Characterizing the Structure of Twitter Network Through Socially-Aware Clustering of User

Eugene Tan[1], Professor Reza Rejaie PHD[1,2]

[1]University of Oregon, [2]Department of Computer and Information Sciences

Introduction

- Online Social networks (OSNs) , such as Facebook and Twitter, have been increasing in popularity in the past 10 years.
- OSNs have many societal, economical, and cultural impacts within today's society.
- Determining the structural characteristics of these OSNs is of great interest (e.g., How users interconnect to form a network).
- Characterization of OSNs is inherently difficult due to their size (e.g., Twitter has hundreds of millions of users making it difficult to capture a complete snapshot of the OSN).
- Characterizing connectivity on an OSN without considering user attributes (e.g., social role, location, etc.) is of limited value.
- Therefore, how can we characterize connectivity of a large OSN in a socially-aware manner?

Background

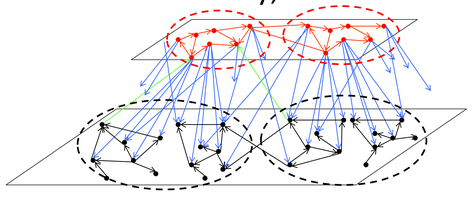
- Twitter is an OSN that has structural connectivity between its user vertices; our goal is to capture this connectivity structure.
- We label 'elite users' to be users who are highly-active or exhibit more influence than regular ('non-elite') users.
  - They compose a core subgraph of the OSN structure.
- Elite users have direct connectivity to a large number of regular users within the OSN.
- Identifying relationship patterns amongst elite users provides insight to the overall structure of the OSN.
- Elite communities are communities of elite users exhibiting a 'social cohesion' around a specific common theme.

Method

- The primary method of capturing the OSN is to capture the "Elite Network" composed of Twitter's elite users to reveal relationship and influence patterns of elite users in the OSN.
- Majority of users in an OSN have low activity or connectivity and do not contribute significantly to the structure of an OSN.
  - Twitter is a large OSN, with millions who exhibit low connectivity, thus we must find a way to capture users of importance: elite users.
- We leverage elite users to characterize the OSN, as these individuals exhibit higher levels of connectivity and are an important component of the OSN's connectivity structure.
  - It is prohibitively expensive to try and leverage regular users to characterize the OSN.
  - The connectivity between elite users is what constitutes the Elite Network.
- We capture the "Elite Network" by crawling the top 10,000 elite users of the OSN from the website SocialBakers.com (OSN statistic site).
  - To independently identify accounts which fit the criteria of an elite user, randomized walks are employed.
  - An efficient technique to identify visible and heavily influential users; ensures master list is in order.
- Elite communities form meaningful components of the OSN.
- Themes of community broadly divided to Country, Language, Cultural Interest, and Business sector as depicted in the figure:

Label	Size	Dens.	Cond.	Theme
US/PoP	2.9K	384	0.26	US celebs/actor/music
Spanish	1.9K	208	0.35	Spanish Speaking
US/Corp	1.3K	242	0.58	US Corporate/Media
Arabic	1K	698	0.13	Arabic Speaking
ID	533	93	0.34	Indonesian
BR	508	162	0.38	Brazilian
PH	475	210	0.46	Filipino
IN	335	185	0.57	Indian
TR	271	87	0.34	Turkish
Unstable	155	268	0.98	Unstable nodes
K-PoP	150	51	0.44	Korean Popstars
TH	28	34	0.63	Thai
Adult	20	57	0.48	Adult/Porn
US/TV	19	541	0.99	US TV channels
GLB/Fun	13	119	0.98	Global Entertainment

- By using elite communities, shadow partitions are formed from grouping users based upon interest with an elite community; viewed as extensions of an elite community.
- Here, directed edges indicate relationships between the regular users and elites.



- Utilizing these methods, it is possible to capture and characterize the Twitter OSN through leveraging the existence of Elite Users and their impact.

Expected Result

- An expected result is a new technique for capturing the Twitter interconnectivity structure using an elite network.
- By gathering elite users and their communities, the project will be able to characterize the Twitter OSN structure and determine structural characteristics of the OSN itself in a socially-aware manner.
- Additionally, by leveraging the elite communities, the research also expects to expand upon the partitions of regular users by clustering regular users based upon relative connectivity to the elite communities found.
- Ultimately, an accurate characterization of the OSN is the result intended to be found. As well as a socially-aware method to navigate the interconnectivity of the OSN.

References

- **Unveiling Social Characteristics of Twitter Elite Network**
  Reza Motamedi, Saed Rezayi, Reza Rejaie and Walter Willinger
  In Proceedings of *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, Spain, August 2018. [Acceptance rate 15%]
- Figures provided are from above reference.

Acknowledgments

I would like to acknowledge Professor Reza Rejaie, Christopher Misa, and Soheil Jamshidi for their extensive guidance and assistance during this study.