

RESEARCH

Open Access



Unlocking the power of Twitter communities for startups

Ana Rita Peixoto^{1,2*}, Ana de Almeida^{1,2,6}, Nuno António^{3,7}, Fernando Batista^{1,4}, Ricardo Ribeiro^{1,4} and Elsa Cardoso^{1,4,5}

*Correspondence:
rita_peixoto@iscte-iul.pt

¹ Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal

² Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, Lisbon, Portugal

³ NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

⁴ INESC-ID Lisboa, Lisbon, Portugal

⁵ CIES-Iscte, Lisbon, Portugal

⁶ CISUC, Coimbra, Portugal

⁷ Desenvolvimento e Inovação em Turismo, CITUR, Centro de Investigação, Faro, Portugal

Abstract

Social media platforms offer cost-effective digital marketing opportunities to monitor the market, create user communities, and spread positive opinions. They allow companies with fewer budgets, like startups, to achieve their goals and grow. In fact, studies found that startups with active engagement on those platforms have a higher chance of succeeding and receiving funding from venture capitalists. Our study explores how startups utilize social media platforms to foster social communities. We also aim to characterize the individuals within these communities. The findings from this study underscore the importance of social media for startups. We used network analysis and visualization techniques to investigate the communities of Portuguese IT startups through their Twitter data. For that, a social digraph has been created, and its visualization shows that each startup created a community with a degree of intersecting followers and following users. We characterized those users using user node-level measures. The results indicate that users who are followed by or follow Portuguese IT startups are of these types: "Person," "Company," "Blog," "Venture Capital/Investor," "IT Event," "Incubators/Accelerators," "Startup," and "University." Furthermore, startups follow users who post high volumes of tweets and have high popularity levels, while those who follow them have low activity and are unpopular. The attained results reveal the power of Twitter communities and offer essential insights for startups to consider when building their social media strategies. Lastly, this study proposes a methodological process for social media community analysis on platforms like Twitter.

Keywords: Twitter data, Communities, Social media, Startups

Introduction

Startups are innovative and typically, when successful, present an above-average increase in customers and revenue (Lugović and Ahmed 2015). Nevertheless, they own limited resources having to reach customers on a restricted budget. In order to succeed, startups must raise funding. Social media platforms can serve as a gateway for various communities, allowing companies to gain customers, obtain funding and grow by achieving their goals (Rizvanović et al. 2023). Undoubtedly, social media has become a fundamental piece of the global information ecosystem, generating large amounts of data. Social media data can provide information about clients, products, and the overall market, helping to improve decision-making processes (Saura et al. 2021). However, social raw

data must be structured, prepared, and interpreted to infer relevant information to support decisions. Understanding social media as a tool can help enhance the company's investment—ROI—while enabling better customer relationship management—CRM. Recent existing studies support this argument and consider social media data-driven projects a strategic business knowledge source (Saura et al. 2021).

Social media platforms offer cost-effective digital marketing opportunities that benefit startups (Ruggieri et al. 2018). Studies have shown that active engagement on social media platforms increases digital engagement and can lead to better funding from venture capitalists or significant success in crowdfunding projects (Zhang et al. 2017; Ko and Ko 2021; Hadley et al. 2018). Lugović and Ahmed (2015) found a positive correlation between the Twitter activity of European startups and the total investment in their country of origin. Additionally, by creating communities relating users and service providers, startups can monitor the market and take advantage of electronic word-of-mouth spread of positive opinions about their products (Ruggieri et al. 2018; Chu and Kim 2011). Social network analysis, also known as SNA, is commonly called the process of monitoring the market and allowing for data-driven marketing strategies based on social media data (Hansen et al. 2019). Several existing studies have explored digital data using this methodology. For example, Ruggieri et al. (2018) focus on finding startup success patterns based on their presence on digital platforms. Hingle et al. (2013) collected Twitter content to analyze dietary behavior, and the authors highlighted that data visualization allowed the identification of relationships between diet-related behavioral factors. Wu et al. (2016) show that visualization methods can help uncover social media analysis results and support data interpretation, leading to a network analysis and visualization process. The authors of Hansen et al. (2019) propose a methodology, Network Analysis and Visualization, or NAV, to act as a design process model for enabling meaningful network analysis and extracting relevant insights.

The primary aim of this study is to determine the degree to which startups use social media platforms to create communities, what distinguishes these communities, and if the individual startup communities intersect. Our findings might highlight the relevance of social networks and their online communities for startups. To the best of our knowledge, none of the existing literature aims to understand how startups create communities on Twitter or if these communities intersect and create a global ecosystem. Therefore, this study aims to contribute to filling the gap in the literature regarding the analysis of social media communities around startups. For this, we selected eight renowned Information Technology (IT) startups founded by Portuguese administrators or headquartered in Portugal from the *Sifted* 2020 Portugal startups list (Sifted 2020). The companies were chosen because they are currently at different stages in their life cycle and are considered active on Twitter. Portugal has created an excellent startup ecosystem by promoting initiatives like the Startup Portugal 200 M fund and several business incubators (Portugal Digital and Startup Portugal 2021). Additionally, Portugal is recognized for forging high-quality engineering talents and showing a very high English proficiency index (Education First 2022). Twitter was selected as the social media data source due to its simple API access. This social media platform is a short text source widely used in academic research to analyze online behavioral patterns and the structure of the formed social graphs (Antonakaki et al. 2021). Furthermore, Twitter differs from other social media platforms because it allows users to

communicate with themselves publicly. It enables businesses to interact, create a community of users, and, instead of working solely as a marketing tool, allows for instant feedback, being crucial to companies that want to stay relevant and make immediate connections with their audience (Tanner 2023). Therefore, Twitter is an ideal platform for small businesses like startups, where they are now massively present (Ward 2020). Under these conditions, this case study intends to provide answers to the following research questions:

- RQ1: Do IT Portuguese startups form their own social communities on Twitter?
RQ2: In the case of community formation, are these disjoint, presenting different (types of) users, or do they overlap?

Towards our goal, and to better perceive community creation, we have applied the NAV process to social media data. Specifically, a social digraph was built, representing the *followers* and *following* communities of the startups under investigation. We utilized a community detection algorithm to determine if the startups form communities on Twitter based on modularity. This algorithm enabled us to visualize the communities in the digraph structure using different colors. The resulting visualization revealed that each startup had indeed formed a community. Furthermore, we were able to identify links between nodes of different communities, indicating that there was some overlap between the communities. To address the second research question, we characterized the communities' users by analyzing their type, popularity, and activity level. This information enables the emergence of social media strategies that can be effective for startups to achieve their proposed goals, either for financial support or for product/service marketing actions.

This work highlights the benefits of using social media platforms for startups to establish user and service provider communities. Through a case study, we present a systematic process that enables the visualization of startup communities and allows for the detection of intersections between these groups, which can help startups monitor similar companies and select relevant users to follow or relevant public to attract. Additionally, we characterized the types of users in the overlap of Portuguese IT startup communities, including profiles related to the IT area or the startup ecosystem. Finally, the results indicate that startups should follow users showing high levels of activity and popularity and are relevant to their field to increase the effectiveness of their social media strategies.

The remainder of the article is organized as follows: In “[Related work](#)” section presents relevant related work, namely, the key role that informed visualizations perform for social media analysis. The following section describes the methodology employed for this research, which is based on the NAV process model (Hansen et al. 2019). After that, we present an analysis and discussion of the results. Lastly, the conclusions of this research are presented, and paths for future work are delineated.

Related work

Social media platforms are essential digital marketing tools for small businesses like startups. This section explains how we can analyze communities created by startups when using social media platforms. We present a literature review of the methods and tools available for mining social media data using visualization, including network analysis and community detection algorithms, focusing on their application to study social

communities on Twitter. Furthermore, we describe the role of social media in facilitating online communities. Lastly, we explain some relationships between startups and their social communities.

Visual analytics in social media analysis

Social media can be a valuable data source for businesses to extract digital marketing knowledge. Additionally, it can serve as a means to interact with their clients and potentially facilitate funding. Social media platforms generate two types of data: content data and interaction data. The content is usually found in a non-structured format, such as text and images. It can be retrieved from tweets and users' comments. On the other hand, the interaction data can be represented by a network structure (a graph). An example of this interaction is the *following* relation, i.e., when a user follows another. Other examples are actions, such as: *likes*, *shares*, *replies*, and *mentions*.

Social media analysis is no more than the process of extracting information from social media data (Hansen et al. 2019). In Serrat (2017), the authors explain that such analysis can either focus on the social actors or their relationships. However, social media platforms generate high amounts of data, which can make it difficult to understand and analyze fully. Visual analytics is a promising approach for dealing with the challenges of understanding complex data (Keim et al. 2008). It aims to explore complex data through visualization using interactive visual interfaces. Visualization techniques can uncover social media patterns and trends and support data interpretation (Wu et al. 2016). Furthermore, it helps gather insights from larger datasets, combining visualization techniques with the human dimension for enhanced data analysis. The NAV methodology arises from the need to combine network analysis with visualization, and it can be applied to each type of social media data to attain different goals.

Recent studies employ social media analysis through visualization regarding the content data. Saura et al. (2023) study tweets and apply topic modeling and sentiment analysis intending to mine the opinion of Twitter users about open innovation. They used a graph-based visualization to unveil the relation between the topics. Hu et al. (2017) also analyzed tweets and performed topic modeling. However, this study's originality lies in the design of a particular technique for visualizing the content of unstructured social media text. Likewise, Smith et al. (2014) developed a new visualization to disclose the relationships between words and topics in topic models applied to unstructured social media data. The creation of novel visualization methods is key since it acts as an enabler to improve data understanding and unveil new insights. Hingle et al. (2013) use Twitter content to extract dietary behaviors and highlight that data visualization helped identify relationships between diet-related behavioral factors.

In the same way, the literature highlights visualization methods and intrinsic data aggregation as tools to understand and extract knowledge from social media interaction data. When a user follows another, this relation is represented by a directed link (edge), with the source being the follower and the sink on the followed one. The users connected through the *following* relation generate the so-called social graph (Gabiolkov and Legout 2012). The visualization of the social graph can reveal network features that can help answer important research questions. As an example, Molla et al. (2014) performed sentiment analysis of user Twitter contents and applied the results to color the

edges of the graph. This visualization highlighted where, in the social graph, negative, neutral, and positive opinions about a company existed. Abdelsadek et al. (2018) applied a community detection algorithm to a social graph, visually revealing the community's structure and related characteristics.

Based on the previous works, we can conclude that visualization is an essential tool in order to extract insights and knowledge from social media data. Additionally, we can improve the visualization by selecting and computing features that will be applied to the structure, such as color or format.

Social graphs and communities detection

As previously mentioned, the social media interaction data derived from the action of "user following another user" is encapsulated into a network structure called the social digraph (Gabelkov and Legout 2012). In the digraph, a node represents a user and an edge represents the user-following-user relation.

Studies regarding social graphs can be performed either at the node level or at the network level. Antonakaki et al. (2021) explains methods that can be applied to node-level studies of social graphs to measure users' activity, popularity, and influence. Activity means how frequently the user interacts. In the case of Twitter, activity is measured by the number of tweets and retweets the user performs. Popularity measures how well a user is recognized, which usually can be estimated by the number of *followers*. A simple popularity measure is the *Structural Advantage* (Cappelletti and Sastry 2012), a ratio between the number of *followers* and of *followings*. Lastly, influence estimates how a user's action influences (the actions of) other users, being the most used metric at the node level for studies involving this type of graph. Influential users are better disseminators of information through social platforms because they are more central in the graph. Consequently, graph centrality measures like *PageRank*, *betweenness centrality*, and *closeness centrality* are applied to evaluate the user's influence (Das et al. 2018). Furthermore, a recent work by Esposito et al. (2022) evaluated the relationship between network centrality measures and a firm's success. The work's results suggest success has a strong positive association with centrality measures of the firm and its large investors.

Regarding the network level, network metrics enable quantitative comparison between graphs and analysis of temporal evolution (Hansen et al. 2019). Between the metrics used, we can find counts of nodes and links, average counts, or the application of concepts such as *density* and *centrality*. Antonakaki et al. (2018) used the average node degree and the average of incident edges to measure the evolution of a Twitter social graph over time. Said et al. (2019) conclude that Twitter communities have unique attributes that may impact the social media usage of their users.

Another way to dissect and extract information from a network is to apply algorithms that output some relevant structure or characteristic in the data. An essential concept in networks is that of the group or community: a set of nodes more densely connected between themselves than to others. The methods that find those groups are called community detectors and work as cluster algorithms (Hansen et al. 2019). These social media communities are essential for business, enabling a fast way to cultivate online brand awareness (Zaglia 2013). The community detection algorithms commonly used in the literature are based on modularity optimization. Modularity measures the strength

of the division of a graph. High values imply the graph has dense connections between the module's nodes and sparser connections between nodes of the different modules (Blondel et al. 2008). The modules represent the clusters and, in this case, the communities. The modularity optimization algorithms will explore every node if the modularity score increases when changing between modules. The specific steps and parameters depend on the algorithm used since, in the literature, many adaptations exist depending on the graph characteristics. Regarding social networks, Devi and Poovammal (2016) performed a complete review of the applicable options. One algorithm that stands out for social media platforms is found in the work of Leicht and Newman (2008), which considers the direction of the edges (connections). An example of an application of modularity optimization in Twitter communities can be found in Cruickshank and Carley (2020). The authors used multi-view modularity clustering to characterize and analyze hashtag COVID-19 pandemic Twitter communities.

In summary, social graphs represent social media relationships formed by users following each other. Analyzing these graphs at the node level provides insights into individual users' activity, popularity, and influence. At the network level, based on modularity, community detection algorithms can identify highly connected community nodes, acquiring information that is valuable for businesses looking to build brand awareness.

Startups and social media communities

Existing literature reveals investigations on the possible connections between startups and social media platforms. Lugović and Ahmed (2015) found a positive correlation between startups' Twitter usage and the total investment in their country of origin. Zhang et al. (2017) analyzed startups' Facebook and Twitter metrics, discovering that active engagement positively correlates with startup crowdfunding success. Ko and Ko (2021) conducted a social media analysis regarding fashion startups using Instagram as the data source. They conclude that the startups presenting a higher number of *followers* showed a higher probability of succeeding in crowdfunding projects, meaning their popularity on Instagram helps raise funds. Hadley et al. (2018) conducted a study regarding startups to analyze how their influence and popularity may affect their funding by combining US-based technology startups with venture capitalists and using Twitter as the data source. The authors found that the more central startups in the network, i.e., the most influential ones, received better funding and presented a more significant revenue. A similar study by Esposito et al. (2022) found that the network centrality of a firm and its large investors positively affects business success.

Ruggieri et al. (2018) aimed to identify trends in thriving startups' digital activity. The study indicates that startups predominantly use digital platforms because of their cost-effective functionality. In fact, social media platforms possess a widespread reach, are easy to access, and incur low operating costs, making them the ideal digital marketing gateway for startups to monitor the market. Furthermore, the study inferred that a community of clients and companies, as service providers, is crucial for business success.

Communities are fundamental for a positive impact on digital platforms on the startup, primarily social media communities, since they provide positive or negative opinions on both the products and the companies. Word-of-mouth is critical in everyday oral communication, creating an impression or idea about a specific subject (Keller

2007). In the realm of digital platforms, opinions are called electronic word-of-mouth (eWOM) (Hennig-Thurau et al. 2004), and social media are ideal tools for eWOM. Chu and Kim (2011) describe that eWOM enables the creation of a large community, which allows for increased digital engagement via social interactions, such as *comments*, *likes*, *shares*, and *followings*. The large quantity of those interactions might help raise a positive feeling in the social media profile (Wolny and Mueller 2013).

Following the literature presented, the relationship between startups and social media platforms provides startups with cost-effective digital marketing opportunities to monitor the market and create communities of users and service providers. These communities help spread positive opinions about the products and companies via electronic word-of-mouth, increasing digital engagement through social interactions. Several studies have found that startups with active engagement on social media platforms have a higher chance of succeeding in crowdfunding projects or receiving better funding from venture capitalists.

Methodology

To visualize each startup's Twitter community, we employed a methodology based on the NAV process model (Hansen et al. 2012). This methodology stresses the need for a heavy interactive process built around the following phases: (1) Define the visualization goal, (2) Collect and structure data, (3) Interpret data, and (4) Report results. Figure 1 illustrates the process pipeline, describing each one of the step phases.

As we aim to understand how the *following/follower* relations create communities, in the first step, we defined our goal as that of constructing an informed visualization of Portuguese IT startups social media communities in Twitter, in alignment with the previously proposed research questions.

The next step consisted of collecting and transforming the social media data extracted via the startups' Twitter accounts into structured data. This study case features the startups in the information technology domain that are active on Twitter, founded by Portuguese, or have headquarters in Portugal. The eight chosen startups are: *attentiveMobile*, *codacy*, *DefinedAi*, *feedzai*, *prodsmart*, *Talkdesk*, *Unbabel*, and *Virtuleap*. The names of the startups are presented using the Twitter account username. *attentiveMobile* is a B2B company that offers a personalized mobile messaging platform; *codacy* is an automated code review platform; *DefinedAi* is a company that develops artificial intelligence training data services and solutions; *feedzai* is an artificial intelligence startup, and its core business is finance risk management; *prodsmart* deals with transforming factories into digital and smart ones by employing automation software to control workflows and production; *Talkdesk* is a platform to support sales teams for customer satisfaction and cost savings; *Unbabel* enables companies to serve customers in their native language with a scalable translation across digital channels; Lastly, *Virtuleap* offers a virtual reality application that promotes brain health, supported by a library of games designed by neuroscientists.

Then, we extracted data from the Twitter accounts of users who follow the companies and users whom the startups follow. The extracted data format corresponds to the Twitter user object, from which the following features have been considered: *id*, *screen_name*, *followers_count*, and *friends_count*. Subsequently, the data was structured into a

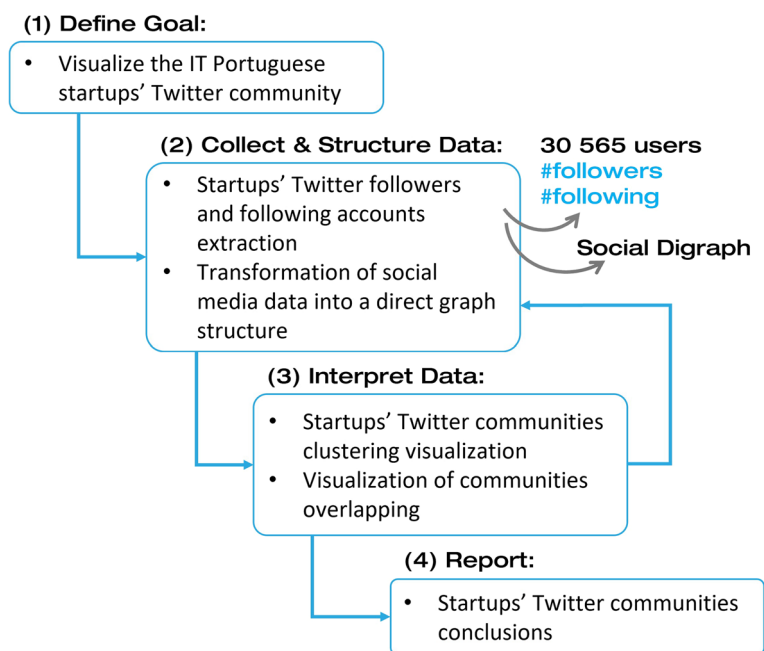


Fig. 1 Current project’s pipeline using the NAV process model (Hansen et al. 2019)

social digraph, with each node representing a user and each link denoting a following relationship, thus achieving a dataset with users that follow or are followed by startups.

After the organization of the data into a digraph structure, we resorted to using different visualizations to interpret the data and enable information extraction. In order to visualize the social graph, we employed *Gephi* (Bastian et al. 2009) and defined a layout and community clustering for data interpretation. As displayed in Fig. 1, steps (2), data structuration, and (3), interpretation, occur in an iterative fashion, where the visualization and respective interpretation may require a different organization of the data to explore emerging insights further. For this case study, this happened mainly when trying to visualize the communities’ overlap. In the related literature, no specific visualization for the overlap between communities in a large social graph has been found, which led to a deeper exploration of possible visualization techniques, such as the ones presented in the coming sections, that in turn required different data organization.

Dataset

The Twitter API was used to extract relevant data, that is, data from the users that follow the eight chosen startups active on Twitter - *followers* - or users that the startups follow - *following*. The extraction occurred on May 31st, 2022, resulting in 30,565 accounts of Twitter users. Table 1 presents the number of *followers* and *following* users for each one of the companies, and Fig. 2 shows an illustration of the respective percentages in terms of the total number of links (edges) for each company using a stacked bar chart.

In terms of descriptive quantities, notably, all startups present a higher number of *followers* than of *followings*, meaning that their communities are mostly composed of Twitter users who follow their accounts. The only exception is *prodsmart*, for which the distribution of *followers* and *following*, although being approximately identical, shows

Table 1 Comparison of the startups counts of *followers* and *following*

Startups	Following	Followers
@attentivemobile	387	6842
@codacy	283	5048
@Definedai	170	1908
@feedzai	920	3132
@prodsmart	1037	905
@Talkdesk	685	7252
@Unbabel	1116	3627
@virtuleap	4	877

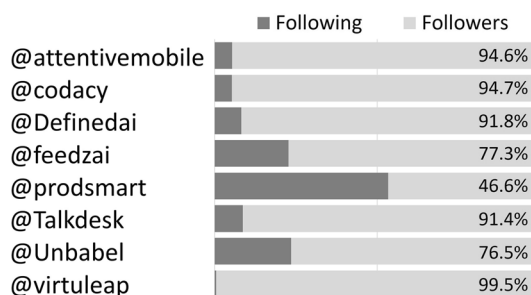


Fig. 2 Distribution of *followers* and *following* users by startup

that this company mostly follows others. Interestingly, while *Virtuleap* is the startup presenting the smaller community, it also presents an expressively higher percentage of *followers* than of *following*, being the startup showing the highest rate of *followers*, 99.5%, closely followed by *attentivemobile* and *codacy*, with 94.6% and 94.7%, respectively. *Talkdesk* is the startup showing the highest number of *followers* (7252), followed by *attentivemobile* (6842), and *codacy* (5048). *Virtuleap* stands out as the startup with the smallest number of links, likely attributed to its relatively recent foundation year: 2018.

Social digraph creation

After extraction, data has been structured into a social digraph, that is, a directed graph, where a node represents a user and a directed edge represents the user-following-user relation. This action resulted in a graph consisting of 30,565 nodes and 34,184 directed links/edges. The graph’s density, 7.32×10^{-5} , indicates that it is a very sparse graph, meaning that it presents very few edges compared to the maximum possible number of edges for this number of nodes. This sparsity was expected since the graph nodes represent mostly users who follow the startups, while information about the other nodes those users may follow or about their followings was not extracted. No weights were used since we have not extracted any quantitative information towards this end.

To enable community visualization, a community detection algorithm was used. As previously mentioned in the related work section, based on the description of the analysis performed by Devi and Poovammal (2016), we chose to use the modularity algorithm for social digraphs created by Leicht and Newman (2008). Modularity measures the density of the connections within a graph’s structure and groups it into

modules. As expected, the results showed eight modules, one for each startup community, which has been validated by discovering the company at the center of each founded community. To evaluate the results, we measured the modularity score, ranging between 0 and 1, with higher values indicating a stronger community structure (McDiarmid and Skerman 2020). Our case study graph achieved a modularity value of 0.768, suggesting a robust community structure. We used the Python library *CDLib* (Rossetti et al. 2019) for the algorithm and evaluation.

For our analysis, we have selected a modularity-based algorithm due to its success in these social network scenarios showing few large communities, even though they usually suffer from the resolution limit problem. However, community detection algorithms are known to be quite unstable, with different algorithms sometimes producing different results. To gauge the stability of our findings, we have also tested an alternate method proposed by Traag et al. (2015) to evaluate the eventual differences that may arise. This alternate algorithm uses asymptotical surprise, a metric that, like modularity, is employed to evaluate the quality of community detection in networks. This metric is a statistical approach that calculates the probability of observing at least a certain number of internal edges within the communities, given the total number of edges in the network. We choose to apply this algorithm because it is nearly unaffected by the resolution limit problem, the modularity optimization primary weakness. The results with the new algorithm are mostly identical to the previous results and with few large communities. We decided to carry on using the results obtained by the algorithm proposed by Leicht and Newman (2008) for the analysis.

Figure 3 shows the number of nodes—users—for the found communities, comparing it with the respective number of *followers* plus *following* users of the startup, i.e., the total count of links of each of the companies. We have numbered each modular class from 1 to 8, representing each one of the startups' communities.

Bearing in mind the quantities presented in Fig. 3, *Talkdesk* and *attentivemobile* present the larger communities and *Virtuleap* the smallest, as expected. Interestingly, the number of nodes for each of the communities is lower than the sum of each startup's number of *followers* and *followings*. This fact indicates that some of the users are shared between communities, meaning that the users who follow or are followed by the startups may intersect. *Virtuleap* appears as an exception, presenting an identical number of linked users and community nodes: 881 linked users and 873 nodes in the community. This means that only eight users are shared with different companies.

Data visualization and interpretation

This section presents visualizations of the social Twitter communities built around the different startups, focusing on the extent and characterization of the overlap between them. Overlap in this context means that a user follows more than one of the startups or is followed by more than one startup. To extract knowledge that may inform the creation of social media marketing strategies, We examined information at the user level for the ones found in an overlap situation to understand their type profiles and characterize the general communities of *followers* and *following*.

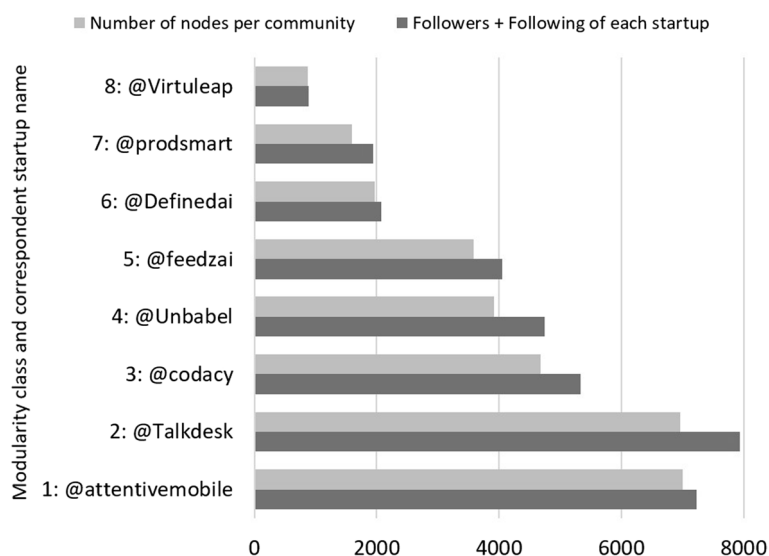


Fig. 3 Communities size distribution

Social digraph visualization

In a first step towards informative visualizations, the *circle pack layout* (Groeninger 2015) was applied to the social digraph to perceive what and how the communities are outputted by the modularity algorithm, as described in the previous section. This layout organizes the network into circles using a selected set of network features. In this case, the only selected feature has been the modularity class, which allowed us to create a visualization where each circle represents one of the communities. To help interpret the graph, the nodes and links were colored using different colors. Since the central (centroid) of each formed cluster was found to be each of the startups, the coloring rule used a color based on the logotype of each company to color the nodes corresponding to their modularity class value. However, the links were colored using a mix of the colors from the source and the target nodes. Figure 4 displays the visualization thus obtained for this social digraph, where each circle represents one community. The name of the correspondent startup community central point, the startup, is also shown.

As expected, the graph shows that the communities varying sizes, with the *Talkdesk* community showing its larger number of members and the *Virtuleap* community the smallest, as seen in Fig. 3. Furthermore, as anticipated, the graph shows a significant overlap between the several communities, with nodes (members) connected to more than one community in the social network graph. However, the exact degree of overlapping cannot be determined from this particular visualization alone, and additional analysis is needed to assess the implications of this overlap for the startups involved. It is essential to understand the degree of overlapping and how it may or may not differ in terms of *followers* and of *followings* since this distinction may have meaningful implications regarding actions in a startup’s social media strategy.

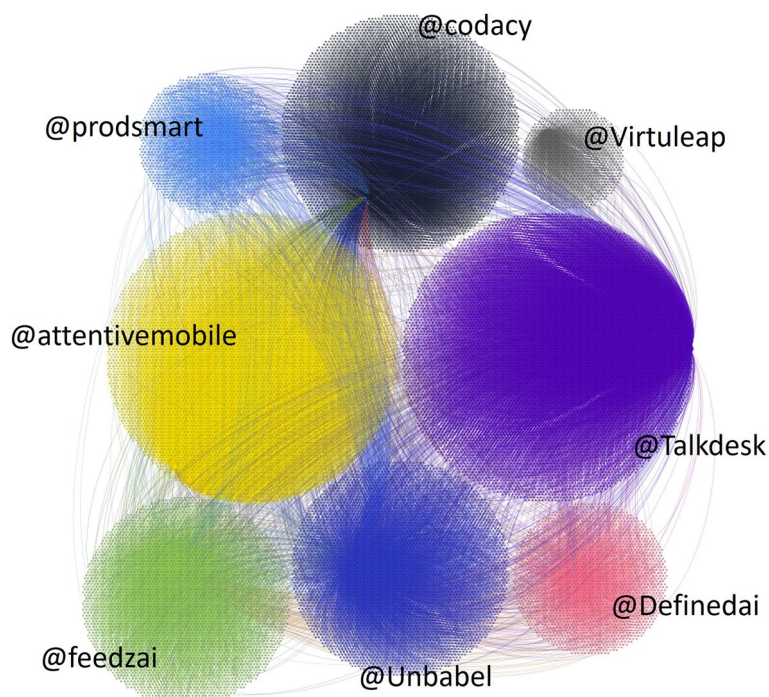


Fig. 4 Social graph visualization: circle pack layout using the modularity class

Overlap of communities

Understanding the communities' common points may help startups expand their networks and gain a competitive advantage in their respective markets. The following section aims to answer this question and grasp how the communities overlap. The overlap between communities in the context of social media refers to the situation where some users in the community around one startup are also part of other communities established around different startups. In other words, as depicted by the edges traversing different communities in Fig. 4, these users follow more than one startup or are themselves followed by more than one startup.

To distinguish between the two different situations—when a startup follows a user or when a user follows a startup—we divided the network in two: one representing only the user-*follow*-startup relation and the other representing the startup-*following*-user relation. The resultant graphs' dimensions show the difference in number between these two features that have been previously noticed in the analysis of Fig. 2: while the graph with the startups' *followers* has 27,682 nodes and 23,270 edges, the one for the users followed by the startups consists of 4306 nodes and 4225 edges.

However, for a better understanding of what is the overlap between startups, the visualizations must be more comprehensive. Data grouping is required since we discovered that some of the users in the overlap are shared between two or more communities. To accomplish this, we created two matrices: one for the *followers* of the startups and the other for the users the startups are *following*. The users in common for each combination of startups were counted, resulting in an overlap with 1289 *followers* (Fig. 5) and one with 249 *following* users (Fig. 6). The Python library *upsetplot* (Lex et al. 2014) was used to visualize the matrices, and the resultant plots represent the overlap in both of the

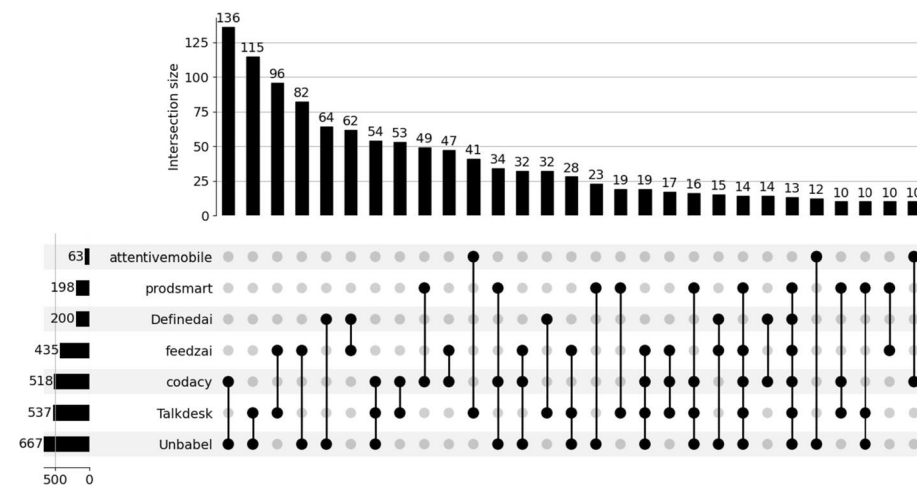


Fig. 5 Followers overlap visualization

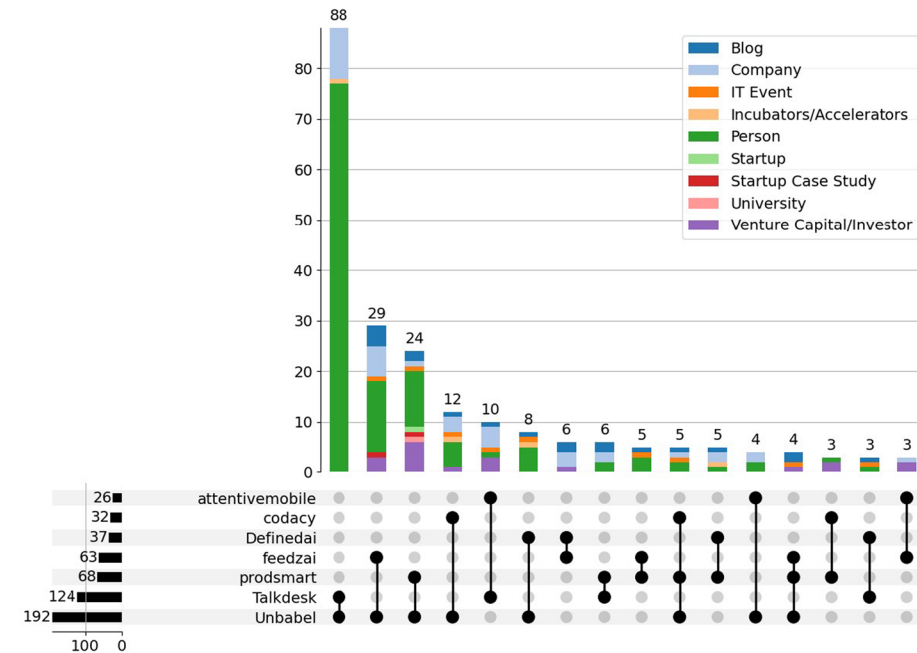


Fig. 6 Following overlap visualization, including user categories

domains: Fig. 5 plots the different combinations of startups having *followers* in common and how many and Fig. 6 regards combination and counts for startups *following* shared users.

In fact, Fig. 5 is a subplot of the total visualization of the overlap with *followers*. Since a considerable number of users was found to be in overlap, for a more effective visualization, the data were filtered by applying a threshold to show values only when the number of users in common was above nine. As seen in the previous section, *Virtuleap* displays only eight users in an overlap situation, and thus *Virtuleap* appears not to be sharing *followers* with other startups in this scenario where the

threshold was applied. Notably, the startup sharing most *followers* with other startups is *Unbabel*, which shares a total of 667 *followers* with all the other six startups, even if under different combinations. The one sharing the least is *attentivemobile*, with a total of 63 shared *followers*. Furthermore, *attentivemobile* only overlaps with the startups showing the three highest numbers for overlapping. Interestingly, all these shares are duets that is these *followers* follow only two companies: *Talkdesk* (41 shared *followers*), *Unbabel* (12 shared *followers*) and *codacy* (10 shared *followers*). Nevertheless, this is an unexpected result since *attentivemobile* displays the biggest counts, both in its community dimension and for the total number of linked users. The pair of startups with more *followers* in common is *codacy* and *Unbabel*, showing an overlap of 136, followed by *Talkdesk* and *Unbabel* with 115, and *Talkdesk* and *feedzai* with 96. The visualization presents eight trios, with the one sharing more *followers* composed by *Unbabel*, *Talkdesk*, and *Codacy*, with a total of 54 users in common. Additionally, we can observe two quartets (with 19 and 16 users), one quintet (with 14 users), and one sextet (with 13 users).

Next, the shared *following* users have been analyzed. Figure 6 displays the overlap of the *following* by the startups, which may express coincident digital marketing options between them. Knowing the users in or not in the overlap can help to direct a digital marketing strategy. Therefore, we decided to manually categorize the 249 users in the overlap. Since the categorization was manual, we chose a set of global and vague categories of the startup ecosystem to facilitate manual categorization. The annotation procedure looked at the user's profile and bio description (usually stating the type of profile) and searched in Google for verification if needed. This annotation resulted in categories and in the colored visualization shown in Fig. 6. Similarly to what has been done with the *followers*' scenario, we applied a threshold and considered values only above two shared *following* relations. Again, *Virtuleap followings* do not appear in an overlap situation since this startup only follows 4 different users.

Both the startups showing the higher and the lowest levels of overlap are still the same: *Unbabel* shares more *following* users (192 users shared in total) and *attentivemobile* shares the least (26, in total). Notice that the latter has exchanged its behavior: while it shared many *followers* with other startups in the previous analysis, it now differentiates by following different users. By the analysis of Fig. 6, we can also conclude that *Talkdesk* and *Unbabel* are the pair presenting the highest number of *followings* in common, 88 users. Interestingly, most of the overlap occurs among Twitter users that are from the category "Person", mostly experts in the core business field of these two startups: applications encompassing natural language processing. The next profile of common *followings* are "Company" and, naturally, "Incubators/Accelerators." The next pairings and groups of startups show much less following in common, as can be noted by the abrupt decrease shown from the second column of Fig. 6. We can see two trios: one consisting of five shared users and the other shares four. The trio sharing more *following* relations—*codacy*, *prodsmart* and *Unbabel*—consists of: the CEO of *codacy* (a "Person"), a Portuguese journalist (a "Person"), the Lisbon Investment Summit (an "IT Event"), *beta-i* (a "Company"), and a Portuguese blog (a "Blog").

Overlapping users characterization

In this section, we study the users that were found in overlapping communities. Understanding who startups follow and who follows them is important to better characterize the overlap.

Regarding the *followers* overlap, we found it appealing to understand what users follow most of the startups. These users are defined as the ones that follow four or more companies, comprising a total of 68 users. Next, node-level metrics were applied to evaluate their activity and popularity levels. These metrics were used both for the 68 users selected in the *followers* and for the 249 users selected in the *following*. Concerning user activity, we retrieved the total number of tweets and retweets in their Twitter profiles. We used a version of the *Structural Advantage* for popularity, the FF ratio (Cappelletti and Sastry 2012), that involves the number of *followers* and of *followings*:

$$FF \text{ Ratio} = \frac{\#followers}{\#followers + \#following}$$

This ratio indicates how popular a user is on the social media platform, with higher ratios indicating that a user has more *followers* than is *following* others. Values between 0 and 0.5 indicate that the user is not particularly popular, following more users than being followed.

Concerning the *followers'* overlap, Table 2 shows the percentages for each type of Twitter user that has been encountered in this set of shared users and also the corresponding average values for the FF ratio and activity of each of the types.

Clearly, more than half of the users are classified as “Person,” accounting for 56.9% of the total, followed by “Blog” (10.8%), “Incubators/Accelerators” (10.8%), “Company” and “Venture capital/Investor,” both showing 6.2%. The categories showing the least are “IT event” and “Startup,” with 4.6%. The type “Incubators/Accelerators” displays the highest average FF ratio (0.65), followed by “IT event” (0.60), “Venture capital/Investor” (0.54), all above-average level of popularity. The remaining types show less favorable FF ratios, especially the type “Person,” which shows an average FF ratio of 0.27.

Regarding the activity levels, “Incubators/Accelerators” shows the highest average levels (3988), indicating that this type of user is more engaged with their *followers* than the remaining user types. Next, we see “Blog,” with 2450, “Person” (2245), and “Company” (2216), all presenting similar levels of activity. Finally, we have “IT Event” (1184), “Venture capital/Investor” (989), and “Startup” (287).

Table 2 *Followers'* overlap: percentage of profiles encountered and node-level measures

User type	n	%	FF ratio (average)	Activity (average)
Person	37	56.9	0.27	2245
Blog	7	10.8	0.44	2450
Incubators/accelerators	7	10.8	0.65	3988
Company	4	6.2	0.38	2216
Venture capital/investor	4	6.2	0.54	989
IT event	3	4.6	0.60	1184
Startup	3	4.6	0.34	287

In terms of the overlap that exists in the *following* relations, that is, the users that startups follow, Table 3 presents the percentage of each type of user in this overlap, as well as the average FF ratio and the activity counts. The majority of the users followed by the startups are “Person”, accounting for 53% of the total, followed by “Company” (18.10%), “Blog” (10.50%), “Venture capital/Investor” (8.50%), “IT Event” (4.40%), “Incubators/Accelerators” (2.80%), and “Startup” (1.20%). In this profiling, we can also encounter the type “University” (0.40%), albeit showing the least number of followings. The specific university is Instituto Superior Técnico, Lisbon, Portugal, from which many of the case study startups are either spin-offs or from where their founders obtained their degrees.

Comparing these users’ activity levels, we see that the type “Blog” shows the highest average activity level (107,924 tweets/retweets), which is consistent with this type’s main function. A “Blog” engages with its audience by regularly sharing meaningful content, which also explains why this is one of the most popular user types found in this overlap. In fact, when comparing this set of blog users with one of the *followers* previously discussed, the averages now are considerably higher than before, which entails that the blogs followed by startups are respected and credited blogs in this ecosystem. The next ranking position in terms of activity is occupied by the type “Person” that, with a 28,999 average count of tweets/retweets, positions itself rather distant from “Blog”. “Company”, showing an average of 26,494 showing an average of 26,494, follows closely. All the remaining types show considerably less activity when compared with any of the previous ones. It should be noticed, however, that all these user types display higher activity levels on average than those found in the *followers* set. Namely, the lowest activity count for the *followings*—3216 average tweets/retweets (Table 3)—is still higher than the highest count for the *followers* (Table 2).

In terms of popularity, we can see that the user types with the highest FF ratios are “Blog” and “University” (both attaining 0.95), followed by the startups of this case study (0.88), whose popularity levels are close to “Company” (0.87), again close to “Venture capital/Investors” (0.84) and “Startup” (0.84) popularity levels. On the other hand, user types “IT Event”, “Incubators/Accelerators,” and “Person” show the lowest FF ratios, but still, all of these users can be classified as popular since any of them shows to have more *followers* than *following* others.

Notably, our case study startups mostly share the action of following persons and companies relevant of the ecosystem. Other *following* types are “Venture Capital/Investor”

Table 3 *Followings’* overlap: percentage of profiles encountered and node-level measures

User type	n	%	FF ratio (average)	Activity (average)
Person	132	53	0.71	28,999
Company	45	18.1	0.87	26,494
Blog	26	10.5	0.95	107,924
Venture capital/investor	21	8.5	0.84	13,563
IT event	11	4.4	0.74	18,759
Incubators/accelerators	7	2.8	0.74	3216
Startup	3	1.2	0.84	8459
Startup case study	3	1.2	0.88	7752
University	1	0.4	0.95	6664

and “Blog.” While both show a relatively high FF ratio, indicating popularity, the average activity levels are quite different. “Venture Capital/Investor” activity is not expressive, which may indicate that these users look upon Twitter more as an observational or promotional tool than with the intent of engaging with other users. “IT Event” and “Incubators/Accelerators” show the lowest average FF ratios and activity levels, suggesting these users albeit being both quite popular, are not as active on this Twitter ecosystem as the remaining types.

Significantly, the categories of users found in the studied overlaps all belong to the startup universe, highlighting the interconnectedness of this ecosystem. Upon examination of the categorization of users in both overlaps, it becomes evident that most of the users who follow startups and those whom startups follow are individuals, and the second most common user type that startups follow is companies. Furthermore, startups follow users showing higher activity levels, with an average of 19,669 tweet/retweet count, which strikingly compares with those users that follow startups, showing expressively lower activity levels. In terms of popularity, startups follow users that are more popular than those who follow them, with an average FF ratio of 0.46 among *followers*, compared to an average of 0.84 among those whom startups follow.

Conclusions

Startups, known for their innovation and limited resources, must raise funding and reach customers on a restricted budget. Social media platforms offer a cost-effective gateway to various communities, enabling startups to achieve their goals and expand their business. Active engagement on social media can lead to better funding and create communities of users and service providers. This work investigates how startups fare on social media platforms, namely on Twitter, and if they create their own communities. Startups can benefit from data-driven projects using social media data, like the one in this study, as a strategic digital marketing knowledge source and unlock the power of social networks.

Our primary research goal was to understand if the follower/following relations on Twitter’s social graph create social communities around startups. Using Portuguese IT startups as a case study, we proceed with collecting and treating the needed Twitter data to create meaningful visualizations, enabling us to extract relevant knowledge about the communities. The case study data, using eight IT startups having some type of connection with Portugal, was organized into a social digraph, representing the users and links between the different users found in the data, resulting in a graph with over 30,000 nodes. Applying a community detection algorithm enabled the identification of communities in the data. Notably, the results showed that the communities were built around our eight chosen startups. By encoding the color of the social graph, the created visualization highlighted each one of the startup’s community of users. Thus, we showed that IT Portuguese startups form their own social communities on Twitter and that these communities heavily relate to the fact that these companies are startups and Portuguese-related. Next, we used other types of visualizations, paired with manual user categorization and node-level metrics, that enabled us to characterize the found communities and find out that, as expected, these communities show an interesting degree

of overlap between them, either from the perspective of the startups' *followers*, as from the perspective of whom the startups are *following*.

We discussed the concept of overlap between communities on social media, which occurs when users belong to multiple communities established by different startups. We presented two graphs, one for *followers* and one for *following*, and analyzed the overlap between the startups. The resultant plots provide a comprehensive understanding of the social digraphs of startups and show the overlap between different startups.

As understanding communities is essential, we did two visualizations representing the overlap, one showing the startups' *followers* and the other showing whom they follow, to analyze the overlap between startups. The resulting graphs fully represent the startups' communities overlap. Then, we manually categorized the users in the overlaps. We discovered that all categories of users belong to the startup universe, highlighting the interconnectedness of this ecosystem. Examining user categories in both overlaps revealed that most users who follow startups and those whom startups follow are persons. Companies represent the second most prevalent user category that startups choose to follow, whereas blogs and incubators/accelerators are the second categories that most follow startups. In addition, startups tend to follow users who post high volumes of tweets and have high popularity levels. On the other hand, those who follow startups have low activity levels and are not popular.

Theoretical and managerial implications

As stated in the related literature, social media platforms offer startups affordable digital marketing opportunities to monitor the market and establish user and service provider communities. This study proposes a methodological process for social media community analysis on platforms like Twitter based on Network Analysis and Visualization (Hansen et al. 2019). The specific process displayed in this study involves five steps. First, select the users intended to be analyzed as the center of the communities. In this case, the center users were the eight chosen startups. Second, collect information on the *follower* and *following* users. Third, perform data transformation, including: user classification and user profile description, using popularity and activity metrics; data structuring into a digraph; creating two distinct tables, one for the *followers* and another for the *followings*. Fourth, visualize the data: use clustering algorithms and color to illustrate the communities formed. Additionally, use visualization tools like *upsetplot* to gain insights into the overlaps between these communities. Finally, the fifth step consists of concluding and data-driven supported strategies gained by analyzing the communities created by the selected central users. This process can be applied in future social media community analysis studies.

This study's first managerial contribution consists of a viable process for the visualization of a startup's community and when in an ecosystem, understanding the existence of shared *followers* or *followings*. Startups can use this method to monitor similar companies, select users to follow, study which users others follow, and perceive their community. They can build marketing campaigns for extending it as needed, facilitating the creation of a wider social media community that might benefit them or benefit from the ecosystem.

The second implication relies on the type of users found in the IT startup's overlapping communities. After manually categorizing 317 Twitter users, we found that the user's profiles in the overlap of the Portuguese IT startup communities are: "Person," "Company," "Blog," "Venture Capital/Investor," "IT Event," "Incubators/Accelerators," "Startup," and, exceptionally, "University." The first three categories found were all profiles of the IT area or of the startups' core business. Furthermore, the last five categories relate to the startup's ecosystem; thus, the ecosystem is mostly considered a closed environment. Examining the overlap and the type of users comprising it enables us to perceive the communities' common points, which may help startups expand their networks and gain a competitive advantage in their respective markets.

The third relevant managerial implication relies on the activity and popularity of the users in the overlap of the communities. Startups seem to follow users who post high volumes of tweets and have high popularity levels, while those who follow them generally show low average activity levels and also, on average, are not considered as popular. Accordingly, the study recommends that startups keep studying their Twitter users' activity and popularity profiles to stay relevant in their field and reach a wider number of other users.

Limitations and future work

Like all studies, the study has limitations that should be addressed in future research. Firstly, this study used eight Portuguese IT startups to develop a proof of concept. Future work should explore wider communities of startups, study startups working in different areas or industries, and be based in different countries to confirm and expand our results. Secondly, we focused only on social media communities created by startups on Twitter. Future research should compare community creation and overlap within other social media platforms. Thirdly, we manually performed the user categorization, enabling us to understand the types of community users overlap. However, if larger in scale, future studies should automatize the categorization process to allow the scalability of the process. Fourthly, this research focuses on the social graph created by the action of following on Twitter. Therefore, future works should consider extending the graph with more social variables, for example, using the number of interactions (likes, replies, and mentions) between two users as edges' weights. Lastly, future research should consider the life cycle phases of startups as described in a related work (Peixoto et al. 2023) and explore the potential influence between those phases and the startups' Twitter activity.

Author contributions

Conceptualization: All; methodology: All; software: AP; validation: All; investigation: All; writing original draft preparation: AP; writing, review, and editing: All; supervision: AA e NA; All authors have read and agreed with the published version of the manuscript.

Funding

This work was partially supported by Fundação para a Ciência e a Tecnologia, I.P. (FCT) namely by UIDB/04466/2020 and UIDP/04466/2020 (ISTAR_Iscte); UIDB/04152/2020 (MagIC/NOVA IMS); UIDB/50021/2020 (INESC-ID); and UIDB/03126/2020 (CIES_Iscte).

Availability of data and materials

The code used to extract the Twitter user accounts and to create the social digraph is available at <https://github.com/arhpo-iscteulpt/TwitterSocialDigraph>.

Declarations

Ethical approval and consent to participate

This declaration is not applicable.

Competing interests

The authors declare no competing interests.

Received: 14 June 2023 Accepted: 13 September 2023

Published online: 20 September 2023

References

- Abdelsadek Y, Chelghoum K, Herrmann F, Kacem I, Otjacques B (2018) Community extraction and visualization in social networks applied to Twitter. *Inf Sci* 424:204–223. <https://doi.org/10.1016/j.ins.2017.09.022>
- Antonakaki D, Ioannidis S, Fragopoulou P (2018) Utilizing the average node degree to assess the temporal growth rate of Twitter. *Soc Netw Anal Min*. <https://doi.org/10.1007/s13278-018-0490-5>
- Antonakaki D, Fragopoulou P, Ioannidis S (2021) A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Syst Appl* 164:114006. <https://doi.org/10.1016/j.eswa.2020.114006>
- Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp*. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Cappelletti R, Sastry N (2012) Iarank: ranking users on twitter in near realtime, based on their information amplification potential. In: 2012 international conference on social informatics, pp 70–77
- Chu S-C, Kim Y (2011) Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites. *Int J Advert* 30(1):47–75. <https://doi.org/10.2501/IJA-30-1-047-075>
- Cruikshank IJ, Carley KM (2020) Characterizing communities of hashtag usage on twitter during the 2020 COVID-19 pandemic by multi-view clustering. *Appl Netw Sci*. <https://doi.org/10.1007/s41109-020-00317-8>
- Das K, Samanta S, Pal M (2018) Study on centrality measures in social networks: a survey, vol 8, no 1
- Devi JC, Poovammal E (2016) An analysis of overlapping community detection algorithms in social networks. *Procedia Comput Sci* 89:349–358
- Education First (2022) Ef epi 2022 – ef english proficiency index. <https://www.ef.com/wwen/epi/>. Accessed 13 June 2023.
- Esposito C, Gortan M, Testa L, Chiaromonte F, Fagiolo G, Mina A, Rossetti G (2022) Venture capital investments through the lens of network and functional data analysis. *Appl Netw Sci* 7(1):42. <https://doi.org/10.1007/s41109-022-00482-y>
- Gabrielkov M, Legout A (2012) The complete picture of the Twitter social graph. In: CoNEXT Student 2012—proceedings of the ACM conference on the 2012 CoNEXT student workshop, pp 19–20. <https://doi.org/10.1145/2413247.2413260>
- Groeninger M (2015) Gephi - circular layout. <https://github.com/mgroeninger/gephiplugins/tree/master/modules/CircularLayout>. Accessed 13 June 2023
- Hadley B, Gloor PA, Woerner SL, Zhou Y (2018) Analyzing VC influence on startup success: a people-centric network theory approach. *Stud Entrep Struct Change Ind Dyn*. https://doi.org/10.1007/978-3-319-74295-3_1
- Hansen DL, Rotman D, Bonsignore E, Milic-Frayling N, Rodrigues EM, Smith M, Shneiderman B (2012) Do you know the way to SNA?: A process model for analyzing and visualizing social media network data. In: 2012 international conference on social informatics, pp 304–313
- Hansen D, Shneiderman B, Smith MA (2019) Analyzing social media networks with nodexl: insights from a connected world. Morgan Kaufmann, Burlington
- Hennig-Thurau T, Gwinner KP, Walsh G, Gremler DD (2004) Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet? *J Interact Mark* 18(1):38–52. <https://doi.org/10.1002/dir.10073>
- Hingle M, Yoon D, Fowler J, Kobourov S, Schneider ML, Falk D, Burd R (2013) Collection and visualization of dietary behavior and reasons for eating using twitter. *J Med Internet Res* 15(6):1–16. <https://doi.org/10.2196/jmir.2613>
- Hu M, Wongsuphasawat K, Stasko J (2017) Visualizing social media content with sententree. *IEEE Trans Vis Comput Graph* 23(1):621–630. <https://doi.org/10.1109/TVCG.2016.2598590>
- Keim D, Andrienko G, Fekete J-D, Görg C, Kohlhammer J, Melançon G (2008) Visual analytics: definition, process, and challenges. *Information visualization*, vol 4950 LNCS. Springer, Berlin, Heidelberg, pp 154–175
- Keller E (2007) Unleashing the power of word of mouth: creating brand advocacy to drive growth. *J Advert Res* 47(4):448–452. <https://doi.org/10.2501/S0021849907070468>
- Ko J, Ko E (2021) What fashion startups should know before launching Crowdfunding projects: focusing on Wadiz reward Crowdfunding. *J Glob Fashion* 12(2):176–191. <https://doi.org/10.1080/20932685.2020.1870521>
- Leicht EA, Newman ME (2008) Community structure in directed networks. *Phys Rev Lett* 100(11):118703
- Lex A, Gehlenborg N, Strobelt H, Vuilleumot R, Pfister H (2014) Upset: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 20(12):1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>
- Lugović S, Ahmed W (2015) An analysis of twitter usage among startups in Europe. In: Infuture2015: e-institutions-openness, accessibility, and preservation proceedings. INFUTURE, pp 299–308
- McDiarmid C, Skerman F (2020) Modularity of erdős-rényi random graphs. *Random Struct Algorithms* 57(1):211–243
- Molla A, Biadgie Y, Sohn KA (2014) Network-based visualization of opinion mining and sentiment analysis on twitter. In: 2014 international conference on IT convergence and security, ICITCS 2014, 2012–2015. <https://doi.org/10.1109/ICITCS.2014.7021790>

- Peixoto AR, de Almeida A, António N, Batista F, Ribeiro R (2023) Diachronic profile of startup companies through social media. *Soc Netw Anal Min* 13(1):52. <https://doi.org/10.1007/s13278-023-01055-2>
- Portugal Digital & Startup Portugal (2021) Portugal, the best place to startup. <https://portugaldigital.gov.pt/wpcontent/uploads/2022/02/Portugalthebestplacetostartup.pdf>. Accessed 13 June 2023
- Rizvanović B, Zutshi A, Grilo A, Nodehi T (2023) Linking the potentials of extended digital marketing impact and start-up growth: developing a macrodynamic framework of start-up growth drivers supported by digital marketing. *Technol Forecast Soc Change*. <https://doi.org/10.1016/j.techfore.2022.122128>
- Rossetti G, Milli L, Cazabet R (2019) Cdlib: a python library to extract, compare and evaluate communities from complex networks. *Appl Netw Sci* 4(1):1–26
- Ruggieri R, Savastano M, Scalingi A, Bala D, D'Ascenzo F (2018) The impact of digital platforms on business models: an empirical investigation on innovative start-ups. *Manag Mark* 13(4):1210–1225. <https://doi.org/10.2478/mmcks-2018-0032>
- Said A, Bowman TD, Abbasi RA, Aljohani NR, Hassan SU, Nawaz R (2019) Mining network-level properties of Twitter altmetrics data. *Scientometrics* 120(1):217–235. <https://doi.org/10.1007/s11192-019-03112-0>
- Saura JR, Ribeiro-Soriano D, Palacios-Marqués D (2021) Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research. *Ind Mark Manag* 98(August):161–178. <https://doi.org/10.1016/j.indman.2021.08.006>
- Saura JR, Palacios-Marques D, Ribeiro-Soriano D (2023) Exploring the boundaries of open innovation: evidence from social media mining. *Technovation*. <https://doi.org/10.1016/j.technovation.2021.102447>
- Serrat O (2017) Knowledge solutions: tools, methods, and approaches to drive organizational performance. <https://doi.org/10.1007/978-981-10-0983-9>
- Sifted (2020) Top Portuguese startups to follow in 2020. <https://sifted.eu/portugal-startups-top-rankings/>. Accessed 24 July 2023
- Smith A, Chuang J, Hu Y, Boyd-Graber J, Findlater L (2014) Concurrent visualization of relationships between words and topics in topic models. In: Proceedings of the workshop on interactive language learning, visualization, and interfaces, pp 79–82
- Tanner J (2023) The ultimate guide to twitter marketing for small businesses and startups. <https://www.socialmediacollege.com/blog/the-ultimate-guide-to-twitter-forstartups-small-businesses>. Accessed 24 July 2023
- Traag VA, Aldecoa R, Delvenne JC (2015) Detecting communities using asymptotical surprise. *Phys Rev E Stati Nonlinear Soft Matter Phys*. <https://doi.org/10.1103/PhysRevE.92.022816>
- Ward S (2020) Why your business should use twitter. <https://www.liveabout.com/top-reasons-why-your-small-business-should-use-twitter-2948523>. Accessed 24 July 2023
- Wolny J, Mueller C (2013) Analysis of fashion consumers' motives to engage in electronic word-of-mouth communication through social media platforms. *J Mark Manag* 29(5–6):562–583. <https://doi.org/10.1080/0267257X.2013.778324>
- Wu Y, Cao N, Gotz D, Tan YP, Keim DA (2016) A survey on visual analytics of social media data. *IEEE Trans Multimed* 18(11):2135–2148. <https://doi.org/10.1109/TMM.2016.2614220>
- Zaglia ME (2013) Brand communities embedded in social networks. *J Bus Res* 66(2):216–223. <https://doi.org/10.1016/j.jbusres.2012.07.015>
- Zhang Q, Ye T, Essaidi M, Agarwal S, Liu V, Loo BT (2017) Predicting startup crowd funding success through longitudinal social engagement analysis. *Proc Int Conf Inf Knowl Manag Part F* 1318:1937–1946. <https://doi.org/10.1145/31332847.3132908>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
