Scalable and explainable user role detection in social media

Johannes Kastner, Peter M. Fischer

Angaben zur Veröffentlichung / Publication details:

Kastner, Johannes, and Peter M. Fischer. 2021. "Scalable and explainable user role detection in social media." In *New Trends in Database and Information Systems: ADBIS 2021 Short Papers, Doctoral Consortium and Workshops: DOING, SIMPDA, MADEISD, MegaData, CAoNS, Tartu, Estonia, August 24-26, 2021, Proceedings*, edited by Ladjel Bellatreche, Marlon Dumas, Panagiotis Karras, Raimundas Matulevičius, Ahmed Awad, Matthias Weidlich, Mirjana Ivanović, and Olaf Hartig, 263–75. Berlin: Springer. https://doi.org/10.1007/978-3-030-85082-1_23.

Scalable and Explainable User Role Detection in Social Media

Johannes Kastner and Peter M. Fischer(✉)

University of Augsburg, Universitätsstr. 6a, 86159 Augsburg, Germany
{johannes.kastner,peter.m.fischer}@uni-a.de

Abstract. While identifying specific user roles in social media -in particular bots or spammers- has seen significant progress, generic and all-encompassing user role classification remains elusive on the large data sets of today's social media. Yet, such broad classifications enable a deeper understanding of user interactions and pave the way for longitudinal studies capturing the evolution of users such as the rise of influencers.

We build on the fundamental role definitions of previous empirical studies and provide a *largely automated*, *scalable detection* of *fine-grained roles*. Our approach clusters users *hierarchically* and explains the salient features. To associate clusters with roles, we use *supervised classifiers*: trained by experts on completely new media, but transferable on related data. Furthermore, we employ the combination of samples in order to improve scalability and allow probabilistic assignments of user roles.

Our evaluation on Twitter indicates that a) *stable* and *reliable detection* of a wide range of *roles* is possible b) the labeling *transfers* well as long as the fundamental properties don't strongly change between datasets and c) the approaches *scale* well with little need for human intervention.

Keywords: Social media · User role detection · Classification

1 Introduction

Automatically identifying user roles in social media at scale and speed promises interesting insights on their prevalence and impact. Furthermore, a stable recognition explains how individual users and communities evolve over time.

We propose a method that combines *unsupervised learning* to discover *fine-grained* classes of users over a wide range of features with *supervised learning* - generalizing expert knowledge from manually labeled reference data to new datasets, mapping role candidates to well-known roles or identifying new roles.

The paper provides the following *contributions*:

- Learning the *structure* of *user groups* as well as assigning *suitable labels*.
- A study on *large, complementary datasets* shows that both *recognizing* and *transferring* roles is feasible over *longer time periods* or *topic variations*.
- The classification hierarchy with salient *feature detection* and the cluster metrics support reviewing, so that *identification* requires *little intervention*.
- *Sample combination* provides *scalability* and *probabilistic role assessment*.

The remainder of this paper is structured as follows: In Sect. 2 we discuss related work. We introduce our methodology in Sect. 3 and provide more details on structure discovery and labeling in Sects. 4 and 5, respectively. After an extensive evaluation (Sect. 6), we conclude the paper.

2 Related Work

Clearly, identifying user roles has been one of the textbook examples of classifier algorithms, yet the application to social networks has been limited to particular aspects. Often, the studies focus on detecting specific roles or describing only a small number of coarse-grained classes. Considering the negative dynamics of many social networks, most researchers focus on identifying malicious users like bots [2] and spammers [9] or aggressors in the context of cyber bullying [1,7]. In contrast, our goal is to comprehensively assign all users to roles. Multi-role approaches such as Varol et al. [11], Rocha et al. [4] and Lazaridou et al. [8] limit themselves to identify a small number (often 3–5) of major, course-grained groups, roughly corresponding the upper levels of our detection hierarchy. Du et al. [3] provide a somewhat higher number of rules (still lower than ours), but only give fairly generic descriptions. All of the previously mentioned methods are constrained on just detecting the structure by unsupervised learning: clustering via K-Means [8] or EM [4] or via topic models [3], leaving the analysis entirely to human experts. Varol et al. [11] fully rely on such human classification, using similarity matrices and handcrafted rules. In contrast, qualitative works like Tinati et al. [10] or Java et al. [6] provide a comprehensive overview on fine-grained roles and their semantics, but provide only general rules on how to detect them. An interesting, complementary direction is the work on content communities/web forum, often exploring complex temporal models, e.g., [5]. It should be noted that all of these works (with the exception of [3] (Weibo, 12K users), [7] (Instagram, 18K users), and [5] (Stack Overflow)) solely rely on Twitter due to the limited availability of data from other services.

3 Approach

In order to classify diverse user roles in large data, we phrase three questions:

1. To which extent can clusters of users be utilized to sensibly *detect* user roles in social media and build a *classifier* to *(semi-)automatically* label them?
2. Can this approach be applied individually over *a wide variety of data sets*, currently stemming from the same social media?
3. Can the *knowledge* on roles be *transferred* to *new data sets*?

As the related work only describes instances of user roles, but not the concept of a role itself, we use the following, *basic definition*: A **user role** is a *set of users* that *share similar feature values* and are *well separated* from other groups. The features *capture salient properties* of users and allow a *meaningful categorization*, typically capturing *behavior* and *position* in the network/media. Groups constitute roles if they are *present* in sufficient number within a data set and *re-occur* over multiple data sets.

Our approach can be applied in *two complementary scenarios*, requiring different quantities of human involvement given the amount reference data:

1) If only **unexplored datasets** are available, we *discover* groups of *similar users* and their *hierarchical relationship*, providing *candidates* for *user roles*. The analyst is then aided by *metrics* and *visualizations* in assigning *role labels*. In turn, these *labels* form the *input* for a *classifier* that captures this *knowledge*.
2) If a **reference dataset** with a *classifier* is available, the labeling process can be cut short by providing *candidate labels*. The user can *evaluate* these labels within the new dataset or *compare* roles across datasets. We explored causes of mislabelings and methods to adapt, yet a full exploration remains future work.

4 Feature Selection and Data Clustering

In this work, we aim to use *features* that cover *significant* and *complementary* aspects of *users* and are well established in the literature [1,4,8]. In addition, they should be feasible to compute in *large scale* so that data is commonly available and incur *moderate cost* to compute. Likewise, we want to avoid a large number of features, as this hurts both algorithm performance and explainability.

Figure 1 highlights our features: **static user properties** express (self-)description: most relevant is the *verified* status, traditionally reserved for VIPs. **User activity** is characterized by the number of original tweets of each user (observed and "offtopic"), the activities on other tweets such as retweets and replies within the topic as well as mentions of other users. Basic **network position** features like the number of *followers* and *followees* (aka degree centrality) of a user as well underpin the potential to exert influence. In turn, the user's ability to actually elicit **reactions from the network** is captured by the *ratio of tweets* to lead to *replies* and *retweets* as well as the frequency of *being mentioned*.

We investigated a wide variety of additional features from these classes, but dropped them as they were *correlated* or had *little discriminative power*. Furthermore, we did not include complex network metrics such as path-based centralities, spatio-temporal features [11] as well as explicit content analyses [1,7]. Even partial social graphs are exceedingly hard to get from any social media (including Twitter), while our *crawling strategy* already provides a *topic focus*.

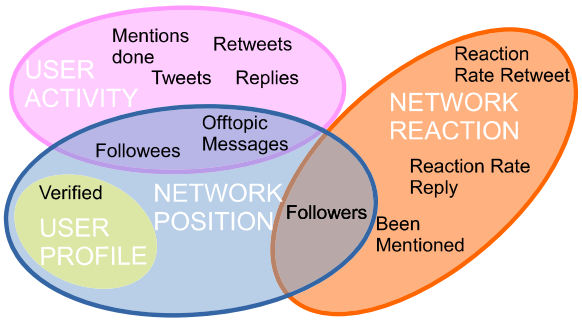

Fig. 1. User feature classification

As most features exhibit *significant skew* and *domain variation*, we *normalize* each dataset using *logarithmic transformation*, followed by a *Min-Max normalization*. This way, we capture the *relative distributions* and *feature drift*.

To *identify* the *structure* and *(sub)-groups* among the user data, we evaluated a broad range of *unsupervised learning* approaches based on centroids (e.g., K-Means), density (like DBScan) and probability distribution (e.g., EM). *Hierarchical clustering* with geometric linkage (Ward) was the best fit: a) it can *capture complex, irregular shapes* without requiring a fixed number of clusters and b) the *hierarchy* serves as an (yet unlabeled) *classification tree*.

To overcome the *limited scalability* ($O(n^2)$ scaling for both CPU and memory, single-threaded execution) and the lack of probabilistic approaches for hierarchical clustering, we chose a *sampling/ensemble-based approach*: Clustering a small number of samples allows us to *quickly discover* the *structure* while *drastically reducing* the *cost* compared to clustering the whole dataset. By incrementally drawing more samples, we see a *linear cost increase* (while allowing parallel execution) and provide a *faithful representation* of the data. With *overlapping cluster* results from several samples for the same user, we can choose to assign to a *majority role* or the probability for specific roles. Likewise, we can determine how *stable* the *role recognition* is. The number of samples becomes a tuneable, trading off the effort of computation (and labeling) with the *coverage of users* and the *amount of support* for the roles on each user or role. If all users need to be covered, we can also minimize the overlap or apply metric-based assignments.

5 User Role Identification

While the cluster structure identifies candidates, it does not provide the actual roles. We now describe the *assignment* and *transfer* of the roles to new datasets.

5.1 Manual Role Assignment

Considering the lack of precise, commonly established roles (see Sect. 2), we apply *complementary methods* to derive *candidates*: 1) *internal cluster quality metrics*, as well as *statistical metrics* and heuristics like the *elbow method* provide

indications on the (approximate) number of clusters (typically around 20–30), but no meaning. 2) analyzing the *dendrogram* and *salient features* of *clusters splits* allows us to match clusters to *fine-grained* user roles. 3) *dimensionality reduction* such as *PCA* or *LDA* simplifies visual inspection, helps with *correlating user roles* across datasets and exposes the *drift/evolution*.

Certain heuristics support this process: 1) Role generalizations tend to occur further up the hierarchy, creating *subtrees* that could be *refined*. 2) some very *distinct roles* tend to show up in most datasets, providing an *"anchor"* for labeling. The process is stopped once we do not gain well-discernible clusters. In some cases, it may be useful to coarsen the roles or combine clusters into a single role.

We match these aspects to well-known role descriptions when possible (e.g., *Star*), but also observed stable, recurring clusters that did not align well with the known role descriptions. In our dataset, among of them were *Rising Stars* (gaining followers via activity, receiving significant retweet reactions, not yet star or semi-stars) and *Loners* (low activity and weak connections). Figure 3 provides an overview in terms of groups, features and frequency.

The same process can be applied *across datasets*, so that we can *track* the user roles and *evaluate concept shifts* and *drifts*, such as their *frequency/probability* of roles or their feature distributions. Typically, we observed around 10–15 class candidates that did show up in varying frequency, sometime disappearing entirely.

While we capture *domain expert knowledge* and produce *well-described role clusters*, we suffer from *limited scalability* and *reproducibility*.

5.2 Classifier

To overcome these issues, we trained an *n-class classifier* that on role samples of one dataset and determined the role labels on clusters in other datasets. While classifying individual users yielded lower runtimes (no clustering step needed), we observed lower classification quality due to the inherent "noise" shown by individual users. Clusters were represented by *aggregate feature values* of all members where *means* tended to provide better *separation* than median, while *pooled Cohens d* seemed to capture more *temporal evolution* than "pure" means. We took samples that showed the best cluster separation to minimize the noise. As initial experiments showed, the original number of dimensions in the data yielded better quality than reduced dimensionality.

As our (clustered) data sets are relatively small and skewed, yet we seek to express a large number of classes, we see little support for some classes. This more or less rules out deep learning. Instead, methods based on *ensembles of decision trees* (Gradient Boosted Decision Trees (GBM), Extremely Randomized Trees (ET)), *multi-class support-vector machines* (SVM) or *k-nearest-neighbor* (kNN) turned out to be most suitable. We utilized the Python implementations of scikit-learn for ET, SVM and kNN as well as XGBoost[1] for GBM.

The setup to build *training sets* utilized *repeated stratified cross validation* with three splits (leave-one out, due to the small amount of data) and three

[1] https://xgboost.readthedocs.io/en/latest/.

repetitions (with different permutations to cater for possibly missing groups). We used *F1-macro* as a metric to *compensate* for *class imbalance* and prevent focus on either precision or recall and applied grid search to tune parameters. All classifiers learn and generalize well, leading to 94–95% score in validation and training set with no obviously stronger or weaker candidates.

When *transferring* the *classification* to new datasets, we compensated for mislabelings by *varying training* and *prediction data* (e.g. cluster number) or choosing *more suitable training sets*. Explicitly including drift models and relevance feedback from the user remain future work.

6 Evaluation

After introducing our concepts, we provide an *evaluation* on *diverse data* from Twitter. We address the *three questions* outlined in Sect. 3, not only on the *technical aspects* but also on the *empirical observations*.

6.1 Datasets and Preparation

While our long-term goal is to *recognize* user roles over a variety of social media, we focused our *initial analysis* on *well-defined data* with a large number of users. As in most of the related work, we relied on Twitter, as it is one for the few social media services which fulfills these criteria.

In order to *transfer user role detection knowledge*, we are looking at several classes (Table 1): *major sports* tend to be *repetitive* and *predictable* with a very large number of messages and users, covering *significant periods of time*. Different types of sports provide a (albeit limited) thematic variance. These datasets are complemented by those of two *major disasters* which also tend to have a strong, yet very different *topic focus* and different *interaction patterns*.

Table 1. Overview on data sets.

Dataset	Messages	Users	Time period	Category
Olympic Games 2012	13.68M	2.27M	August 2012	Sport event
Olympic Games 2014	14.58M	1.96M	February 2014	Sport event
Olympic Games 2016	38.05M	4.76M	July/August 2016	Sport event
FIFA World cup 2014	109.00M	10.40M	June/July 2014	Sport event
2015 paris attacks	6.77M	0.74M	November 2015	Tragic incidence
NFL Superbowl LIV 2020	8.89M	0.89M	2. March 2020	Sport event
2016 Berlin truck attack	0.66M	0.15M	19. December 2016	Tragic incidence

Our datasets had each been *recorded* using the *Twitter Streams* and *Search API* using commonly proposed *hashtags*. We only considered users that were *active* at least twice to cater for aggregate metrics. Generally speaking, the *relative feature distributions* after *normalization* showed only minor changes from 2012 until today: The *verified* status is more prevalent. Overall activity increased moderately, while users tend to move into *"reactive"* behavior of *forwarding*.

6.2 Initial Dataset: 2012 Olympics

The first step focuses on a *single dataset* with *uniform feature usage* and *role stability* due to the relatively short period of time. The analyses provide insights to which extent such as clustering, user roles detection and automated labeling are feasible, as stated in Q1.

The *benefits* of sampling are shown in Table 2. The numbers were generated using `scipy.cluster` on an 8-core partition of an AMD Epyc 7401. A small dataset like Berlin 2015 may still be clustered completely, yet a sample can be generated almost instantly. For large datasets, *full clustering* is clearly impossible, while *samples* fit well. The cost can almost entirely be attributed to creating the *linkage matrix*, so *refinement/exploration steps* are interactive in all variants. Combining them (Fig. 2) shows how *coverage* and *certainty* of *roles* (number of role assignments per user) improve, while *cost scales linearly*. The decreasing "no majority" part gives insights on user that are not well identified - which is dataset-dependent, but often includes Spammer, Loners, etc.

Table 2. Runtime and memory of samples, full data sets and approximated (*).

	Oly12 5%	Oly12 10%	Oly12 100%	Berlin16 10%	Berlin16 100%
Runtime	19 min	136 min	226 h*	10 s	38 min
Memory	94 GB	375 GB	375 TB*	1.2 GB	184 GB

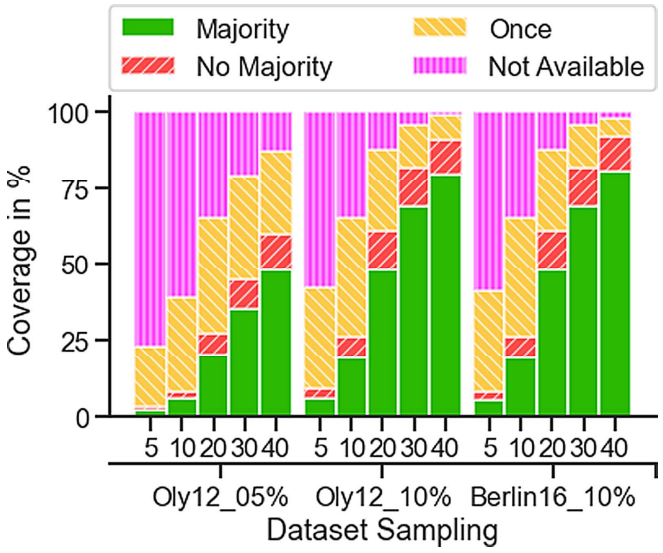

Fig. 2. Coverage and overlap of sampling.

After clustering, we manually labeled clusterings of the samples to get a ground truth as training and test data which can be done incrementally.

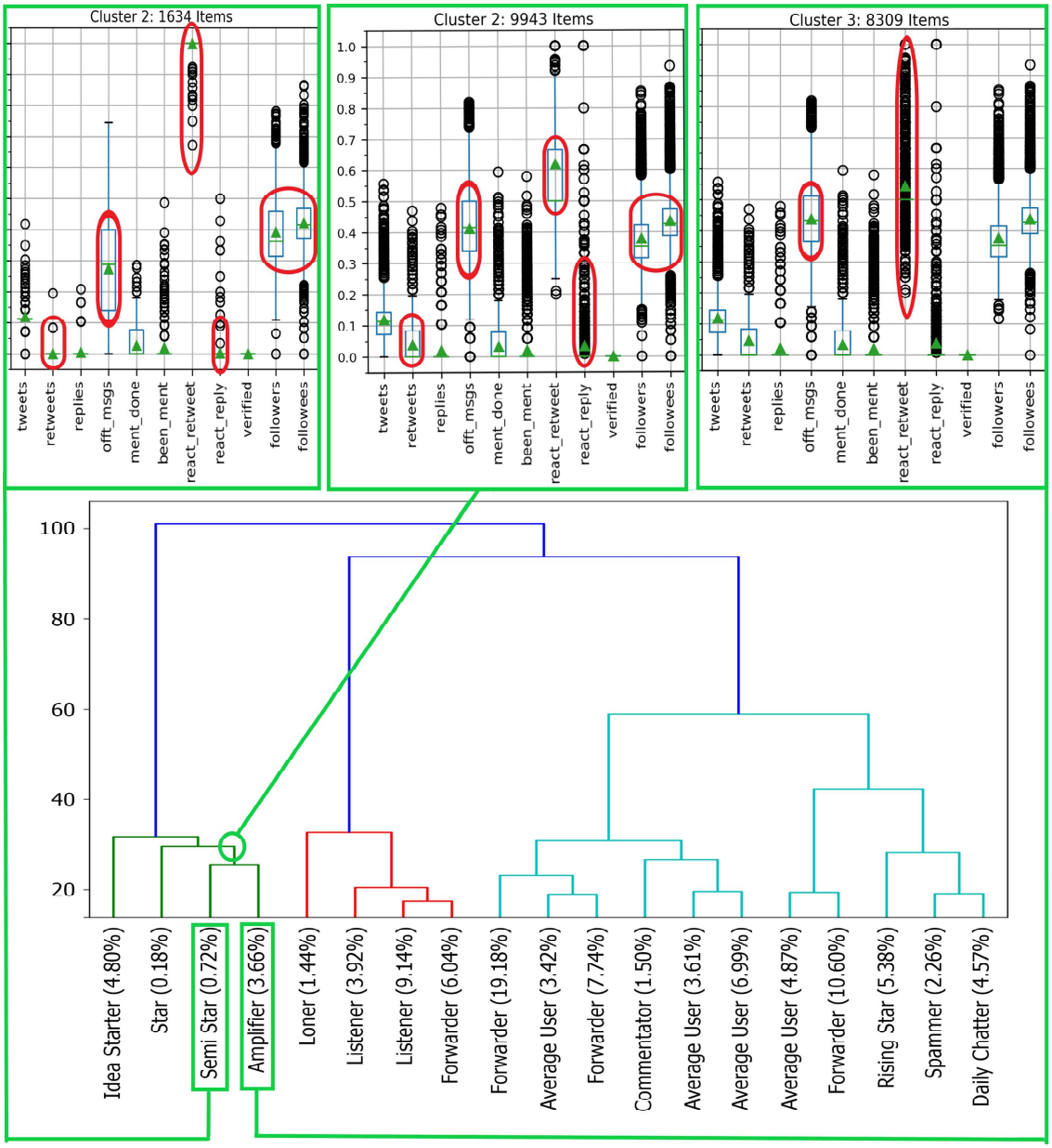

Fig. 3. Olympics 2012 10% sample Dendrogram with salient features.

In particular after applying PCA (see Fig. 4), we can *identify* a *number of well-separated clusters*. Despite some minor variances, the *dendrograms* (see Fig. 3) and the *features* over the set of samples exhibit a *high degree of similarity* that has become a part of our overall classification (Table 3, leftmost column): there are between 3 and 5 *subtrees* representing *major groups*: The first (green) shows users that are able to *trigger strong reactions* (*retweets, replies, being mentioned*), the second (red) shows *passive user* with fairly *weak positions in the network*, while the group(s) in between show various degree of *moderate activity and impact*. Even further down the tree, (as shown on the boxplots), we see a strong motivation for *fine-grained roles*. While the cluster sizes are often small, there are *salient feature differences* (which we can detect using *statistical*

tests like *cohens d*) that explain the existence and semantics of this group. In the example one can see how *Semi Stars* and *Amplifiers* split on (among others) *retweet activities* and *reactions*. Overall, we determined 12 roles in the Olympics 2012 data set that are described in Table 3. Some *characteristics* are shown in the second column, in particular stronger *deviations* from the *average* as well as (broadly) similar user groups.

Table 3. User roles and their characterization: ↓/↑ denote feature deviation from whole dataset, ≈ closeness to other roles, ↘ / ↔ / ↗ changes over time

	Role	Characteristics	Freq./Trend
action triggering	Star	followers > followees, verified, ↓ activity, ↑ mentioned	0.2–0.8 ↗
	Semi Star	≈ Stars, ↓ followers, mentioned, ↑ react. (re)tweet, retweets, replies	0.2–1.4 ↘
	Idea Starter	≈Semi Star, ↓followers, ↑reactions	1–4 ↔
	Amplifier	≈ Idea Starters, Semi Stars, ↑ followers, followees	0.5–5 ↘
intermediates	Rising Star	≈ Semi Star, Idea Starter, Amplifier, ↑ followers, (re)tweets, replies	1.5–5.5 ↘
	Daily Chatter	≈ Average User, Spammer, ↓(re)tweets, offtopic	5–15 ↔
	Commentator	↑ replies, offtopic, reations	0.3–2 ↘
	Spammer	↑ (re)tweets, replies, offtopic ↓ followers, followees, reactions	1–7 ↔
passive	Average User	offtopic > tweets, retweets	8–30 ↓
	Forwarder	retweets > tweets, ↑ offtopic, followers, followees. ↓ reactions	25–65 ↑
	Listener	↓ (re)tweets, reactions	6–20 ↗
	Loner	↓↓ tweets, offtopic, followers	0–1.5 ↘

We evaluated the clustered and labeled samples (in total 507 clusters) with the *classifiers* mentioned in Sect. 5 and achieved nearly perfect results, as the leftmost data points in Fig. 5 show. There are only very few misclassifications between *Average User*, *Daily Chatter* and *Listener*, respectively - which are also *close in feature space* and *low in certainty*. The strong variance in the feature distribution (Fig. 3) also shows why training and classifying individual users instead of clusters yields inferior results.

Overall, the results show that both *clustering* and *classification* work well. *Expert knowledge* is needed to interpret the *dendrogram* and *assign roles*, but already within a single dataset, the *knowledge* can be transferred to additional samples and their clusters.

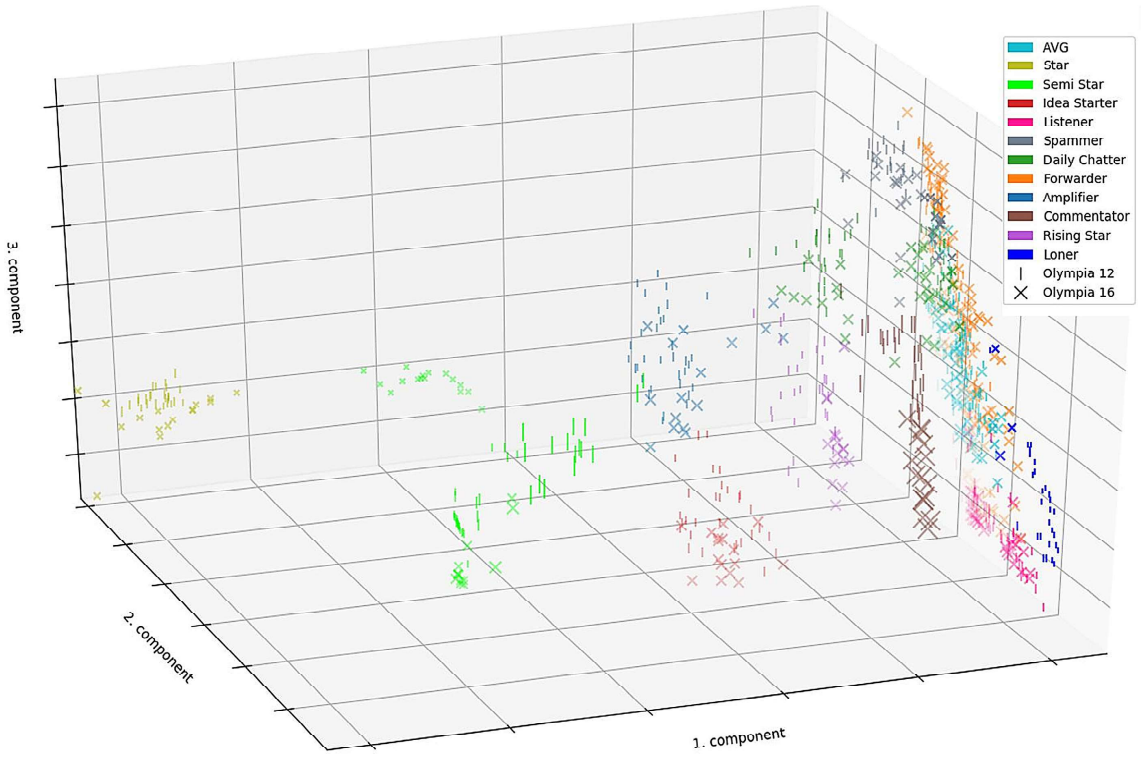

Fig. 4. PCA of clustered samples from Olympics 2012 vs 2016.

6.3 Multiple Individual Datasets

We now analyze several *datasets individually* to understand if the approach is more *widely applicable* (see Q2). Furthermore, this will show if the *same* or *similar roles* are *present* on how they *evolve* in *frequency* and *features*.

The 12 user roles identified on the *Olympics 2012* dataset are also *present* and *well-separated* in the other datasets, though -as the rightmost column of Table 3 shows- the frequency (in percent) varies over datasets (and over time):

In the *Olympics 2014* (278 clusters) and *FIFA World Cup 2014* (193 clusters) data very few changes can be observed: *Average User* and especially *Loner* occur less frequently, while *Forwarder* and *Listener* occur more frequently.

Significant changes occur in *Olympics 2016* (355 clusters). The PCA in Fig. 4 shows a *salient concept drift*, in particular for *Semi Stars* that tends to also cover a space much closer to *Stars*, as the *"verified"* status was more freely distributed by Twitter. The trends on the *Average Users* /*Loner* and *Forwarders* strengthens, and continues for the *Superbowl 2020* (345 clusters), which is otherwise (despite the difference in sports and time) similar to *Olympics 2016*.

The *2015 Paris Attacks* (160 clusters) covers a very different topic and distinct interactions (fewer offtopic messages, more retweets). Some user roles are not present (*Commentator*, *Loner*), yet most of the overall trends match the picture of the "sports events": *forwarding* instead of *content creation* becomes more dominant. In turn, *"influencer"* roles become pronounced, to the point where the *Semi Star* may have to split into two separate sub-roles.

The only exception where we could not apply our methodology was the random *Sample Stream*, as features based on topics lose their usefulness.

Overall, we see the *same features*, leading to *consistently recognizable user roles* that we can *correlate across data sets* to *trace shifting distributions*. Yet, at this step, labeling samples of each dataset manually is a limiting factor.

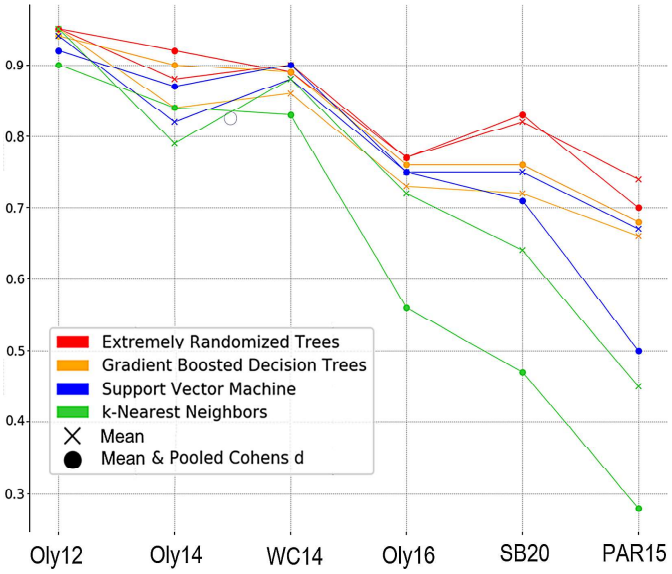

Fig. 5. Information retrieval measure F1 for classifiers.

6.4 Applying Models on New Datasets

In the third step, we *classify new datasets* with the *models* gathered from *reference data* to assess the *quality* and *effort* involved, answering Q3. We further study the *impact of variation* and *drift* to understand the limitations.

Figure 5 shows the *weighted F1 scores* when classifying the dataset based on the *Olympics 2012* as the reference, as it provides the longest prediction period. Overall, one can see a *gradual degradation* over time on the sport events, as the classification methods do not explicitly capture for the drifts, but still can generalize the roles over time. Still, the best methods achieve a 0.85 F1 score for "late" sport events. The *2015 Paris Attacks* data set sees the largest degradation, showing topic and interaction differences have a more profound impact than time. When comparing all these results to the slightly worse "macro" values, one can see that *small groups* are *captured well*, while larger clusters tend to be somewhat "blurry".

kNN and SVC keep up well for *short time intervals*, but tend to lose ground on longer distances. ET holds a small edge over GBM, while the latter stays still *competitive* and incurs much lower runtime cost. Both benefit from *enriching* the datasets with the *pooled Cohens d values*.

Roles that were either not well separated in the *Olympics 2012* data or drifted significantly are most affected. Yet, these *misclassifications* often leads to *adjacent roles*, e.g., *Average Users* as *Listener* and *Forwarder*, thus the F1-scores actually understate the result quality.

We added the data set of the *2016 Berlin Truck Attack* (Christmas market) that was not evaluated in the previous stages and provides *topic similarity* to *2015 Paris Attacks*, while being close to the *Olympics 2016* in time. This data set provides a good opportunity to assess the impact of *different training sets*: in addition to baseline of the *Olympics 2012* and close sets (*Olympics 2016*, *2015 Paris Attacks*) and *Superbowl 2020* as a small, recent dataset, we tested two combinations: As Table 4 shows, these combined data sets provide the best results, *matching manual classification* or producing *misclassifications* to *close roles*. *2015 Paris Attacks* by itself seems to be too small to provide a sufficiently general model, but is able to *boost* the *full time range model*.

The experiments show that a *transfer* of *labeling knowledge* is effective with certain limitations: *large topic differences* or *very long time differentials* diminish the *usefulness*, yet a *good choice* of *reference data* can mitigate this effect.

Table 4. Classification of Berlin 2016 data set. Comb1: Oly12 & SB 20, Comb2: Oly12 & SB 20 & Par15.

Classifier	Oly12	Oly16	Par15	SB 20	Comb1	Comb2
XGB	0.58	0.59	0.51	0.70	0.78	0.92
ET	0.74	0.63	0.56	0.73	0.77	0.82

7 Conclusion and Future Work

In this paper we proposed a method to determine *user roles* in *large-scale social media data*. It combines *unsupervised learning* (i.e., hierarchical clustering) to *discover* and *explain* such roles over a *wide range of features* with *supervised learning* to generalize the *knowledge* from *manually labeled* smaller data.

Our analysis on a range of large data sets from Twitter shows that *well-separated roles* can consistently be *recognized* and *transferred*. The labeling achieves *high accuracy* not only within the same data set, but also on *new data sets* from *different event types* and/or *years* apart. *Scalability*, *incremental evaluation* and *probabilistic assignment* are achieved by *combining samples*.

For *future work*, we see a number of interesting directions: As the *quality* of *classification* begins to deteriorate over *longer time frames*, we plan to address *evolution*, considering both *temporal models* (for long-term studies of snapshots) and *stream clustering* (for short-term, continuous analyses). They may also pave the way for *longitudinal studies* of user groups and user mobility among groups. Likewise, adapting our model to cope with *topically non-related* or even *topically*

unconstrained data sets poses a new set of challenges. Initial experiments show that the method should generally work, but *significant challenges* remain. In either case, testing our method on a *wider range of data sets* from Twitter or *other social networks* would be highly interesting.

References

1. Chatzakou, D., Kourtellis, N., Blackburn, J., Cristofaro, E.D., Stringhini, G., Vakali, A.: Mean birds: detecting aggression and bullying on Twitter. In: WebSci (2017)
2. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S.: Who is tweeting on Twitter: human, bot, or cyborg? In: ACSAC (2010)
3. Du, F., Liu, Y., Liu, X., Sun, J., Jiang, Y.: User role analysis in online social networks based on Dirichlet process mixture models. In: 2016 International Conference on Advanced Cloud and Big Data (CBD), pp. 172–177 (2016)
4. Edgar, R., Alexandre, P.F., Caladoa, P., Sofia-Pinto, H.: User profiling on Twitter. Semant. Web J. (2011)
5. Fu, C.: Tracking user-role evolution via topic modeling in community question answering. Inf. Process. Manage. **56**(6), 102075 (2019)
6. Java, A., Song, X., Finin, T., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: WebKDD/SNA-KDD (2007)
7. Kao, H.T., Yan, S., Huang, D., Bartley, N., Hosseinmardi, H., Ferrara, E.: Understanding cyberbullying on instagram and Ask.Fm via social role detection. In: WWW 2019 Companion (2019)
8. Lazaridou, E., Ntalla, A., Novak, J.: Behavioural role analysis for multi-faceted communication campaigns in Twitter. In: WebSci (2016)
9. Li, H., et al.: Bimodal distribution and co-bursting in review spam detection. In: WWW (2017)
10. Tinati, R., Carr, L., Hall, W., Bentwood, J.: Identifying communicator roles in Twitter. In: WWW 2012, rel MSND Workshop (2012)
11. Varol, O., Ferrara, E., Ogan, C.L., Menczer, F., Flammini, A.: Evolution of online user behavior during a social upheaval. In: WebSci (2014)