The Slashdot Zoo:
Mining a Social Network with Negative Edges

Preliminary Presentation
(Based on the existing paper)

Siva Satvik - CS15BTECH11023
Goutham Meka - CS15BTECH11025
Bhanu Prakash - CS15BTECH11037
Vishwak Srinivasan - CS15BTECH11043

Paper Deconstruction

What is Slashdot (Zoo)?

- **Slashdot** (founded in 1997) is a **social news website** which features news stories on science, technology and politics.
- The **Slashdot Zoo** feature (added on 2002) allowed Slashdot users to tag other users are **friends** or **foes**.
- The **Slashdot Zoo corpus** is as a graph dataset containing **77,985 users** (as nodes) and **510,157 relationships** (as edges).

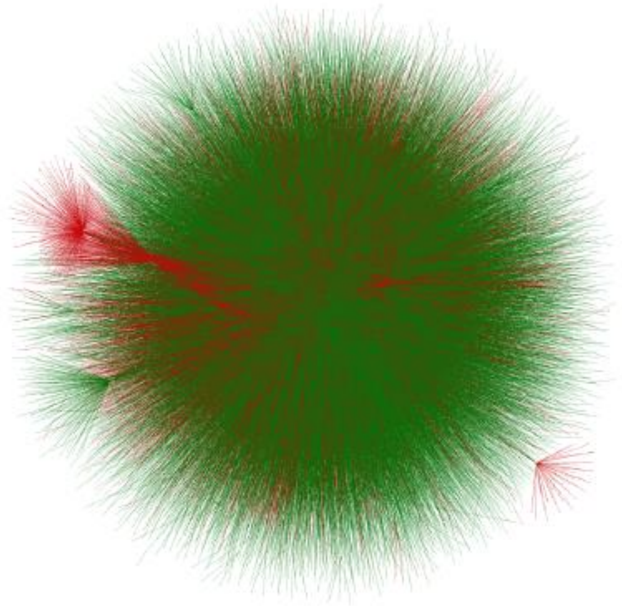

Figure 1: The Slashdot Zoo corpus represented as a graph, where nodes represent users and edges indicate relationships. The network contains 77,985 nodes and 510,157 edges. "friend" relationships are shown in green edges and "foe" relationships in red; the orientation of edges is not shown. The graph is centered at user *CmdrTaco*, founder and editor of the site.

Introduction

- Social network analysis’ are usually based on **positive edge weight** values. We attempt to analyse **inherently negative relationships** (distrust and dislike) with **negative edge weights**.
- To prove our assumption of **multiplicative transitivity**. Reflected by the statement, ***“the enemy of my enemy is my friend”***.

The Slashdot Zoo Corpus

- The Slashdot Zoo corpus we consider in this paper contains **77,985 users** and **510,157 links** (endorsements).
- Endorsements can be either positive (“**friend**") or negative (“**foe**").
- A user is always the **“fan”** of his friends and the **“freak”** of his foes.
- Slashdot generally limits the number of friends and foes to 200 users and to 400 users for subscribers. (Anomalies exist in the dataset)
- The Slashdot Community is known for both having **popular** and rather **unpopular users (trolls**).

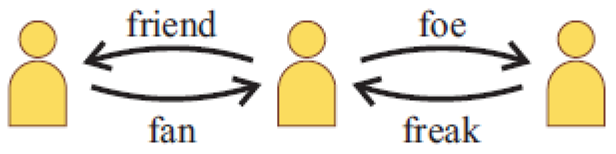

Figure 2: The two types of links allowed in the Slashdot Zoo (friend and foe) give rise to four kinds of relationships: friends, fans, foes and freaks. A user is the fan of his friends and the freak of his foes.

Definitions

- **n** – the number of users
- **u**, **v** – are specific users
- **A ∈ {–1, 0, +1}$^{n\times n}$** – is the adjacency matrix with values A_{uv} = +1 when user u marked user v as a friend and A_{uv} = –1 when user u marked user v as a foe. A is sparse, square and asymmetric.
- $\overline{\boldsymbol{A}}$ – the absolute adjacency matrix defined by $\overline{A_{ij}} = |A_{ij}|$
- $\boldsymbol{B}$ = A + A^{T} – the symmetric asymmetric matrix
- $\overline{\boldsymbol{B}} = \overline{\boldsymbol{A}} + \overline{\boldsymbol{A}^{T}}$ – the absolute symmetric asymmetric matrix
- $\overline{\boldsymbol{D}}$ – the absolute diagonal degree matrix defined by $\overline{D_{ii}} = \sum_j |A_{ij}|$
- $\overline{\boldsymbol{E}}$ – the absolute symmetric diagonal degree matrix defined by $\overline{E_{ii}} = \sum_j |B_{ij}|$

Corpus Statistics

Table 1: Statistics about the Slashdot Zoo corpus. The mean friend count and mean fan count are necessarily equal, as do the mean foe and freak counts.

Users	77,985
Links	510,157
Friend links	388,190
Foe links	121,967
Sparsity	0.000083884
Mean link count	6.542
Mean friend/fan count	4.978
Mean foe/freak count	1.564
Median links	3
Median friend count	1
Median foe count	0
Median fans count	1
Median freaks count	1

Table 2: The Slashdot Zoo's graph diameter, radius and mean shortest-path distance. The sign and direction of edges was ignored in the calculation of these values. In parentheses, we show the average distance in a random graph, as defined by Watts and Strogatz.

Diameter	6
Radius	3
Average distance	3.86 (5.82)

Corpus Statistics (Cont.)

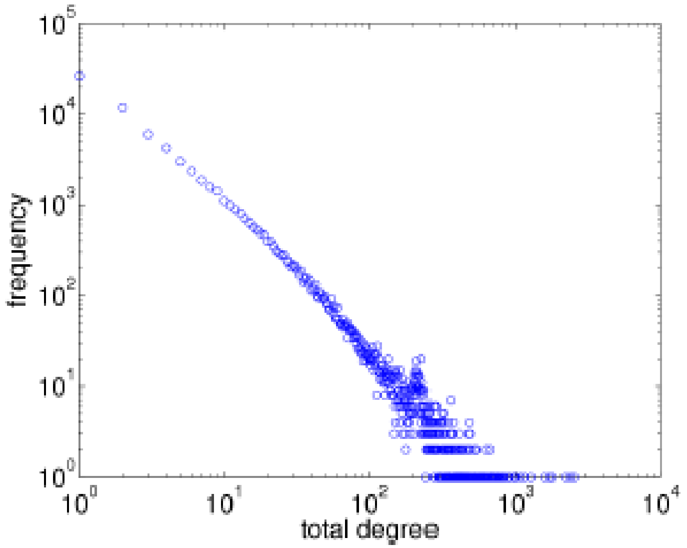

Figure 3: Logarithmic plot of the degree distribution showing that the degree distribution in the Slashdot Zoo follows a power law. The limit of 200 friends and foes is visible.

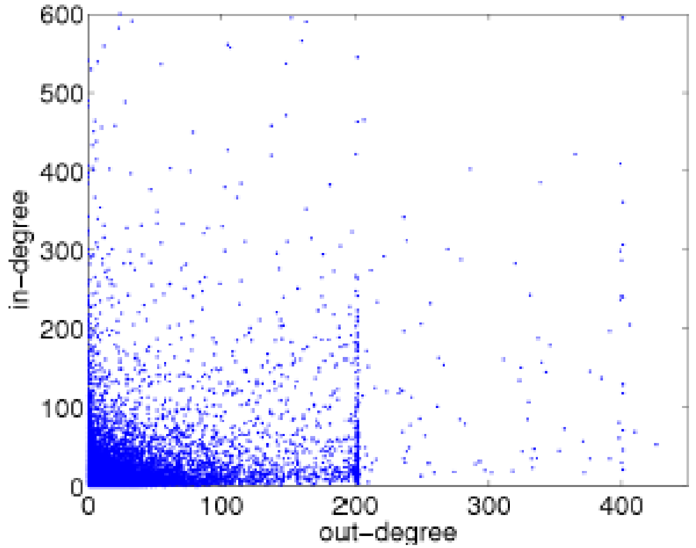

Figure 4: The in-degree plotted against the out-degree. The limits of 200 and 400 friends and foes are visible.

Clustering Coefficient

- To test the hypothesis of multiplicative transitivity, we study the whole network to see the extent of applicability of the multiplication rule.
- We bring the concept of signed clustering coefficient, and relative clustering coefficient to see if our multiplicative transitivity assumption is justified.
- **Multiplicative Transitivity:** A signed network exhibits multiplicative transitivity when any two incident edges tend to be completed by a third edge having as a weight the product of the two edges' weights.

Popularity & Centrality

- This part of the analysis helps us understand nodes that are central to the graph.
- **Various Popularity & Centrality Measures:**
  - Fans Minus Freaks (FMF)
  - PageRank (PR)
  - Signed Spectral Ranking (SR)
  - Signed Symmetric Spectral Ranking (SSR)
  - Negative Rank (NR)

Link Prediction

- We focus on prediction of the **sign** of the edges between two nodes, given that is the most relevant information.

- **Baseline Algorithms:** The following three baseline algorithms are tested.
  - **(1) –** Always predict a positive edge.
  - **(A^T)** – If there is an edge in the opp. direction, predict the sign of the edge. Else, a positive edge.
  - **(A^2)** - Use the squared adjacency matrix for prediction. Makes use of the multiplicative transitivity logic.

Table 5: The three baseline algorithms for link sign prediction. The accuracy is measured on a scale from −1 to +1. Greater values denote higher prediction accuracy.

1	0.517
A^T	0.536
A^2	0.552

Link Prediction: Algebraic Similarity Measures

- Corpus size makes A^2 and A^3 computationally expensive. Thereby, we look at alternative **algebraic similarity measures**.
- **Dimensionality Reduction (A):** The matrix A can be reduced dimensionally by performing a sparse singular value decomposition, resulting in a low-rank approximation of the original matrix:

$$A \approx A_k = U_k D_k V_k^T$$

- **Symmetric Dimensionality Reduction (A sym):** We apply dimensionality reduction to the symmetric matrix A+AT . In the case of symmetric matrices, we use the eigenvalue decomposition:

$$A + A^T \approx U_k D_k U_k^T$$

Link Prediction: Algebraic Similarity Measures Graph

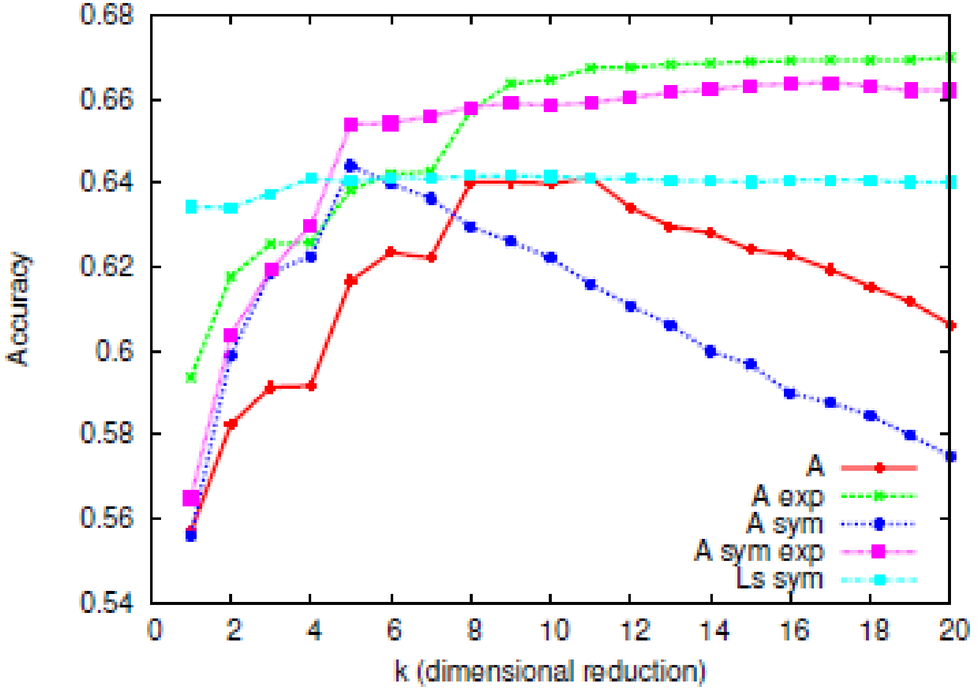

Figure 11: The accuracy of the various algebraic similarity measures on the link sign prediction task. Each similarity measure is tested using a varying dimensional reduction parameter k. Greater values denote higher prediction accuracy.

Link Prediction: Experimental Evaluation

- We split the edges into a **training set (70%)** and **testing set (30%).**
- Accuracy is predicted based on matching parity.
- A thorough study was performed on multiple kernels based on the this accuracy score and the best kernel was reported.

Conclusion

- Our analysis on social network graphs with negative weights is broken down to three levels:
  - On the global level, we defined the signed clustering coefficient and the relative signed clustering coefficient.
  - On the node level, we defined Negative Rank. A new popularity measure that can identify trolls.
  - On the link level, using various signed spectral similarity measures, we studied for the task of link sign prediction. This could be useful in implementation as troll detection.
- We concluded that the network exhibits multiplicative transitivity, as described by ***“the enemy of my enemy is my friend”***.
- We showed that these methods for analyzing a network bring out results that common, unsigned techniques can’t.

Prospective Approach

Tentative Plan

- Hope to complete coding the paper and get results by a couple of weeks.
- We are thinking about extending the work to incorporate more social network analysis methods spanning various centrality measures and not just PageRank.
- We could also propose a more robust kernel for the sake of prediction.

Thank you.