FITNet: Identifying Fashion Influencers on Twitter

JINDA HAN, University of Illinois at Urbana-Champaign, USA
QINGLIN CHEN, University of Illinois at Urbana-Champaign, USA
XILUN JIN, University of Illinois at Urbana-Champaign, USA
WEIKAI XU, University of Illinois at Urbana-Champaign, USA
WANXIAN YANG, University of Illinois at Urbana-Champaign, USA
SUHANSANU KUMAR, University of Illinois at Urbana-Champaign, USA
LI ZHAO, University of Missouri, USA
HARI SUNDARAM, University of Illinois at Urbana-Champaign, USA
RANJITHA KUMAR, University of Illinois at Urbana-Champaign, USA

The rise of social media has changed the nature of the fashion industry. Influence is no longer concentrated in the hands of an elite few: social networks have distributed power across a broader set of tastemakers. To understand this new landscape of influence, we created FITNet — a network of the top 10k influencers of the larger Twitter fashion graph. To construct FITNet, we trained a content-based classifier to identify fashion-relevant Twitter accounts. Leveraging this classifier, we estimated the size of Twitter's fashion subgraph, snowball sampled more than 300k fashion-related accounts based on *following* relationships, and identified the top 10k influencers in the resulting subgraph. We use FITNet to perform a large-scale analysis of fashion influencers, and demonstrate how the network facilitates *discovery*, surfacing influencers relevant to specific fashion topics that may be of interest to brands, retailers, and media companies.

CCS Concepts: • **Human-centered computing → Social network analysis**; • **Information systems → Social networking sites**.

Additional Key Words and Phrases: fashion; influencers; Twitter

ACM Reference Format:
Jinda Han, Qinglin Chen, Xilun Jin, Weikai Xu, Wanxian Yang, Suhansanu Kumar, Li Zhao, Hari Sundaram, and Ranjitha Kumar. 2021. FITNet: Identifying Fashion Influencers on Twitter. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 153 (April 2021), 20 pages. https://doi.org/10.1145/3449227

1 INTRODUCTION

The rise of social media has transformed the creation and diffusion of fashion influence [26]. New social media networks distribute fashion authority across a broad set of tastemakers known as *influencers*: entities with large followings who can inform and guide the buying decisions of their

Authors' addresses: Jinda Han, University of Illinois at Urbana-Champaign, Champaign, IL, USA, jhan51@illinois.edu; Qinglin Chen, University of Illinois at Urbana-Champaign, Champaign, IL, USA, qchen50@illinois.edu; Xilun Jin, University of Illinois at Urbana-Champaign, Champaign, IL, USA, xjin12@illinois.edu; Weikai Xu, University of Illinois at Urbana-Champaign, Champaign, IL, USA, weikaix2@illinois.edu; Wanxian Yang, University of Illinois at Urbana-Champaign, Champaign, IL, USA, wyang52@illinois.edu; Suhansanu Kumar, University of Illinois at Urbana-Champaign, Champaign, IL, USA, skumar56@illinois.edu; Li Zhao, University of Missouri, Columbia, MO, USA, zhaol1@illinois.edu; Hari Sundaram, University of Illinois at Urbana-Champaign, Champaign, IL, USA, hs1@illinois.edu; Ranjitha Kumar, University of Illinois at Urbana-Champaign, Champaign, IL, USA, ranjitha@illinois.edu.

2573-0142/2021/4-ART153 $15.00
https://doi.org/10.1145/3449227

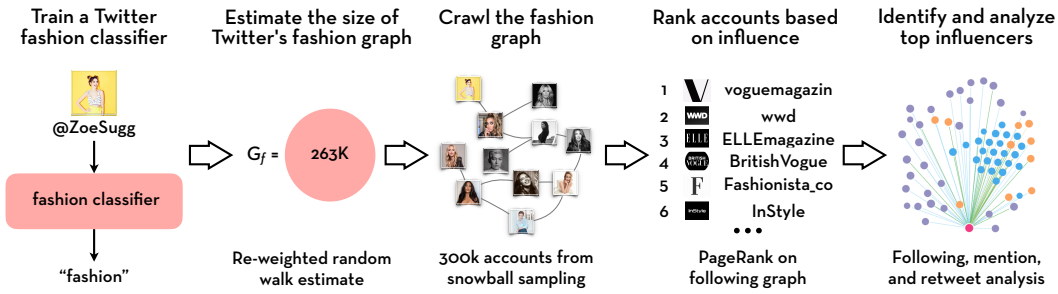

Fig. 1. To construct FITNet, we trained a classifier to predict a Twitter account's fashion relevance; leveraged the classifier to estimate the size of Twitter's fashion subgraph and snowball sample more than 300k fashion-related accounts; and ran PageRank over the resulting subgraph, identifying the top 10k influencers and capturing the interactions between them.

audience. Fashion brands, retailers, and media conglomerates in turn strive to engage with these influencers, who may hold significant sway over a company's target market [61].

To better understand the relationship between fashion companies, influencers, and social networks, this paper presents FITNet[1], a network of the 10k most influential Twitter accounts within the larger Twitter fashion graph. FITNet is the first large-scale analysis of fashion influencers on social media, exposing the fashion categories (e.g., *individual*, *brand*, *retailer*, *media*) of its accounts, the *following* relationships between them, as well as their *retweet* and *mention* interactions between January 1st, 2018, and February 1st, 2019.

To construct FITNet, our research team manually identified 11.5k fashion-related Twitter accounts, and used this dataset to train a classifier with 92% accuracy for predicting fashion relevance based on Tweet content and account metadata (Figure 1). Leveraging this classifier and a re-weighted random-walk sampling algorithm [31], we estimated that Twitter's fashion-related subgraph comprises 260k accounts. To approximate this subgraph, we snowball-sampled the follower-following graph of our training set until our classifier identified 300k fashion accounts. Running PageRank [52] over this subgraph yields an influence ranking over accounts, and we define FITNet to be the 10k highest-ranked of these.

Once identified, we demonstrate how FITNet can be used to answer questions about fashion influencers, including how influential they are, who they influence, who they are influenced by, and how influence is geographically distributed. Our analysis reveals that fashion influencers are bloggers (54%), editors (14%), stylists (12%), designers (11%), and celebrities (7%), mostly concentrated in fashion hubs like *New York* and *London*, but also distributed across the globe. FITNet highlights the *homophily* of fashion influencers: celebrities most often mention other celebrities, and bloggers predominantly mention other bloggers.

Lastly, we show how FITNet can facilitate *discovery*. A fashion company wishing to identify new influencers who resonate with their customers' interests need only identify a single competitor or target influencer to start, and then leverage the homophily exhibited by the *follow*, *mention*, and *retweet* graphs to discover other on-brand influencers in the network.

[1]FITNet is available for download at http://fashioninfluence.net/fitnet.

2 BACKGROUND AND RELATED WORK

This work is closely related to three types of social network analysis. Researchers have previously analyzed influencers and their behaviors on different social media platforms [16, 74]; developed methods for identifying domain-specific Twitter subgraphs [21, 63, 65]; and examined fashion on social media [29, 43, 44].

2.1 Social Network Influencers

Influencers in a social network are individuals who have significant reach over other members of the community. The *influence* of an individual, or node, in the network, is often computed as the node's total number of incoming edges, or *indegree*, divided by the total number of nodes in the graph [67]. Indegree does not always accurately capture influence, and algorithms such as PageRank [52] often perform better. For example, according to PageRank, an account that is followed by many influencers in the network has a higher rank than an account that has many followers from the ordinary population. PageRank has been extended in many areas and applied to different contexts [69, 73]. For example, TwitterRank extends PageRank by accounting for both link structure and topical similarity between accounts [69]. Given that all the accounts in our graph are fashion-related, we do not need to use TwitterRank [69]; instead, we use the basic formulation of PageRank to evaluate the relative influence of fashion accounts in our network.

An *influencer* on social media refers to a specific persona: an individual who is generally followed by a large number of other accounts and can impact the buying decisions of their followers. Therefore, social media influencers are often paid to post content that promotes third-party products and services. Researchers have studied influencers on social media platforms such as Twitter [16, 32, 36, 51, 68], Instagram [9], Facebook [31], Pinterest [30], Tumblr [19], and GitHub [23].

Yang et al. [74] identified 18.5k influencers on Instagram and studied how they *mention* brands in their Instagram posts. Instagram does not support a public API for mining *following* relationships, or allow users to *repost* other users' content. Therefore, Yang et al.'s influencer analysis is limited to studying *mention* interactions. Social influence is often closely tied to the concept of *homophily*. Chang et al. [18] explored how homophily influences *following* and *repinning* interactions on Pinterest. This paper analyzes three types of relationships among fashion influencers on Twitter: who they *follow*, *mention*, and *retweet*.

2.2 Domain-Specific Twitter Subgraphs

Creating domain-specific Twitter subgraphs and tweet repositories is a well-studied problem in the literature. This type of work is challenging because it often requires a large number of annotated inputs and domain knowledge. Some work has focused on building efficient data Extract-Transform-Load (ETL) architectures that leverage Twitter's streaming API to create tweet repositories focused on specific topics or events [10, 63]. These tweet datasets are often created to support future machine learning research in a domain. For example, Conforti et al. mined tweets related to mergers and acquisitions operations to create a dataset for studying stance detection [21].

2.3 Fashion and Social Media

Social media is an important petri dish for studying fashion. Therefore, many researchers have created and published large fashion datasets from mining social media platforms. Fashion10000 comprises 32k fashion images and their corresponding social metadata mined from Flickr [41, 42]. The Fashionpedia dataset contains 48,825 images of people and their accompanying human-annotated apparel segmentation masks; the images are harvested from Flickr and other free-license photo websites [35]. Similarly, NetiLook contains 355,205 photographs and their associated comments

mined from Lookbook.nu, a social fashion platform where users post photos of themselves wearing their favorite outfits [40]. Fashionista contains 158,235 photographs and their corresponding social metadata mined from Chictopia.com, another social fashion platform similar to Lookbook.nu [72]. StreetStyle is a dataset of 12k photos from Instagram annotated with clothing attributes [48]. GeoStyle [45] builds on StreetStyle, and combines it with photos from the Flickr 100M dataset [59]. Given the visual nature of fashion, there are not many fashion repositories containing only text [17].

These datasets have been leveraged to perform large-scale analyses about fashion. For example, using vision-based techniques and StreetStyle, Matzen et al. produced a large-scale analysis of fashion trends based on geographic location [48]. Similarly, Al-Halah et al. developed a vision-based technique in conjunction with the GeoStyle dataset [45] to understand how fashion styles are influenced across the world [11].

In addition to using fashion-related data from social media to run large-scale analyses about fashion, researchers frequently leverage this type of data to power fashion applications [11, 11, 29, 39, 43, 44, 75]. Images are often mined from fashion-related accounts and used to train vision-based systems for recommending apparel [75], forecasting clothing trends [29, 45], and even predicting the success of fashion models [53]. Similarly, other systems use textual metadata to automatically extract fashion knowledge from social media platforms such as Instagram [43, 44]. By enabling new types of data-driven fashion interactions, these large-scale repositories are reshaping the fashion industry itself [76].

While many large-scale analyses have been done on many aspects of fashion, prior works on studying fashion influencers have only been completed on a small-scale. Hund determined that the visual aesthetic of three top Instagram fashion influencers — as ranked by industry publications [34] — is often reserved and conservative in many ways [33]. Similarly, Çukul et al. studied ten well-known fashion brands and their attitudes on Instagram to discover how brands communicate to consumers [22]. Manikonda et al. focused on 20 top fashion brands to analyze how they target their customers on both Instagram and Twitter [46, 47]. Farinosi et al. studied four young fashion bloggers and 20 older fashion influencers (over the age of 70) on Instagram to show that older influencers still have a considerable impact on the industry [27]. Finally, Abidin [9] studied three Instagram influencers and their 12 followers in Singapore through mentions and hashtags such as #OOTDs (Outfit Of The Day). In contrast to these small-scale studies, this paper presents a large-scale analysis of more than 4k fashion influencer accounts — the first of its kind.

This paper extends our initial work on identifying fashion influencers on Twitter, which only presented a content-based classifier for identifying fashion-related Twitter accounts [38]. This paper leverages the trained classifier — described in the next section — to first estimate the size of Twitter's fashion graph, and then to perform a snowball sampling to discover over 300k distinct, fashion-related accounts.

3 THE FASHION CLASSIFIER

To mine Twitter's fashion subgraph, we first need to identify Twitter accounts related to fashion: we train a classifier that determines whether a Twitter account is fashion-relevant based on the account profile's metadata (e.g., profile description, follower count, following count) and the content of its recent tweets. For the purposes of training this classifier, we define *fashion* as industry related to how people dress and style themselves; topics that are related to the industry include the goods it produces — clothing, accessories, beauty products, — and how those goods are manufactured, marketed, bought, and sold.

3.1 Training Dataset

We seeded the classifier's training dataset with four main types of Twitter fashion accounts: individuals (e.g., models and bloggers), brands, retailers, and media. We mined fashion individuals from online lists of top fashion influencers [1, 3–5, 28, 37, 49, 56, 57] and Wikipedia [70, 71]; brands and retailers from fashion ecommerce and aggregator websites (e.g., *Farfetch.com*, *Net-a-Porter.com*, and *Lyst.com*); and media companies from Amazon's bestsellers in women's and men's fashion magazines [7, 8] and highly trafficked fashion websites [6, 60].

We had to manually map the mined entities to Twitter accounts. From this diverse set of sources, we identified 11,546 distinct English fashion Twitter accounts. We then randomly sampled roughly the same number of Twitter accounts to serve as the set of negative training examples, yielding an overall training dataset with 23k Twitter accounts.

3.2 Feature Selection and Preprocessing

For each account in the training dataset, we extracted its metadata — such as username, user id, profile description, follower count, following count, geo-location, account creation time, verification status — and computed content-based features on its 200 most recent tweets. To preprocess the users' profile description and users' tweets, we lowercased all words, and leveraged Natural Language Toolkit (NLTK) [24] to remove stop words, punctuation, and non-English Twitter accounts. Furthermore, we used scikit-learn's TfidfVectorizer library [54] to tokenize all tweets and to compute term frequency-inverse document frequency (TF-IDF) feature vectors for each account.

3.3 Classification and Results

We trained multiple classifiers to predict whether a Twitter account is fashion-relevant. During the training process, we compared different classification models: logistic regression, deep neural networks (DNNs), naive Bayes, random forests, and support vector machines (SVMs). We used a 70/30 train/test split, and computed true positive and true negative rates to determine sensitivity and specificity, respectively.

To train the classifiers, we used `scikit-learn` libraries [54]. We trained a logistic regression model using the `liblinear` solver with L2 regularization and balanced weights. We trained a DNN using a multi-layer perceptron with three hidden layers of sizes 1024, 256, and 32, and a 10% validation set. The network converged in 30 epochs with a validation loss of 0.004895. For naive Bayes, we used Laplace smoothing for regularization in rare cases. The random forest classifier was trained with 1k trees using eXtreme Gradient Boosting (XGBoost) [20]. Finally, we trained an SVM with an RBF kernel ($\gamma = 0.05$) and regularization constant $C = 4$ to control the degree-of-freedom of the decision boundary. We used 10-fold cross-validation to perform a hyperparameter grid search.

	True Positive	True Negative
Logistics Regression	0.92	0.81
Random Forest	0.92	0.79
Support Vector Machine	0.92	0.81
Naive Bayes	0.89	0.81
Neural Network	0.88	0.82

Fig. 2. True positive (sensitivity) and true negative (specificity) rates of the trained fashion classifiers.

Figure 2 demonstrates that the trained logistic regression and SVM models outperform the other models with respect to true positive and negative rates. We chose to employ the logistic regression model to mine Twitter's fashion subgraph since it is a simpler and more interpretable model than an SVM.

The trained logistic regression model reveals that content-based features computed over profile descriptions and mined tweets have the highest weights. The presence of terms (and their weights) such as *fashion* (7.1), *dress* (3.5), *beauty* (3.3), *collection* (3.2), and *wearing* (3.0), and hashtags such as *#fashion* (2.7), *#ootd* (2.0) — outfit of the day — and *#nyfw* (1.62) – New York fashion week — is highly predictive of fashion relevant Twitter accounts. The model also indicates that fashion accounts often mention other fashion accounts in their tweets, including magazines *@marieclaireuk* (1.3), luxury fashion brands *@gucci* (1.1), and blogs *@WhoWhatWear* (1.0). Conversely, geo-location is one of the least discriminative features, which is interesting given that fashion is often thought of as being concentrated in specific, metropolitan cities.

4 ESTIMATING THE SIZE OF THE FASHION SUBGRAPH

Leveraging the trained classifier for identifying fashion-relevant Twitter accounts, we can start mining Twitter's fashion subgraph. However, it is challenging to design stopping criteria for the crawl without knowing the approximate size of the subgraph. Estimating the fashion subgraph by sampling data from Twitter is a challenging task. Not only is Twitter's network large, but it is also dynamic: new accounts are being formed, accounts become inactive, and the accounts that users follow can change over time.

Researchers have developed methods for estimating the properties of domain-specific subgraphs via sampling. Gjoka et al. [31] obtained unbiased estimators using the Metropolis-Hastings random walk (MHRW) and a re-weighted random walk (RWRW) to produce a representative sample of Facebook users. Berry et al. [13] proposed a new unbiased method for estimating group properties of online social networks. These methods offer strategies for quantifying "a not known and must be crawled" new domain network, such as fashion.

To estimate the size of the fashion subgraph, we must compute an *unbiased sample* of Twitter: that is, a sample where the proportion of fashion nodes is the same as in the complete Twitter graph. Although there are several unbiased samplers, including RWRW and MHRW, we employed RWRW because of its convergence properties [31]. While RWRW provides an unbiased sample, it is possible that we may miss some accounts, given the large size of the Twitter network and the fact that our crawl is finite.

We considered using Twitter's streaming API to estimate the fashion subgraph's size, but discarded this idea after investigation. The streaming API serves the most recent tweets from active Twitter users. However, the API produces a *biased* sample susceptible to daily variation. For example, significant events in the fashion world (e.g., runway shows) that coincide with a crawl could bias the estimate and skew the sample of tweets. Moreover, Twitter subsamples the data in a non-uniform way [50, 55, 66].

4.1 Random Walk Based Estimation

We can create a sampled subgraph G over Twitter accounts by performing a random walk on the directed Twitter graph as described by Wang et al. [64]. This random walk treats each Twitter account as a vertex with edges extending to its follower and following accounts, but ignores the direction of edges to craw an undirected graph.

We randomly picked five seed accounts to initiate five such random walks, including two fashion accounts — (*@blackbirdlondon* and *@TheSource*) — from the 11.5k fashion accounts we identified,

and three non-fashion accounts — (*@LamWill*, *@EhhYoWIL*, and *@ShowoffMadeThis*) — from the non-fashion accounts in the training dataset.

First, given a seed, we use the *Tweepy* [62] library to crawl the vertex's following and follower neighbors. Consider a vertex v_i: the probability that the crawler moves to a neighboring vertex v_j is $1/d(v_i)$ if (v_i, v_j) is an edge in G and $d(v_i)$ is the degree of vertex v_i; otherwise, the probability is 0. Second, we randomly select a new vertex connected to the previously crawled vertex. To determine our estimate, we repeated these two steps until we had crawled 200k nodes for each run.

An ordinary random walk on a graph is biased toward high-degree nodes: the probability that the random walk visits a node v is proportional to its degree $d(v)$, i.e., $p(v) \propto d(v)$. We use the Hansen-Hurwitz estimator to debias the walk, re-weighting the visited nodes by their inverse degree [31]:

$$f_a = \frac{\sum_{x \in F_a} \frac{1}{d(x)}}{\sum_{x \in F} \frac{1}{d(x)} + \sum_{y \in \bar{F_a}} \frac{1}{d(y)}},$$

where f_a is the fraction of active[2] English fashion accounts, and F_a and $\bar{F_a}$ refer to the set of active English accounts labeled as *fashion* and *non-fashion*, respectively.

4.2 Estimation Results

We found that fashion percentage for all five re-weighted random walks stabilizes after crawling 50k accounts (Figure 3), and 60% of accounts in the RWRW sample were active. The average active English fashion percentage across all the random walks was 0.231%. As of 2018, Twitter had 335 million monthly active users in the world [58], and 34% of them are English accounts based on the latest data we could find from 2013 [2]. Therefore, given 113.9M English Twitter accounts, we

[2]We define an account to be *active* if it has posted two tweets at least six hours apart, in the past month. We determined that six hours of separation between social media activity is indicative of distinct sessions.

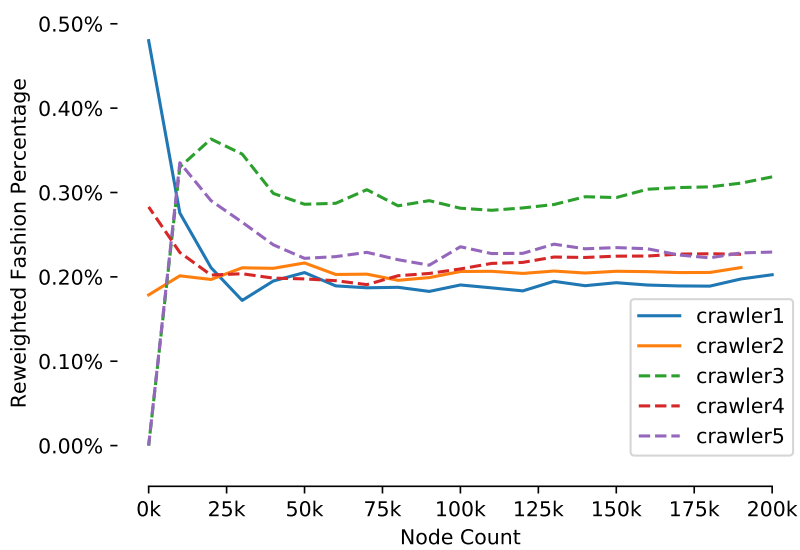

Fig. 3. The figure shows the estimated fraction of English fashion accounts using five re-weighted random walks crawled over five months. We crawled around 1M Twitter nodes. Crawlers 1 and 2 started the random walk from fashion accounts, while crawlers 3, 4, and 5 started from non-fashion accounts.

estimate that the active English fashion subgraph comprises roughly 263k (0.231%) accounts. In the future, we will try to estimate the ground-truth properties of the fashion group using the classifier as previously done by Berry et al. [13].

Across all five re-weighted random walks, we crawled 990k distinct accounts in total, from which we identified 9,747 fashion accounts. At this point, after deduplication with our 11.5k seeds, we had identified 20k fashion accounts, which we used as the seed set for the larger crawl of the fashion subgraph.

5 CONSTRUCTING FITNET

To construct FITNet, we need to identify the top influential, fashion-related accounts. Similar to prior work, we employ PageRank [52] to measure an account's influence within a network defined by Twitter's *following* and *follower* relationships [15, 32, 36, 69]. To create a fashion *following* graph, we leverage *homophily*: the tendency of individuals in a social network to connect with people similar to them [25]. We hypothesize that prominent fashion accounts on Twitter often follow other important fashion accounts.

Starting from a network seeded with 20k fashion accounts, we snowball sampled the *following* graph, using the fashion classifier to identify new fashion-related accounts to add to the network. As we crawled more accounts and grew the fashion subgraph, we computed PageRank over the network of the *following* relationships. We demonstrated that the set of top 10k accounts converge as the fashion subgraph approached 300k accounts. Finally, we recruited 55 undergraduates in a fashion-related major to validate and further categorize this ranked list of Twitter accounts until 10k fashion accounts had been identified. This validated, ranked, and categorized set of 10k fashion-related accounts comprise FITNet. To explore what FITNet accounts post about and how they interact with each other's content, we mine their tweets from a fixed year-long time window.

5.1 Crawling and Ranking a Fashion Subgraph

The random walk sampling done to estimate the size of the fashion subgraph produced a seed list of 20k fashion accounts. We used this seed list to initialize a fashion subgraph, and computed PageRank over the network of the following relationships. To grow this fashion network, we

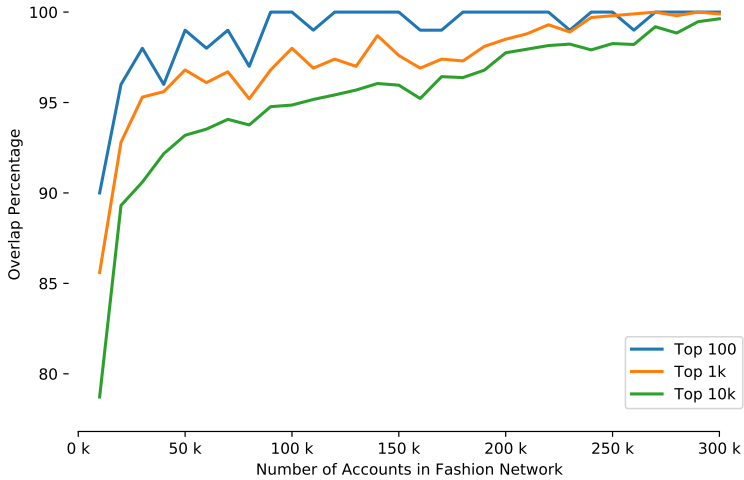

Fig. 4. As the fashion network grew to 300k accounts, the set of top 100, 1k, and 10k accounts under PageRank converged.

snowball sampled the *following* graph starting from the seed list; used the fashion classifier to identify new fashion-related accounts; and added these new account nodes to the graph, along with all of their *following* and *follower* edges to existing network accounts.

We recomputed PageRank over the network for each batch of 10k fashion-related accounts added to it. Figure 4 demonstrates that the set of top 10k accounts converges as the subgraph approaches 300k accounts. Given the classifier's false positive rate of 8%, the number of true positives can be estimated to be 276k, which is close to the 263k size estimate of the fashion subgraph. Therefore, we terminated the crawl after identifying 300k fashion-related accounts from snowball sampling over 70M Twitter accounts.

5.2 Validating and Categorizing the Influencers

To validate and further categorize the top-ranked fashion accounts, we recruited 55 undergraduate students majoring in Textile and Apparel Management at a large public university in the United States. We built a web interface that facilitated the validation and labeling process. Starting from the top of the ranked list, the interface presented accounts one at a time to participants until 10k fashion accounts had been validated and categorized.

The interface first asked participants to validate whether an account was fashion-related, and also determine whether an account was still *valid*: not deactivated and the most recent post being within the last two years. If participants deemed that an account was both fashion-related and still valid, then they were asked to further categorize the account. The interface provided 22 fashion categories (Appendix A) organized into four groups: *influencers* (e.g., model and blogger), *brands* (e.g., luxury and fast-fashion), *retail* (e.g., department store and ecommerce), and *media* (e.g., magazine and blog). Participants could assign multiple categories to each account, and they were allowed to write in additional categories if necessary.

For each account, the interface collected labels from two different participants. If participants disagreed about whether an account was related to fashion or valid, a member of the research team reviewed the account and broke the tie. Overall, participants labeled 10,919 accounts: 821 accounts were marked *non-fashion*, and 98 *invalid*. Note that the empirical true positive rate (92%) is consistent with the classifier's true positive rate. The remaining set of valid, fashion-related 10k accounts comprise FITNet. The FITNet accounts are distributed across the fashion categories in the following way: 4,188 (42%) *influencers*, 2,066 (20%) *brands*, 1,356 (14%) *retailers*, and 3,105 (31%) *media*.

5.3 Mining the Tweets, Retweets, and Mentions

To explore what FITNet accounts post about and how they interact with each other, we used Twitter's API to mine all of their tweets between January 1st, 2018, and February 1st, 2019. On average, we captured 2,923 posts per account. We index this repository of tweets by *hashtags*, *mentions*, and *retweets*. This allows us to identify FITNet accounts that post about specific topics, and understand how accounts interact with each other's content by visualizing *mention* and *retweet* networks.

6 RESULTS: UNDERSTANDING INFLUENCERS

This paper presents the first large-scale analysis of fashion influencers on social media. FITNet comprises 4,188 *influencer* accounts. Among them, there are 2,248 bloggers (54%), 593 editors (14%), 458 designers (11%), 485 stylists (12%), 307 celebrities (7%), 291 models (7%), 137 photographers (3%), and 79 casting directors (2%). Therefore, more than half of the influencers in FITNet are categorized as bloggers.

Celebrities	Rank	Designers	Rank	Models	Rank	Editors	Rank	Bloggers	Rank
victoriabeckham	23	RachelZoe	20	cocorocha	80	HilaryAlexander	50	susiebubble	37
KimKardashian	35	themarcjacobs	120	KendallJenner	113	ninagarcia	69	bryanboy	70
ladygaga	41	LaurenConrad	122	kourtneykardash	127	evachen212	95	CathyHoryn	90
rihanna	51	RebeccaMinkoff	135	tyrabanks	132	DerekBlasberg	101	LibertyLndnGirl	107
ninagarcia	69	KarlLagerfeld	136	KylieJenner	167	LibertyLndnGirl	107	byEmily	150
taylorswift13	79	Zac_Posen	142	heidiklum	175	annadellorusso	160	fashionfoiegras	153
KendallJenner	113	JasonWu	181	DitaVonTeese	193	Edward_Enninful	164	ChiaraFerragni	177
LaurenConrad	122	xoBetseyJohnson	190	ZooeyDeschanel	214	francasozzani	165	theglamourai	182
OliviaPalermo	123	whitneyEVEport	192	CindyCrawford	220	tavitulle	187	_tommyton	186
kourtneykardash	127	masato_jones	211	jessicastam	226	SBRogue	206	Sartorialist	195
44% Influence		35% Influence		32% Influence		31% Influence		29% Influence	

Fig. 5. The top ten list for five influencer subcategories: Celebrities, Designers, Models, Editors, Bloggers.

6.1 How influential are they?

We measured the influence of every single account by computing the total number of incoming *following* links (also known as *indegree centrality*) divided by the total number of nodes in the fashion subgraph (300K accounts). The top ten influencers under PageRank have 46% influence, meaning that nearly half of the fashion subgraph that we have identified are following these ten influencers. Similarly, the top 100 influencers have 65% influence, and all of the influencers in FITNet have 91% influence.

We also computed the influence based on the influencer categories. The top ten celebrities have the most influence (44%), followed by the top ten designers (35%), models (32%), editors (31%), bloggers (29%), stylists (26%), photographers (18%), and casting directors (8%). The top 100 celebrities also have the most influence (62%), followed by bloggers (51%), designers (48%), models (47%), editors (36%), stylists (36%), photographers (27%), and directors (6%).

Taken together, all the bloggers on FITNet have the most influence (78%), followed by celebrities (65%), designers (58%), models (48%), editors (47%), stylist (44%), photographers (28%), and directors (6%). So in aggregate, bloggers in FITNet have the most influence, but each individual celebrity still has more influence on average. In fact, the top ten celebrities, designers, models, and editors are all more influential than the top ten bloggers. Figure 5 presents the top ten accounts in each influencer subcategory along with their corresponding influence.

6.2 How is influence geographically distributed?

To study the geographical distribution of fashion influencers, we used the Geocoder library [14] and Google geocoding service [12] to retrieve the coordinates of the influencers' accounts based on location data (e.g., cities) specified in their profile. Based on the retrieved coordinates, we created Figure 6 to illustrate geographical influencer hotspots. For each influencer, Figure 6 presents their number of followers on both Twitter (*left*) and Instagram (*right*).

While a large portion of bloggers are concentrated in fashion hubs such as *New York*, *London*, and *Paris*, others are widely distributed. Some highly-ranked influencers live in states such as North

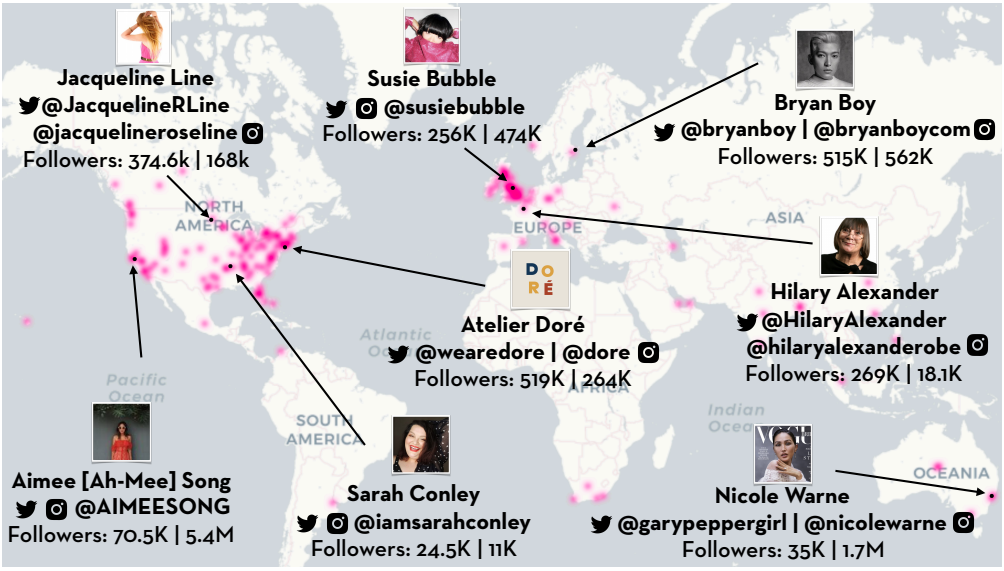

Fig. 6. A heat map visualizing the geographical distribution (in pink) of FITNet's influencers. Although many influencers are still concentrated in well-known fashion hubs such as *New York*, *London*, and *Paris*; social media has also distributed fashion influence more broadly.

Dakota *(@JacquelineRLine, PageRank 108)* and Arkansas *(@imsarahconley, PageRank 408)*, which are traditionally not considered as fashion-centric in the U.S. Similarly, some other top influencers live in large cities that are also not considered as fashion-centric. For example, *Bryan Boy (@bryanboy, PageRank 70)* is from Stockholm, Sweden, and *Nicole Warne (@garypeppergirl, PageRank 613)* lives in Sydney, Australia.

6.3 Who influences them?

To understand how influencers influence each other, we measured reciprocity in the *follow* relation network (e.g., how many connections in the graph are two-way edges). Only 7% of the 302,217,549 edges in FITNet are reciprocal. However, reciprocity is more prevalent among influencer accounts. Overall, influencer reciprocity within FITNet is 15% and 18% within the top 100 influencers accounts. This phenomenon could be ascribed to homophily: top influencers in the community follow one another and form social clusters.

Figure 7 shows that bloggers follow everyone — including brands, retailers, and media accounts — probably to be in the know. In comparison, celebrities often follow a very small number of accounts when compared to the number of followers they have.

We also analyzed the tweets that were mined to understand who influencers interact with on Twitter through *mentions*. Figure 8 shows *homophily* in action: influencers tend to mention other people in their own categories. Bloggers mention other bloggers more than any other fashion category, and similarly, celebrities mention other celebrities more than any other fashion category. Editors and stylists mention indiscriminately across all the categories. Bloggers also mention ecommerce sites, fast-fashion brands, and beauty brands more than luxury and premium brands. In addition to other celebrities, celebrities mention magazines in their tweets more than other fashion categories.

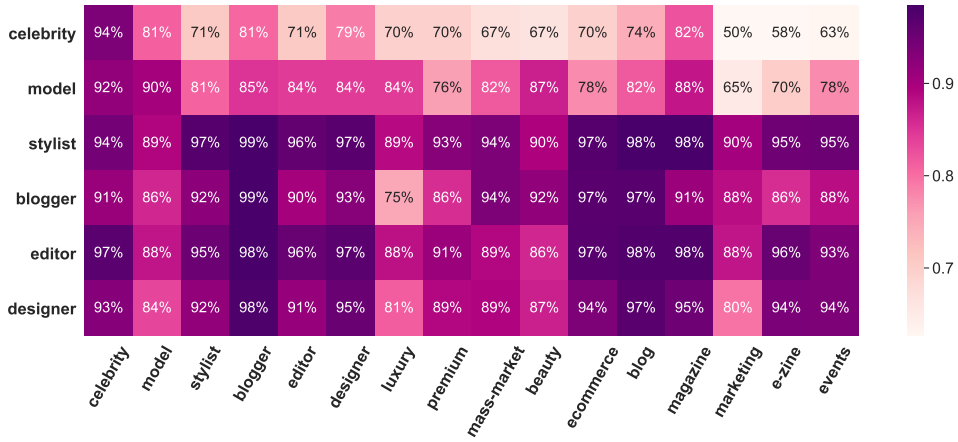

Fig. 7. Percentage of influencer accounts that *follow* other fashion accounts, by category.

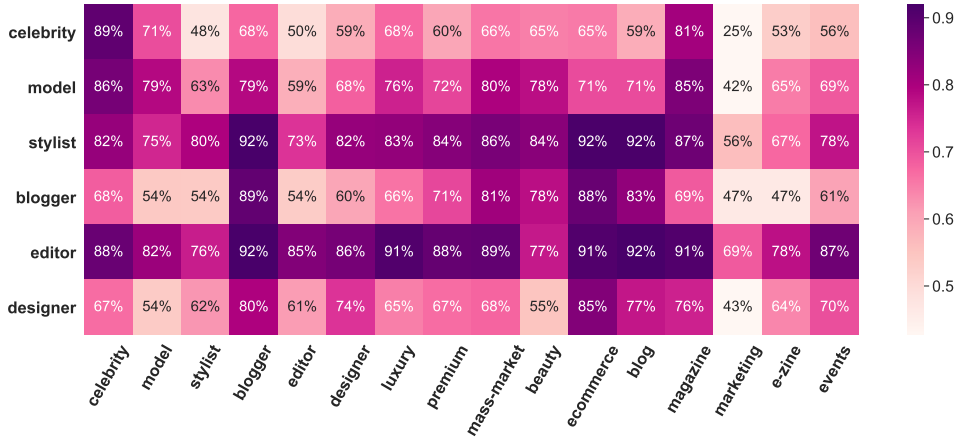

Fig. 8. Percentage of influencer accounts that *mention* other fashion accounts, by category.

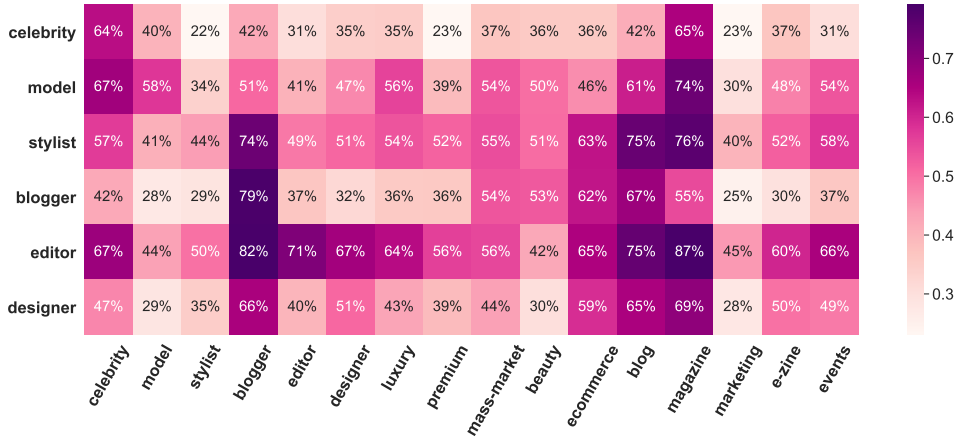

Fig. 9. Percentage of influencer accounts that *retweet* other fashion accounts, by category.

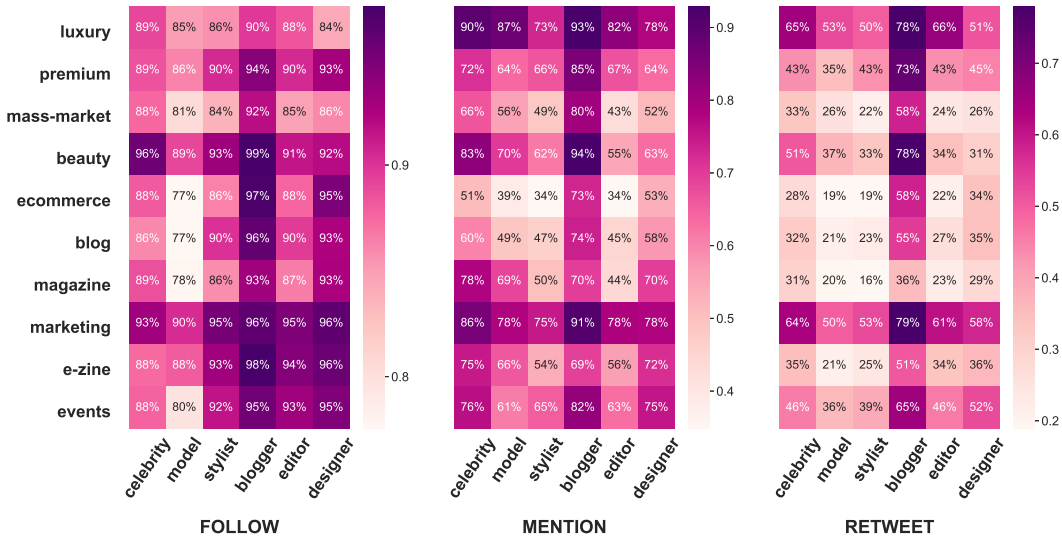

Fig. 10. Pecentage of other accounts that *follow*, *mention*, and *retweet* influencers, by category.

Finally, we analyzed the tweets that were mined to understand who influencers interact with on Twitter through *retweets*. Figure 9 shows that the retweet behavioral patterns generally mirror the mention ones. For example, bloggers mostly retweet bloggers/blogs as well as ecommerce sites; celebrities retweet other celebrities as well as magazines. Overall, blogs and magazines are retweeted more than brands, perhaps because they often generate interesting, original content that is not always self-promotional.

6.4 Who do they influence?

In addition, we also analyzed how other fashion categories — brands, retail, and media — follow, mention, and retweet influencers (Figure 10). Of all the non-influencer categories, luxury fashion brands, PR/Marketing agencies, and beauty brands interact with influencers the most. Although PR/Marketing agencies engage heavily with influencer accounts, influencers do not reciprocate: they hardly mention or retweet PR/Marketing agency accounts.

Of all the influencers, bloggers are the most influential category: they are heavily followed, mentioned, and retweeted by other categories. The next tier of influence comprises designers and celebrities, who are often followed and mentioned, but not heavily retweeted. Finally, brands, retail, and media accounts engage the least with models, stylists, and editors.

7 DISCUSSION: SUPPORTING INFLUENCER DISCOVERY

We designed FITNet's data representation with *discovery* in mind. Fashion companies often want to discover influencers who post about topics that resonate with their brand image or their customers' interests. To support this type of topic-based exploration, we index tweets by hashtags, enabling us to retrieve the set of accounts in FITNet that previously tweeted about a specific hashtag. However, just using a hashtag a few times does not signify that an account cares deeply about a specified topic. We demonstrate how the structure of hashtag-related interaction subgraphs reveals influencers who champion specific fashion topics.

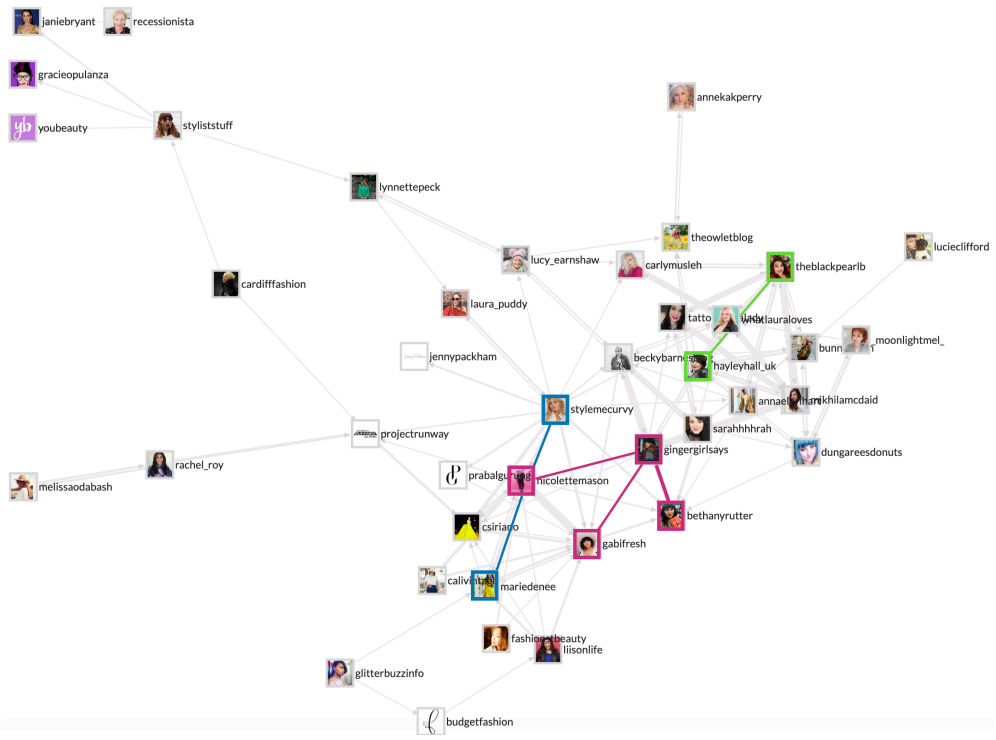

Fig. 11. The largest connected component of the *#plussize mention* graph reveals clusters with strong interaction ties related to *plus-size* fashion.

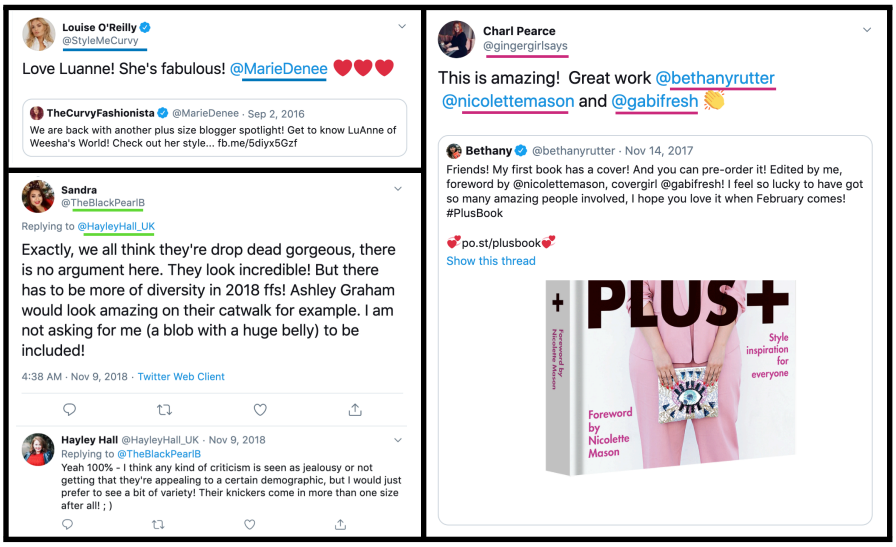

Fig. 12. Tweets illustrating *mention* interactions between FITNet influencers relating to *plus-size* fashion.

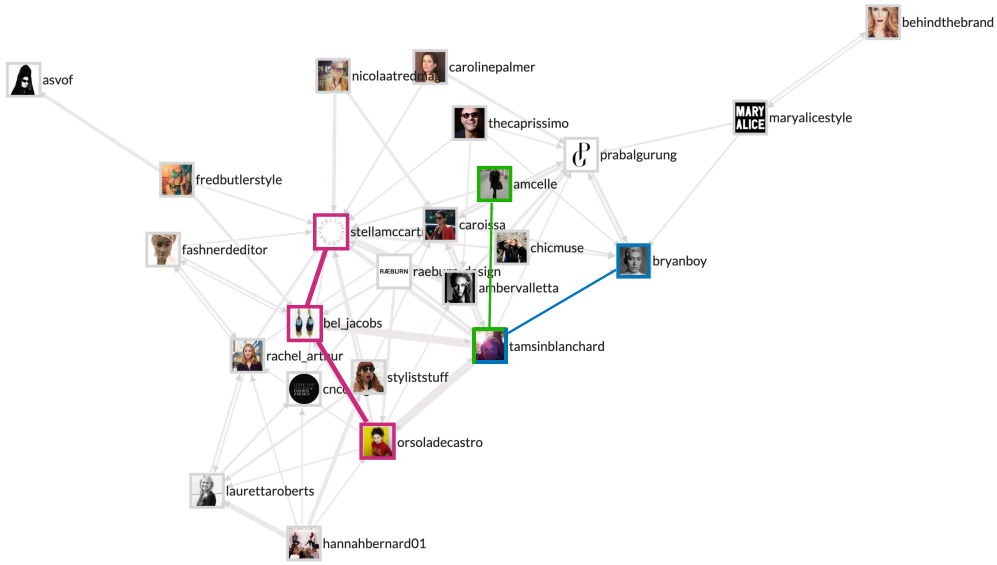

Fig. 13. The largest connected component of the *#sustainability mention* graph reveals clusters with strong interaction ties related to *sustainable* fashion.

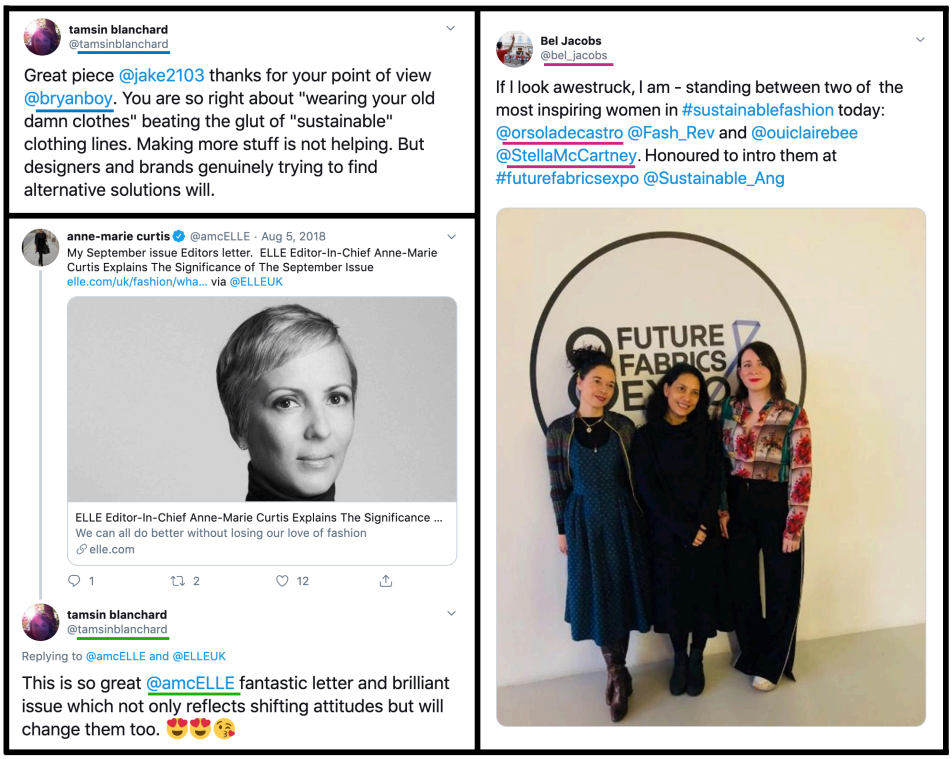

Fig. 14. Tweets illustrating *mention* interactions between FITNet influencers relating to *sustainable* fashion.

For example, if a fashion brand is trying to expand its range of sizes to include plus-sizes, they might want to find influencers on social media to promote their new product line. They could query FITNet with the *#plussize* hashtag to retrieve all the accounts that have previously used that hashtag in a tweet.

There are 70 influencers in FITNet who have used *#plussize* in their tweets, and they are a close-knit group. The *#plussize follow* network forms one connected component, and on average, one account follows 9.3 other accounts (13.3%) in the network. The *mention* network comprises two connected components, and 41 (58.6%) accounts mention at least one other account in that group. Finally, the *retweet* network comprises five connected components, and 51 (72.9%) individuals retweet at least one other account in that subset.

By visualizing the largest connected component of 38 influencers in the *mention* graph, we can identify clusters with strong interaction ties, which surface influential accounts in this topic network (Figure 11). Figure 11 highlights some of these clusters, and Figure 12 provides evidence of *#plussize* tweets where individuals in these clusters mention each other. Therefore, based on the interactions uncovered in the *#plussize mention* graph, a brand could reach out to bloggers such as *@nicolettemason*, *@bethanyrutter*, *@gabifresh*, *@gingergirlsays*, *@MarieDenee*, *@StyleMeCurvy*, *@TheBlackPearlB*, and *@HayleyHall_UK* to promote their plus-size line launch.

By employing a similar method, we demonstrate how to use FITNet to discover *sustainability* influencers. Sustainability is increasingly becoming an important topic in the fashion industry; there are 49 influencers in FITNet who have used *#sustainability* in their tweets. The *#sustainability follow* network forms one connected component, where one account, on average, follows 6.2 other accounts (12.7%) in the subgraph. The *mention* graph comprises three connected components, where 28 (57.1%) individuals mention at least one other account in that group. Finally, the *retweet* graph comprises three connected components, where 30 (61.2%) individuals retweet at least one other account in that subset.

By visualizing the largest connected component of 23 influencers in the *mention* graph, we can again identify clusters with strong interaction ties with respect to the topic of sustainability (Figure 13). Figure 14 provides evidence of *#sustainability* tweets where influencers from the clusters mention each other. Therefore, we can identify that *@tamsinblanchard*, *@bryanboy*, *@amcELLE*, *@bel_jacobs*, *@orsoladecastro*, and *@StellaMcCartney* would be good influencer candidates to promote fashion sustainability causes.

8 CONCLUSION AND FUTURE WORK

This work is a first step toward studying the diffusion of fashion influence on social media. Future work can leverage FITNet to build systems that monitor fashion influencers: the content they post, their impact on the fashion industry, etc. To comprehensively capture an influencer's digital footprint, it would be necessary to mine their activity on other popular fashion social media such as Instagram and YouTube. Since many fashion influencers use consistent screen names across social media platforms, FITNet's list of Twitter screen names can be used to bootstrap mining efforts on platforms where APIs are not readily available. Since fashion influencers are the tastemakers of the industry, such monitoring systems could also enable researchers to study the evolution of fashion trends, capture them as they happen, and possibly even predict future ones.

9 ACKNOWLEDGEMENTS

We thank the reviewers for their helpful feedback and suggestions; Kedan Li, Yuan Shen, Vivian Hu, Doris Jung-Lin Lee, Wenxin Yi for their assistance in implementing different parts of this work; and the participants who completed the validation and categorization tasks. This work was supported in part by an Amazon Research Award.

REFERENCES

[1] 2012. Top 20 Fashion Bloggers to Follow on Twitter . http://www.fashiontrendsdaily.com/events/top-20-fashion-bloggers-to-follow-on-twitter.

[2] 2013. Only 34% of All Tweets Are in English. https://www.statista.com/chart/1726/languages-used-on-twitter/.

[3] 2016. Meet The Top 10 Fashion Influencers of 2016 . www.izea.com.https://izea.com/2017/01/05/top-fashion-influencers/.

[4] 2017. Retail Fashion: Top 300 Influencers, Brands and Publications. http://www.onalytica.com/blog/posts/retail-fashion-top-300-influencers-brands-publications/.

[5] 2018. 15 Fashion Influencers to Follow. https://en.vogue.fr/fashion/fashion-websites/diaporama/15-fashion-influencers-to-follow/51325.

[6] 2019. Alexa - Top Sites by Category: Arts/Design/Fashion. https://www.alexa.com/topsites/category/Arts/Design/Fashion.

[7] 2019. Best Sellers in Men's Fashion Magazines. https://www.amazon.com/Best-Sellers-Magazines-Mens-Fashion/zgbs/magazines/637728.

[8] 2019. Best Sellers in Women's Fashion Magazines. https://www.amazon.com/Best-Sellers-Magazines-Womens-Fashion/zgbs/magazines/637730/.

[9] Crystal Abidin. 2016. Visibility labour: Engaging with Influencers? fashion brands and# OOTD advertorial campaigns on Instagram. *Media International Australia* 161, 1 (2016), 86–100.

[10] Rasha Al Bashaireh, Mohamed Zohdy, and Vian Sabeeh. 2020. Twitter Data Collection and Extraction: A Method and a New Dataset, the UTD-MI. In *Proceedings of the 2020 the 4th International Conference on Information System and Data Mining*. 71–76.

[11] Ziad Al-Halah and Kristen Grauman. 2020. From Paris to Berlin: Discovering Fashion Style Influences Around the World. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10136–10145.

[12] Google Maps APIs. 2017. GeocodingAPI. https://developers.google.com/maps/documentation/geocoding/start.

[13] George Berry, Antonio Sirianni, Nathan High, Agrippa Kellum, Ingmar Weber, and Michael Macy. 2018. Estimating group properties in online social networks with a classifier. In *International Conference on Social Informatics*. Springer, 67–85.

[14] Denis Carriere. 2017. geocoder 1.8.0. https://pypi.python.org/pypi/geocoder/1.8.0.

[15] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. 2010. Measuring User Influence in Twitter: The Million Follower Fallacy.. In *AAAI Conference on Weblogs and Social Media*, Vol. 14.

[16] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, P Krishna Gummadi, et al. 2010. Measuring user influence in twitter: The million follower fallacy. *Icwsm* 10, 10-17 (2010), 30.

[17] Rupak Chakraborty, Senjuti Kundu, and Prakul Agarwal. 2016. Fashioning Data-A Social Media Perspective on Fast Fashion Brands. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. 26–35.

[18] Shuo Chang, Vikas Kumar, Eric Gilbert, and Loren G Terveen. 2014. Specialization, homophily, and gender in a social curation site: findings from pinterest. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 674–686.

[19] Yi Chang, Lei Tang, Yoshiyuki Inagaki, and Yan Liu. 2014. What is tumblr: A statistical overview and comparison. *ACM SIGKDD explorations newsletter* 16, 1 (2014), 21–29.

[20] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 785–794.

[21] Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. Will-They-Won't-They: A Very Large Dataset for Stance Detection on Twitter. *arXiv preprint arXiv:2005.00388* (2020).

[22] Dilek Çukul. 2015. *Fashion marketing in social media: using Instagram for fashion branding*. Technical Report. International Institute of Social and Economic Sciences.

[23] Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1277–1286.

[24] NLP developers. 2018. Natural Language Toolkit. https://www.nltk.org/book/ch03.html.

[25] David Easley, Jon Kleinberg, et al. 2012. Networks, crowds, and markets: Reasoning about a highly connected world. *Significance* 9 (2012), 43–44.

[26] Bonnie Lee English. 2007. *A cultural history of fashion in the 20th century: from the catwalk to the sidewalk*. Berg Publishers.

[27] Manuela Farinosi and Leopoldina Fortunati. 2020. Young and Elderly Fashion Influencers. In *International Conference on Human-Computer Interaction*. Springer, 42–57.

[28] Alberto Ferrari. 2017. Top Fashion Influencers. http://www.topfashioninfluencers.com.

[29] Vijay Gabale and Anand Prabhu Subramanian. 2018. How To Extract Fashion Trends From Social Media? A Robust Object Detector With Support For Unsupervised Learning. https://arxiv.org/pdf/1806.10787.pdf.

[30] Eric Gilbert, Saeideh Bakhshi, Shuo Chang, and Loren Terveen. 2013. I need to try this?: a statistical overview of pinterest. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2427–2436.

[31] Minas Gjoka, Maciej Kurant, Carter T Butts, and Athina Markopoulou. 2010. Walking in facebook: A case study of unbiased sampling of osns. *Infocom, 2010 Proceedings IEEE. Ieee, 2010*. (2010).

[32] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. Wtf: The who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 505–514.

[33] Emily Hund. 2017. Measured Beauty: Exploring the aesthetics of Instagram's fashion influencers. In *Proceedings of the 8th International Conference on Social Media & Society*. 1–5.

[34] Lauren Indvik. 2016. The 20 Most Influential Personal Style Bloggers: 2016 Edition. *Fashionista. Np* 14 (2016).

[35] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath Hariharan, Hartwig Adam, and Serge Belongie. 2020. Fashionpedia: Ontology, Segmentation, and an Attribute Localization Dataset. *arXiv preprint arXiv:2004.12276* (2020).

[36] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a social network or a news media?. In *Proceedings of the 19th international conference on World wide web*. ACM, 591–600.

[37] Ralph Lauretta. 2016. Men's Fashion Twitter Accounts You Should Be Following. https://sallauretta.com/mens-fashion-twitter-accounts/.

[38] Doris Jung-Lin Lee, Jinda Han, Dana Chambourova, and Ranjitha Kumar. 2017. Identifying Fashion Accounts in Social Networks. *KDD 2017 Machine Learning Meets Fashion Workshop* (August 2017).

[39] Sungjae Lee, Yeonji Lee, Junho Kim, and Kyungyong Lee. 2019. RnR: Extraction of Visual Attributes from Large-Scale Fashion Dataset. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 5043–5047.

[40] Wen Hua Lin, Kuan-Ting Chen, Hung Yueh Chiang, and Winston Hsu. 2018. Netizen-style commenting on fashion photos: dataset and diversity measures. In *Companion Proceedings of the The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 395–402.

[41] Babak Loni, Lei Yen Cheung, Michael Riegler, Alessandro Bozzon, Luke Gottlieb, and Martha Larson. 2014. Fashion 10000: an enriched social image dataset for fashion and clothing. In *Proceedings of the 5th ACM Multimedia Systems Conference*. ACM, 41–46.

[42] Babak Loni, Maria Menendez, Mihai Georgescu, Luca Galli, Claudio Massari, Ismail Sengor Altingovde, Davide Martinenghi, Mark Melenhorst, Raynor Vliegendhart, and Martha Larson. 2013. Fashion-focused creative commons social dataset. In *Proceedings of the 4th ACM Multimedia Systems Conference*. ACM, 72–77.

[43] Yunshan Ma, Lizi Liao, and Tat-Seng Chua. 2019. Automatic Fashion Knowledge Extraction from Social Media. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2223–2224.

[44] Yunshan Ma, Xun Yang, Lizi Liao, Yixin Cao, and Tat-Seng Chua. 2019. Who, Where, and What to Wear? Extracting Fashion Knowledge from Social Media. In *Proceedings of the 27th ACM International Conference on Multimedia*. 257–265.

[45] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. 2019. Geostyle: Discovering fashion trends and events. In *Proceedings of the IEEE International Conference on Computer Vision*. 411–420.

[46] Lydia Manikonda, Ragav Venkatesan, Subbarao Kambhampati, and Baoxin Li. 2015. Evolution of fashion brands on Twitter and Instagram. *arXiv preprint arXiv:1512.01174 v1 [cs. SI]* (2015).

[47] Lydia Manikonda, Ragav Venkatesan, Subbarao Kambhampati, and Baoxin Li. 2015. Trending chic: Analyzing the influence of social media on fashion brands. *arXiv preprint arXiv:1512.01174* (2015).

[48] Kevin Matzen, Kavita Bala, and Noah Snavely. 2017. Streetstyle: Exploring world-wide clothing styles from millions of photos. *arXiv preprint arXiv:1706.01869* (2017).

[49] Megan Mayer. 2014. 57 Twitter Accounts You Need To Follow This Fashion Week. https://www.huffingtonpost.com/2014/09/03/twitter-fashion-week-_n_4718795.html.

[50] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M Carley. 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. *arXiv preprint arXiv:1306.5204* (2013).

[51] Victoria Nebot, Francisco Rangel, Rafael Berlanga, and Paolo Rosso. 2018. Identifying and classifying influencers in twitter only with textual information. In *International Conference on Applications of Natural Language to Information Systems*. Springer, 28–39.

[52] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[53] Jaehyuk Park, Giovanni Luca Ciampaglia, and Emilio Ferrara. 2016. Style in the age of instagram: Predicting success within the fashion industry using social media. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. ACM, 64–73.

[54] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[55] Jürgen Pfeffer, Katja Mayer, and Fred Morstatter. 2018. Tampering with TwitterâĂŹs sample API. *EPJ Data Science* 7, 1 (2018), 50.
[56] Kerry Pieri. 2018. The Fashion Blogger Instagrams to Follow Now. https://www.harpersbazaar.com/fashion/street-style/g3957/best-blogger-instagrams/.
[57] Hannah Stacey. 2017. Top Fashion Twitter Accounts: 10 Must-Follow Fashion Industry Insiders. https://blog.ometria.com/top-fashion-twitter-accounts.
[58] Statista.com. 2018. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 2nd quarter 2018 (in millions). *The Statista Portal* (2018). https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.
[59] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.
[60] Gabrielle Thuy, Marissa Evans, and Roderick Chow. 2011. What websites in the fashion space have the highest traffic? https://www.quora.com/What-websites-in-the-fashion-space-have-the-highest-traffic.
[61] Avery Trufelman. 2016. The Trend Forecast. http://99percentinvisible.org/episode/the-trend-forecast/.
[62] Tweepy. 2017. An easy-to-use Python library for accessing the Twitter API. http://www.tweepy.org/.
[63] Stefanie Urchs, Lorenz Wendlinger, Jelena Mitrović, and Michael Granitzer. 2019. MMoveT15: A Twitter Dataset for Extracting and Analysing Migration-Movement Data of the European Migration Crisis 2015. In *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE, 146–149.
[64] Tianyi Wang, Yang Chen, Zengbin Zhang, Peng Sun, Beixing Deng, and Xing Li. 2010. Unbiased sampling in directed social graph. In *Proceedings of the ACM SIGCOMM 2010 conference*. 401–402.
[65] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. 1031–1040.
[66] Yazhe Wang, Jamie Callan, and Baihua Zheng. 2015. Should we use the sample? Analyzing datasets sampled from Twitter's stream API. *ACM Transactions on the Web (TWEB)* 9, 3 (2015), 1–23.
[67] Henry Ward. 2017. The Shadow Org Chart. https://medium.com/@henrysward/the-shadow-org-chart-cfcdd644575f.
[68] Jeremy D Wendt, Randy Wells, Richard V Field, and Sucheta Soundarajan. 2016. On data collection, graph construction, and sampling in Twitter. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 985–992.
[69] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. TwitterRank: Finding Topic-sensitive Influential Twitterers *(WSDM '10)*. ACM, New York, NY, USA, 261–270.
[70] Wikipedia contributors. 2020. List of fashion designers. https://en.wikipedia.org/wiki/List_of_fashion_designers [Online; accessed 2017-2020].
[71] Wikipedia contributors. 2020. List of Victoria's Secret models. https://en.wikipedia.org/wiki/List_of_Victoria%27s_Secret_models [Online; accessed 2017-2020].
[72] Kota Yamaguchi, M Hadi Kiapour, Luis E Ortiz, and Tamara L Berg. 2012. Parsing clothing in fashion photographs. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3570–3577.
[73] Yuto Yamaguchi, Tsubasa Takahashi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. 2010. Turank: Twitter user ranking based on user-tweet graph analysis. In *International Conference on Web Information Systems Engineering*. Springer, 240–253.
[74] Xiao Yang, Seungbae Kim, and Yizhou Sun. 2019. How do influencers mention brands in social media? sponsorship prediction of Instagram posts. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 101–104.
[75] Yin Zhang and James Caverlee. 2019. Instagrammers, fashionistas, and me: Recurrent fashion recommendation with implicit visual influence. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1583–1592.
[76] Li Zhao and Chao Min. 2018. The Rise of Fashion Informatics: A Case of Data-Mining-Based Social Network Analysis in Fashion. *Clothing and Textiles Research Journal* (2018), 0887302X18821187.

A FASHION CATEGORIES

The following options were provided to participants for categorizing Twitter accounts:
Influencers (8): celebrity, model, stylist, blogger, photographer, editor, designer, casting director
Brands (4): luxury fashion brand, premium fashion brand, mass-market/fast fashion brand (e.g.

Zara, Gap, Uniqlo), beauty (e.g. Estee Lauder, Clinique)
Retailers (3): department store (e.g. Nordstrom, Macy's), ecommerce site (e.g. Farfetch, 6pm), other retailers (e.g. Target, Kohls)
Media (7): blog, newspaper, magazine, PR/marketing agency, e-zine (digital magazine only and social media website), fashion week or other fashion events, social networking site (e.g. Instagram, Twitter, Pinterest)