

The Slashdot Zoo

Mining a Social Network with Negative Edges

Jérôme Kunegis, Andreas Lommatzsch & Christian Bauckhage
DAI-Labor, Technische Universität Berlin, Germany
18th Int. World Wide Web Conference, Madrid 2009



CC IRML
Information Retrieval
& Machine Learning



Outline

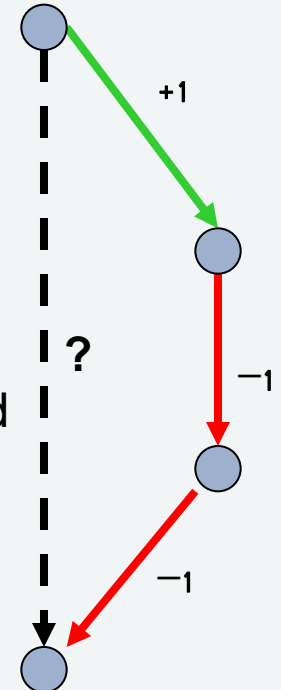
- Signed networks and the multiplication rule
- The Slashdot Zoo

- Analysis at global level: Signed clustering coefficient
- Analysis at node level: Trust and troll detection
- Analysis at edge level: Link sign prediction

- Conclusion

Signed Networks and the Multiplication Rule

- Signed social networks have **enemy** relations in addition to **friend** relations
- Assumption: *The enemy of my enemy is my friend*
 - See e.g. (Hage&Harary 1983)
 - Assumption of structural balance (Harary 1953)
- Mathematical formulation: if edges are weighted by ± 1 , relationships between unconnected nodes may be predicted by multiplying the weights along a path
- Study a social network with negative edges
 - At the global level
 - At the node level
 - At the edge level



- Technology news website founded in 1997 by Rob Malda (a.k.a. CmdrTaco)
- Powered by Slash (http://slashcode.org/)
- Features: user accounts, threads, moderation, tags, journals and the **zoo** (and more)

The screenshot shows the Slashdot homepage with a dark green header. The main navigation bar includes 'Log In', 'Create Account', 'Help', 'Subscribe', and 'Firehose'. A sidebar on the left lists various sections like 'Main', 'Apple', 'AskSlashdot', 'Backlash', 'Books', 'Developers', 'Games', 'Hardware', 'IT', 'Idle', 'Index', 'Interviews', 'Linux', 'Mobile', 'Politics', 'Science', 'Technology', 'YRO', 'Help', 'FAQ', 'Bugs', 'Stories', 'Old Stories', and 'Old Polls'. The main content area features several article snippets, including one about Sony recalling 73,000 Vaio laptops and another about insects and dinosaur extinction.

This screenshot shows a thread titled 'Chrome iPhone' with a score of 5 and 'Funny' tag. The post is by user 'gldhack' on Wednesday, September 03, at 09:55 PM. The content of the post is partially visible, mentioning 'Stick Chrome with iPhone and you can run them stories to fill up...'. There is a 'Reply to This' button below the post.

This screenshot shows a reply thread titled 'Re:Firefox Damage Control Is More Than Enough' with a score of 5 and 'Funny' tag. The post is by 'Anonymous Coward' on Wednesday, September 03, at 10:45 AM. The content includes the text 'Forget the iPhone. The AMount of dAmage conTROL a the Net.' and 'There. You said it all there.' There are 'Reply to This' and 'Parent' buttons.

This screenshot shows a user profile page for 'Friends of turg (19864)'. The page has a dark green header with navigation links for 'Info', 'Journal', 'Firehose', 'Friends', 'Fans', 'Foes', 'Freaks', 'Tags', and 'Bookmarks'. Below the header, there are sections for 'Friends of Friends' and 'Foes of Friends'. The main content area is a table with columns for 'User', 'User's...', and 'Last Journal'. The 'User' column lists various usernames like 'Helix1', 'AB3A', 'Abm0raz', etc. The 'User's...' column shows relationships like 'friends' and 'fans'. The 'Last Journal' column shows dates and times of journal entries.

The Slashdot Zoo

Slashdot Zoo: Tag users as **friends** and **foes**

You are the fan of your friends and the freak of your foes.

Graph has **two** types of edges: friendship and enmity

Your Relationship with eldavojohn (898314) ●

Info | **Relation** | Journal | Firehose | Friends | Fans | Foes | Freaks

[Friends of Friends](#) | [Foes of Friends](#)

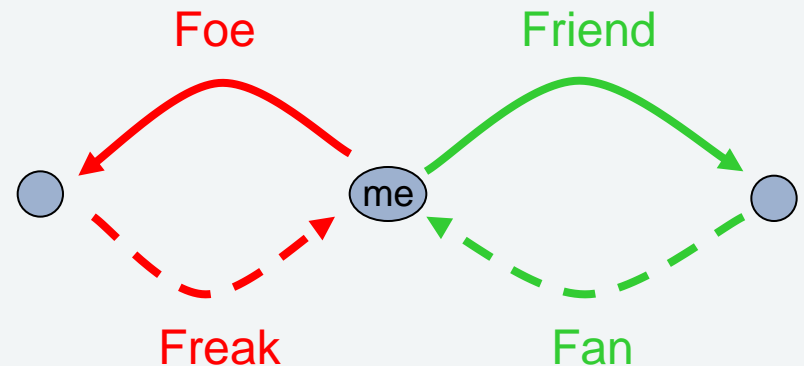
Friend

Neutral

Yup, I'm positive

Foe

Change this?

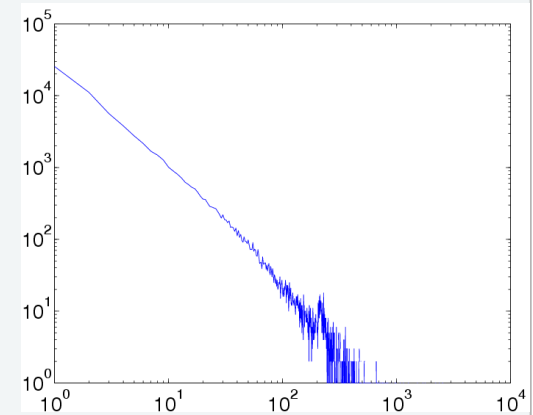


The resulting graph is sparse, square, asymmetric and has signed edge weights.

Statistics about the Slashdot Zoo

Statistics about the **giant connected component**:

- 77,985 users
- 510,157 endorsements (388,190 friends / 122,967 foes)
- 75.9% of all endorsements are positive
- Sparsity: 0.00839% of all possible edges exist
- Mean links per user: 6.54 (4.98 friends / 1.56 foes)
- Median number of links: 3
- Diameter = 6, Radius = 3
- Degree distribution: power law



Famous (and Popular?) Slashdotters

From left to right:

- CmdrTaco (Rob Malda, founder/editor)
- John Carmack (Quake, Doom, etc.)
- Bruce Perens (Debian, Open Source Definition)
- CleverNickName (Wil Wheaton, Star Trek)

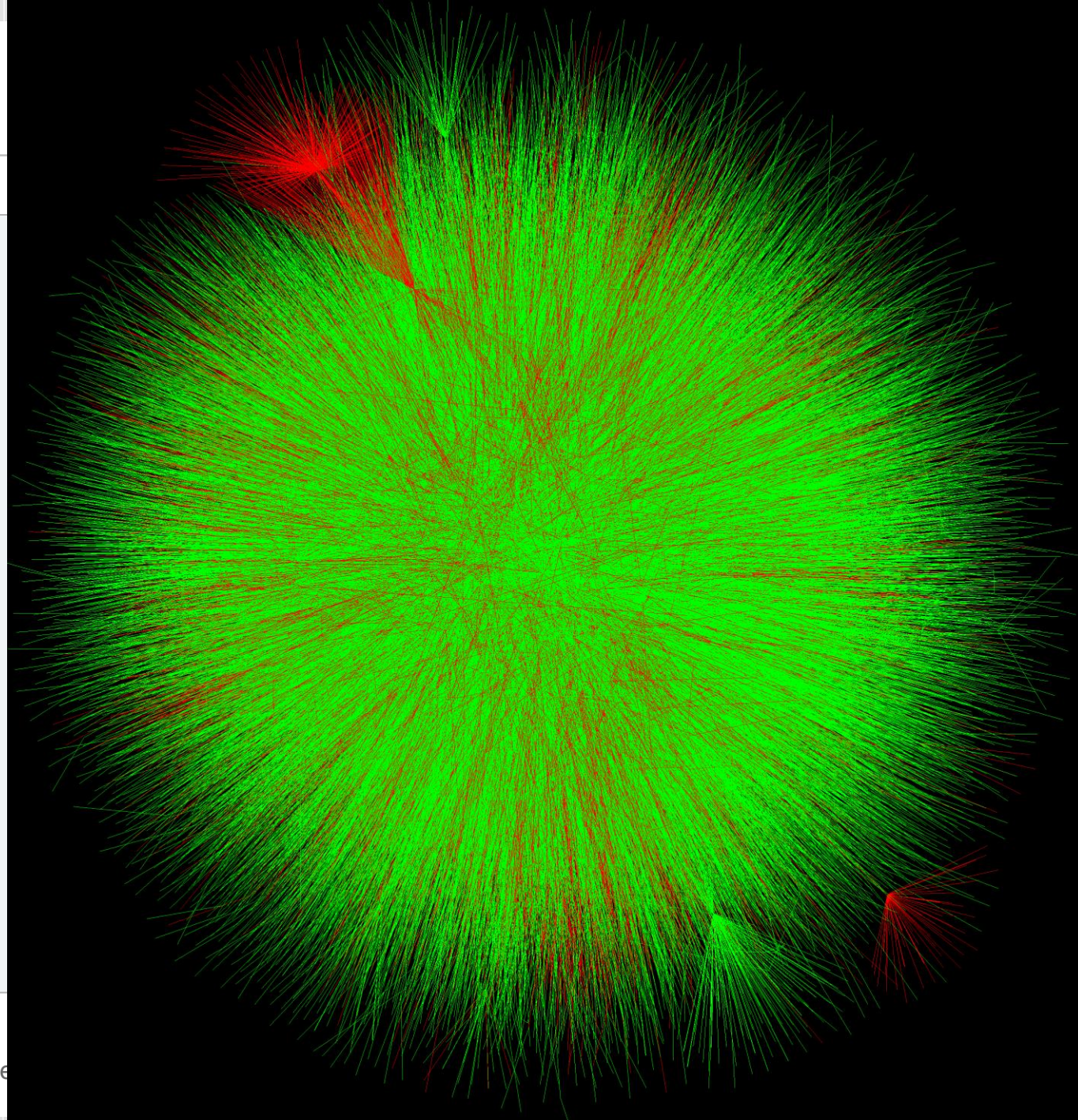


The Slashdot Zoo

GREEN: friend link

RED: foe link

Centered at
CmdrTaco



Analysis at the Global Level: The Clustering Coefficient

Characteristic number of a network, $0 \leq C \leq 1$ (Watts & Strogatz, 1998)

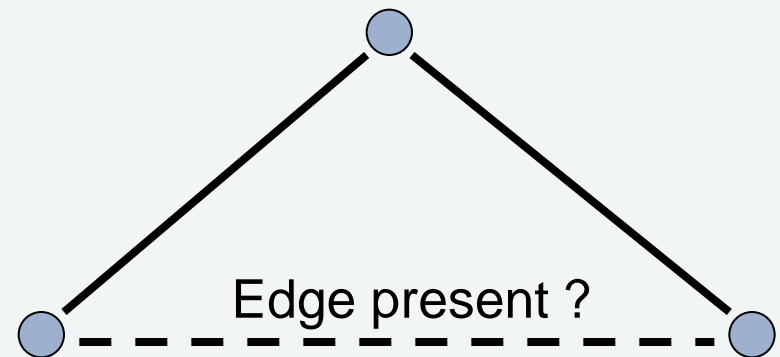
Def.: Percentage of incident edge pairs completed by an edge to form a triangle.

$$C = |\hat{A} \circ \hat{A}^2|_+ / |\hat{A}^2|_+ \quad \hat{A} = \text{abs}(a)$$

High clustering coefficient: clustered graph with many cliques. (Graph is clustered when the value higher than that predicted by random graph models.)

Slashdot Zoo has $C = 3.22\%$
(vs. 0.0095% random)

Also works for directed graphs.



Signed Clustering Coefficient

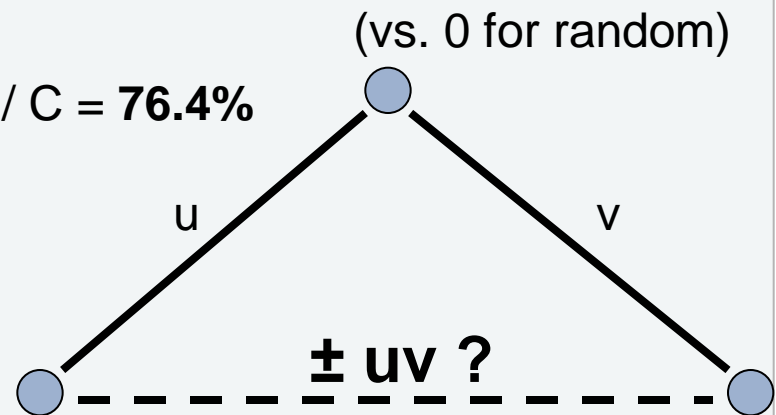
“The enemy of my enemy is my friend”

→ multiplication rule

- Denote the amount to which the network is balanced by counting “wrongly” signed edges negatively

$$C_S = |A \circ A^2|_+ / |\hat{A}^2|_+$$

- Range: $-1 \leq C_S \leq +1$
- Slashdot Zoo has $C_S = +2.46\%$
- Relative signed clustering coefficient: $C_S / C = 76.4\%$



Analysis at the Node Level: Centrality

Measures that apply to single nodes: centrality, trust, reputation, etc.

Simple functions:

- Fan count minus freak count

Algebraic function:

- PageRank, ignoring edge sign (Page&Brin 1998)
- Signed variants of PageRank, e.g. (Kamvar 2003)

PageRank and the Multiplication Rule

- Let A be the network's adjacency matrix
- Compute PageRank by iterated multiplication of a vector with normalized, A (+ extra node for teleportation)
- Result: Dominant eigenvector of matrix A given by repeated multiplication with A
- Implicit assumption: powers of A denote relations in the network

Matrix multiplication:

$$(AA)_{ij} = \sum_k A_{ik} A_{kj}$$

Observation: Matrix multiplication relies on edge weight products

Thus: **Methods based on matrix multiplication assume the validity of the multiplication rule.**

Top Users

- For each user score, show the top 6

	#1	#2	#3	#4	#5	#6
Fans minus Freaks	CleverNickName	Bruce Perens	CmdrTaco	John Carmack	NewYorkCountryLawyer	\$\$\$\$\$exyGal
PageRank	FortKnox	SamTheButcher	Ethelred Unraed	turg	Some Woman	gmhowell
Signed PageRank	FortKnox	SamTheButcher	turg	Some Woman	Ethelred Unraed	gmhowell

Conclusion:
Fans minus Freaks denotes prominence,
PageRank denotes community.

Detecting Trolls

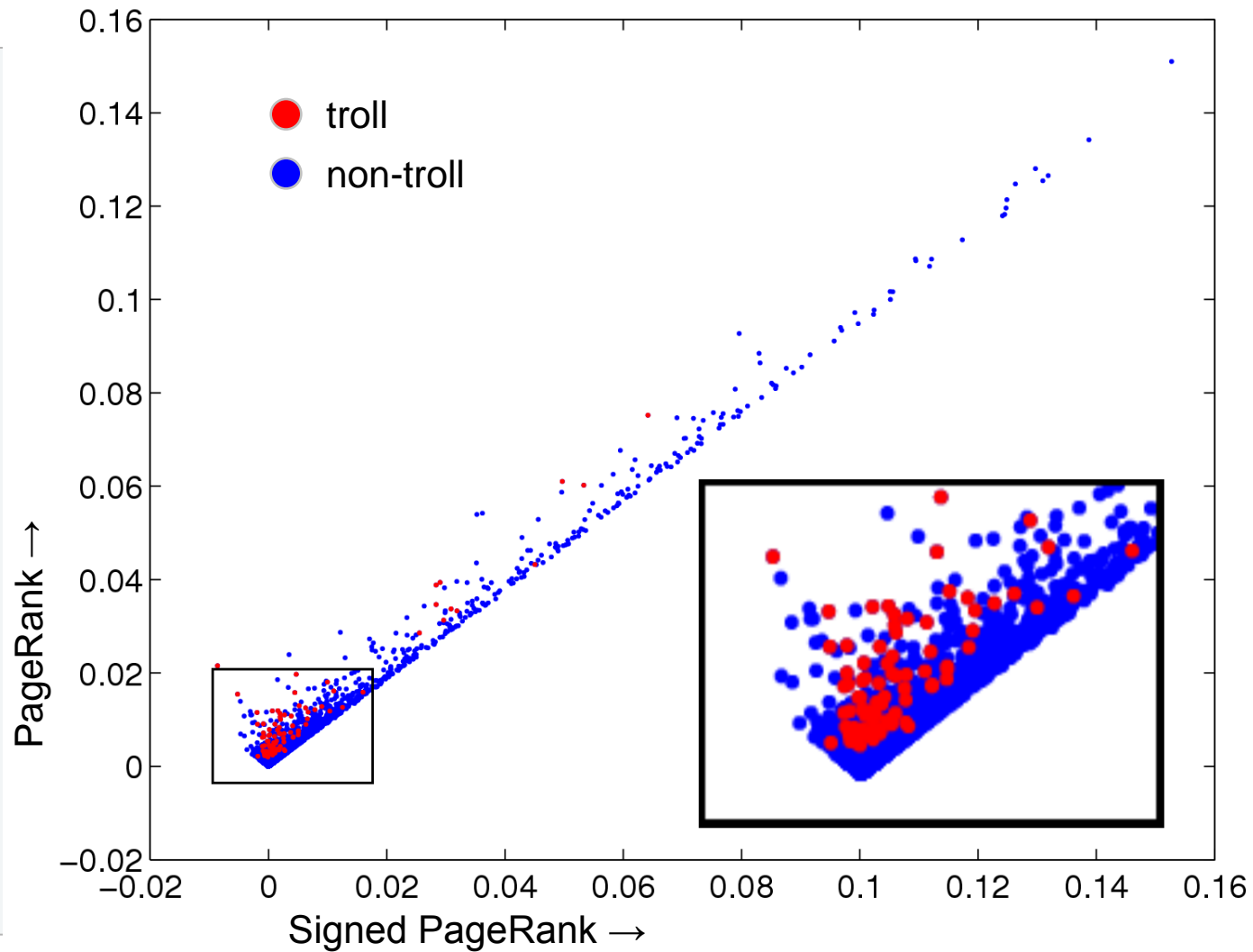
trolling, n. posting disruptive, false or offensive information to fool and provoke readers

- Slashdot is known for its **trolls**
- Task: Predict foes of blacklist “No More Trolls” (162 names^[1])

PhysicsGenius
Profane Motherfucker
Klerck ObviousGuy CmderTaco
YourMissionForToday
\$\$\$\$\$exyGal IN SOVIET RUSSIA
SexyKellyOsbourne BankofAmerica_ATM
jOnkatz CmdrTaco (editor) spintgaked
TrollBurger CmdrTaco (troll)
Twirlip of the Mists

[1] See <http://slashdot.org/~No+More+Trolls/foes/>

PageRank for Trolls



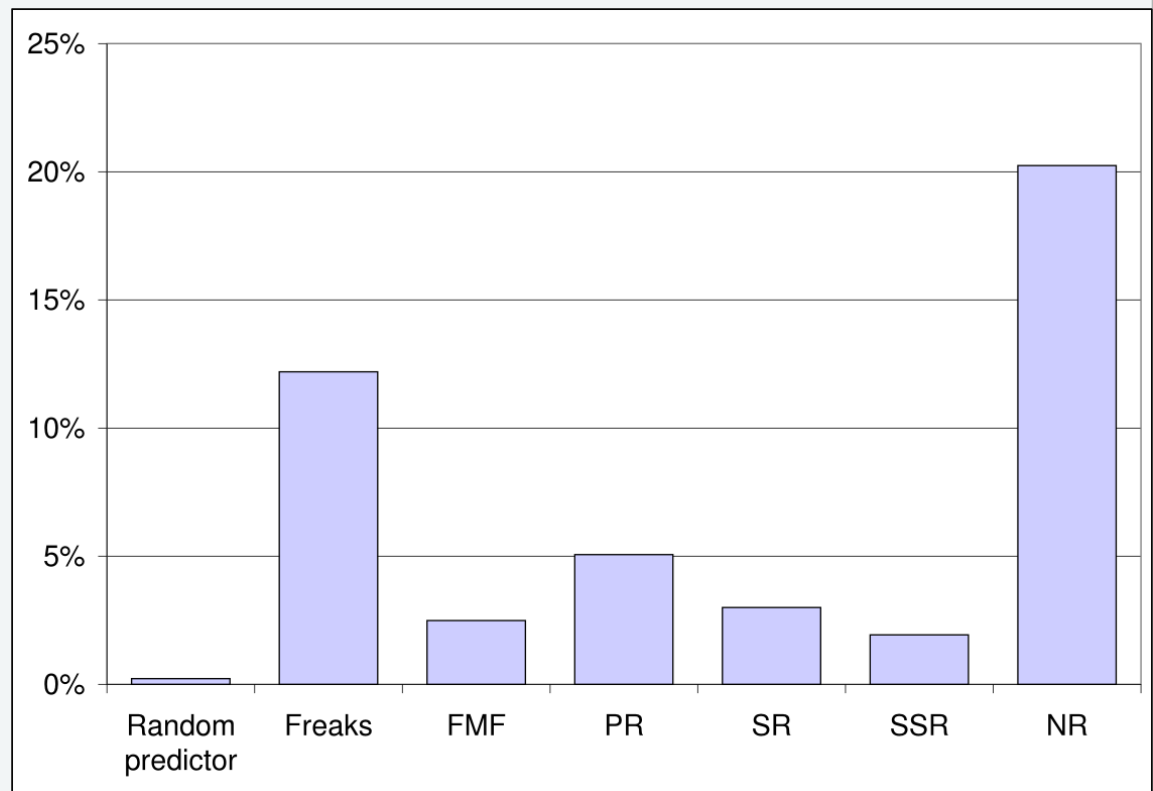
Negative Rank

- Observations:
 - PageRank and signed PageRank are almost equal for most users
 - For trolls, signed PageRank is less
- Conclusion:
 - Define $\text{NegativeRank} = \text{Signed PageRank} - \text{PageRank}$

How does Negative Rank perform at troll prediction?

Performance at Prediction

- Mean average precision (MAP) at troll prediction
- Negative Rank works best!



Analysis at the Edge Level: Link Sign Prediction

Task: Predict the *sign* of links

- Use the adjacency matrix $A \in \{-1, 0, +1\}^{n \times n}$
- Powers of A implement the multiplication rule

Simple algorithms

- Mutual friendship (A^T)
- Signed triangle completion (A^2)

Algebraic algorithms

- Rank reduction (A)
- Symmetric dimensionality reduction (A sym)
- Matrix exponential (A exp)
- Symmetric matrix exponential (A sym exp)
- Inverted signed Laplacian (L s sym)

Algebraic Link Prediction Algorithms

Compute the rank-reduced eigenvalue decomposition:

$$A = U_k D_k U_k^T$$

Matrix exponential:

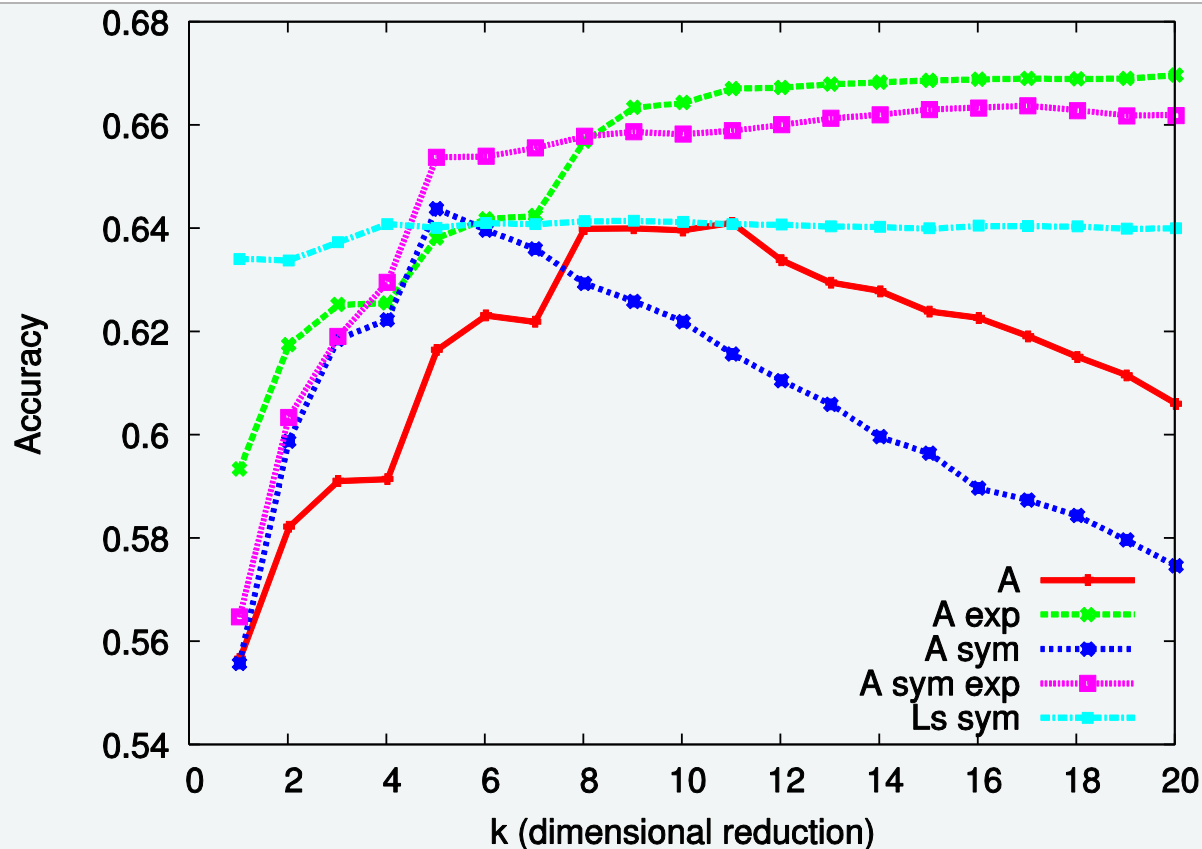
(multiplication rule)

$$\begin{aligned} \exp(A) &= U_k \exp(D_k) U_k^T \\ &= I + A + 1/2 A^2 + 1/6 A^3 + \dots \end{aligned}$$

Inverted signed Laplacian (Kunegis 2008):

$$L^+ = (D - A)^+$$

Evaluation Results



Accuracy is measured on a scale from -1 to +1.

1	0.517
A^T	0.536
A^2	0.552

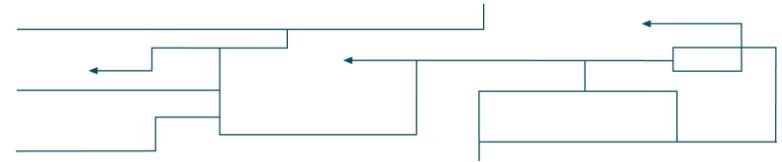
Best link sign prediction: matrix exponential, confirms multiplication rule

Ongoing and Future Work

- More datasets
 - Essembly.org, Epinions.com, LibimSeTi.cz
- Study Negative Rank in detail
- Social networks with semantic relationships (more than two types)
- Other networks that can be extended to negative values
 - Folksonomies with negative tags (e.g. tags like *!funny* on Slashdot)

Conclusions

- Multiplication rule confirmed at global, nodal and relational scale
- The foe relationship can be used for trust computation and link sign prediction in social networks
- New concepts:
 - Signed clustering coefficient
 - Negative Rank
 - Link sign prediction in signed networks



Thank You



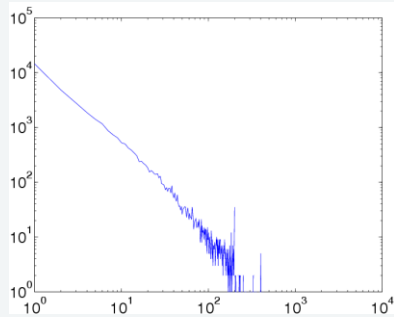
CC IRML
Information Retrieval
& Machine Learning



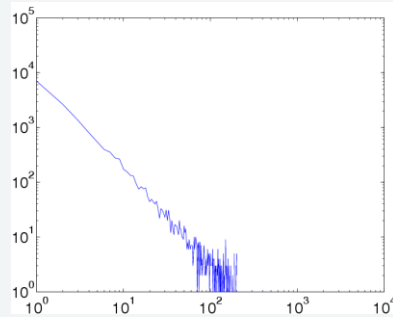
References

- S. Brin and L. Page. *The anatomy of a large-scale hypertextual Web search engine*, Proc. Int. Conf. on World Wide Web, pages 107–117, 1998.
- P. Hage and F. Harary. *Structural models in anthropology*, Cambridge University Press, 1983.
- F. Harary. *On the notion of balance of a signed graph*, Michigan Math. J., 2:143–146, 1953.
- S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. *The EigenTrust algorithm for reputation management in P2P networks*, Proc. Int. Conf. on World Wide Web, pages 640–651, 2003.
- J. Kunegis, S. Schmidt, C. Bauckhage, M. Mehlitz and S. Albayrak. *Modeling Collaborative Similarity with the Signed Resistance Distance Kernel*, Proc. Eur. Conf. On Artificial Intelligence, pages 261–265, 2008.
- D. J. Watts and S. H. Strogatz. *Collective dynamics in ‘small-world’ networks*, Nature 393(6684):440–442, 1998.

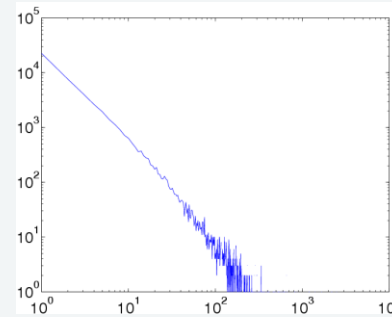
Appendix -- Degree Distributions



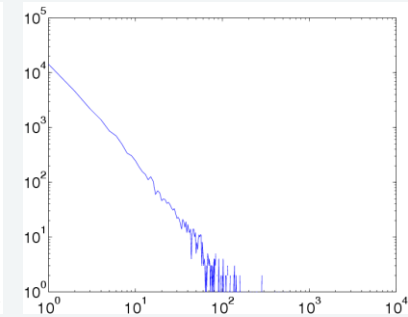
Friends



Foes

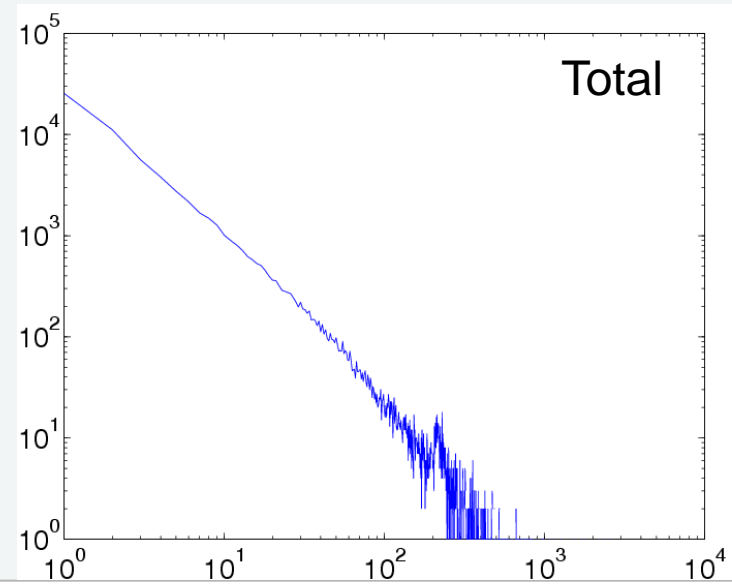


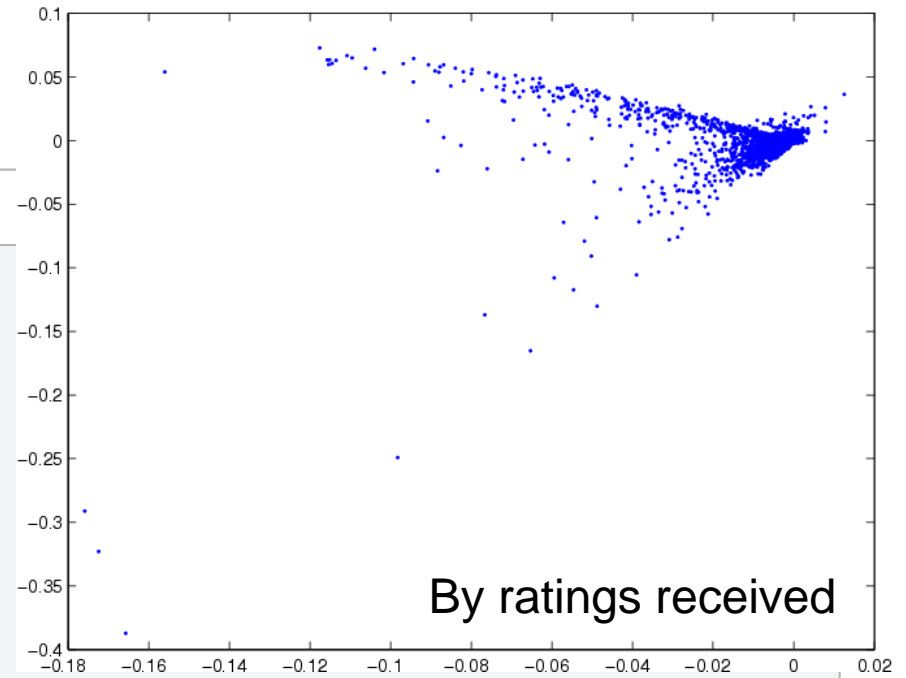
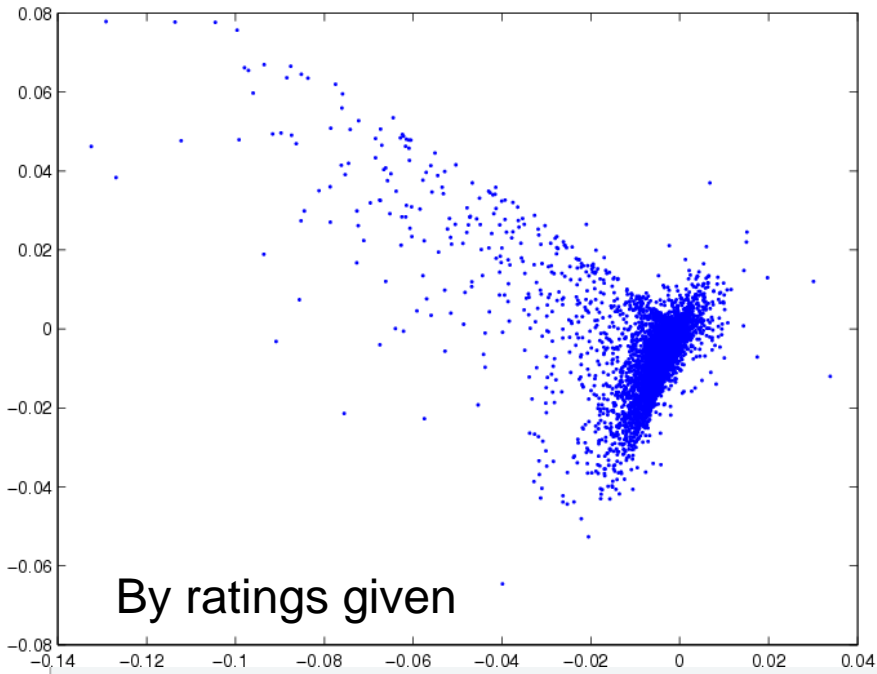
Fans



Freaks

• **Observation: power laws for all four relationship types**





Appendix -- Principal Component Analysis

