

How Influential Are You: Detecting Influential Bloggers in a Blogging Community

Imrul Kayes, Xiaoning Qian,
John Skvoretz*, Adriana Iamnitchi

Department of Computer Science and Engineering, USF

*Department of Sociology, USF



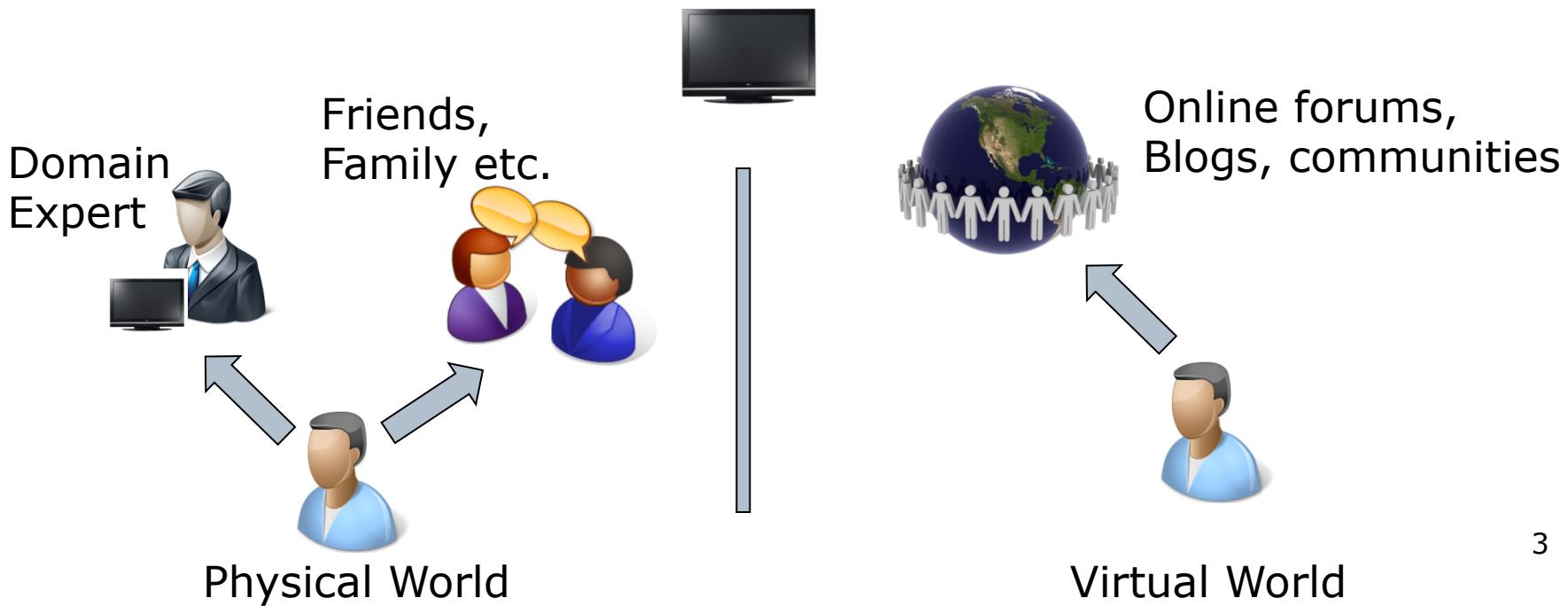
Internet Blogs

- Citizen journalism of Bloggers: Arab spring, Tsunami, Katrina
- WordPress:
 - used by 14.7% of “top 1 million” websites
 - manages 22% of all new websites
- Bloggers quit jobs

The screenshot shows the ProBlogger website. At the top, there's a navigation bar with links for HOME, BLOG, FORUM, WORKBOOKS, BOOK, JOBS, MAKE MONEY, and ARCHIVES. Below the navigation, a post is displayed. The post is written on 6/13/2011 at 6:05 am by [Guest Blogger](#). The title of the post is **How I Quit My Job for Blogging**. Below the title, it says "Filed Under: [Blogging for Dollars](#)". There are social sharing buttons for Facebook ("Like" button with 34 likes) and Twitter ("Tweet" button with 267 tweets). A note below the sharing buttons states, "This guest post is by Joshu Thomas of [NapIncome.com](#)." The main content of the post begins with the sentence, "Blogging is no more a mere hobby; it has become a full-time profession for many bloggers, including myself. Today I will share my journey as a blogger and how it helps me to earn a full-time income blissfully working from anywhere I like." At the bottom of the post, there's a section titled "How it started".

Bloggers Matter

- 68% claimed to be influenced by other blogs
- 38% talk about brands on their blogs
- 64% companies shift focus to blogging



Measuring Influence of Bloggers

- Is online influence measureable?
- Are active bloggers influential?
- Are influential bloggers active?
- Does user location play any role?

How to use location(centrality) to measure potential influence of a user in the network?

Centrality Aggregation Technique

- Aggregation of different representative centrality metrics into a final influence score
$$\text{Influence Score} = f(\text{Network Centralities})$$
- Influence score is the average centrality scores over centrality metrics
- Representative centralities
 - Degree
 - Betweenness
 - Closeness
 - Eigenvector
 - Hub
 - Communicability

BlogCatalog

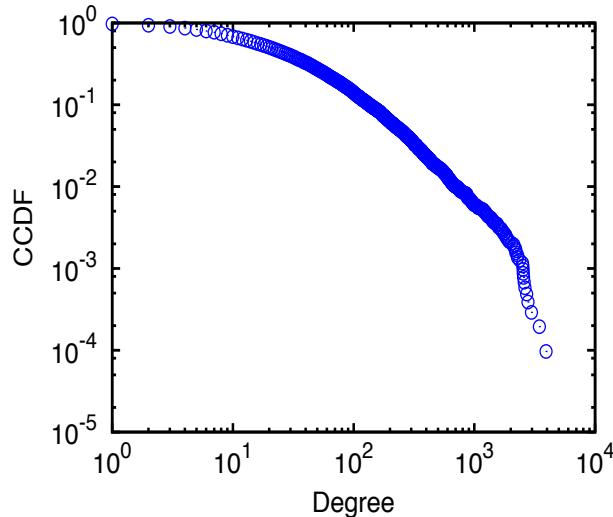
- Blogging service
 - Online profiles, blog posts
 - Friendship declaration (OSN)
 - Bloggers receive blogging updates from their friend
- OSN: 10,312 nodes and 333,983 edges

Comparison of the BlogCatalog Dataset

Network	Type	Nodes	Avg. Path Len.	Radius	Diameter	CC
BlogCatalog	Blog	10,312	2.38	3	5	0.460
Erdos-Renyi	Random	10,312	2.65	3	3	0.006
LiveJournal	Blog	5,284,457	5.88	12	20	0.330
Orkut	Social	3,072,441	4.25	6	9	0.171

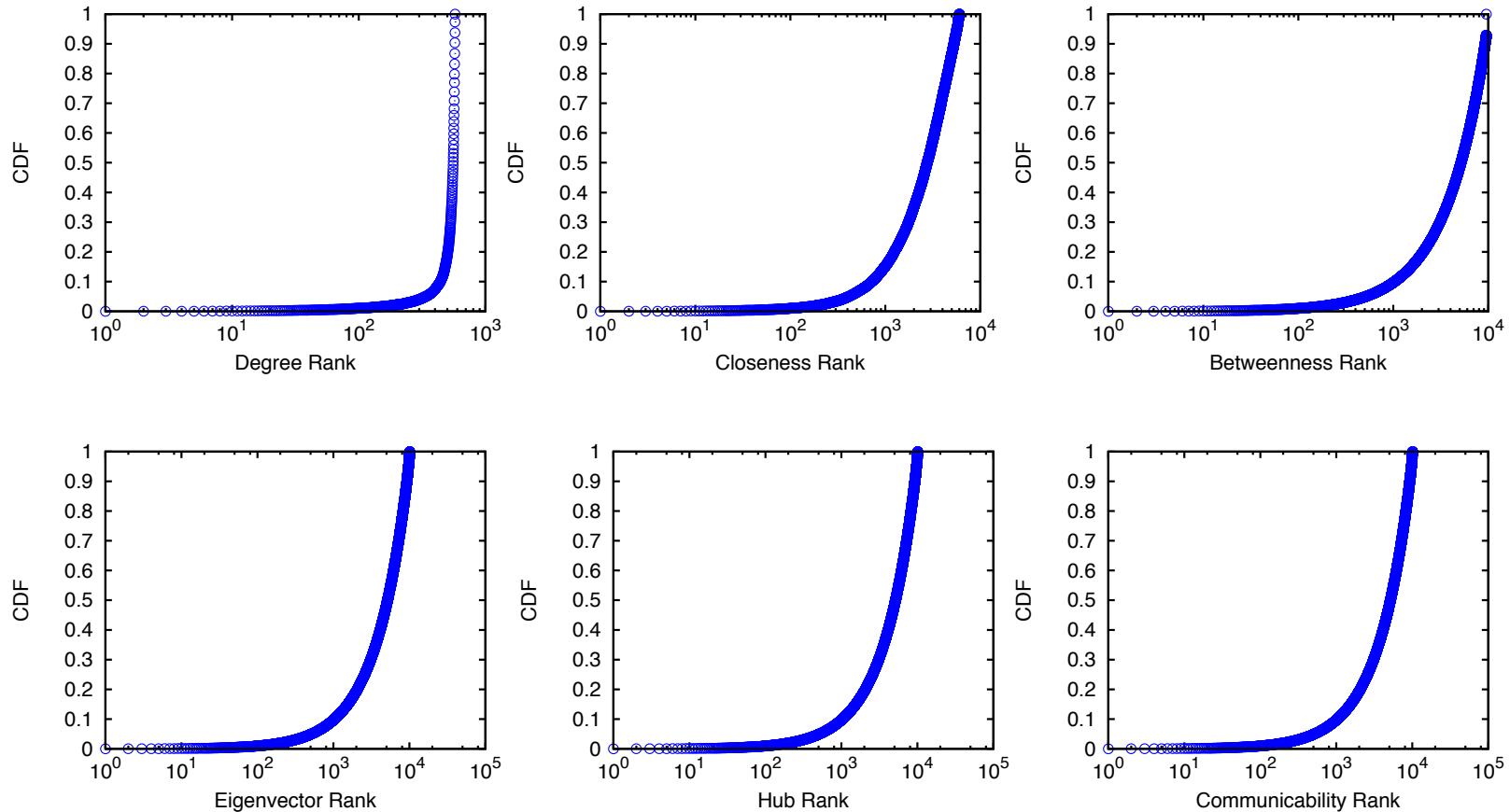
- High clustering coefficient in BlogCatalog and LiveJournal Blogs
- The average path length is small (2.38)
- BlogCatalog network is a small-world graph

BlogCatalog is Scale-free



- Degree distribution fits a power-law distribution with exponent $\alpha = 2.52$.
- Hubs control the “connectedness” of the network
 - **Do influential bloggers act as hubs?**

Centralities are granular

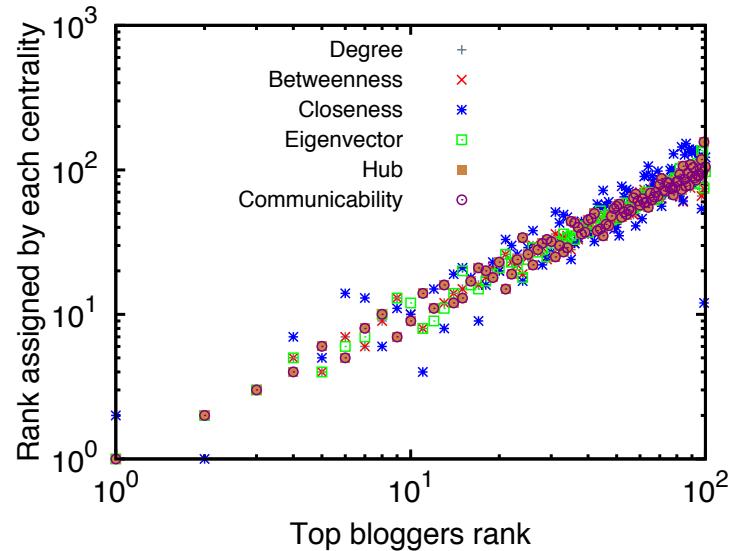
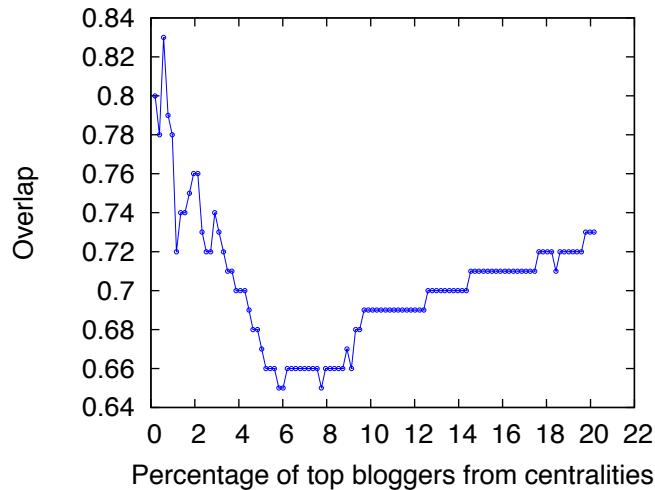


Top Bloggers are Common among Centralities

DC	4839	176	4374	8157	1226	4997	4984	8859	645	446	7098	7806	3198	2521	667
BC	176	4839	4374	8859	8157	645	1226	7806	233	446	3198	1932	4997	4984	7098
CC	4839	176	4374	8157	1226	4984	4997	8859	7098	645	7806	446	3198	2521	233
EC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669
HC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669
CoC	4839	176	4374	1226	4984	8157	3198	4997	446	645	7098	2521	667	8859	4669

Bloggers' ID	Average Rank
4839	1.17
176	1.83
4374	3.00
1226	4.83
8157	5.17
4984	7.00
4997	8.33
645	9.17
3198	9.17
446	9.83

Correlation of Centralities

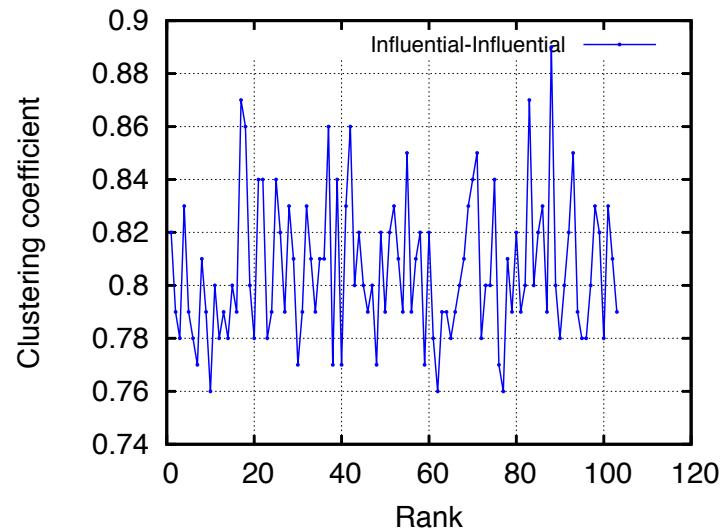
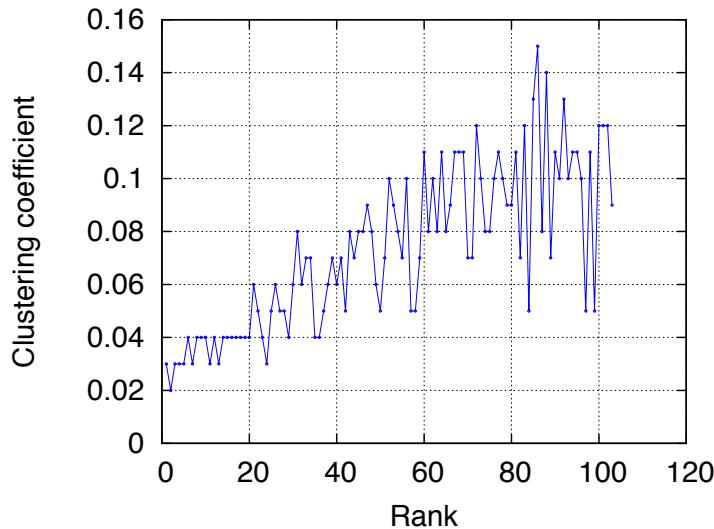


- High fraction (0.65~0.83) of top bloggers are common among centralities
- Centralities tend to rank the same bloggers in the top

Correlation Coefficient Matrix

	Degree	Betweenness	Closeness	Eigenvector	Hub	Communicability
Degree	1.00	0.67	0.65	0.68	0.68	0.67
Betweenness	0.67	1.00	0.85	0.89	0.89	0.88
Closeness	0.65	0.85	1.00	0.98	0.98	0.97
Eigenvector	0.68	0.89	0.98	1.00	1.00	0.98
Hub	0.68	0.89	0.98	1.00	1.00	0.98
Communicability	0.67	0.88	0.97	0.98	0.98	1.00

Central Bloggers are Themselves Highly Connected



- ❑ The average clustering coefficient of the top 1% most central bloggers **0.07**
- ❑ Considering central -central ties Average CC **0.81**
- ❑ Network average CC **0.46**

Central Bloggers Tend to Connect to Central Bloggers

Subnetwork Definition	Assortativity
Entire network	-0.25
Subnetwork of bloggers excluding top 1% most central bloggers	-0.67
Subnetwork of top 1% most central bloggers	-0.16
Subnetwork of top 1% central bloggers, considering ties among them	+0.07

Summary

- Influence measurement
 - Centrality aggregation technique to measure relative influence of bloggers
- Learning from BlogCatalog
 - The six network centrality metrics are highly correlated in this community
 - Core-periphery structure of the network
 - ❖ Central bloggers form a densely connected core
 - ❖ Non-Central bloggers remain at the periphery, less likely to connect to each other.

Thank you!

Questions
?