A Network Analysis for Assessing Similarities between Micro-Influencers and Their Followers in Music

Andreea Soran
University of Twente
Enschede, The Netherlands
a.soran@student.utwente.nl

ABSTRACT

Current ranking methods of influencers in social media are purely quantitative. This paper explores a qualitative alternative based on social network analysis (SNA) relying upon the novel content similarity score (CSS) between an influencer and a follower. In this scope, an AI Natural Language Processing (NLP) and data scrapping techniques are leveraged for CSS computation, while the network visualization tool Gephi contributes to SNA. Music micro-influencers from Instagram are considered in this paper because of the lack of research on the topic. By seeing through the similarity perspective, we can identify groups of reflective and non-reflective followers, which along other findings, help to reveal fruitful opportunities for influencers and marketers.

KEYWORDS

social network analysis, Gephi, CSS, social media influencer, micro-influencer, NLP, Instagram, marketing

ACM Reference Format:
Andreea Soran. 2022. A Network Analysis for Assessing Similarities between Micro-Influencers and Their Followers in Music . In *Proceedings of ACM Conference (TScIT 37).* TScIT, Enschede, The Netherlands, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Social media influencers are becoming increasingly popular, making influencer marketing an important part of many companies' strategies. As influencer marketing continues to grow in popularity, more and more marketers are incorporating it into their media plans, either by working with influencers directly or by increasing their use of influencer-generated content [16]. When developing and executing successful influencer marketing campaigns, it is essential for marketing practitioners to consider the deep connections between influencers and their followers. The most important qualities of a social media influencer are authenticity, relatable, and having the ability to curate content [8]. The fact that social media influencers can curate content implies that consumers perceive them not only as sources of information but also as role models who inspire them through their lifestyle choices and competent content curation. Some of those choices are visible through the online activity of the community making it resemble the influencer. Thus, relying on the number of likes and followers is not enough to select the best influencer and marketing strategy.

The current methods of ranking potential social influencers on platforms such as Instagram are based on shallow metrics such as likes and followers [6]. Those metrics by themselves do not provide valuable insights anymore as it became easy to artificially inflate these numbers through bots or other means. However, there is a need for more precised formulas that consider network characteristics, position within the network, and more elaborate metrics.

Considering the aforementioned, we propose a new metric for assessing the influential position of influencers, naming the degree of similarity between them and their community. Moreover, the similarity alone is not sufficient without looking at the overall network characteristics of influencers. In this scope, we aim to observe and analyze influencer-community similarity through social network analysis (SNA). SNA is a powerful tool for observing and analyzing social networks. It can help us understand the relationships between influencers and their audiences, and to identify patterns of similarity between them.

Due to the high interest of brands in micro-influencers [11], the study will be limited to those types of influencers. To focus even further and provide valuable research, the domain of influence of those influencers will be limited to the music category of Instagram. The top three most popular Instagram categories are media, fashion, and music, however, no studies with a focus on music influencers have been conducted [13]. Therefore, we limit the study to music micro-influencers that act on the Instagram platform.

This paper aims to answer the following research question:

RQ: What is the similarity between micro-influencers and their followers in the Instagram music category?

But first, we need to answer the sub-research question:

sRQ: How can the influencer-community similarity be computed?

The similarity mentioned in the sub-question refers to which the behavior between two parties resemble each other. On social media, behavior represents among others, posts or content created. Thus, similarity computation can be achieved by the amount of influencer-specific content reflected on its followers compared to the overall influencer content. Data collection and data preprocessing will be used in answering this research question.

The overall similarity within the music category is meant to be estimated by analyzing similarity characteristics - outlined by CSS - of multiple influencer communities. For this purpose, we

TScIT 37, July 2022, Enschede, The Netherlands

selected at random a dataset of ten music influencers and their communities. To answer the main research question, we conduct the following processes: data collection, data preprocessing, and network analysis.

2 RELATED WORK

Social media influencers are increasingly prevalent on social media platforms such as Facebook, YouTube, and Instagram. They have acquired large audiences of followers, which gives them the ability to influence a vast number of people with their messages. Social media influencers can be classified according to the size of their following: mega, macro, mid, micro, and nano. Papers have shown that micro-influencers, who have between 5,000 and 50,000 followers, are often the best choice for brands looking to market their products, as they are more effective and trustworthy than celebrity influencers and have a great impact on their niche audience [11][9][15][4].

One study challenged the common metrics correlated to the influential position of influencers and emphasized the relevance of content-based metrics [6]. The analysis revealed that in the presence of additional metrics, the counter-metrics weighed less in the effectiveness of influential positions. Another study has found that focusing on counter-metrics like the number of likes and followers can be risky, as it can lead to influencers paying for followers and bots [1]. However, one study shows that in limited cases the number of followers an influencer has can affect how popular the influencer is perceived to be, which in turn can affect the influencer's perceived opinion leadership [3].

Moreover, a paper from 2020 on the role of similarity in purchasing intentions concludes that there is a positive correlation between feeling similar to an influencer and trusting them, as well as between an influencer's credibility and trust [2].

3 METHODOLOGY

The research proposed aims to investigate and analyze the similarity behavior within influencers' communities that are not sufficiently researched as well as a frame of reference to calculate those similarities. Hence, we perform exploratory research through a Social Networks Analysis (SNA) that will enable conclusive results to be generated. SNA investigates social structures by looking at the relationships between different nodes in a network. It uses graph theory to understand how different elements in a network are connected, and to identify patterns and trends in that data.

Considering the amount of data and time efficiency factors, those processes are done automatically by appropriated tools. We use PyCharm as the main development environment and Python libraries specifically design for Instagram scraping data in the process of data collection. Data will be preprocessed and transformed accordingly in the same environment as well as in Excel. Finally, the analysis will be aided by a network visualization tool named Gephi [14]. Those processes will be further explained in this section.

3.1 Data collection

The process of data collection requires gathering and storing public data that can be used to make observations and infer information. For the purpose of this paper, data must be collected from the Instagram platform. This can be possible using Python data scraping libraries like instaloader [5] and instagramy [12].

3.2 Data preprocessing

After data collection, the next step is data preprocessing. This is the process of transforming raw data into a format that can be understood and analyzed by different tools. For the sake of network analysis, data must be transformed into nodes (user accounts) and edges (relationships between nodes). This transformation is done by creating CSV files of nodes and edges, which can then be uploaded to Gephi to visualize the network. Additionally, data is transformed into a suitable format to calculate the similarity.

3.3 Content Similarity Score (CSS)

We define the similarity between an influencer and their community by the extent to which the behavior between the two parties resembles. There is no metric available to calculate the behavior similarity between an influencer and its community, hence we propose a novel metric in this regard, called Content Similarity Score (CSS):

$$CSS = \frac{CC}{SC}$$

CC - amount of influencer-follower common content
SC - amount of influencer specific content

The Content Similarity Score (CSS) measures how similar a follower's content behavior is to that of their influencer. This score is calculated by looking at the content posted by the influencer and the follower and comparing the two. For comparison, the posts need to be transformed into captions, a written format of the posts.

In order to achieve a suitable conversion of the data (data preprocessing), we use an AI system that processes images into captions. Natural Language Processing is the branch of AI that makes it possible [10]. The conversion with the use of the AI was possible through a Python library called spaCy [7] which is an open-source library for advanced Natural Language Processes.

An example of captions gathered from an influencer across their most recent posts (last 5 posts) can be seen in **Table 1**. Since the SC represent the amount of influencer-specific content, in this example SC equals 6. The CC is calculated by removing any captions that are not shared by both influencers and followers.

Table 1: Captions of an influencer

Captions
a cartoon
an illustration
flower
nature
art
dog

Finally, having the two variables, the CSS can be finally calculated. It is vital to highlight that the new measure calculates a one-to-one similarity between one influencer and one follower. Hence, the similarity of a whole influencer community needs to be inferred from the CSS of all influencer-follower's connection and network characteristics.

3.4 Network analysis

The CSS formula is used to compute the similarity between one influencer and one follower in a network. This indicates that the relationship between nodes in a network can be given by this similarity (if it exists). However, the CSS formula alone cannot be used to generalize the similarity to an overall Instagram category because it only reflects the relationship between one influencer and one follower. Therefore, in order to assess the similarity of the music category, we will firstly analyze a sample dataset of influencers and their communitites. The average observations of the dataset could be used to reflect the similarity of the overall category.

The sample dataset will consist of ten influencers and their followers. In order to overcome time problems due to computer power limitations, we will only retrieve data from 500 followers per influencer. Additionally, both public and private accounts will be part of the data retrieved. Apart from this, the data collection process will remain the same as the CSS process.

As mentioned above, data processing transforms raw data into formats that can be used to construct networks. More precisely, CSV files of nodes and edges must have the following structure:

Table 2: CSV struture of the nodes

Id	Label	Score

The nodes file must have the following columns: Id, Label, and Score. The Id column is unique for every node, so we chose to use the username of every account. The Label column is a common caption list, and the Score column is the CSS calculated. The score column is used to change nodes' size and color accordingly to offer visibility on the important nodes. (See **Table 2**)

Table 3: CSV struture of the edges

Id	Source	Target	Score

The edges file must have the following columns: Id, Source, Target, and Score. The Id column must match the Id column in the nodes file. The Source column indicates the influencer node, and the Target column indicates the follower node. The Score column indicates the CSS between the two nodes. (See **Table 3**).

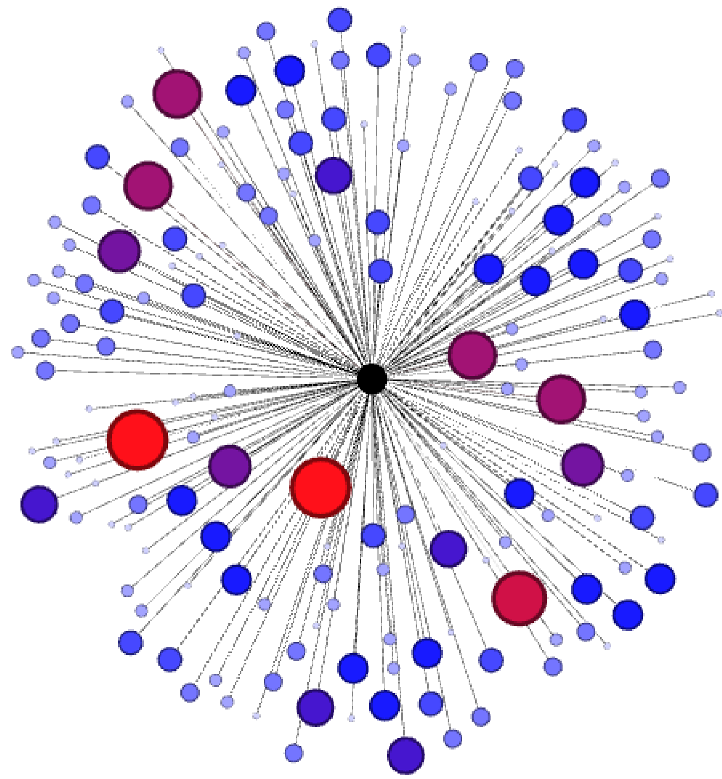

Figure 1: Network of an influencer with CSS as node degree.

The final influencers' graph highlights the followers with the highest CSS and the percentage of different similarity scores. An example of what an influencer's network looks like can be seen in **Figure 1** as well as the proportions of different CSS within the network in **Figure 2**.

0.0	(65.6%)
0.09	(9%)
0.18	(8.8%)
0.27	(6%)
0.36	(4%)
0.45	(3.6%)
0.55	(1%)
0.73	(0.8%)
0.64	(0.6%)
0.91	(0.4%)
0.82	(0.2%)

Figure 2: CSSs and their distribution percentage in the network.

For the overall network consisting of 10 influencers in the music category, we only take into consideration and present in the network only those nodes with CSS above the threshold of 0.5, in order to avoid overcrowding the visualization. Those graphs and their results will be discussed in the next section.

4 RESULTS AND DISCUSSIONS

In the music category, we have found three interconnections between audiences of different influencers. In this section, we will explore and discuss them. The overall graph can be found in **Appendix A**.

4.1 First interconnection

There is a high degree of overlap between the audiences of the two influencers, with 12 common followers and a fair number of remaining followers from both sides (See **Figure 3** and its extension in **Appendix A**). Among those 12 followers, 4 of them have the same CSS with respect to both influencers. This means followers perceive influencers as equally representative figures for themselves. Brands can benefit from this observation by using either influencer to reach a large number of potential customers.

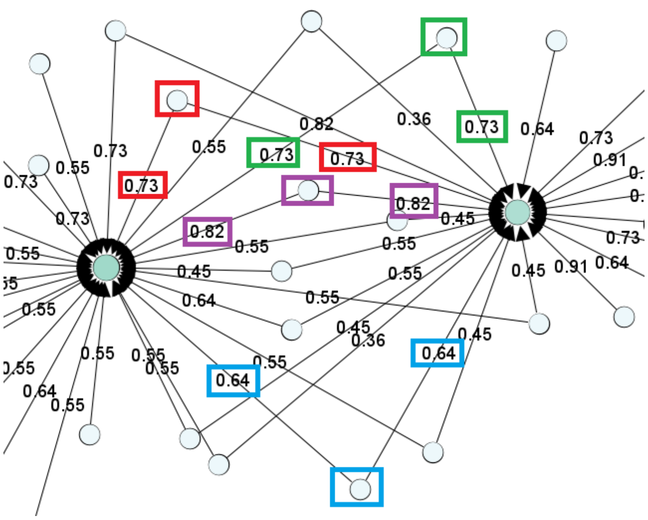

Figure 3: First interconnection.

There are only a few marketing strategies that brands could use on those common followers. One strategy would be to use influencer marketing, and have both influencers promote the brand to their shared audience. Another strategy would be to use targeted advertising, and specifically target the shared audience of both influencers. There are a few reasons why these strategies could be effective. Firstly, the shared audience of both influencers is likely to be interested in the brand if they are already following both influencers. Secondly, both influencers already have a relationship with the shared audience, so they are more likely to trust them when they promote the brand.

Moreover, when considering the individual networks of both influencers, it is clear that they have a close high CSS (0.82; 0.91), slightly the same network dimension (27; 24), and even similar content (See **Table 4**).This benefits brands, as they have the liberty to choose either influencer, in case they do not dispose of many resources but still want to cover a large audience, or use both influencers as described above.

Table 4: Captions comparison of two influencers.

Influencer 1	Influencer 2
2 people	2 people
1 person	1 person
people	people
3 people	3 people
a musical instrument	5 people
musical instruments	a musical instrument
guitar	musical instruments
indoor	guitar
standing	indoor
crowd	standing

From an influencer perspective, the similar characteristics between individual networks can be interpreted as either competition or an alliance opportunity. An alliance between them can lead to cross-influence of the audience, reach their message and help build trust and credibility.

It's worth mentioning that these two influencers follow each other, and this might be why they have such a high degree of overlap in their followers. This is an interesting thing that could be investigated further.

4.2 Second interconnection

Looking at each network separately there is a visible observation that the left influencer has a smaller audience (17 nodes) that resembles his content compared to the right one (39 nodes). Moreover, the latter reaches a high CSS of 0.75 whereas the other one has a maximum of only 0.58. The audience intersection of two influencers displays four followers, each of them having a close CSS towards both influencers; the difference does not exceed 0.09 (See **Figure 4** and its extension in **Appendix A**).

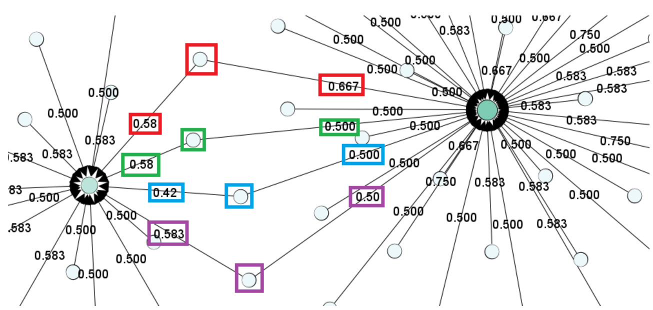

Figure 4: Second interconnection.

Considering bridge nodes and their similarities in regard to each influencer no clear decision can be made to aid brands in finding a suited influencer as their differences are minor. However, network dimensions and higher CSS have more weight in decision-making process.

4.3 Third interconnection

The final audience intersection found in the music category is between three influencers: *A*, *B* and *C*. Taking each network separately, *B* and *C* have no followers with a CSS greater than 0.5 so they are not included in their own network visualization.

A on the other hand has the biggest audience(95 nodes) in this category and a high CSS of 0.83. Among this vast audience, there can be found a part of *B* followers. Those followers fail to get a score above the threshold even as part of *A* audience. The same happens with a part of *C*'s audience following their intersection with *A* audience. (See **Figure 5** and its extension in **Appendix A**)

Those followers maintain a low similarity profile even when surrounded by a network of high similarities. We can call them non-reflective followers because they are not influenced by content creators or do not reflect influencers' average content. This information can help influencers consider adapting or improving their content in such a way the non-reflective followers will be more likely to engage with it.

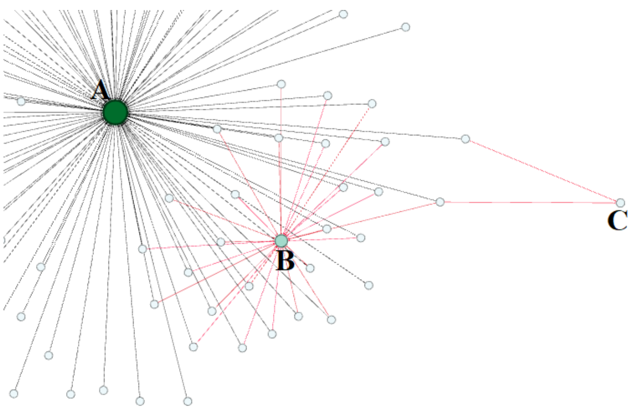

Figure 5: Interconnection of the three influencers.

From a brand perspective, it is best to avoid influencers who are sustained by non-reflective followers. This is because their audience can be hardly persuaded by content or trends.

5 CONCLUSION

For the sake of answering the sub-question relating to similarity computation, we have proposed and created a novel metric in this regard. This metric called CSS is based on the extent to which the influence-specific content appears on each of its followers' content. In this scope, AI methods such as NLP were used to transform content into a more suitable data format fit for the CSS formula.

The data from influencers shows that most communities have a fair number of followers with similarities that pass the 0.5 CSS threshold, and in some cases even reach 0.9. Network characteristics also show that there are similarities between influencers, through their common followers. In addition, these common followers have high similarities for both influencers of the communities they are part of. Therefore, it is clear that similarity in the music category of Instagram exists both in terms of metric formulas and network characteristics and has the potential to extrapolate valuable information for influencer marketing as well as influencers.

However, we must include the limitations that out of the whole community, we analyze just 500 followers that might have their profiles public or private. In the case of private profiles, their content could not be accessed to properly calculate the CSS, so they scored zero by default. If only public profiles would have been included, the statistical representation would have been of higher quality. Despite the limitations of including private profiles, the collected data was rich enough to reflect representative, interesting and valuable facts about music micro-influencer networks.

REFERENCES

[1] Colin Campbell and Justine Rapp Farrell. 2020. More than meets the eye: the functional components underlying influencer marketing. *Business Horizons*, 63, 4.
[2] Zainab Al-Darraji, Zahra Al Mansour, and Shilan Rezai. 2020. Similarity, familiarity, and credibility in influencers and their impact on purchasing intention. (2020).
[3] Marijke De Veirman, Veroline Cauberghe, and Liselot Hudders. 2017. Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude. *International journal of advertising*, 36, 5.
[4] Tian Gan, Shaokun Wang, Meng Liu, Xuemeng Song, Yiyang Yao, and Liqiang Nie. 2019. Seeking micro-influencers for brand promotion. In *Proceedings of the 27th ACM International Conference on Multimedia*.
[5] Alexander Graf and André Koch-Kramer. 2016-2020. Instaloader. https://github.com/instaloader/instaloader. (2016-2020).
[6] Jan-Frederik Gräve. 2019. What kpis are key? evaluating performance metrics for social media influencers. *Social Media+ Society*, 5, 3, 2056305119865475.
[7] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. DOI: 10.5281/zenodo.1212303.
[8] Chung-Wha Ki, Leslie Cuevas, and Heejin Lim. 2019. Identification of key attributes of social media influencers: an exploratory study. In *International Textile and Apparel Association*.
[9] Gloria HW Liu, Ying Li, and Neil Chueh-An Lee. 2021. Size does matter: how do micro-influencers impact follower purchase intention on social media? In *Conference on Electronic Business*. Vol. 21.
[10] 2020. Nlp, machine learning amp; ai, explained. (June 2020). https://monkeylearn.com/blog/nlp-ai/.
[11] Jiwoon Park, Ji Min Lee, Vikki Yiqi Xiong, Felix Septianto, and Yuri Seo. 2021. David and goliath: when and why micro-influencers are more persuasive than mega-influencers. *Journal of Advertising*, 50, 5.
[12] YOGESHWARAN R. 2021. Instagrammy. https://github.com/yogeshwaran01/instagramy. (2021).
[13] Kate Santoro and Olivia Adie. 2021. Top instagram categories and their average engagement rates: heepsy. (Sept. 2021). https://blog.heepsy.com/top-instagram-categories-and-their-average-engagement-rates/.
[14] [n. d.] The open graph viz platform. (). https://gephi.org/.
[15] Shaokun Wang, Tian Gan, Yuan Liu, Li Zhang, JianLong Wu, and Liqiang Nie. 2021. Discover micro-influencers for brands via better understanding. *IEEE Transactions on Multimedia*, 24.
[16] Steven Woods. 2016. # Sponsored: the emergence of influencer marketing.

A MUSIC GRAPHS RESULTS