Horizon 2020

Understanding Europe's Fashion Data Universe

Surveys Design and Crowdsourcing Tasks

Deliverable number: D3.3

Version 2.0

Funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 732328

Project Acronym:	FashionBrain
Project Full Title:	Understanding Europe's Fashion Data Universe
Call:	H2020-ICT-2016-1
Topic:	ICT-14-2016-2017, Big Data PPP: Cross-sectorial and cross-lingual data integration and experimentation
Project URL:	https://fashionbrain-project.eu

Deliverable type	Report (R)
Dissemination level	Public (PU)
Contractual Delivery Date	31/12/2019
Actual Delivery Date	30/12/2019
Number of pages	32, the last one being no. 26
Authors	Alessandro Checco, Soufiana Mekouar - USFD Inès Arous - UNIFR
Peer review	Ying Zhang - MDBS

Change Log

Version	Date	Status	Partner	Remarks
1.0	30/06/2018	Final	USFD	Rejected
2.0	30/12/2019	Final	USFD	

Deliverable Description

The tangible result of task 3.3 will be models generated by means of crowdsourcing which will be used to address our use-cases in WP5 and WP6.

Abstract

This deliverable present the FashionBrain research towards understanding the fashion influencer ecosystem, the way they are related to each other and to the public. We developed a set of tasks and questionnaires that can run on a focused sample of the crowd population.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

AMT	Amazon Mechanical Turk (www.mturk.com), micro-task crowdsourcing platform
API	Application Programming Interface
AUC	Area Under the Curve
EM	Expectation Maximization
RFC	Random Forest Classifier
TF-IDF	Term Frequency-Inverse Document Frequency

1 Introduction

Fashion influencers have in important role in shaping web advertising, contributing to create fashion trends and affecting marketing strategies. FashionBrain broader goal is to develop a set of techniques that will allow to continuously probe social media to keep the knowledge on fashion trends updated. To achieve this goal, it is necessary to understand the way influencers are connected to each other and to the public, i.e. explore the social graph of existing influencers and understand how users start to follow new accounts.

We start in Section 2 by studying the crowdsourcing environment with respect to fashion, and the state-of-the-art of influencer detection.

In Section 3, we first describe the crowdsourcing task used to collect candidate fashion influencers, then, we present Open Crowd, a crowdsourcing aggregation framework designed to infer the real fashion influencers from a set of candidates given by workers, and evaluate it against state of the art aggregation methods.

In Section 4, we design a crowdsourcing experiment where workers actions are recorded while they explore social media feeds to identify new fashion influencers: the corresponding dynamical social graph will be studied and used to develop Machine Learning (ML) techniques (using Flair trained on the FashionTweets dataset of Deliverable D3.4) with humans-in-the-loop to automatically explore a social graph and extract candidate fashion influencers.

Moreover, we will use some traditional statistical techniques to provide estimates on the size of the unknown population of (still) undiscovered emerging fashion influencers.

1.1 Scope of this Deliverable

This Section contributes to the findings of T3.3 "Focused sampling: Crowdsourcing fashion data source search" and contributes to the Core Technology 3: "Crowdsourcing interfaces and quality metrics". It uses findings from T3.1, T3.2, and T3.4. In particular, it uses some models and datasets from D3.4 (Flair model trained on our *FashionTweets Dataset*), and assess their capability to generalise to different contexts. This deliverable is the result of collaboration of UNIFR (Open Crowd), USFD (Fashion Influencer) and Zalando (Flair for fashion content detection).

2 Pilot Experiments and Ecosystem Assessment

In this section we report our pilot studied necessary to understand both the crowdsourcing environment with respect to fashion, and the state-of-the-art of influencer detection.

2.1 Crowdworkers' Fashion Expertise

We first run an experiment to assess crowdworkers expertise in fashion, to then successfully devise a framework to probe social networks for new trends. The first question to answer is: are European crowdworkers able to correctly recognise fashion influencers? Are crowdworkers familiar with them and with the most important social networks used for fashion?

To answer these questions, we used the dataset kindly provided by Fashwell. This dataset contains 118 Instagram fashion influencers, with profile picture, biography and 100 posts each, together with some basic metrics as number of comments and likes per post.

We then have built a task asking the crowd to assess each Instagram account, as shown in Figures 2.2-2.4. The questions are build so that we could assess crowd worker proficiency in the social network, as well as their ability to recognise famous and less famous influencers. We asked the set of questions to three different workers for each influencer.

From a preliminary analysis we can draw the following observations:

- Workers are able to identify that these influencers work in fashion easily, as shown in Figure 2.1. We believe that the crowd can be used successfully to recognise fashion influencers from a larger pool of Instagram accounts.
- Only in 47 of 336 cases the worker was able to recognise the influencer.
- Only in 3 of those 47 cases the worker is actually familiar with the influencer and following them on Instagram.

From these preliminary results we can conclude that a human-in-the-loop solution is possible, but very arduous if not first preceded by a target recruiting/targeted training of the workers.

For this reason, we will first focus on expert recognition to clearly characterise fashion influencers. Expert recognition have been extensively studied in the data mining community [14, 15].

We can divide the methods used to identify experts into two main approaches: a graph based approach and a feature based approach. In the graph based approach, the community is represented as a graph and experts are identified with algorithms such as PageRank, HITS and their extensions [12]. In the second one, the expertise dimensions are learned with supervised machine learning methods and used to identify experts. In this work, we assess workers expertise in fashion. We see this as a first step towards using the crowd network to identify fashion influencers. The plan is to design a task where we optimize the routing between workers to eventually reach those who can reveal emerging fashion influencers.

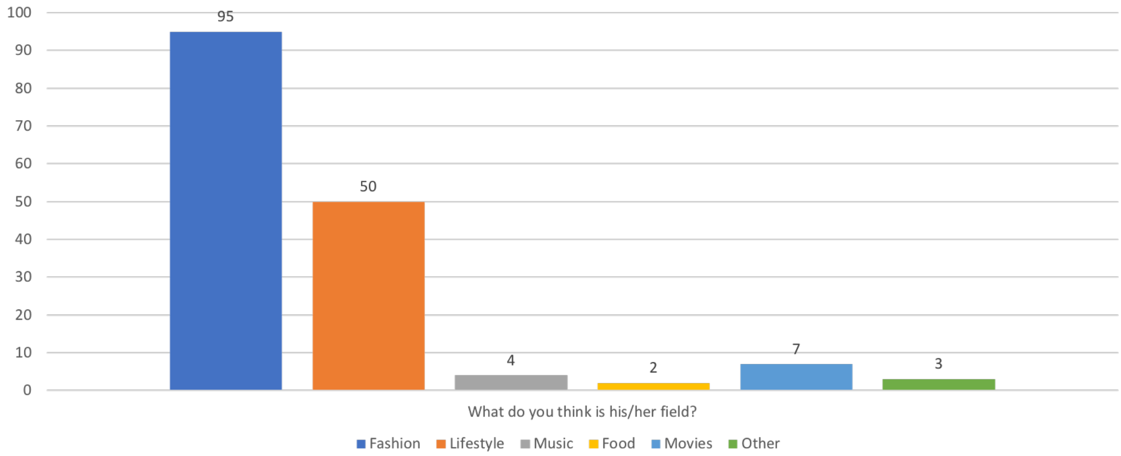
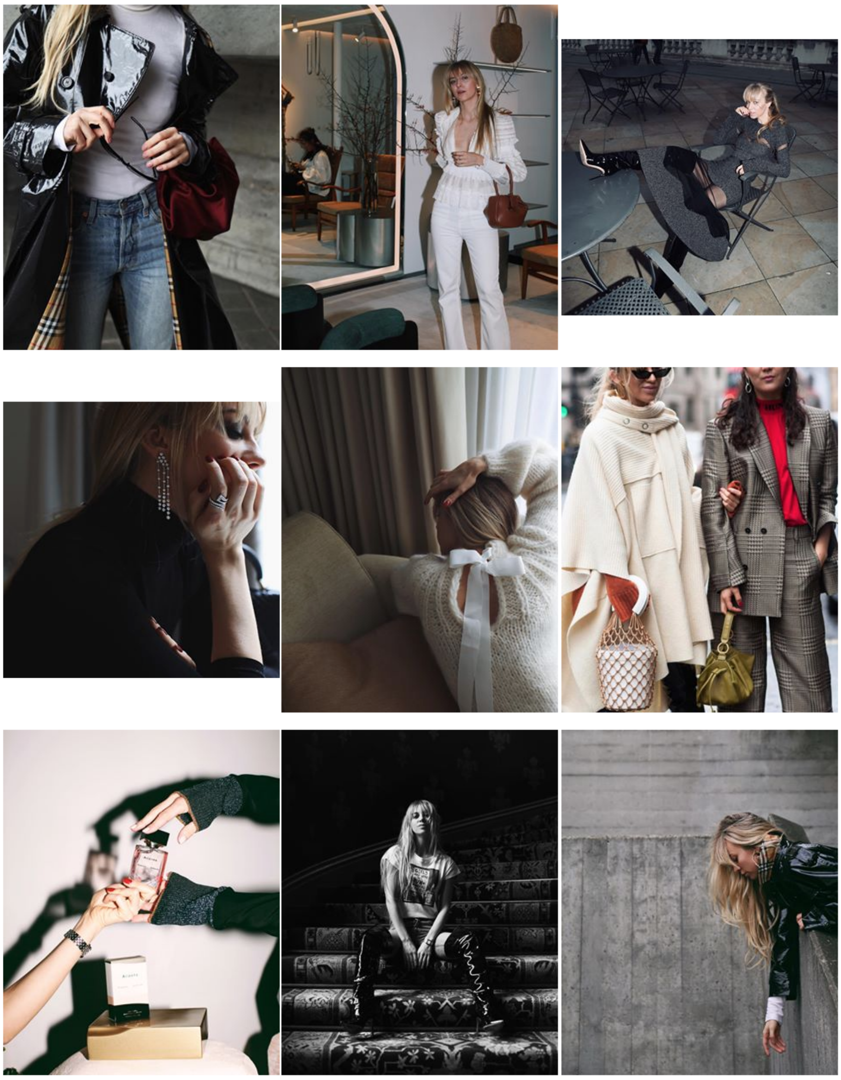


Figure 2.1: Crowdworkers' recognition of field of work of Instagram accounts.

Do you use instagram? (required)
- Yes
- No

Please have a look at the following Instagram account:

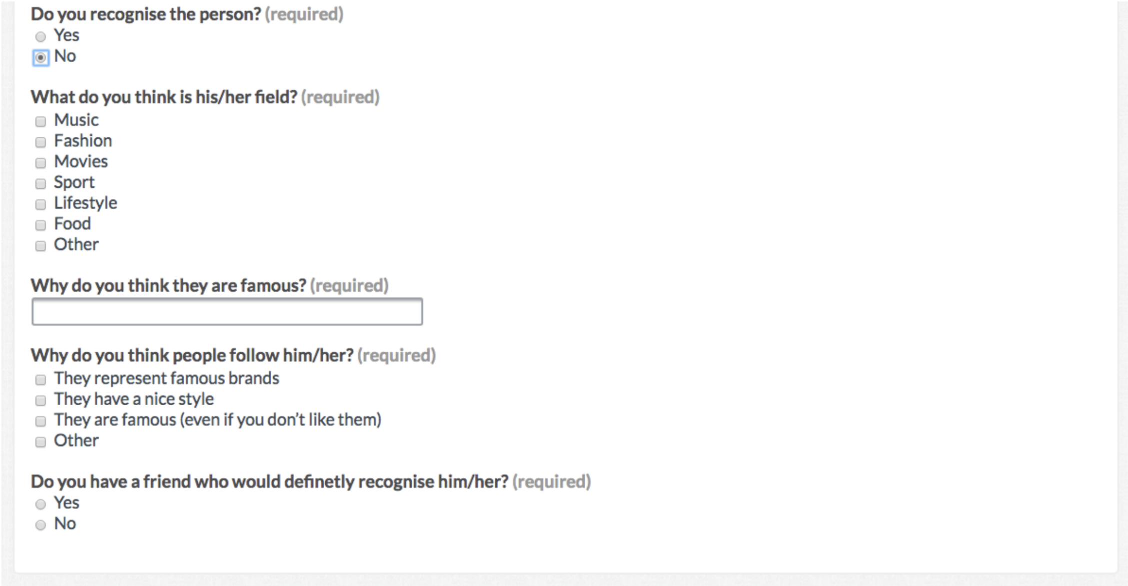


Do you recognise the person? (required)
- Yes
- No

What is his/her name (required)

What is his/her field? (required)
- Music
- Fashion
- Movies
- Sport
- Lifestyle
- Food
- Other

Figure 2.2: Crowdsourcing task for Instagram accounts (part 1).



Do you recognise the person? (required)
- Yes
- No

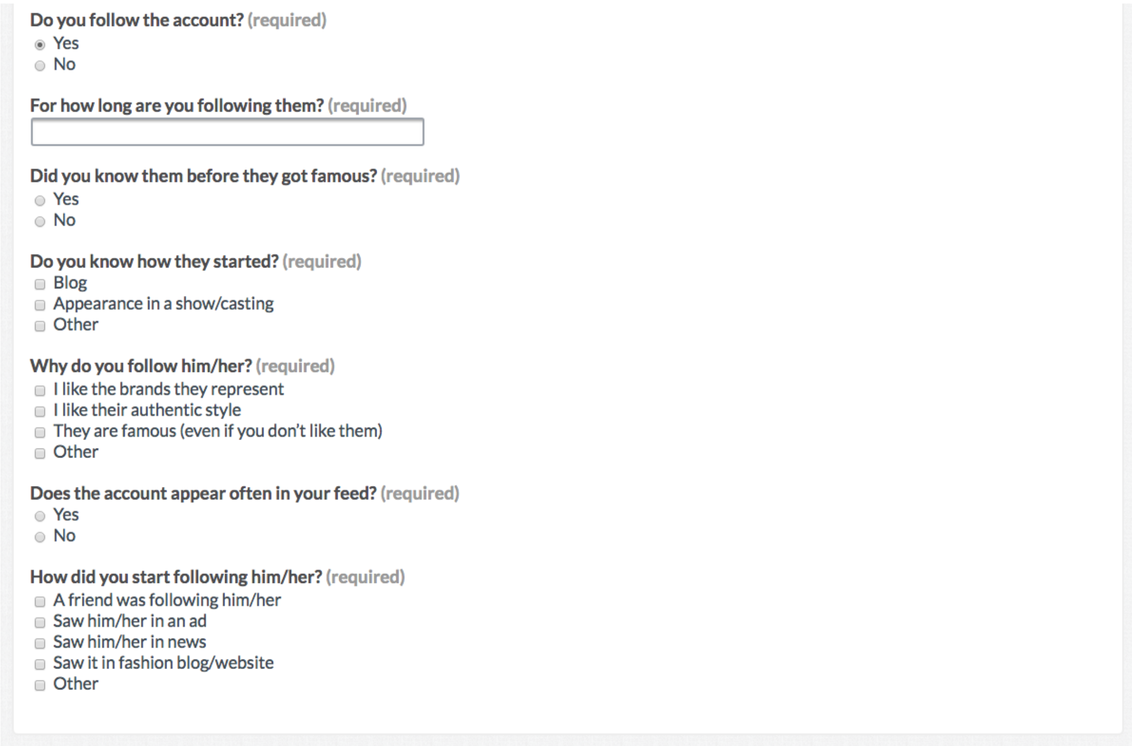
What do you think is his/her field? (required)
- Music
- Fashion
- Movies
- Sport
- Lifestyle
- Food
- Other

Why do you think they are famous? (required)

Why do you think people follow him/her? (required)
- They represent famous brands
- They have a nice style
- They are famous (even if you don't like them)
- Other

Do you have a friend who would definetly recognise him/her? (required)
- Yes
- No

Figure 2.3: Crowdsourcing task for Instagram accounts (part 2).



Do you follow the account? (required)
- Yes
- No

For how long are you following them? (required)

Did you know them before they got famous? (required)
- Yes
- No

Do you know how they started? (required)
- Blog
- Appearance in a show/casting
- Other

Why do you follow him/her? (required)
- I like the brands they represent
- I like their authentic style
- They are famous (even if you don't like them)
- Other

Does the account appear often in your feed? (required)
- Yes
- No

How did you start following him/her? (required)
- A friend was following him/her
- Saw him/her in an ad
- Saw him/her in news
- Saw it in fashion blog/website
- Other

Figure 2.4: Crowdsourcing task for Instagram accounts (part 3).

2.2 Crowdsourcing Reproducibility and Repeatability

We investigated the repeatability (over time) and reproducibility (over multiple platforms) capabilities of crowdsourcing experiments in fashion (reported in detail in [16]): from this analysis, we confirmed that crowdsourcing tasks repeated over time are reliable, but it is necessary to consider some platform biases, as explained in detail in Deliverable D3.4 Section 3.3, where we provide a tested method to mitigate such biases.

2.3 Analysis of the Influencers Market Ecosystem

In order to understand how to reliably and efficiently detect new trends in fashion, it is necessary to first examine the state-of-the-art from the industrial side. We reached out and interviewed three influencers experts working for three European companies: *Collabary*, *Influencer-Check.ch* and *Reachbird.io*, asking specific questions on how the influencers market works and what are the techniques used to detect fashion influencers.

2.3.1 Fashion Influencers Ecosystem

The fashion influencer ecosystem is made essentially of three parties: The fashion brands, the influencers and a platform connecting the two. Every party benefit from one another.

Brands and retailers have adopted fashion influencers to promote new trends and reach a wider audience. In fact, fashion influencers help brands to present their new products to consumers in an authentic way. Therefore, their impact on spreading awareness of these new products is more relevant than paid advertisement. In fact, 65% of fashion and retail brands launched campaigns with influencers over the past year and 74% of those experts found that influencer marketing was effective at driving sales in 2016 [1].

Fashion influencers need to continuously create content in their social media accounts (Instagram, Youtube, etc.) and have an engaged relation with their audience. By definition, influence is the "act or power of producing an effect without apparent exertion of force or direct exercise of command". In fact, the fashion influencers followers look for an authentic content therefore influencers need to "stay true to their style" and have the right collaborations.

This is where it comes the role of the platforms that connect brands with influencers. For example Collabary offers "a marketplace that give access to all relevant players and hence enables the brand to reach their audience in an authentic way [...] Collabary covers the campaign creation, the discovery of influencers and the management of their participation in the campaign." These platforms also play a role in creating new campaigns and then provide "extensive reporting on the campaign performance".

2.3.2 Profile of Fashion Influencers

Understanding whether a person is a fashion influencer can be rather difficult in the fashion marketing world: celebrities that have other professions, like models, actors, and athletes, can sometimes be considered influencers, because celebrities also start to boost their own social media channels and some Influencers that started off with blogging as a hobby, now are considered celebrities, e.g. Chiara Ferragni [19]. They both get either paid by a brand or genuinely like it and tell the world with either themselves or their social personality [8].

However, we can identify three key differences between influencers and other fashion actors:

- Celebrities core profession is related to an industry – singers, actors, professional sportsmen, politicians – they can of course be brand ambassadors, but their main professional activity is not being a full-time influencer. On the other hand, fashion influencers have this as profession (full time) – they focus heavily on creating and curating content for their Social Media accounts, that is in line with their persona/brand and cater to their community [19].
- Influencers are closer – one could say they have a personal relationship – to their followers, while celebrities might have a greater and worldwide following, but not as close relationship to them [19].
- There is a stronger proclivity among influencer to actively engage in the creative process, as in contrast to many celebrities they draw their credibility directly from the content. Content creation is at the heart of their business model and not just one way of monetization as for e.g. an athlete [8].

Influencers are usually tied to specific brands, and their "specialisation skills" (such as lifestyle, fashion, beauty, food, etc.), their geographic location and the ones of their followers are important factors taken into consideration when looking for a fit between an influencer and a brand for a collaboration [8, 19].

The companies we contacted consider five main characteristics an influencers should have:

Authenticity: the ability to stay true to their style/brand and community when communicating and deciding on collaborations.

Communication: the ability to engage with one's audience and the relevant influencer community (get to know other Influencers in real life and support colleagues, even planning co-creation sessions), as well as being professional (responsive) in the communication with brands during collaborations.

Dedication: the ability to manage their account as a full-time job, meaning, continuously creating and curating content (postings/video/stories) for their accounts as well as being active on social media.

Branding: the ability to treat and work on their social media account as a brand, meaning, the ability to find and keep a consistent and unique style, imagery (feed) and tone of voice.

Mission: the ability to generate value either for society in general or their community.

2.3.3 Influencer Detection – State-of-the-art

We now investigates what are the techniques used to detect fashion influencers. Usually [13], a weighted average of the following indicators is used:

Average engagement rate: the ratio of number of comments/likes to number of followers.

Comments/like ratio: the ratio of number of comments to number of likes.

Followers/followed ratio: the ratio of number of followers to the number of the following accounts.

Mentions: the number of mentions of the influencer.

Ad/No-Ad ratio: the ratio of ads to the number of posts without ads.

Follower growth rate: change on the number of followers within one month.

Sentiment of the comments: analysis of the mood in the comments.

Klout score: A numerical score from 1 to 100 that measures the size of an account's social media network and correlates the content created to measure how other users interact with that content [17].

However, often these metrics are not enough to properly detect new influencers, because of the plague of bought followers, and for the difficulty or properly predict authenticity and engagement rates. Because of that, multiple solutions are taken into consideration.

Manual screening: the use of experts that will manually screen the influencers posts, taking attention to imagery quality, feed consistency etc. Experts can decide to onboard accounts based on exceptional results even when the metrics are below the established thresholds, for example because of exceptional engagement rate/imagery or a recent fast-growing follower community [19].

The shift towards micro influencers, where screening is easier and potential manipulations are easier to spot [8].

2.4 Lessons Learned

The information gathered from our pilot analysis and our experts interviews make clear that influencer detection can be a candidate problem to be solved in a human-in-the-loop fashion: some features are easily detected automatically (number of followers or metrics related to understand activity level like number posts over time), while others require more manual annotations (like authenticity or quality of content). This is why a hybrid crowdsourcing approach using automatic metrics together will manual screening is a solution that FashionBrain will pursue. However, we discovered that a preliminary targeted recruiting phase is necessary to select the most competent crowdworkers in this niche field.

3 Influencer Detection Framework

As discussed in the previous chapter, workers in crowdsourcing platforms can identify fashion influencers. They possess in an aggregated level a broader knowledge of fashion influencers than individual experts. As an example, while it is generally difficult for an expert to come up with a long list of fashion influencers in a short period, it is much easier to obtain such a list by asking online workers. Therefore, we design a crowdsourcing task where workers are asked to name as many candidate fashion influencers as possible. By aggregating these answers, we can identify the identities (e.g., usernames on Twitter) of a large number of real social influencers efficiently and cost-effectively. In this section, we first describe the crowdsourcing task used to collect the usernames of candidate fashion influencers. Then, we present Open Crowd, a crowdsourcing aggregation framework designed to infer the real fashion influencers from a set of candidates given by workers. Finally, we evaluate Open Crowd against state of the art aggregation methods.

3.1 Task Description

We published a question-answering task on Figure Eight[1], asking workers to name fashion influencers they know on Twitter. A snapshot of the crowdsourcing task is shown in Figure 3.1. To set the context and promote workers to reflect on their experience, we asked workers to assess their domain-specific knowledge (five-point scale), estimate how often do they read social media posts from influencers (never, rarely, sometimes, always), and describe how they got to know the influencers. Workers name candidate influencers by providing their Twitter usernames. The task took 2 minutes to complete on average. Workers who completed the task were paid with an initial amount of 30 cents (USD), and with an additional bonus: they were paid ten additional cents (and up to 50 cents) for every social influencer they provided after naming three influencers. Through this task, we collected 890 candidate fashion influencers named by 250 workers. We relied on the guidelines discussed in Section 2.3.2 to label a sample from the collected candidates. Our primarily analysis revealed that 30.64% are real fashion influencers which confirms our hypothesis that crowdsourcing can help us find fashion influencers.

[1] https://www.figure-eight.com

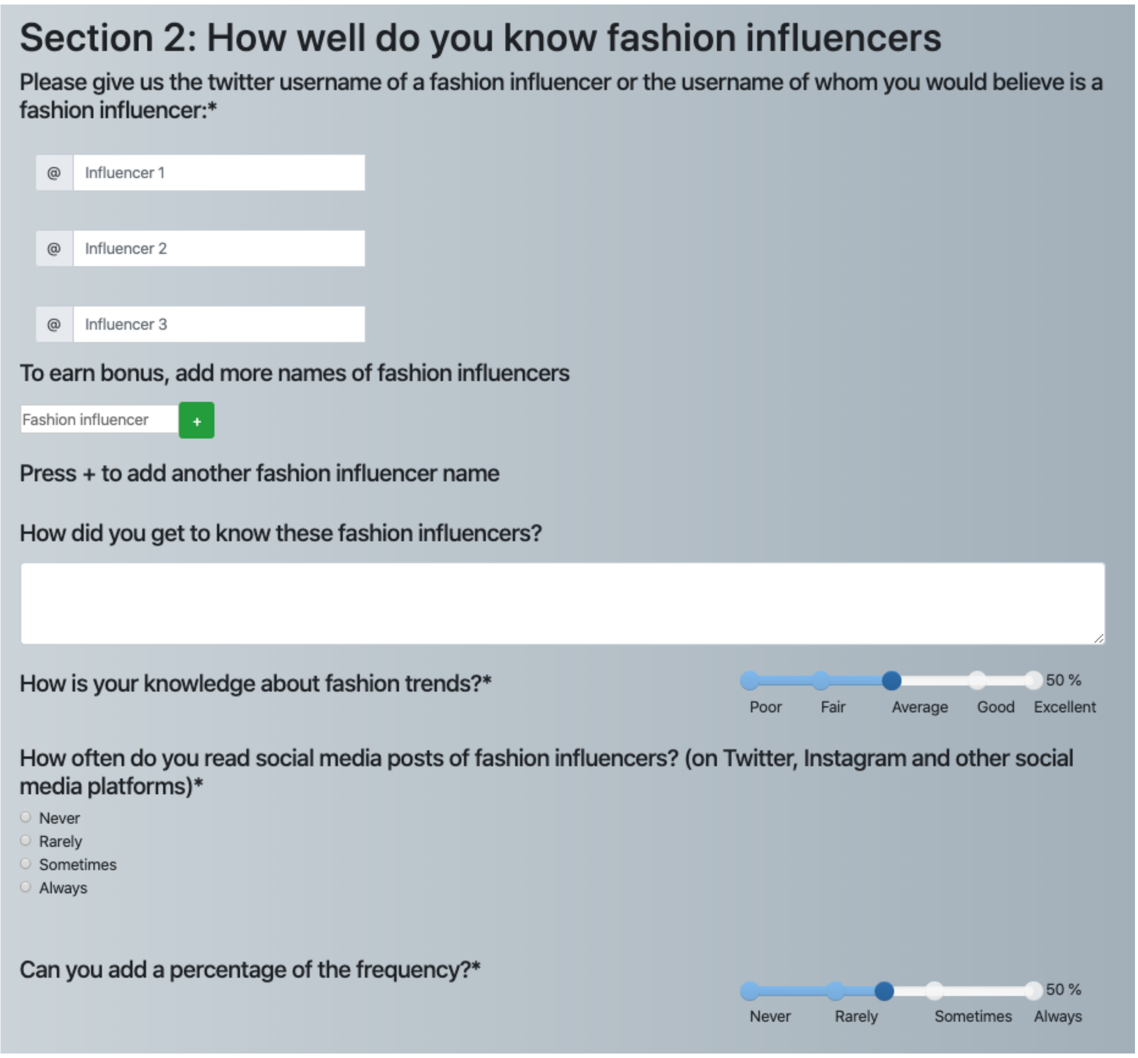


Figure 3.1: Crowdsourcing task to find social influencers.

3.2 Open Crowd: A framework to identify fashion influencers

The answers collected from our crowdsourcing task are in the form of free text. These answers are unknown in advance and we don't expect workers to provide us with the same answers. This type of task is known as an open-ended task. In the A to Z of Methodology of Cambridge[2], an open-ended task is defined as "tasks to which there is not a single absolutely correct answer or where a variety of answers are possible." This type of task is very popular in crowdsourcing, however, none of the existing aggregation methods is designed to handle this type of tasks. We propose Open Crowd, a framework for finding social influencers through open-ended answers aggregation.

[2] https://www.cambridge.org/elt/ces/methodology/openendedtasks.htm

Our framework is human-AI collaborative approach that integrates both machine learning and crowdsourcing for aggregating open-ended answers. It models the true label of a candidate influencer as dependent on both the features of the candidate and the reliability of the workers who named the candidate. To infer the truth, we leverage a small number of expert labels to bootstrap the inference process. Open Crowd then jointly learns a feature-based model for the quality of the answers and the reliability of the crowd workers. The model parameters and worker reliability are updated in an iterative manner, allowing their learning processes to benefit from each other until an agreement on answer quality is reached. We formalize such a learning process with a principled optimization algorithm based on variational expectation-maximization.

3.3 Experiments

We compared our method against state of the art aggregation methods discussed in Deliverable D3.2. These methods include 1) ZenCrowd [6], an expectation-maximization (EM) method that estimates worker reliability as a model parameter; 2) Dawid-Skene [5], an EM method that learns worker reliability as a confusion matrix; 3) GLAD [22], an EM method that simultaneously learns worker reliability and task difficulty; and 4) LFC [18], an EM method that incorporates priors in modeling worker reliability. The results are reported in Table 3.1.

Open Crowd achieves the best performance among all answers aggregation methods under comparison: it improves the state of the art by 6.94% accuracy and 62.06% AUC. This significant improvement clearly demonstrates the effectiveness of our framework in open-ended answers aggregation.

Method	Metric	Fashion				
		50%	60%	70%	80%	90%
DS	Accuracy	0.689	0.716*	0.703	0.688	0.711
	AUC	0.191	0.169	0.242+	0.244	0.263
GLAD	Accuracy	0.697	0.716*	0.724	0.700	0.688
	AUC	0.183	0.189	0.229	0.224	0.263
ZenCrowd	Accuracy	0.701	0.686	0.733*	0.702*	0.688
	AUC	0.157	0.175	0.203	0.239	0.287+
LFC	Accuracy	**0.721**	0.694	0.718	0.691	0.755*
	AUC	0.203+	0.203+	0.225	0.264+	0.277
Influencer Detection Framework	Accuracy	0.708*	**0.740**	**0.751**	**0.769**	**0.889**
	AUC	**0.304**	**0.350**	**0.350**	**0.452**	**0.495**

Table 3.1: Performance (accuracy and AUC) comparison of aggregation techniques with supervision degree from 50% to 90%. The best performance is highlighted in bold; the second best performance is marked by '*' for accuracy and by '+' for AUC.

4 Influencer Explorer

The existing theory of graph exploration is based on a rather static, connectivity-based paradigm, where the indication on how two users are connected in the social graph is given by the friendship status[1]. The identification of most influential users in the network have been investigated in many research work mainly focusing on graph structure and connectivity as the main ingredient to design effective ranking measure according to centrality measures from network science [23, 21, 2, 10].

However, modern social media are heavily dynamic, and they are dominated by ephemeral topics (e.g. trending hashtags) and by the platform recommender systems (suggestion on whom to follow, trending accounts etc.). In other words, often a user will see the content of an account they do not follow, simply because that content is relevant in that specific moment according to the user interests and browsing behaviour.

This makes traditional graph theory not ideal for our goal of understanding how new fashion influencer are discovered by users: this is especially true in the exploration phase, when a user is actively browsing to discover new fashion trends and accounts.

We will test these hypotheses and see how they affect the user behaviour in the fashion exploration phase, by designing a crowdsourcing experiment where workers actions are recorded while they explore social media feeds to identify new fashion influencers. Moreover, we will use some traditional statistical techniques to provide estimates on the size of the unknown population of (still) undiscovered emerging fashion influencers. Finally, we will analyse how these findings can be used to develop Machine Learning (ML) techniques (using Flair trained on the FashionTweets dataset of Deliverable D3.4) to automatically explore a social graph and extract candidate fashion influencers.

4.1 Data Collection

We prime crowd workers to start from a fashion influencer account (one of the 125 accounts obtained from the Open Crowd experiment described in Section 3) and explore the Twitter feed with the intent of organically discovering new influencers. We allow the workers to freely use the Twitter webpage, and collect their actions up to 5 new account discoveries (5 hops). Workers were allowed to pass through non-influencers during the exploration phase, as for example fashion hubs/magazines or other accounts that might be connected to emerging fashion influencers. The

[1] or similarly, by the not necessarily mutual following status.

rationale of this kind of "in the wild" experiment is that relying on crowd workers expertise and memory alone is not enough to properly understand the way new influencers are discovered.

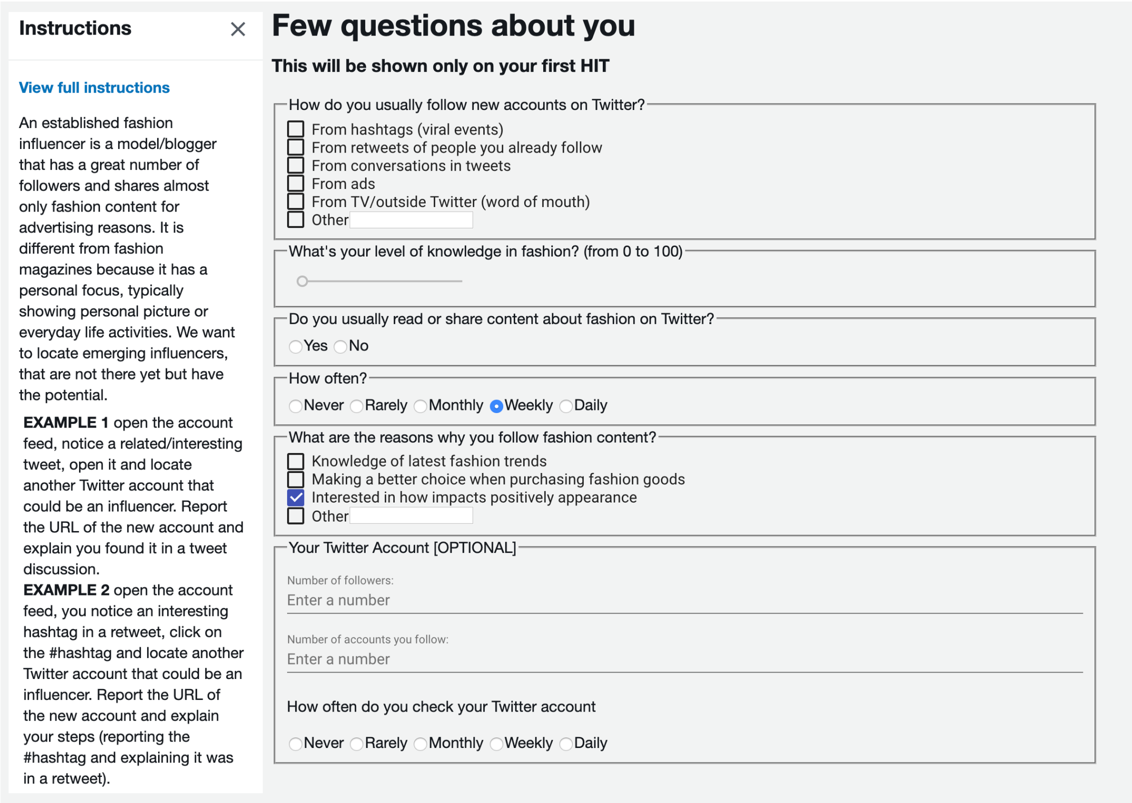


Figure 4.1: AMT task for Influencer Explorer (part 1).

In Figures 4.1-4.3 we show the instructions and reporting mechanism used in Amazon Mechanical Turk (AMT): workers were allowed to freely use Twitter starting from a specific account, and had to report their actions each time they landed on a new account. The range of actions is rather vast, from clicking on tweets and then a new account, to follow Twitter trends or hashtags.

We collected 375 exploration sessions, having 3 workers starting an exploration session for each of the 125 fashion influencers, and a maximum of 5 new accounts (hops) recorded per session. In total, we collected 653 exploration steps (hops), from 176 unique workers: workers were allowed to have multiple sessions, but without starting from the same account. They have also been encouraged to discover a new influencer at each session.

After a pilot session and a qualification session as explained in D3.4, workers received a fixed amount of \$0.6, plus a bonus of \$0.2 for each new account discovered. Moreover, at the end of the collection a time analysis has been performed to provide additional payment to workers that have been slower than the estimated median time, to guarantee a payment of UK minimum wage.

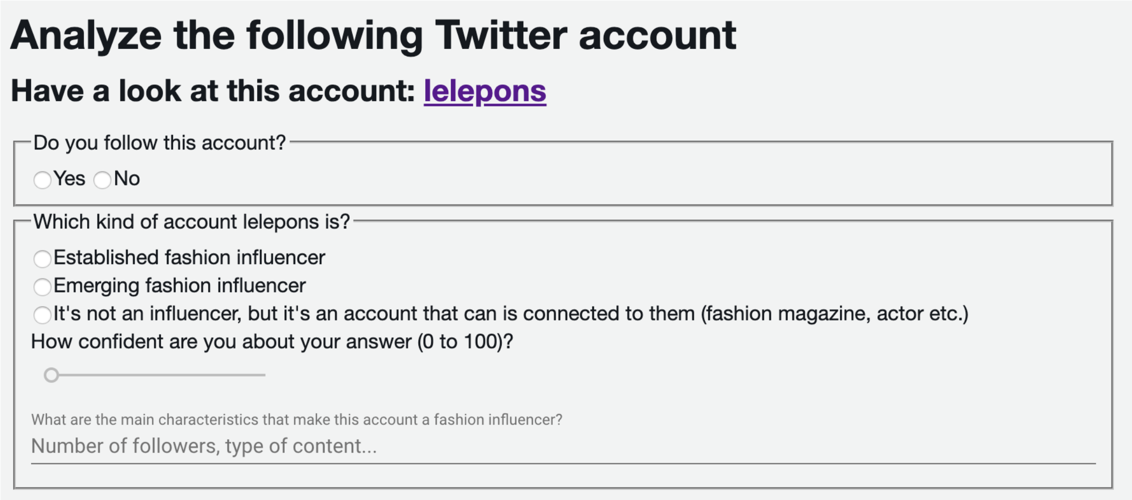
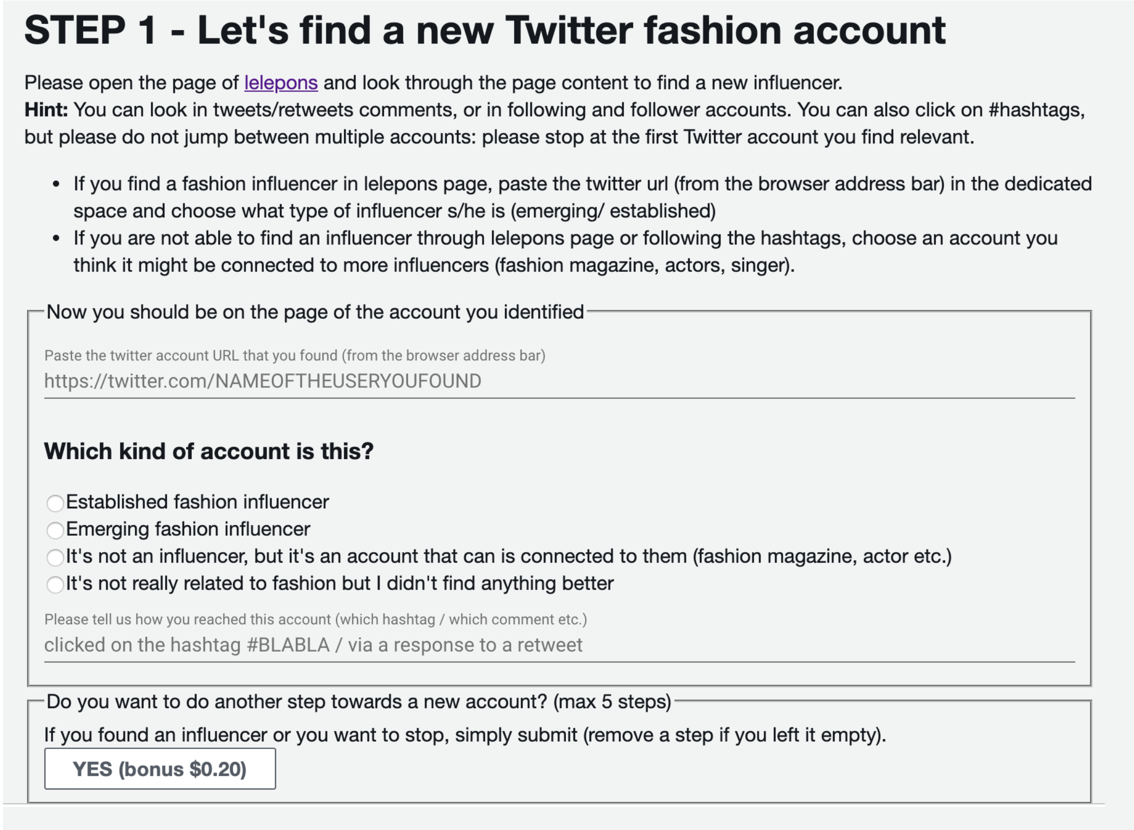


Figure 4.2: AMT task for Influencer Explorer (part 2).

This experiment also extends the Open Crowd collection: here we do not only rely on worker memory to name candidate influencers, but rather we follow them during the discovery phase. This extension also allows us to develop a statistical analysis to estimate the number of undiscovered fashion influencers, as shown in Section 4.2.



STEP 1 - Let's find a new Twitter fashion account

Please open the page of lelepons and look through the page content to find a new influencer.
Hint: You can look in tweets/retweets comments, or in following and follower accounts. You can also click on #hashtags, but please do not jump between multiple accounts: please stop at the first Twitter account you find relevant.

- If you find a fashion influencer in lelepons page, paste the twitter url (from the browser address bar) in the dedicated space and choose what type of influencer s/he is (emerging/ established)
- If you are not able to find an influencer through lelepons page or following the hashtags, choose an account you think it might be connected to more influencers (fashion magazine, actors, singer).

Now you should be on the page of the account you identified

Paste the twitter account URL that you found (from the browser address bar)
https://twitter.com/NAMEOFTHEUSERYOUFOUND

Which kind of account is this?

- Established fashion influencer
- Emerging fashion influencer
- It's not an influencer, but it's an account that can is connected to them (fashion magazine, actor etc.)
- It's not really related to fashion but I didn't find anything better

Please tell us how you reached this account (which hashtag / which comment etc.)
clicked on the hashtag #BLABLA / via a response to a retweet

Do you want to do another step towards a new account? (max 5 steps)
If you found an influencer or you want to stop, simply submit (remove a step if you left it empty).
YES (bonus $0.20)

Figure 4.3: AMT task for Influencer Explorer (part 3).

4.2 Estimating the Number of Undiscovered Influencers

While the number of undiscovered fashion influencers is unknown, it is possible to use the fashion explorer experiment to estimate this number, at least in the vicinity of the sample taken. We use a traditional mark and recapture method, commonly used in ecology to estimate an animal population's size where it is impractical to count every individual. A portion of the population is captured, marked, and released. In our case, the captured population is our 125 starting influencers in the fashion explorer experiment. Later, another portion is captured and the number of marked individuals within the sample is counted (in our cases this is the number of influencers signaled by our workers that were already in the list). Since the number of marked individuals within the second sample should be proportional to the number of marked individuals in the whole population, an estimate of the total population size $\hat{N}$ can be obtained by dividing the number of marked individuals by the proportion of marked individuals in the second sample. We use a Lincoln–Petersen estimator [20]

$$\hat{N} = \frac{Kn}{k},$$

where n is the number of influencers obtained from the Open Crowd experiment, K is the number of accounts signaled in the fashion explorer experiment, and k is the subset of those that were also in the Open Crowd experiment (recaptured accounts). This method assumes that the study population is "closed": no accounts are assumed to have disappeared between the first visit (Open Crowd experiment) and the second.

In order to estimate the confidence interval of this estimator, we perform bootstrapping: starting from the empirical distribution function of the observed data, we construct 10,000 resamples with replacement of the observed dataset (and of equal size to the observed dataset) and evaluate the confidence interval [7].

From this analysis we obtain an estimated number of undiscovered fashion influencers of 300.03, with a bootstrapped confidence interval equal to $(277.5, 320.2)$.

4.3 Analysis of Fashion Exploration Behaviour

After analysing the collected data, we observed the following salient behavioural characteristics of Twitter users for fashion influencer detection.

4.3.1 Motivation

In Figure 4.4, we show the main motivation of following influencers from our sampled population. Often the users chose indicated multiple options, with almost all of them indicating both the need of discovering new trends and of improving their own appearance. Regarding the "other" field, the majority of them indicated as main motivation for following a fashion influencer curiosity and the desire to get new ideas.

4.3.2 Discovery Methods

We report here the main ways new accounts have been discovered. We notice that the following figures may not sum to 100%, as some workers have performed multiple operations, or operations that were not reported correctly. We indicate as *current account* the account they are currently browsing.

Handles/Conversations

The relative majority of sessions (33.1%) reached the next hop through conversational elements: handles/mentions to other accounts (using "@") from the current account, replies and likes to other tweets. This can suggest that users tend to be engaged when there are meaningful interactions between the influencers and other accounts.

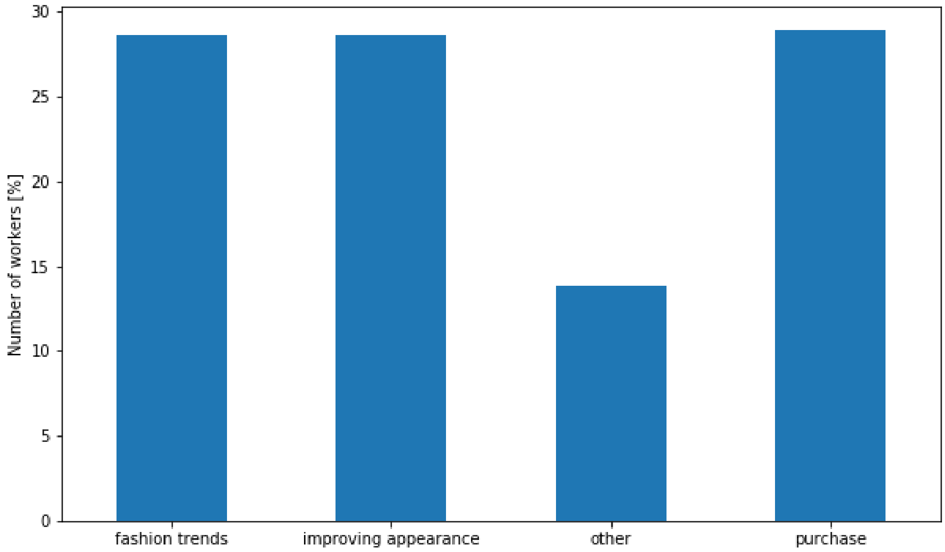


Figure 4.4: Fashion influencer following motivation on Twitter.

Hashtags

Clicking on a hashtag (#) accounted for 23.7% of the new discovery. These hashtags may have appeared on the main feeds, but also on the trending tweets section. The vast majority of them, however, were hashtags used by the starting influencer. This is important, because clicking on a hashtag can create a "jump" in the social graph to a completely different section of the graph, making it hard to predict using traditional connectivity analysis.

Retweets

For 16.1% of the workers, a retweet of the current account is what made them engaged to follow another account.

Followers

Only 16.8% of workers used the traditional connectivity, i.e. looked at new influencers from the list of followers of the current account.

Recommended by Twitter

A small but significant 7.8% of workers followed Twitter recommender engine to find new accounts.

Search

While they were discouraged to do so, a small percentage (1.5%) of workers decided to use the search engine (typically with queries like "fashion influencer" or searching topics related to the feed they were browsing).

Fashion Related Accounts

During the exploration, about 30% of the sessions went through a (worker-reported) "hub"[2], i. e., an account that is not a fashion influencer, but it is connected to many of them, like bloggers, fashion magazines etc.

4.3.3 Discussion

From these numbers, we can conclude that a traditional connectivity analysis of the social graph is not appropriate in modern social media. A more sophisticated approach is suggested, where a crawler of the social media feed should perform a random walk, weighted with the proportions shown in this section, to collect a set of candidate fashion influencers.

Such a crawler would need a hybrid human-machine system to filter out the parts of the graph that are less likely to contain fashion influencers: In Section 4.5, we will analyse how efficiently such hybrid techniques can flag candidate fashion influencers.

4.4 Exploration Graph Analysis

The collected data forms a graph, where two accounts are connected if a worker moved from one account to an other during the exploration phase, as shown in Figure 4.5.

We performed some traditional centrality measures to understand the way workers interacted with these accounts. It is important to notice that these measures are just a partial snapshot of the whole connectivity graph, that in any case is dynamic in nature. We focused on the following local and global centrality measures:

- Indegree: the number of followers.
- Outdegree: the number of followees.

[2] Not to be confused with the centrality measure of standard graph theory.

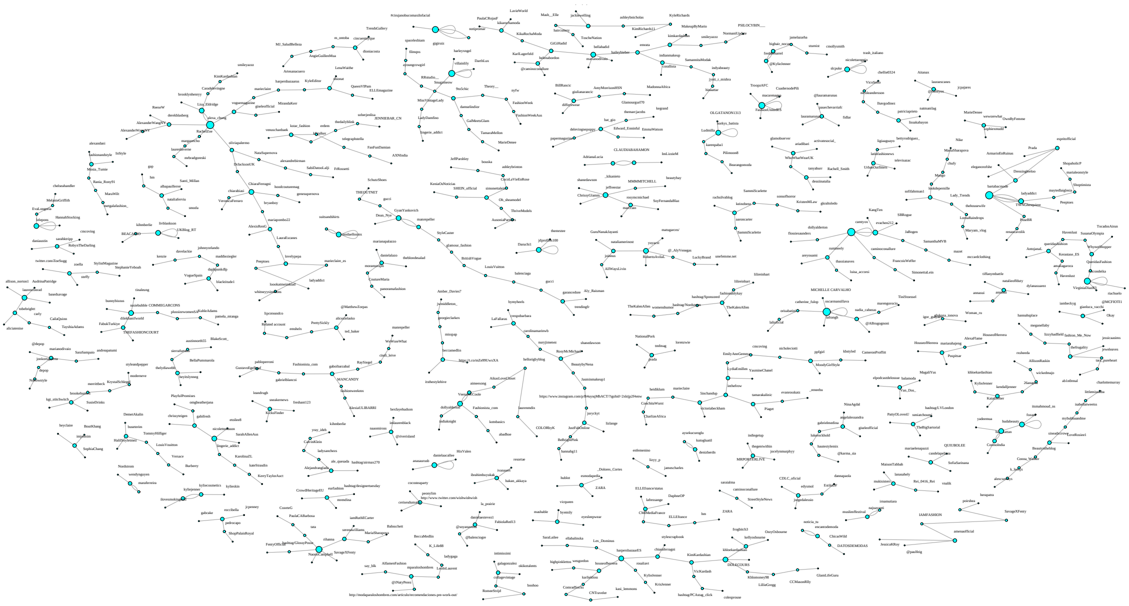


Figure 4.5: Twitter fashion influencers graph obtained from workers navigation with 717 fashion candidates and 650 links, mean degree 1.8, mean closenness 1.97e-06, mean betweenness 1.43.

- Alpha centrality (a variant of Katz centrality [3]): a generalization of eigenvector centrality. It is a combination of the adjacency matrix and relative importance of the endogenous versus exogenous factors.
- Betweenness [9] is the number of shortest paths that fall within a fashion candidate influencer.
- Closenness [9]: counts how many steps are needed to reach every other fashion candidate from a given fashion influencer.
- Hub [11]: is a global centrality measure that can be calculated using the eigenvector of the adjacency matrix. A hub is a property of scale-free network and referring to the node with the highest number of outdegree that exceeds the average (high outdegree).
- An authority value is computed as the sum of the scaled hub values that point to that account.
- PageRank [4]: relies on the fact that more important fashion candidates are more likely to be linked to other fashion candidates. It mainly combines with number and quality of links that receive a fashion candidate influencer.

4.4.1 Discussion

The results suggest the following observations:

- The indegree are positively correlated with fashion magazines/bloggers.

- Outdegree, closeness and hub measures are positively correlated with established fashion influencers.
- Authority, alpha-centrality, and PageRank find a mix of fashion candidate influencers including established fashion influencers, fashion bloggers, retailers, and actors/singers.

These results corroborate the intuition that a crawler should multiply its efforts around fashion magazines/bloggers (because they can be connected to many fashion influencers).

4.5 Human-in-the-loop Fashion Influencer Discovery

We now investigate how well a ML algorithm, and then a hybrid human-in-the-loop algorithm, can flag candidate fashion influencers. To do this we train a Random Forest Classifier (RFC) using information extracted using the Twitter API[3]: number of followers, followees, number of posted tweets, profile description, location, creation time of Twitter candidate fashion influencer, and account type which refers if the account is verified or not.

Given a candidate account, we will evaluate how well the classifier can distinguish fashion influencers from other accounts.

To test the effect of augmenting this algorithm with the crowd, we use the information we collected during the exploration phase, as described in Section 4.1. In Table 4.1 we describe the features, distinguishing which ones are obtained programmatically, and which ones with AMT.

The dataset is split into training and test set, with training set size consisting of 726 observations and test set size of 312 observations, where categorical variables like location, verified, worker ID, and time of posting the tweet are encoded using label encoding. TF-IDF is used for test and training set separately on textual features: the TF-IDF encoding permits to evaluate the importance of each term in the corpus compared to the other terms. For each observation, a sum of terms importance is carried out to determine the weight for each set of terms that characterise a fashion candidate influencer account.

Alternatively, we also investigate a more advanced textual encoding approach, where for textual information we use Flair fashion content detector (majority consolidation model, trained on the FashionTweets dataset) built in Deliverable D3.4 instead of using TF-IDF.

We repeat the evaluation with and without human-generated features, to understand the impact a human-in-the-loop approach can have in the classification.

In Table 4.2 we present the performance of the different approaches. Adding AMT feature increases the performance of the model. Moreover, using Flair dedicated

[3] http://docs.tweepy.org/en/latest/

features	features source
Worker ID	AMT
Worker confidence about fashion candidate influencer	AMT
workers expertise in fashion	AMT
reasons of worker to follow fashion accounts	AMT
worker belief of fashion candidate influencers membership	AMT
Follower Count	Tweepy
Friend Count	Tweepy
Location	Tweepy
Verified	Tweepy
created_at	Tweepy
how workers in AMT follow others	AMT
frequency of checking content by workers	AMT
frequency of worker checking own account	AMT
the way workers follow new account	AMT
fashion candidate influencer main characteristics	AMT
fashion candidate influencer profile description	Tweepy
worker steps to reach new fashion candidate account	AMT
tweets count	Tweepy

Table 4.1: Features description used throughout both experiments.

fashion content classifier on the textual features instead of TF-IDF improved significantly the classification performance.

Text Encoding		RFC	RFC+AMT
TF-IDF	Accuracy	0.79	0.78
	Precision	0.77	0.74
	Recall	0.79	0.78
	F1-score	0.74	0.73
Flair	Accuracy	0.96	**0.99**
	Precision	0.96	**0.99**
	Recall	0.96	**0.99**
	F1-score	0.95	**0.99**

Table 4.2: Performance (macro averages) of the different approaches, for the Random Forest Classifier with and without human-in-the-loop features.

5 Conclusions

The information gathered from our pilot analysis and our experts interviews make clear that influencer detection can be a candidate problem to be solved in a human-in-the-loop fashion: some features are easily detected automatically (number of followers or metrics related to understand activity level like number posts over time), while others require more manual annotations (like authenticity or quality of content).

The results of the crowdsourcing task used to collect candidate fashion influencers has been aggregated using Open Crowd, a crowdsourcing aggregation framework designed to infer the real fashion influencers from a set of candidates given by workers, and evaluated against state of the art aggregation methods. Open Crowd achieved the best performance among all answers aggregation methods under comparison.

We then designed a crowdsourcing experiment where workers actions are recorded while they explore social media feeds to identify new fashion influencers and estimate the number of undiscovered ones. The corresponding dynamical social graph has been studied: we concluded that a traditional connectivity analysis of the social graph is not appropriate in modern social media. We thus developed an hybrid ML algorithm to allow automatic exploration and detection of fashion influencers: using a human-in-the-loop approach together with a Flair model trained over fashion tweets has achieved excellent performance.

Acknoledgments

We thank Collabary (https://www.collabary.com/), Influencer Check (www.influencer-check.ch), and Reachbird (www.reachbird.io) for their valuable insights.

Bibliography

[1] The state of influencer marketing 2017 report. URL https://www.launchmetrics.com/resources/whitepapers/the-state-of-influencer-marketing-2017.

[2] Zeynep Zengin Alp and Şule Gündüz Öğüdücü. Influence factorization for identifying authorities in twitter. *Knowledge-Based Systems*, 163:944–954, 2019.

[3] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201, 2001.

[4] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[5] P. Dawid, A. M. Skene, A. P. Dawidt, and A. M. Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied Statistics*, pages 20–28, 1979.

[6] Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. Zencrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 469–478, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1229-5. doi: 10.1145/2187836.2187900. URL http://doi.acm.org/10.1145/2187836.2187900.

[7] Thomas J DiCiccio and Bradley Efron. Bootstrap confidence intervals. *Statistical science*, pages 189–212, 1996.

[8] Martin Faltl and Cyrill Hauser. personal communication.

[9] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.

[10] Ahmed Ibnoulouafi, Mohamed El Haziti, and Hocine Cherifi. M-centrality: identifying key nodes based on global position and local degree variation. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(7):073407, 2018.

[11] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[12] Amy N Langville and Carl D Meyer. *Google's PageRank and beyond: The science of search engine rankings*. Princeton University Press, 2011.

[13] Emily Linnér, Sitav Taha, and Johan Carlsson. What characterizes an influential instagram fashion influencer?: A descriptive research, 2018.

[14] Aditya Pal, Rosta Farzan, Joseph A Konstan, and Robert E Kraut. Early detection of potential experts in question answering communities. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 231–242. Springer, 2011.

[15] Aditya Pal, Shuo Chang, and Joseph A Konstan. Evolution of experts in question answering communities. In *sixth international AAAI conference on weblogs and social media*, 2012.

[16] Rehab Qarout, Alessandro Checco, Gianluca Demartini, and Kalina Bontcheva. Platform-related factors in repeatability and reproducibility of crowdsourcing tasks. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 135–143, 2019.

[17] Adithya Rao, Nemanja Spasojevic, Zhisheng Li, and Trevor Dsouza. Klout score: Measuring influence across multiple social networks. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2282–2289. IEEE, 2015.

[18] Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo Hermosillo Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from crowds. *Journal of Machine Learning Research*, 11(Apr):1297–1322, 2010.

[19] Viktoria Roussina. personal communication.

[20] George Arthur Frederick Seber et al. *The estimation of animal abundance and related parameters*, volume 8. Blackburn press Caldwell, New Jersey, 1982.

[21] Seyed Mohammad Taheri, Hamidreza Mahyar, Mohammad Firouzi, Elahe Ghalebi, Radu Grosu, and Ali Movaghar. Hellrank: a hellinger-based centrality measure for bipartite social networks. *Social Network Analysis and Mining*, 7 (1):22, 2017.

[22] Jacob Whitehill, Ting fan Wu, Jacob Bergsma, Javier R. Movellan, and Paul L. Ruvolo. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2035–2043, 2009.

[23] Xiaohui Zhao, Fang'ai Liu, Jinlong Wang, Tianlai Li, et al. Evaluating influential nodes in social networks by local centrality with a coefficient. *ISPRS International Journal of Geo-Information*, 6(2):35, 2017.