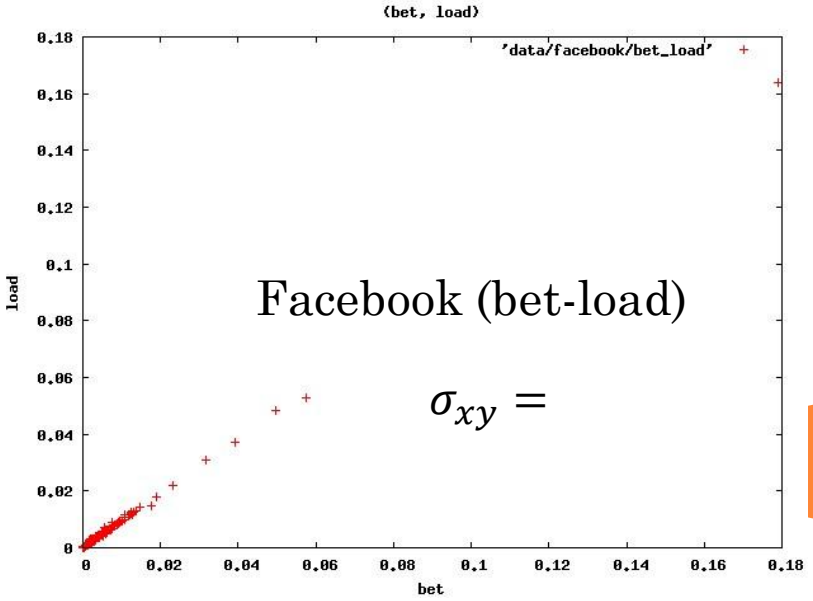
Facebook (bet-deg)

$\sigma_{xy} =$



AS (bet-load)

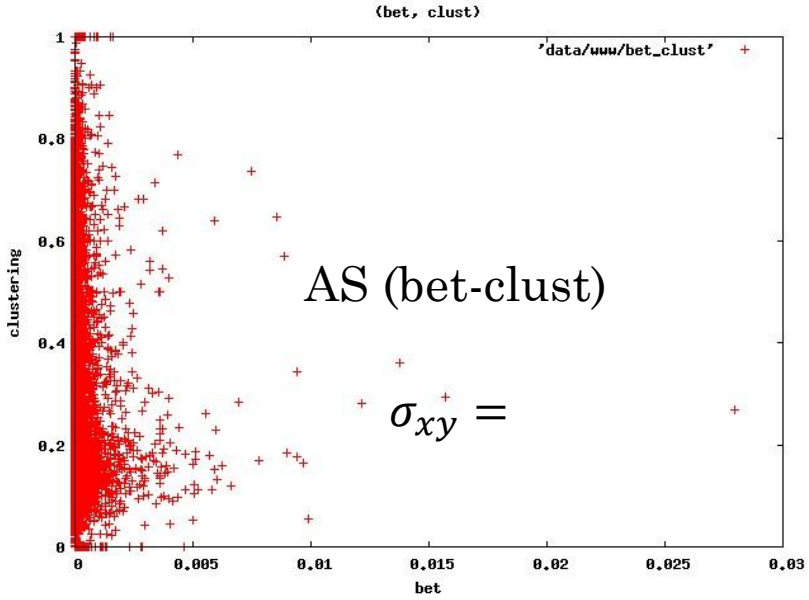
$\sigma_{xy} =$



Facebook (bet-load)

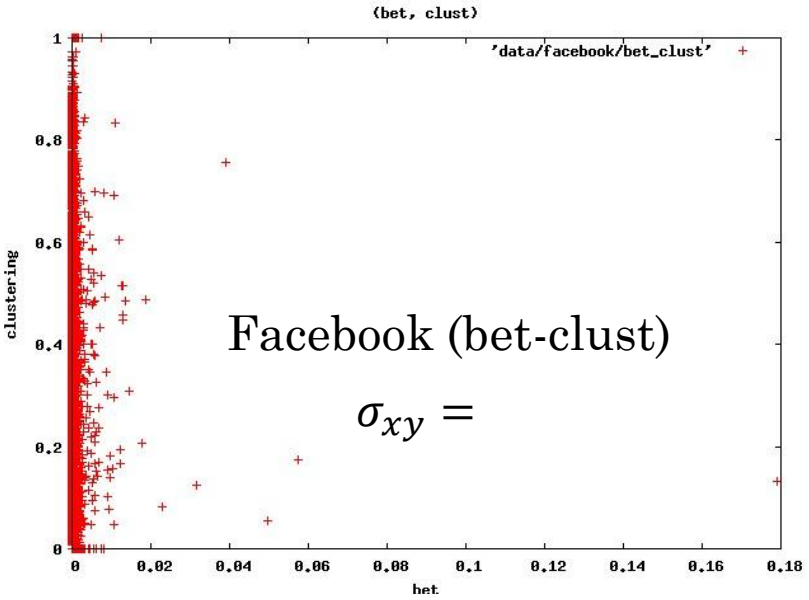
$\sigma_{xy} =$

Co-relation



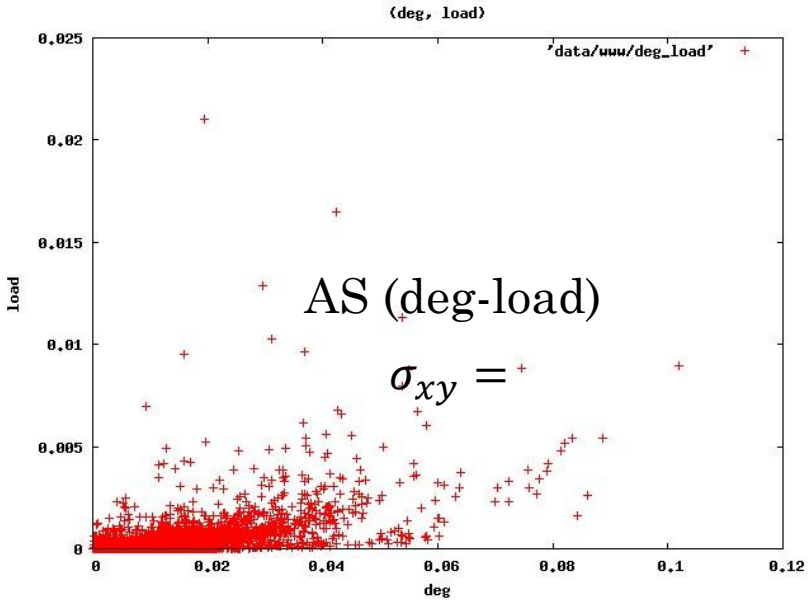
AS (bet-clust)

$\sigma_{xy} =$



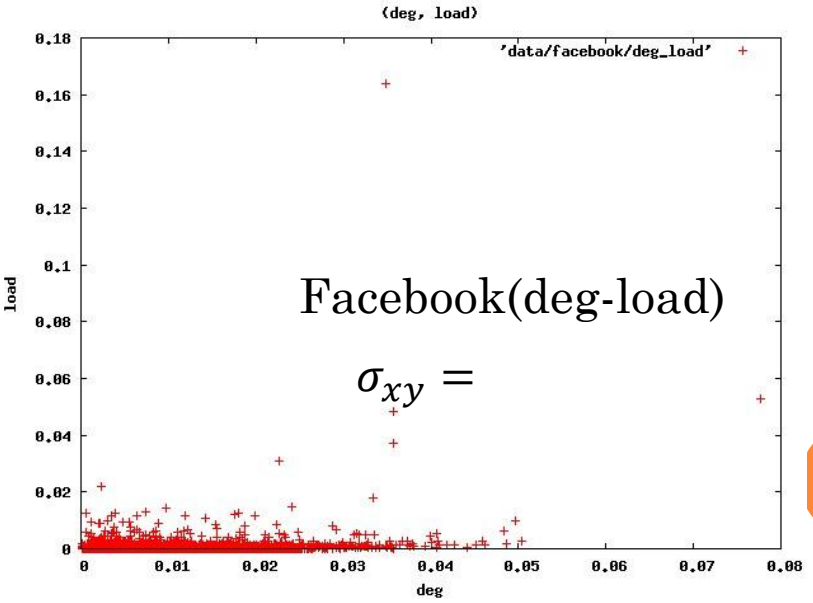
Facebook (bet-clust)

$\sigma_{xy} =$



AS (deg-load)

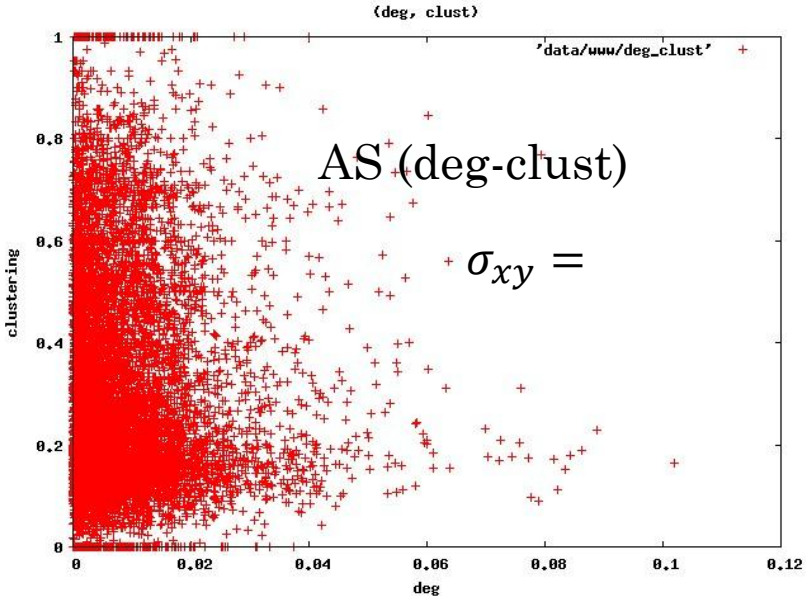
$\sigma_{xy} =$



Facebook(deg-load)

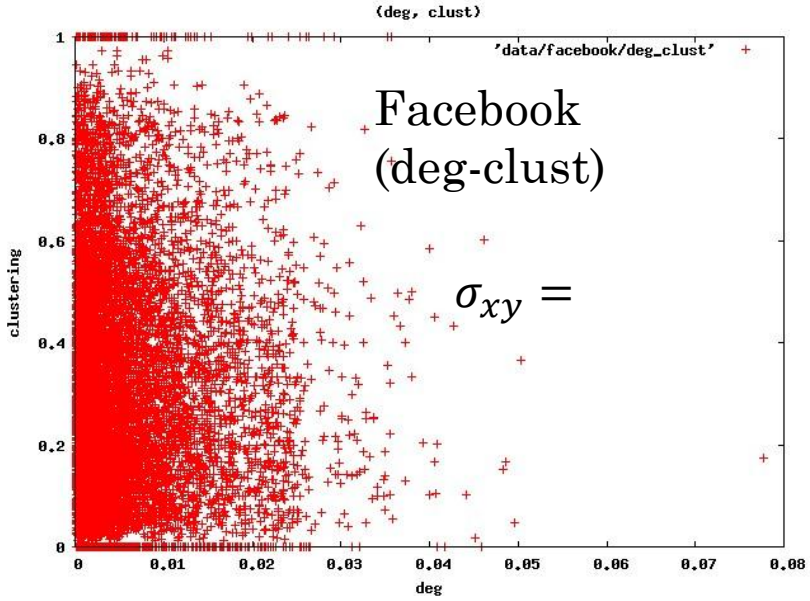
$\sigma_{xy} =$

Co-relation



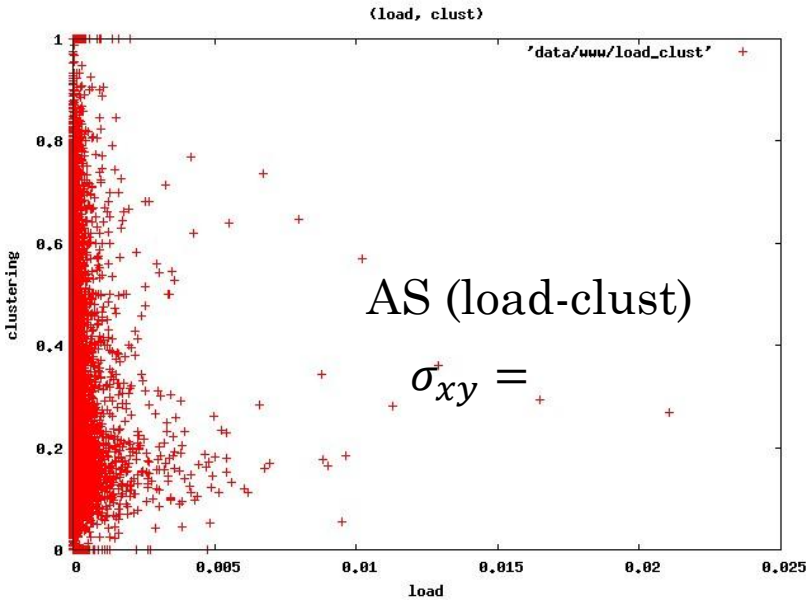
AS (deg-clust)

$\sigma_{xy} =$



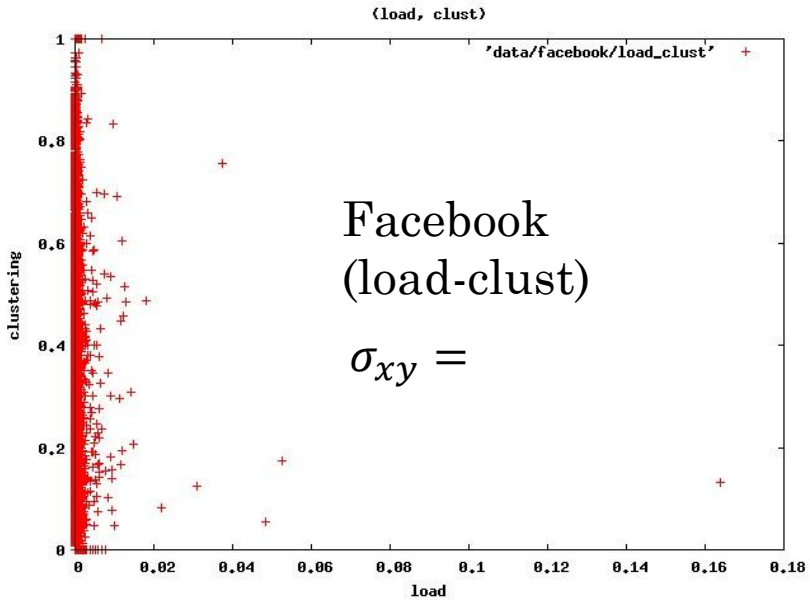
Facebook (deg-clust)

$\sigma_{xy} =$



AS (load-clust)

$\sigma_{xy} =$



Facebook (load-clust)

$\sigma_{xy} =$

Pearson Co-relation Matrix

	BET	DEG	LOAD	CLUST
BET	1.0	0.3179	0.8307	-0.0225
DEG		1.0	0.3204	-0.0079
LOAD			1.0	-0.0229
CLUST				1.0

AS network

	BET	DEG	LOAD	CLUST
BET	1.0	0.1073	0.9641	-0.0052
DEG		1.0	0.1067	-0.0027
LOAD			1.0	-0.0052
CLUST				1.0

Facebook network

Correlation (AS vs Facebook)

- AS network has considerably higher correlation between betweenness and load centrality and degree centrality
- Means that nodes with high degree have higher betweenness and closeness (typical of a router network)
- Facebook has correlation between the load, betweeness centrality and degree centrality but it is lower than AS networks

Correlation (Contd…)

- In social network context degree does not dictate the closeness of a node from other nodes.
- In both cases Load centrality and Betweenness centrality are highly correlated, more so in the case of Facebook.
- In both the cases, negligible negetive correlation with clustering coefficients

Future Works

- Perform the experiments on twitter dataset
- Perform clustering coefficient based node removal
- Study the effect of attacks on network diameter
- Compare the results obtained for the three data sets
- Simulate experiments with bimodal networks

References

[1]. S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, and A. N. Samukhin, "Spectra of complex networks," Phys. Rev. E, vol. 68, no. 4, p. 046109, Oct 2003.

[2]. E. Estrada, Spectral Theory of Networks: From Biomolecular to Ecological Systems., Jun 2009.

[3]. A. N. Samukhin, S. N. Dorogovtsev, and J. F. F. Mendes, "Laplacian spectra of complex networks and random walks on them: Are scale-free architectures really important?" Jun 2007.

[4]. http://snap.stanford.edu/data/as-skitter.html (Internet Data)

[5]. Kwak, Haewoon and Lee, Changhyun and Park, Hosung and Moon, Sue. "What is Twitter, a Social Network or a News Media?". http://an.kaist.ac.kr/traces/WWW2010.html (Twitter data set)

[6]. On the Evolution of User Interaction in Facebook. (Facebook Data Set)

THANKS