Detecting Social Media Influencers of Startup Accelerators Using Social Network Analysis on Twitter

Damian Drewnik, University of Twente, The Netherlands

The study focuses on detecting social media influencers and their power networks within the American technology startup Y Combinator accelerator. It aims to detect the major influencers amongst the people taking part in Y Combinator activities such as startup founders and accelerator employees. It uses social network analysis with a degree of centrality, betweenness of centrality, out-degree, in-degree, and eigenvector centrality as the prime variables. The main objective is to discover how people related to the accelerators are connected, using the Twitter social media platform. The research is expected to contribute to the scientific body of knowledge by presenting experimental results based on the Social Network Analysis approach. It will detect users that could be considered the most influential on the Twitter social media site in the Y Combinator Twitter space. It will also contribute to the startup accelerator's body of knowledge by discovering the social media structure of startup accelerators.

Additional Key Words and Phrases: social network analysis, social media analysis, social media influence, Twitter, startup accelerator, Y combinator

1 INTRODUCTION

Nowadays, social media and the internet are common mediums that can reach the majority of the global population. According to Simon Kemp in the Digital 2022 April Global Statshot, there are 7.93 Billion People in the world and out of those, 5 billion people (around 63% of the population) have access to the internet and 4.65 billion (around 58.7% of the population) are active social media users [10]. The number of social media users is steadily growing from 3.81 billion people in April 2020 to 4.65 billion people in April 2022 which increases the need to study the relations and behaviors on those platforms. This study is going to use Twitter in the context of the Y combinator startup accelerator influence network.

Startup accelerators are institutions that help people to develop cost-effective startups faster in the modern economic context [14]. According to [4], accelerators help startups determine and build initial products, recognize interesting customer segments, and provide resources such as employees and money. Those accelerator programs are limited in time that lasts around three months. They often offer a small amount of capital and working space. On top of that, networking, educational, and mentorship opportunities are offered with peer accelerator participants and mentors who can be accelerator alumni, entrepreneurs, venture capitalists, angel investors, or corporate executives. Startup accelerator programs end with a “demo day” where startups pitch to a great audience of potential investors.

Paul Graham founded the first startup accelerator - Y Combinator in 2005 with the purpose to improve the way new ventures are created [4, 13,14]. Due to the newness of the startup acceleration concept, there is not a lot of research done in this space and there is still a relevant research gap that can be filled [4, 5, 14, 16].

Currently, relevant research about the influencer power network on the startup accelerator social media is lacking. Taking into account that at the time of writing the biggest American startup accelerators like Y Combinator, 500 Startups and Techstars have respectively 1.2 million, 665 thousand, and 351 thousand Twitter followers, the need to investigate the connections within the startup accelerator's social media structure is present.

This study is going to outline a way to find influencers that are involved in the startup accelerator influence power networks on Twitter. The approach of analysis will be based on calculating the following Social Network Analysis metrics: degree of centrality, betweenness of centrality, out-degree, in-degree, and eigenvector centrality which will be further explained in the Methods of Research section.

2 RESEARCH QUESTION

This paper will analyze the social network structure of the biggest American startup accelerator [12] and propose a method to detect social media influencers in such networks. This statement will lead to the following research question:

How to detect social media influencers within startup accelerators using social network analysis on Twitter?

3 RELATED WORK

In this section related work will be discussed. In order to collect the relevant literature regarding research, Scopus and Google Scholar were used. With search terms such as “influence marketing”, “Twitter”, “social network analysis”, “influence”, “startup accelerator”, “startup”, “Y combinator”, “NodeXL” and “Gephi” multiple documents were found that can meaningfully contribute to the current research.

TScIT 38, February 3, 2022, Enschede, The Netherlands

3.1 Social Network Analysis on Twitter

In terms of using social network analysis on Twitter, a lot of research has been done [1, 2, 7, 17]. Those papers that use Twitter and social network analysis can be used as a guideline for what is feasible to do in the research.

Identification of nodes that are the most influential in the complex social network structure is one of the key problems in information retrieval and data mining [1]. The second important issue that emerges from the social network analysis domain is community formation [1]. Those two problems are going to be the core of the social network analysis conducted in the current research paper.

A potent approach to analyzing social networks is to gather the data from the Twitter social networking site, preprocess it to get rid of the less interesting information, and run the SNA metrics that explain the network in more detail [1, 2, 17]. Those metrics are the degree of centrality, in-degree, out-degree, eigenvector centrality, and betweenness centrality.

3.2 Startup Accelerator Studies

Even though the literature on Startup Acceleration is lacking, there are some interesting studies in that space and the knowledge gap in this space is narrowing [9]. Therefore, because the number of accelerators that are opening is growing rapidly, there is a need to investigate this phenomenon from different angles [4, 5, 13].

There are a couple of papers focused solely on investigating the efficiency of startup accelerators [4, 14]. Researchers investigate what metrics to assess startup success can be used and to what degree startup accelerators help entrepreneurs with founding successful companies. The proposed ranking metrics that are important for valuing startup success are Valuation, Qualified Exit, Qualified Fundraising, Survival, and Founder Satisfaction [14].

3.3 Social Network Analysis of the Startup Industry

In my current research, the context of startup accelerators in the Social Network Analysis is important. Influencers in those spaces may have certain characteristics that are relevant to the startup accelerator power network structure. In [7], researchers investigate how venture capitalists play a central role in the success of a startup. The power position of startups and their investors is thoroughly investigated. In general, venture capitalists have a great centrality, because they can be perceived as keepers of both funding and information. Therefore, they have great power and influence in the ecosystem of technology startups. In my current research, I want to discover how the people related to the Y Combinator Twitter network are located and what power they have in the network.

3.4 Recommendation systems based on Social Network Analysis

There are already a few papers that tackle the problem of discovering information from social media using social network analysis.

In [8] researchers are looking for Tweets with content that contain URLs in order to find some relevant information that is going to improve open education. They are using Social Network Analysis as an enabler for their recommendation system. They investigate the most influential nodes whose publications are considered interesting in a group of users that interact in this domain to improve their system. My research is going to improve upon this approach and instead of looking for URLs to find useful content, I will be looking for the accounts of founders that can meaningfully improve the knowledge of the startup industry.

In my research, my method is going to be specifically targeted at the founders that would like to get to know more about the startups and startup industry in general. According to [15] there are no studies that identified concrete information for the entrepreneurs. Researchers in [15] investigate where entrepreneurs get useful information about leading the business. They discovered that beginner entrepreneurs follow more Twitter accounts from wider sources, while more advanced entrepreneurs are focused on fewer accounts. My study is going to provide a method for a person to discover new accounts that can be used for intelligence gathering. The method will be applicable to beginner and more advanced founders that clearly use Twitter for gathering knowledge and inspiration.

4 METHODOLOGY AND APPROACH

This section is going to provide an outline of the steps that were taken to answer the research question.

Firstly, the literature research on social media, social media influencers, social network analysis, and startup accelerators was conducted to provide relevant context and background to the research.

4.1 Data Gathering

The next step was data gathering using NodeXL Pro [11] and Twitter Developer API. I used the 'Import from Twitter Search Network' option to extract the relevant Social Network data. The dataset contains a list of tweets combined with their metadata. A network will be created consisting of nodes expressed by user accounts and edges expressed by interactions between users via retweets and mentions. The list is based on an account search of Y Combinator employees and the Y Combinator account itself. A query used to capture all the tweets can be found in the appendix. The Twitter data snapshot comes from the period from 26.12.2022 to 03.01.2023. The dataset consists of 10423 nodes and 19203 edges.

4.2 Data Preparation

After the data was gathered, the processing needed to be done to extract the relevant interactions. The dataset was exported from NodeXL Pro in GraphML format. This file was then loaded into Gephi [3] as a directed graph. The next step was executing all the statistics that are relevant for further social network analysis. Those are

1. 'Average Degree' in order to get the degree numbers for the whole dataset. This includes metrics such as degree of centrality, in-degree, and out-degree.
2. 'Network Diameter' to get betweenness centrality and closeness centrality metrics
3. 'Modularity' with parameter 'Resolution' set to 1.5 to find communities
4. 'Eigenvector Centrality' with directed nodes and a number of iterations equal to 100

After the execution of those statistics, the size of the nodes was set by the degree of centrality, with a minimum size of 10 and a maximum size of 150. Then the nodes were colored according to the previously computed 'Modularity' metric, to show the communities clearly on the graph. After that, the layout algorithm 'Force Atlas 2' with parameters 'Scaling' 0.05, 'Stronger Gravity' turned on, 'Gravity' 0.05, 'Prevent Overlap' turned on and 'Edge Weight Influence' 1.0 was run. This layout algorithm helps to sort out the network and clearly see the clusters that are closely connected. Finally, in the 'Data Laboratory' section, a column was added to indicate which Twitter aliases were used to create this social network graph. From this point, they will be called 'Network Creators'. Those accounts can be found in the appendix section with the title 'Query Used for data network creation'.

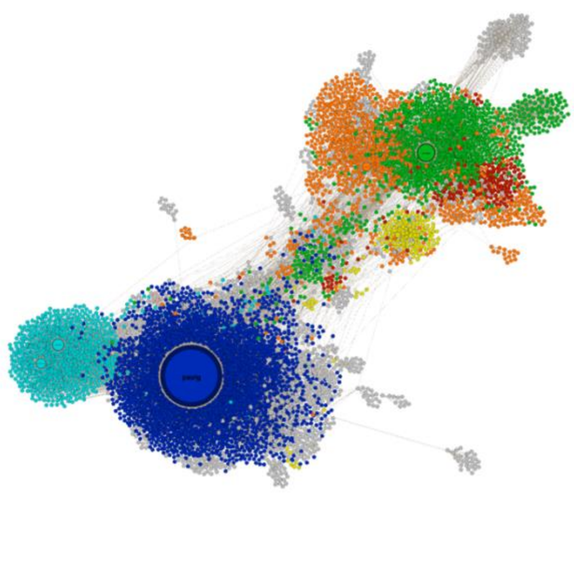

Figure 1: Initial Social Network colored to show the communities and their relations

4.3 Brief metric explanation

Metrics that are relevant for further analysis are briefly explained here:

Degree of centrality: The number of edges connected to the node

Betweenness centrality: Graphs centrality based on on the shortest path between the nodes

Out-degree: The number of outgoing edges to other nodes

In-degree: The number of incoming edges from other nodes

Eigenvector centrality: Measures node's importance while taking into account the importance of its neighbors

5 RESULTS

5.1 Network Structure Analysis

Analyzing the structure of the network one particular insight may be observed about one of the 'Network Creators'. @paulg is Paul Graham – the person who founded Y Combinator startup accelerator. He is clearly the most influential node in the entire network. Paul Graham already retired and is not actively taking part in Y combinator operations. It is clear however that some of the connections to people engaged in the Y Combinator activities are still there. From the network, it can be inferred that he is now positionally closer to the Twitter owner - Elon Musk. It means that even though he still remains an important person in the Y Combinator related networks, he talks about other topics that are less related to the Y Combinator network.

The second big find is the clustering of other 'network creators'. The 3 biggest ones by eigenvector centrality - Y Combinator, Garry Tan, and Michael Seibel are clustered closely together and have smaller communities built around them. This one big cluster will be called the 'Startup cluster'.

The third pattern that can be seen there is that cluster revolved around Elon Musk has almost no connections to the 'Startup cluster'. 'Figure 1' shows that this cluster is connected directly to Paul's Graham cluster almost exclusively.

The nodes between the Y Combinator Focused Cluster and Paul Graham Focused cluster are in between communities. It is not completely clear from the picture who they may be or what interesting value can they bring.

5.2 Network Influencers Analysis

After the initial network structure analysis is done, it is possible to look for more patterns regarding the particular nodes.
The first thing I did to the network was sort nodes by the degree of centrality and mark down the first 10 nodes.Then I did the same action for other metrics (in-degree, out-degree, betweenness centrality and eigenvector centrality). This action

helped me with further analysis of the most influential nodes in the network. After all those people are identified, they can be further studied in more detail. Consequently, I created a ranking of those nodes including every metric. The structure of the network that involves only those nodes is located in 'Figure 2'.

From the data, it can be seen that Paul Graham is indeed the most influential node of the entire network. He is first in all the metrics except for out-degree in which he is ranked second. Garry Tan that is currently the president of Y Combinator is the second more influential person. He is ranked second in all of the metrics except for out-degree in which he is 23rd. Surprisingly, Elon Musk is 3rd the most influential node in the network. He scored a little bit lower in the betweenness centrality metric in which he is 17th. It is clear that he is influential, but as mentioned previously, his cluster is almost exclusively related to Paul Graham only, which means he is not connected that well to the rest of the network. Y Combinator is the node that scores well in all of the metrics
and can be considered an influential node in the entire network. It is ranked 5, 3, 5, 5, and 4 in all relevant metrics. Three of the nodes - Michael Seibel, Tim Urban, and Richard Dawkins follow the same pattern. They are high in eigenvector centrality, degree centrality, and in-degree, but low in betweenness centrality and out-degree. They are important for the network but the flow of information is not impacted when they are gone.

Lastly, one emerging pattern should be underlined. Most people from the top 10 of out-degree are ranked far in other metrics. It means that there is a lot of traffic going out of the node, but not a lot of traffic goes in. Those nodes can be considered diffusers of information and are not relevant when it comes to production.

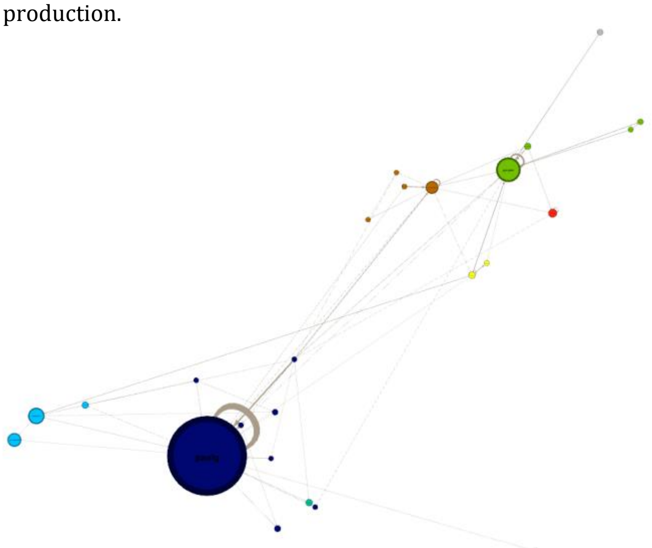

Figure 2: Social Network Structure created from nodes that belong to at least one TOP 10 of the metric

Metric	Founders	Investors	Accelerators
Betweenness Centrality	37	5	1
Eigenvector Centrality	30	5	0
In-Degree	24	4	1
Degree	21	4	0
Followers	15	5	1
Out-Degree	10	1	0

Table 1. New startup-related accounts found by relevant metrics in their 'TOP 100' rankings

5.3 Proposed Method

The method proposed in this study will help with an initial analysis of the startup accelerator network and finding potentially interesting accounts that can be followed.

The method involves a few steps that need to be done on the dataset, on top of prior preprocessing explained in the 'Methodology and Approach' section.

The first step is to sort the network of nodes by eigenvector centrality metric from top to bottom values. Then for the first 100 nodes qualitative research and labeling of the account are done. Labels include: Founder, Investor, Accelerator, and Other. In order to determine which label is the most appropriate, firstly the bio of the Twitter account is searched for 'founder, 'investor' and 'accelerator' keywords. As entrepreneurs have often Twitter bios that do not explain their work status, extra qualitative research is needed. This research was done by looking for every Twitter account that needed to be labeled and then looking for the same person on the internet using mostly LinkedIn and Google searches. One important remark regarding the Investor label is, that if someone is an investor now, but was a founder in the past, he is labeled with a 'Founder' label.

This step is then repeated for every relevant metric (degree of centrality, in-degree, out-degree, betweenness centrality, and followers count). There are many duplicate nodes between the 'Top 100' of each metric, which makes the labeling process faster for further accounts.

Then the results of each metric are compared and analyzed.

The metrics that are the most helpful for finding new founders, investors, and accelerators are betweenness centrality and eigenvector centrality. The first metric gives us 37 new founders, 5 investors, and 1 accelerator while the second metric gives us 30 new founders and 5 investors. There are only 11 duplicate people across those two metrics which gives us potentially 67 new interesting people to follow in the startup accelerator Y combinator space if we use those two metrics combined.

The in-degree metric shows only 2 founders that cannot be found in the betweenness centrality combined with eigenvector centrality. It also had many duplicates with eigenvector centrality in general but gives worse numbers so it can be discarded. A similar situation is with the degree of centrality metric. It has only 1 new value that cannot be found in in-degree metric and it has many duplicates. This metric can also be discarded when looking for new interesting accounts.

Another case is an out-degree metric which gives us 6 new accounts that cannot be found in betweenness centrality and eigenvector centrality metrics combined. In my opinion, it is not worth considering this metric as it only found 10 founders out of 100 accounts which show a bad hit rate.

The last metric that was used in comparison was the count of followers. It shows promise as there are not a lot of duplicates. It gives us however only 15 hits out of 100 which is not promising. The relevance of those results is also limited as the accounts here do not have a great correlation with the analyzed network. In this metric ranking accounts such as Bill Gates, Mark Cuban, Kim Kardashian, and The Rock could be found.

My method then suggests that when looking for new interesting founders, investors, and accelerators to follow in the Y Combinator Social network, one should use betweenness Centrality and Eigenvector Centrality as potentially the best and most effective way to search for them.

In my study, I arbitrarily took the 'top 100 ranking' number. It is possible that taking fewer or more accounts give better results using those metrics. The method proposed in this study is experimental and should be further investigated on the other datasets. It would be interesting to discover the best parameters and optimize this method in the future.

6 CONCLUSION

For the purpose of answering the research question, Social Network Analysis was performed. Firstly, general structure of the social network was analyzed, and the initial findings were made.

The second step was the analysis of the most important nodes based on relevant metrics. By using the approach to take the top 10 results based on every relevant metric the initial validations of the general network analysis were confirmed and the major influencers of the entire network were investigated. As it appeared, the out-degree metric has shown that many of the players in the network who are diffusers of the information are not publicly known. Therefore, there are multiple people in the network just looking for the information shared by Y combinator-related entrepreneurs.

The last step for the analysis was the proposal of the method that could be further used for the recommendation of accounts that could be useful to follow. The method has shown that most founders, accelerators, and investors in the Y combinator Twitter network can be found by using eigenvector centrality and betweenness centrality metrics combined. The method needs to be further investigated on other datasets to confirm its usefulness.

7 DISCUSSION

7.1 Limitations

The main limitation of this research was the Twitter API for Developers. The problem is that it does not allow for accessing all the historical Tweets. It only allows for 7 days of history which is not ideal. It also focuses on relevance, rather than completeness. This access was enough for the purpose of this study but the research could be enhanced by using the Researcher API. Academic Research access is more comprehensive and allows for searching historical data. In order to get access, a person needs to be a graduate or Ph.D. student with a clear research objective and apply to get it [6].

The second limitation is that my method was only checked on one dataset due to the lack of time available. The validity and usability of the method should be further investigated by using it on another dataset related to startup accelerators. The method could be also checked on the dataset created in the same way as in this research, but in a different time span.

7.2 Future Work

Firstly, this study may be enhanced to other social media networks like LinkedIn, Instagram, or TikTok in the future. However, the lack of APIs for those social networks may pose a problem, as web scrapping may be ethically problematic in some areas.

The second issue is that the process of looking for interesting accounts is not automated. The steps that someone needs to take using my method may cause trouble and may not be easy to perform at the beginning. There is a need for further studies that handle this problem. I would suggest building an application that automatically handles searching for like-minded people based on your parameters such as Twitter aliases or hashtags that you use to create the network you are interested in.

8 REFERENCES

[1] Abid Mahmood Malik, H. (2022). Complex Network Formation and Analysis of Online Social Media Systems. Computer Modeling in Engineering & Sciences, 130(3), 1737–1750. https://doi.org/10.32604/cmes.2022.018015

[2] Aguilar-Gallegos, N., Klerkx, L., Romero-García, L. E., Martínez-González, E. G., & Aguilar-Ávila, J. (2021). Social network analysis of spreading and exchanging information on Twitter: the case of an agricultural research and education centre in Mexico. The Journal of Agricultural Education and Extension, 28(1), 115–136. https://doi.org/10.1080/1389224x.2021.1915829

[3] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

[4] cohen, susan and Hochberg, Yael V., Accelerating Startups: The Seed Accelerator Phenomenon (March 30, 2014). Available at SSRN: https://ssrn.com/abstract=2418000 or http://dx.doi.org/10.2139/ssrn.2418000

[5] Del Sarto, N., Cruz Cazares, C., & Di Minin, A. (2022). Startup accelerators as an open environment: The impact on startups' innovative performance. *Technovation*, *113*, 102425. https://doi.org/10.1016/j.technovation.2021.102425

[6] Getting Started with the Twitter API. (n.d.). Docs | Twitter Developer Platform. https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api

[7] Hadley, B., Gloor, P. A., Woerner, S. L., & Zhou, Y. (2018). Analyzing VC Influence on Startup Success: A People-Centric Network Theory Approach. Studies on Entrepreneurship, Structural Change and Industrial Dynamics, 3–14. https://doi.org/10.1007/978-3-319-74295-3_1

[8] Lopez-Vargas, J., Piedra, N., Chicaiza, J., & Tovar, E. (2015). OER Recommendation for Entrepreneurship Using a Framework Based on Social Network Analysis. *IEEE Revista Iberoamericana De Tecnologias Del Aprendizaje*, *10*(4), 262–268. https://doi.org/10.1109/rita.2015.2486387

[9] Ortmans, J. (2016). A Hard Look At Accelerators. The Kauffman Foundation, Policy Dialogue on Entrepreneurship. Retrieved from: http://www.kauffman.org/blogs/policy-dialogue/2016/april/a-hard-look-at-accelerators.

[10] S. Kemp, "Digital 2022: April Global Statshot Report," DATAREPORTAL, 2022. [Online]. Available: https://datareportal.com/reports/digital-2022-april-global-statshot. [Accessed: 06-May-2022].

[11] Smith, M., Ceni A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., Dunne, C., (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016 from the Social Media Research Foundation, https://www.smrfoundation.org

[12] Statista. (2022, January 11). Leading U.S. startup accelerators by investments as of July 2021. https://www.statista.com/statistics/1249900/leading-startup-accelerators-investments-u sa/

[13] Mahmoud-Jouini, S. B., Duvert, C., & Esquirol, M. (2018). Key Factors in Building a Corporate Accelerator Capability. *Research-Technology Management*, *61*(4), 26–34. https://doi.org/10.1080/08956308.2018.1471274

[14] Mishigragchaa, B. (2017). Accelerators as a Tool to Support Startup Ventures: Assessing Their Performance and Success Factors. Literature Review. Studia i Prace WNEiZ , 48 , 9–19. https://doi.org/10.18276/sip.2017.48/1-01

[15] Motoyama, Y., Goetz, S., & Han, Y. (2018). Where do entrepreneurs get information? An analysis of twitter-following patterns. *Journal of Small Business &Amp; Entrepreneurship*, *30*(3), 253–274. https://doi.org/10.1080/08276331.2018.1435187

[16] Torun, M. (2016). Business Accelerators and Their Difference From Incubators. Working Paper. Research Gate. Retrieved from: https://www.researchgate.net/profile/Musta-fa_Torun2/publication/311810125_Business_Accelerators_and_Their_Differences_from_Incubators/links/585bca1908aebf17d3864896.pdf.

[17] Wibisono, A. I., & Ruldeviyani, Y. (2021). Detecting Social Media Influencers of Airline Services through Social Network Analysis on Twitter: A Case Study of the Indonesian Airline Industry. 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT). https://doi.org/10.1109/eiconcit50028.2021.9431876

A Query Used for data network creation

@tlbtlbtlb OR @paulg OR @jesslivingston OR @robertrmorris OR @richaberman OR @gustaf OR @t_blom OR @paultoo OR @daltonc OR @umurc OR @dessaigne OR @aaron_epstein OR @jdotjdotf OR @bradflora OR @snowmaker OR @sdianahu OR @koomen OR @_puneetkumar OR @dflieb OR @gralston OR @arnavsahu341 OR @mwseibel OR @garrytan OR @lizwessel OR @ycombinator