Social Network Influencer Ranking Based on Link Analysis

Cheulsoon Baek, Amirhossein Razavi
Myeongwook Lee, Lu Zhuang

LASSONDE SCHOOL OF ENGINEERING, YORK UNIVERSITY | 2021

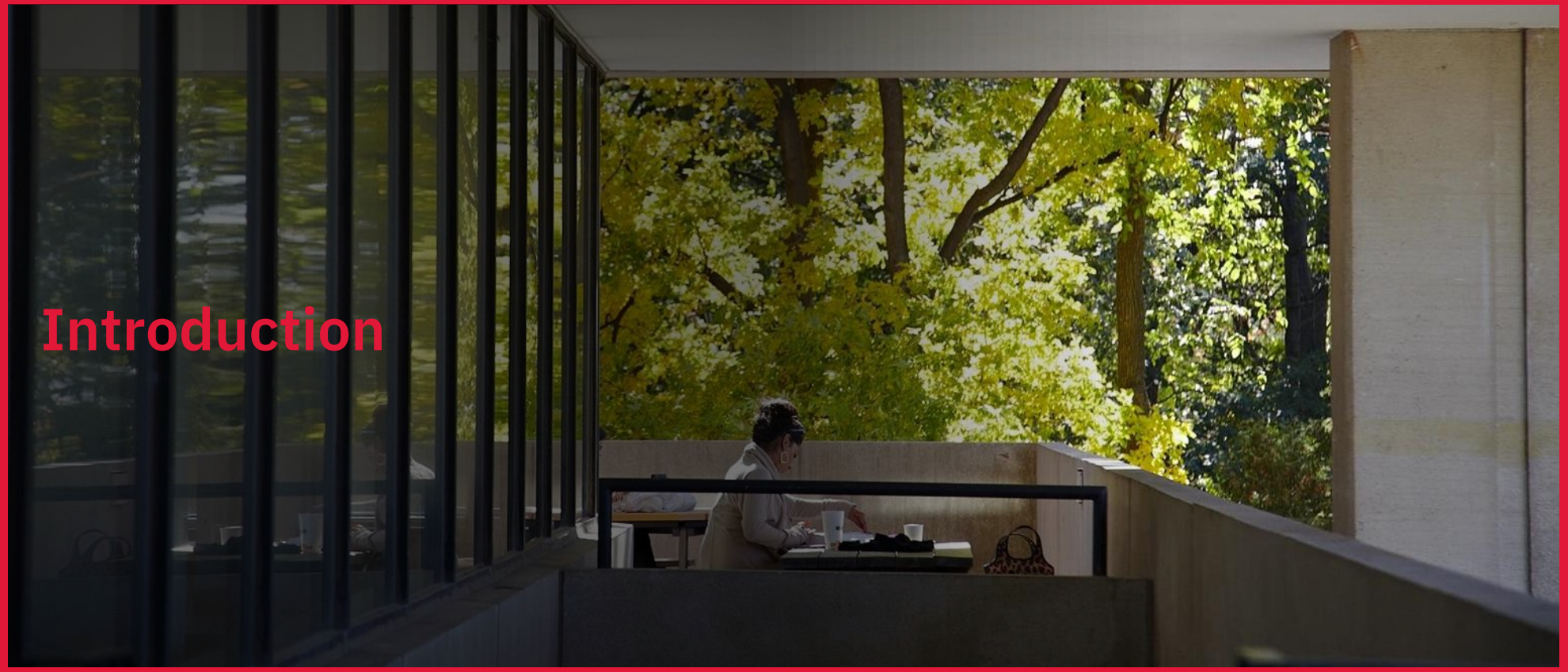
YORK U

Outline

- Introduction
- Problem Definition
- Framework
- Methodology
- Evaluation
- Conclusion
- References

YORK U

Introduction



Introduction

- What is this project about?
  -Ranking people based on their influential power
- How does this project help in the real world?
  -Marketing Industry
- What are potential applications of this project?
  -Advertising, Politics, news, etc.
- What are some of the questions that this project answers?
  - Are the number of followers always proportional to the actual influence level?
  - What kind of impact will be caused if we only considered the influencer rankings by using a single factor “followership”

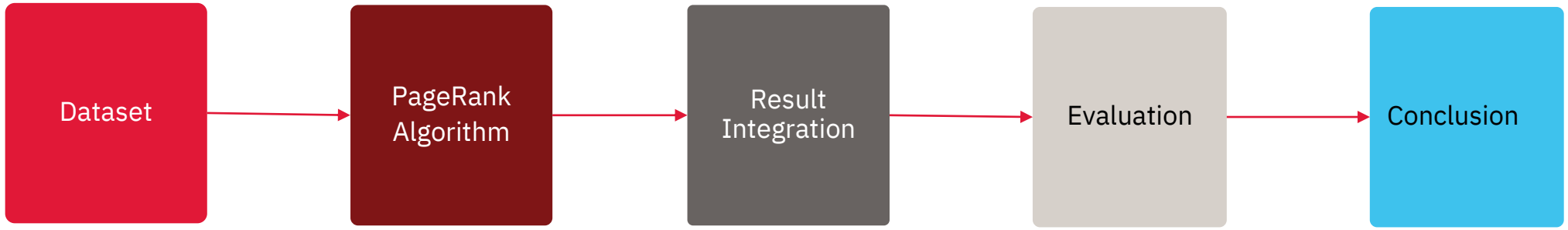
Problem Definition

Problem Definition

- **Problem 1:** Given a network from SNS services, rank the nodes based on their influential power.
- **Problem 2:** The influence of a user cannot be defined directly from the social network; instead, the other factors such as retweet, mention, and reply network should be considered.
- **Problem 3:** Given multiple networks, the method to integrate each network into one final result should be required.


Methodology

Framework


Dataset
PageRank Algorithm
Result Integration
Evaluation
Conclusion

Methodology

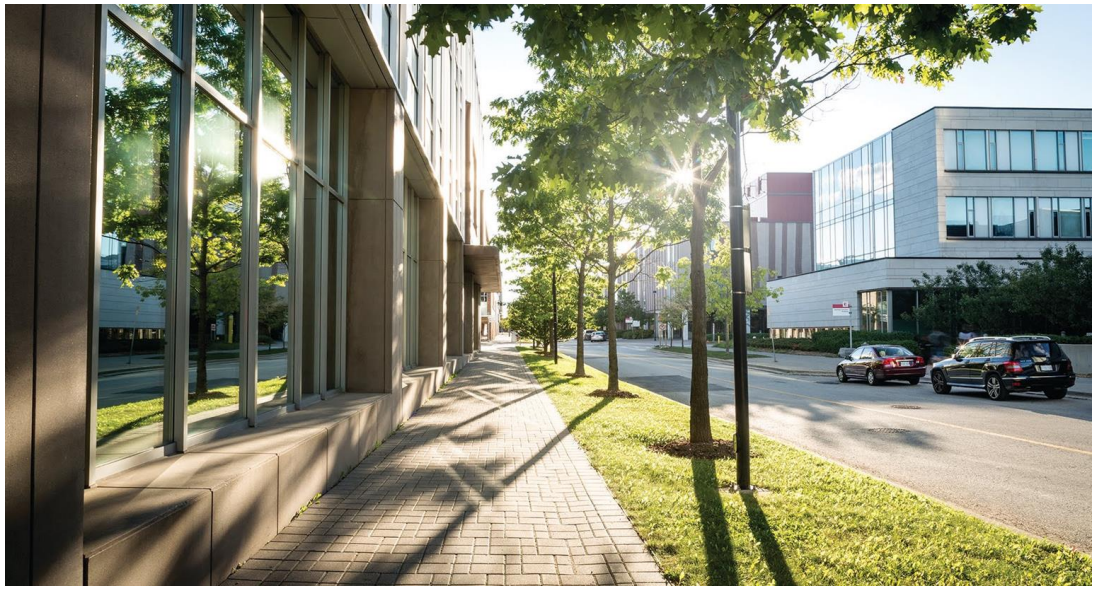
Dataset

- We chose to use a dataset called "Higgs Twitter Dataset" provided by Stanford University which is a snapshot of the Twitter network. The dataset includes Social Network (followership), Retweet Network, Reply Network and Mention Network.
- The Networks are all directed and weighted except the Social Network which is only directed.
- All user IDs have been replaced with a randomly assigned numeric ID.

Methodology

PageRank Algorithm

- What does it do?
  –PageRank Algorithm ranks the nodes based on their importance.
- PageRank Algorithm from NetworkX library was applied to the 4 networks.
- Generating 4 different dictionaries.



Methodology

PageRank Result Integration

- Each of the dictionaries was weighted so that they all could be integrated into one.
- The focus of this project was to check how the retweet, reply, and mention data affect to the influence ranking of the users.
- In total 3 types of weight distribution was used, each type using different weights for networks.

Methodology

PageRank Result Integration

- The types are as follow:

Type	Social Network	Mention Network	Retweet Network	Reply Network
1	0.6	0.1	0.1	0.2
2	0.5	0.15	0.15	0.2
3	0.4	0.2	0.2	0.2

Evaluation

Evaluation

- Influence levels calculated using just the number of followers can possibly include influencers who are not active at the time.
- We can filter these influencers and extract the list of active influencers by considering interaction factors.

Evaluation

Rank	Social Network	Mention Network	Retweet Network	Reply Network
1	1	88	88	677
2	88	677	14454	88
3	1503	2417	677	220
4	138	59195	1988	3549
5	220	7533	283	317
6	317	383	349	349
7	206	3998	68278	3369
8	352	1988	6948	7690
9	667	3369	3571	1988
10	301	11792	3549	16460

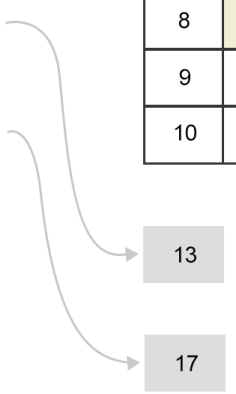
Evaluation

Followership

Rank	Node ID
1	1
2	88
3	1503
4	138
5	220
6	317
7	206
8	352
9	677
10	301

Integrated: Type 1

Rank	Node ID
1	1
2	88
3	677
4	220
5	317
6	138
7	1988
8	3549
9	1503
10	206



YORK U

Evaluation

Followership

Rank	Node ID
1	1
2	88
3	1503
4	138
5	220
6	317
7	206
8	352
9	677
10	301

Integrated: Type 2

Rank	Node ID
1	88
2	1
3	677
4	220
5	1988
6	2417
7	7533
8	138
9	3549
10	317
14	
15	
18	
19	

Followership

Rank	Node ID
1	1
2	88
3	1503
4	138
5	220
6	317
7	206
8	352
9	677
10	301

Integrated: Type 3

Rank	Node ID
1	88
2	677
3	1
4	220
5	1988
6	2417
7	14454
8	7533
9	3549
10	206
14	
15	
16	
22	
23	

YORK U

\

Conclusion

Conclusion

- Influencer rankings calculated simply using a single factor of “followership” would bring us divergent rankings from actual influence levels of users.
- We calculated more accurate rankings by using other factors as well.
- We repeatedly observed that the number of followers are not always proportional to the actual influence level.
- Followership could be used to filter out fake or inactive influencers.
- A possible future work could be about working on the weight distributions for calculation.
  - For viral marketing where the spread of information via sharing feature is the key, a higher weight on retweet factor or anything equivalent would be ideal to find the most fitting influencers.

Hope you enjoyed our presentation!

- Any Questions?

References

[1] R. S. Achrol and P Kotler. 2012. Frontiers of the marketing paradigm in the third millennium. Journal of the Academy of Marketing Science 50, 1 (2012), 35–52. https://doi.org/10.1007/s11747- 011- 0255- 4

[2] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[3] F. P. Sánchez. 2010. Competitivity groups on social network sites. Mathematical andComputerModelling52,7-8(2010),1052–1057. https://doi.org/10.1016/j.mcm.2010.02.031

[4] W. Xing and A. Ghorbani. 2004. Weighted PageRank algorithm. Proceedings. Second Annual Conference on Communication Networks and Services Research (2004), 305– 314. https://doi.org/10.1109/DNSR.2004.1344743