Finding Influential Authors in Brand-Page Communities

[1]Hemant Purohit, [2]Jitendra Ajmera, [2]Ashish Verma, [2]Sachindra Joshi, [1]Amit Sheth

[1]Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis), Wright State University, USA

[2]IBM Research Lab, India

Motivation

- Enterprises are increasingly using social media forums to engage with their customers online a phenomenon known as *Social Customer Relation Management (Social CRM)*.

- In such a brand-page community, **who are the 'influential' users for an enterprise to engage with, on priority basis?**

Whom to engage with first??

Experimental Network: Twitter Microblogging Service

A big challenge is the sparsity in explicit user connections.

Data Collection

Dataset	Crawl Duration	#Tweets	#Brand related tweets	#Authors	#Isolated authors	Explicit Connectivity	#Non-mutual edges
Dateset1	Nov27-Dec17, 2011	2.3M	26K	14k	5k	.05%	49k
Dataset2	Dec27-Dec30, 2011	3.3M	38K	19k	6k	.017%	45k

Evaluation Criterion

If 'X follows Y' and 'Y does not follow X', then X is less influential than Y.

Related Work

- **Static Link Analysis Methods** (PageRank/HITS) [Jeng et. al, WSDM'10]
- **User Attribute based Mining** [Pal et. al, WSDM'11]
- **Quantitative Measures** (#Indegree, #Retweets, #Mentions) [Cha et. al, ICWSM'10]

Proposed approach

1. Implicit Network of brand-page community users

- Create Retweet, Reply and Mention based interaction edges
- Weight of an edge as per *tf-idf* like function
- PageRank and HITS on the generated network

Challenges in Network Analysis methods alone:

- Capture author/node specific attributes effectively (e.g., #tweets written, #followers)
- Modeling the behavior between two nodes outside the network (e.g., other topic)

2. Decision Tree Modeling

- Using all of the above mentioned features (author centric, interaction) in one model

Figure shows different colours for users with different implicit (e.g., Reply) connections in an explicit network of brand-page community.

Users may be implicitly connected without an explicit edge!

Accuracy to find influentials

Set-up / Feature	Dataset1	Dataset2
Author profile feature baseline		
Number of followers	**79.8**	83.5
Activity	36.3	16.7
Klout score	68.0	**84.0**
Explicit network analysis baseline		
PageRank	77.0	90.5
Authority score (HITS)	**77.3**	92.6
Hub score	70.2	82.6
Implicit (Interaction) Network Analysis		
PageRank	76.6	87.6
Authority score	**80.0**	91.4
Hub score	72.1	73.9
Decision Tree based Analysis		
3 interaction features	79.0	90.0
Only first 2 author profile features	81.2	94.7
Interaction and author profile features	85.1	94.7
Interaction, author profile features and authority score	85.0	94.7

*Bold: Best in experiment category
*Highlighted: Best across the experiment categories

Conclusions

- **Implicit Network analysis can be a good proxy for explicit network analysis in sparse communities**
  - Positive correlation between rankings of explicit networks and implicit networks analyses
  - Comparable accuracy to find influentials

- **Combination of interaction and author profile features is a better approach to find influentials**

- **User activity and its influence does not correlate**

- **Authority score (HITS analysis) is suitable metric than the PageRank score in current context**

Applications

Any evolving community scenario where explicit connections are going to be sparse. (e.g., during a disaster event)