Popularity Metrics' Normalization for Social Media Entities

Hiba Sebei[1,2], Mohamed Ali Hadj Taieb[3,2] and Mohamed Ben Aouicha[3,2]
[1]*Computer Science Department, Faculty of Economics and Mangement, Sfax, Tunisia*
[2]*Multimedia, InfoRmation Systems and Advanced Computing Laboratory, Sfax, Tunisia*
[3]*Computer Science Department, Faculty of Sciences, Sfax, Tunisa*

Keywords: Popularity, Social Networks, Social Entity, Social Features, SPI.

Abstract: With the spread of online social media websites, a huge amount of online content is continuously provided. However, some contents gain an important attention from users while other contents are completely ignored. This highlights the analysis of popularity relative to different social content. The popularity is expressed through measures and features that act as factors expressing and influencing the popularity. Those features vary from an online social media website to another as it depends on the type of social entity. This paper tries to create a normalized view of the popularity metrics independent of the online social media and in relation with specific social entities that are user and media content (i.e. text, image, and video). We propose a Service Provider Interface (SPI) as a contract between users. The SPI offers a variety of interfaces for implementing services related to the quantification of social entities popularity independently of the online social media they belong to.

1 INTRODUCTION

Over the last decade, online social media websites have seen an exponential growth of the number of active users as 313 million monthly active users on Twitter[1], over 1 billion users on YouTube[2] and about 1.94 billion monthly active users on Facebook[3]. This supports the explosion of the amount of user-generated data on those websites. In fact, the statistics reveal that about 52 million photos shared every day on Instagram[4], 300 hours of videos uploaded in every minute on YouTube[5] and about 58 million tweets every day on Twitter[6] .

This flood of data did not get the same attention from users, as mentioned by (Lerman and Hogg, 2010) among 1600 new stories submitted on Digg[7] only a handful of them gather thousands of votes while others are completely ignored by users. This encourages the emergence of the notion of popularity related to each social media content entity as video, photo, and text. Where, the popularity represents the corresponding amount of attention from users to the content (Quan et al. 2012; Jiang et al. 2014).

Studying the popularity of social entities is a beneficial task for both social media data consumers and producers. Most efforts made on the popularity of social entities focus on the analysis of popularity evolution and the prediction of popularity that help to avoid the information overload by introducing for users the most popular content as well as giving the opportunity for companies to boost their business strategy.

Through the study of social entities' popularity researchers try to find responses to some questions as how we can boost social items popularity? Will the studied item be popular or not? If, yes how much the item will be popular in near or long time future? Can the popularity of an item be quantified before its creation?

To study the popularity of a social entity, researchers have shared three requirements: popularity measures, popularity features and methods. However, for a specific type of social entity the popularity metrics vary from an online social media websites to another as it corresponds to

[1] https://about.twitter.com/company
[2] https://www.youtube.com/yt/press/en-GB/statistics.html
[3] https://newsroom.fb.com/company-info/
[4] http://www.statisticbrain.com/instagram-company-statistics/
[5] http://www.statisticbrain.com/youtube-statistics/
[6] http://www.statisticbrain.com/twitter-statistics
[7] http://digg.com

Sebei, H., Hadj Taieb, M. and Ben Aouicha, M.
Popularity Metrics' Normalization for Social Media Entities.
DOI: 10.5220/0006693505250535
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 525-535
ISBN: 978-989-758-298-1

likes on Facebook, diggs on Digg, views on YouTube and even in the same website popularity can be measured in different manners as it can correspond to the number of views or can combine both the number of views and the number of comments.

So, we aim through this paper to normalize the features expressing the popularity of social media entities independently of the social media instance to which it belongs. In addition, we propose through this paper a service provided interface offering the services that can be implemented via the APIs of social networks for gathering the existing features to quantify the popularity degree independently of the social media websites. Figure 1 illustrates the problem and the aim of this study.

The rest of this paper is structured as follows: Section 2 introduces the notion of popularity in relation to different social entities (text, video, photo and user) as well as the related terminology. Section 3 categorizes the studies established according to the social entities and presents the metrics used to define popularity. In addition, it highlights the variety of metrics used through different studied social media websites (Facebook, Twitter, Flickr, YouTube and Digg), therefore our proposal to normalize these metrics. The normalization is treated in section 4 based on the analysis of social entities' popularity metrics introduced by the already established researches. Moreover, this normalization is presented in a hierarchical way to show the different categories of popularity metrics. Section 4 introduces, also, the materialization of the proposed normalization under an implemented Service Provider Interface (SPI). The final section is devoted to presenting our conclusions and recommendations for future research.

2 RELATED WORKS: SOCIAL MEDIA POPULARITY

In this section, we introduce the notion of popularity and its terminology related to social entities that structure the content generated in different online social media websites.

2.1 Popularity Notion

Several efforts focus on studying popularity related to social entities. Some of them as (Figueiredo 2013; Li et al. 2016; Hong et al. 2011) are motivated by the information overload coming from online social media data, so they try to predict the popularity of social entities in order to help users receive the important events and digital content. While others as (Khosla et al. 2014; Jiang et al. 2014; Chatzopoulou et al. 2010) are motivated by the act that being popular on social media becomes essential for companies and even for people, so they try to understand and figure out the properties of social items that make an item popular than other in order to help people boosting the popularity of their content. Also, we find studies as (Ma et al. 2013) focus on the improvement of marketing strategies

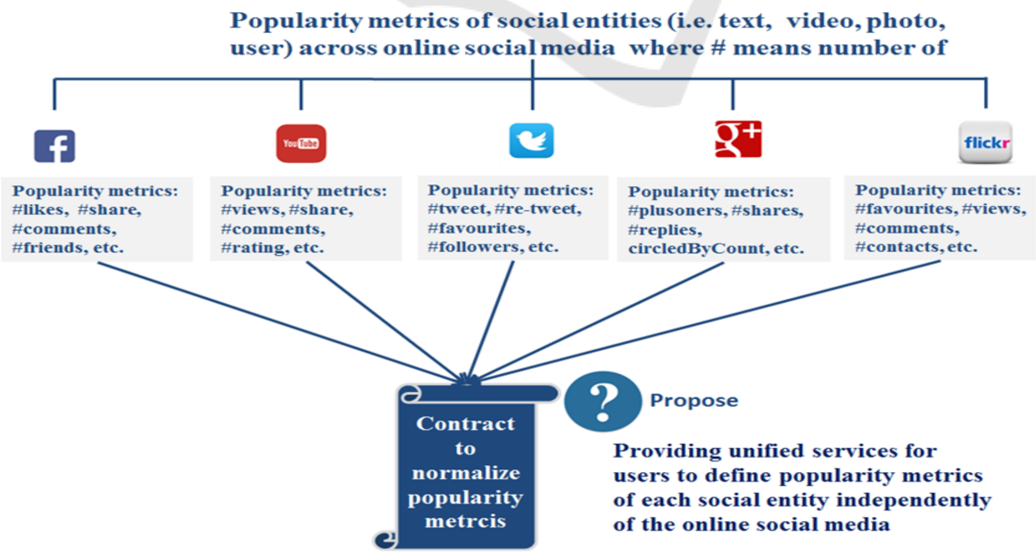

Figure 1: Illustration of the problematic corresponding to the variety of social entities' popularity metrics across different online social media websites and the necessity of popularity metrics normalization.

and the development of diffusion strategies by predicting real-world outcomes as the case of movie revenue estimation (Jiang et al. 2014), predicting popular items is also useful for websites owners as mentioned by (Quan et al. 2012) in order to provide the accurate resources as popular content leads to the traffic increase that should be handled before getting a technical problem.

There are a variety of studies that focus on the popularity of social entities. Li et al. (2016) divided these studies into two main categories: the first includes those focusing on the popularity prediction in microblogs as Twitter and the second one is devoted to the prediction of popularity in media sharing-websites as YouTube. For the first category popularity is related to textual entities as a tweet on Twitter and for the second category popularity is relative to the media content as videos and photos.

The state of the art shows that the researches works related to the popularity social media content cover several social entities such as the textual content as (Hong et al. 2011; Ma et al. 2013; Gao et al. 2014; Lakkaraju and Ajmera 2011), videos as (Jiang et al. 2014; Chatzopoulou et al. 2010), photos as the case of (Khosla et al. 2014; McParlane et al. 2014) and finally there are some few works that discuss the popularity of users in their social media websites as (Jiang et al. 2014; Couronné et al. 2010).

As previously mentioned popularity identifies the amount of attention from user to the content. So, for analysing and predicting the popularity of each social media entity among the already mentioned ones (i.e. text, video, photo, and user) it is required to identify the metrics of popularity, the features and establish a link between those metrics using some algorithms and methods for providing a quantification.

2.2 Terminology

Popularity measures: are metrics to define the popularity and varied from a study to another (Khosla et al. 2014).

Popularity features: they present different factors related to the target social entity and that can affect its popularity (Khosla et al. 2014; Hong et al. 2011) as measuring social entities popularity is a difficult task due the existence of variety of factors that influence the quantification of popularity (Cappallo et al. 2015).

Methods: they are the process used to figure out the correlation between popularity measures and features as factors that influence social media entities popularity. Li et al. (2016) focus on popularity prediction task and categorized the approaches into three main groups: regression-based approach, classification-based approach and model-based approach.

2.3 Social Entities

During, the first age of social media, the content generated by users focus on the text (blogs) then by the integration of the web 2.0 technology as a platform for building social media websites the user-generated content takes additional forms such as video, photo and audio. Each one of these types is considered as a social entity. Also, we consider the user presented by profile as a social entity. As, several researchers focus on studying the popularity relative to each type of user-generated data, in the next section, we classify the related works according to the studied social entities (i.e. text, video, photo and user).

3 SOCIAL ENTITIES: POPULARITY QUANTIIFICATION

The state of the art shows that the identification of popularity measures and features for a specific social entity varies in the same online social media websites and across different websites. These points are presented and discussed in following sections according to each social entity.

3.1 Text

Among the studies focusing on popularity analysis of Twitter messages as textual entities, Hong et al. (2011) define the popularity measure as the number of re-tweets related to the textual entity and they take into consideration the message content, temporal information features, metadata of messages and users, as well as structural properties of the users' social graph as features that influence the popularity of messages on Twitter. Some other studies focus on specific textual entities as a hashtag on twitter messages, Ma et al. (2013) predict the popularity of a hashtag by presenting the number of users who post at least one tweet containing the hashtag within the given time period as the popularity measure. Then they specify two main categories of popularity features: content features and contextual features. Where the content features refer lexical data derived from the hashtag and from the tweet containing the hashtag (e.g. number of segment words from a hashtag). For the contextual features, they are related to data derived from the social graphs formed by Twitter users (e.g. the number of tweets containing the hashtag). Lerman and Hogg (2010) worked on the news as a textual entity they try to predict the popularity of news and through the Digg website. They express the news popularity as the number of votes a story accumulates on Digg. While Wu and Shen (2015) use the number of re-tweets that the news tweet gathers from users on Twitter.

3.2 Image

McParlane et al. (2014) studied the popularity of image in Flickr. They define several popularity measures as the number of views related to the image and he considers three main features image's context (e.g. time, day, size, flash, orientation), visual appearance related to information extracted from the image's pixel (e.g. color, faces, etc.) and user context (e.g. gender, account, contacts, etc.). While Khosla et al. (2014) studies popularity of photos on Flickr by considering the number of views as a measure for popularity and combines both image content features (e.g. color, objects in the image, vision, etc.) and social context features (e.g. user's contact, users' groups, mean view, title, description, etc.) as features for studying image popularity. Gelli et al. (2015) studied the popularity prediction of images based on Flickr photos by considering the number of views on Flickr as a popularity metric including three main features: user features (i.e. metadata related to the author of the image), visual features (e.g. color) , and context features (i.e. tags and description related to the image).

3.3 Video

Several related studies focus on studying popularity of YouTube videos as (Chatzopoulou et al. 2010) that defines the popularity based on the number of views and considers the number of comments, ratings and favorites as features to understand the evolution of YouTube video popularity. Figueiredo (2013) considers the number of views as a popularity measure and classifies the features in three main classes features the first class is related to video content (e.g. video category, upload date, etc.), the second class refers to link features as (e.g. referrer first date and referrer number of views) and the third class refers to popularity features that are measured during a defined period of time (e.g. number of views, number of comments, number of favorites, etc.). Jiang et al. (2014) also exploit the number of views as a measure of popularity but they define different popularity features to study viral YouTube videos as the video metadata (e.g. id, title, text description, category, number of raters, number of likes, number of dislikes), the user metadata who uploaded the video (e.g. user ID, name, profile view count, etc.), the historic of view (e.g. comments, likes and dislikes), the number of inlinks in other social media, and the comments related to the video. Trzcinski and Rokita (2017) focused their research on popularity prediction of videos. They exploit two datasets: one is from YouTube and the second is from Facebook. For YouTube video, popularity metrics are expressed via the number of views, comments, favorites and ratings while for Facebook video, popularity metrics correspond to the number of shares, likes and comments.

3.4 User

Couronné et al. (2010) studied the popularity evolution of online social media user in MySpace which is considered as an online social media. Two popularity measures are identified: the audience of the contents and the user's authority. The first one identifies the figurenumber of visits to the artist's page while the second one defines the number of people recommending the artist by linking to him. The author takes into account two features: music features (e.g. the number the visits of the profile, the number of comments visitors have left on the profile, etc.) and the search variables that define the number of Twitter post containing the artist name in the last month, the number of results of the Yahoo! search engine when searching the artist's name. Zafarani and Liu (2016) discuss the variation of user popularity across sites as individual join multiple sites and quantifies user's popularity based on his number of friends.

Table 1 categorizes the popularity related works for each type of social entity. In addition, it summarizes the features and metrics used in each study.

3.5 Discussion

This study leads to two main results: firstly, then lack of specific metrics to express popularity and secondly, the popularity metrics are expressed differently from a social media to another.

- Lack of specific metrics to express popularity metrics: For a specific social entity (i.e. text, video, photo, and user) a variety of metrics are used by the researcher to express popularity. The variety of these metrics inside the same social entity type is reflected in the same online social media website. As to study popularity of image on Flickr McParlane et al. (2014) defines three main set of features image's context, visual features and user context. While, Khosla et al. (2014) defined other sets image content and social context features. It is worth to mention that despite the difference in the nomination of the sets of metrics between the two works there is an overlap between the sets as the user context set considered by (McParlane et al. 2014) which holds user metadata as well as the social context features set considered by (Khosla et al. 2014).

Table 1: Categorization of popularity related works based on the type of social media entities.

Social entity	Popularity measures and features		
	Reference	Measures	Features
Text	(Hong et al. 2011)	The number of re-tweets	Message content, temporal information features, metadata of messages and users, structural properties of the users' social graph
	(Ma et al. 2013)	The number of users who post at least one tweet containing the hashtag	The number of users who post at least one tweet containing the hashtag
	(Lerman and Hogg, 2010)	The number of votes	Story metadata, historic of votes, the list of friends of the top-ranked users
	(Wu and Shen, 2015)	The number of re-tweets	Metrics related to the topology of the re-tweet propagation (e.g. date of creation, number of direct followers receiving update, number of followers viewed the news, etc.)
Image	(McParlane et al. 2014)	The number of views and number of comments	Image's context (e.g. time, day, size, flash, orientation), Visual appearance (e.g. color, faces, etc.) and user context (e.g. gender, account, contacts, etc.)
	(Khosla et al. 2014)	The number of views	Image content features (e.g. color, objects in the image, vision, etc.), Social context features (e.g. user's contact, users' groups, mean view, title, description, etc.)
	(Gelli et al. 2015)	The number of views	User features (i.e. metadata related to the author of the image) and visual features and Context features (i.e. tags and description related to the image).
Video	(Chatzopoulou et al. 2010)	The number of views	The number of comments , ratings and favorites
	(Figueiredo 2013)	The number of views	Video content (e.g. video category, upload date, etc.), Link features (e.g. referrer first date and referrer number of views) and Popularity features (e.g. number of views, number of comments, number of favorites, etc.)
	(Jiang et al. 2014)	The number of views	Video metadata (e.g. id, title, text description, category, number of raters, etc.), user metadata (e.g. user ID, name, profile views count, etc.), historic of view (e.g. Comments, likes and dislikes), the number of in-links and the video comments
	(Trzcinski and Rokita, 2017)	YouTube: the number of views, comments, favorites and ratings Facebook : the number of shares, likes and comments	Visual features (e.g. Video characteristics, color, etc.), Temporal features: refer to the number of views and number of social interactions (e.g. number of shares, likes and comments)
User	(Couronné et al. 2010)	The audience of the contents: number of visits of the artist's page User's Authority: number of people recommending the artist by linking to him.	Music variables (e.g. the number the visits of the profile, the number of comments visitors have left on the profile) Search variables the number of the Twitter post containing the artist name in the last month, the number of results of the Yahoo! search engine when searching the artist's name

- A Variety of popularity metrics across different online social media websites: Hong et al. (2011) considered a textual entity use case. In fact, the popularity of message on Twitter as a textual entity is expressed based on the number of re-tweets while authors in (Lakkaraju and Ajmera, 2011) measure the popularity relative to a post made by a brand page based on the number of comments gathered by the target post. Also, Trzcinski and Rokita (2017) measured the popularity of video in YouTube by considering the number of views, comments, favorites and ratings. For Facebook, they count the number of shares, likes and comments as popularity measures. Khosla et al. (2014) highlight the variety of popularity metrics. Indeed, they cite that for an image the popularity can correspond to "the number of likes on Facebook, the number of pins on Pinterest[8] or the number of diggs on Digg".

Two main questions arise in relation to the variety of parameters used to analyze the relative popularity of a particular social entity: The first, question that arises is how to evaluate the subjective parameters that express the popularity as mentioned by (Cappallo et al. 2015)? The second question is how to break away from the specific parameters of each social media website in order to express the popularity?

In this paper, we are interested in answering the second question by proposing a factorization of the different popularity metrics relative to each type of online social entity independently of the social media website source. This factorization will be expressed via a Service Provider Interface (SPI) offered the concluded services from the study according to the already discussed social entities: text, video, image and user.

4 NORMALIZED POPULARITY METRICS AND THE PROPOSED SERVICE PROVIDER INTERFACE

In this section we propose a normalization of different social entities' popularity metrics based on the state of the art presented in the previous section as well as the different popularity metrics extracted from a number of online social media websites (e.g. Twitter, Facebook, YouTube, Google+ and Flickr) relative to each social entity.

4.1 Normalization of Popularity Metrics across Online Social Media Websites

Based on the study made in the previous section, we consider the popularity related to the social entities: text, video, photo and user. So, we distinguish between two main categories user entity and media entity.

User: The variety of purposes behind using social network websites reveals a variety of self-presentation on those websites. Social networks are used by simple individuals to establish social or business relationships, by organizations and companies to promote a marketing purpose, by a community of individuals to group people with common social or professional interest or by non-physical individuals such as the presentation of an event or a channel. So, different entities exist to identify the user across the network such as profiles, groups, pages and events. It is worth to distinguish between popular user and influence user. A popular user does not imply that he is influential. The difference between popularity and influence is discussed in (Kwak et al. 2010) where authors adapt the number of followers related to a Twitter user as a popularity measure but they prove its inefficiency regarding the quantification of the user's influence.

Describing the user popularity, is treated through three main categories of metrics: user profile metadata that refer to metadata created during the creation of the profile, (e.g. name, gender, age, member duration, etc.), user activities metadata reflect how much the user is active in his network (e.g. number of posts, number of posted media) and profile's connectivity metadata reflects user's relationships in the network (e.g. number of contacts, number of friends, number of followers, etc.).

Media: refers to the different type of online social media user-generated content: text, video and photo.

Presentation: refers to image, video and textual entities. It is worth to mention that the textual entity can refer to a tweet on Twitter, a Facebook post or an activity posted on Google plus. The textual entity can embed media entities.

Normalization: based on the related works, it is clear that the metrics correspond to the popularity of each social entity define two main categories of metrics: metrics related to the content of the target entity and metrics related to the context of the target entity.

[8] https://fr.pinterest.com/pinterestfr/

- The content metrics: correspond to parameters extracted from the content of the target entity (i.e. video, image, text). These metrics are obtained based on advanced techniques as sentiment analysis, clustering and natural language processing (Khosla et al. 2014) applied to textual entities as made by (Ma et al. 2013) who derives lexical parameters from hashtag as content features. The task becomes harder when the content feature is derived from media objects (Khosla et al. 2014; Cappallo et al. 2015), advanced techniques as computer vision and machine learning are used by (Khosla et al. 2014) in order to extract content features from an image. The content features of media items (i.e. image and video) correspond to visual features as colors, objects in images (e.g. people faces) (Khosla et al. 2014; McParlane et al. 2014), also visual sentiment features as mentioned by (Gelli et al. 2015).
- The contextual metrics: they refer to parameters relative to the target social entity; they do not require the use of complicated algorithms and techniques to get them. Actually, these metrics are directly extracted from online social networks as a category of social media (e.g. Facebook, Twitter, Google+, etc.) using the application provided interface (API) offered by those websites. These metrics vary from a social network to another. The parameters, associated with popularity metrics and related to different social entities (e.g. textual entity from Twitter and Google plus, using the Twitter Search API and Google plus REST API respectively, etc.), are extracted.

The results are summarized in Table 2 that presents some instance of social entities from different social

Table 2: Social entities instance and its related popularity metrics across different social media websites.

Social entity	Social entity instance	Social media	Extracted metrics	API
Text	Tweet	Twitter	FavoriteCount, HashtagEntities, id, retweetCount, text, user, CreatedAt, etc.	Twitter Search API[9]
	Activity	Google Plus	Id , Activity author, Activity publishedAT, Activity Title, Activity URL, Activity content, Activity replies, etc.	Google+ API[10]
	Comment	YouTube	AuthorChannelUrl, AuthorName, ViewerRating, LikeCount,, Text, publishedAt	YouTube API[11]
	Post	Facebook	Id, shares, admin_creator, created_time, description, link, message, place, picture, source, etc.	Facebook Graph API[12]
Video	Video	YouTube	ChannelId, description, PublishedAt,, title, Url, ViewCount, CommentCount, DislikeCount, FavoriteCount, etc.	YouTube API
	Embedded video	Twitter	URL, id , sizes (e .g large, medium, etc.), duration_millis, Video formats, video aspect ratios, updated_at, title, etc.	Twitter Search API
	Video	Facebook	ad_breaks, backdated_time, created_time, id, description, from, length, place, source, title.	Facebook Graph API
Photo	Photo	Flickr	Owner (id, name, etc.), title , description, number of comments, tags, URL, number of favorites	Flickr API[13]
	Embedded photo	Twitter	URL, id, sizes (e .g large, medium, etc.)	Twitter Search API
	Photo	Facebook	Id, album, backdated_time, created_time, from, icon, height, link, name, place, etc.	Facebook Graph API
User	Page	Facebook	About, created time, number of likes, number of fans, name, picture, id, and category	Facebook Graph API
	Profile	Twitter	Id, Name, Screenname, createdAT, StatusesCount, Description FavoritesCount FollowersCount, FriendsCount, User Tweets: list of tweets	Twitter Search API

[9] https://developer.twitter.com/en/docs
[10] https://developers.google.com/+/web/api/rest/
[11] https://developers.google.com/youtube/v3/docs
[12] https://developers.facebook.com/docs/graph-api
[13] https://www.flickr.com/services/api/misc.overview.html

media websites and presents the related popularity metrics.

We focus on the contextual metrics in order to normalize them independently of the online social media websites. So, we distinguish between two main categories of contextual metrics: media contextual metrics and media author contextual metrics.

Media Contextual Metrics: refers to the metadata of the target media. It is divided between media metadata and user feedback metadata. Where the media metadata refer to two sets: firstly, a set describes metadata generated by end users during the upload of the media entity and devoted to describe the entity (e.g. a video description, tags, date of the upload, etc.), secondly, a set of metadata generated after the upload of the media (e.g. accumulated comments, related media, etc.). Then, the user feedback metadata refer to metrics resulted from user activities related to the media this metadata can express either a simple feedback from user (i.e. does not require an explicit activity from the user) as the number of views which is counted as soon as the user just visit the media or it can refer to an explicit feedback accumulated after an explicit activity from the user as sharing a media, rating a video, like or dislike a post from the execution of these activities a number of popularity metrics are generated (e.g. number of likes, number of favorites, number of ratings, etc.). The user feedback metrics are also characterized by their dynamics as they evaluate during the time.

Media Author Metrics: several researchers as (Khosla et al. 2014; Quan et al. 2012; Szabo and Huberman 2010) discuss the impact of the connectivity of the user who uploaded the popularity of the target entity. So, they use the metadata related to the author of the media entity.

The author contextual metrics are those defining the user popularity discussed in the previous paragraph and referring to user's profile metadata as the gender of the user that can be extracted directly using the social network API or based on their names on the target social network as the case of (McParlane et al. 2014). It includes, also, the user activities metadata and user connectivity metadata.

Figure 2 defines the media entity popularity metrics in a hierarchical manner in order to present the different factorization levels. In addition, it illustrates also the popularity of the user entity via the media author popularity (the part framed in red).

This hierarchy is materialized by implementing an extensible application that provides to its users a set of unified services allowing the definition of popularity instances related to social entities and independently of online social media websites.

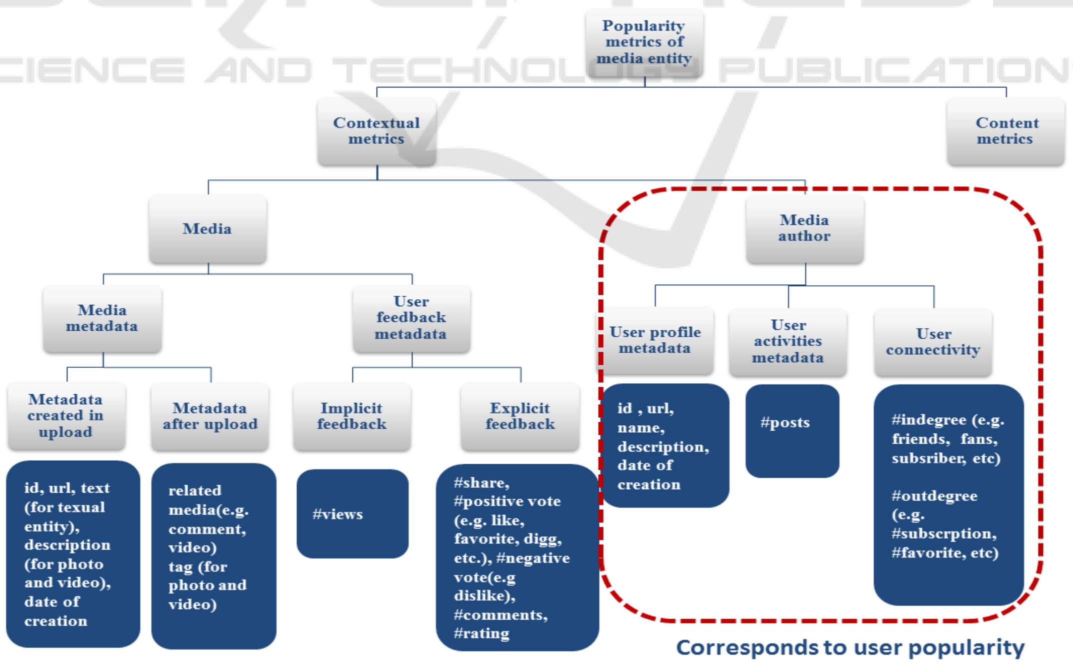

Figure 2: Hierarchical presentation of the media entity popularity metrics with common metrics across online social media websites.

4.2 Proposed Service Provider Interface

Based on the study made in previous sections, we aim to implement the proposed normalization of popularity metrics related to each social entity independently of online social media websites.

In this context, we propose the normalization in the form a Service Provider Interface (SPI). The SPI is considered as a contract between users to define in a unified way the popularity metrics correspond to the different social entities (i.e. text, video, photo and user) independently of the online social media to which they belong. In addition, this SPI allows users to create extensible applications. Because it defines a set of public interfaces and abstract classes that a service defines.

These interfaces are implemented to allow the creation of extensible applications based social media entities popularity. We cite as examples the prediction and detection of online trending topic that aims to define the most trending topic across the online community independently of the social network, the detection of most popular brand sales in online communities and the identification of the most popular users on their networks.

All these applications require the identification of the most popular social items across several social media. So in order to avoid the heterogeneity of metrics across social networks, the SPI gives the opportunity for end-users to define the popularity metrics of each social entity by simply implementing the abstract provided method. The creation of the contract of the social entities' popularity normalization is made through the implementation of an SPI composed of two main interfaces: the media popularity interface and the user popularity interface.

- Media popularity interface: defines the SPI specification of the media popularity service. It includes methods that define the media entity metadata, the media's author metadata and the user feedback metrics given the URL of the social entity.
- User popularity interface: refers to the SPI of the user popularity service. It provides methods to define user' metadata, activities and connectivity that used to study popularity.

Besides, the proposed solution for the normalization provides a set of service provider classes that present the implementation of services

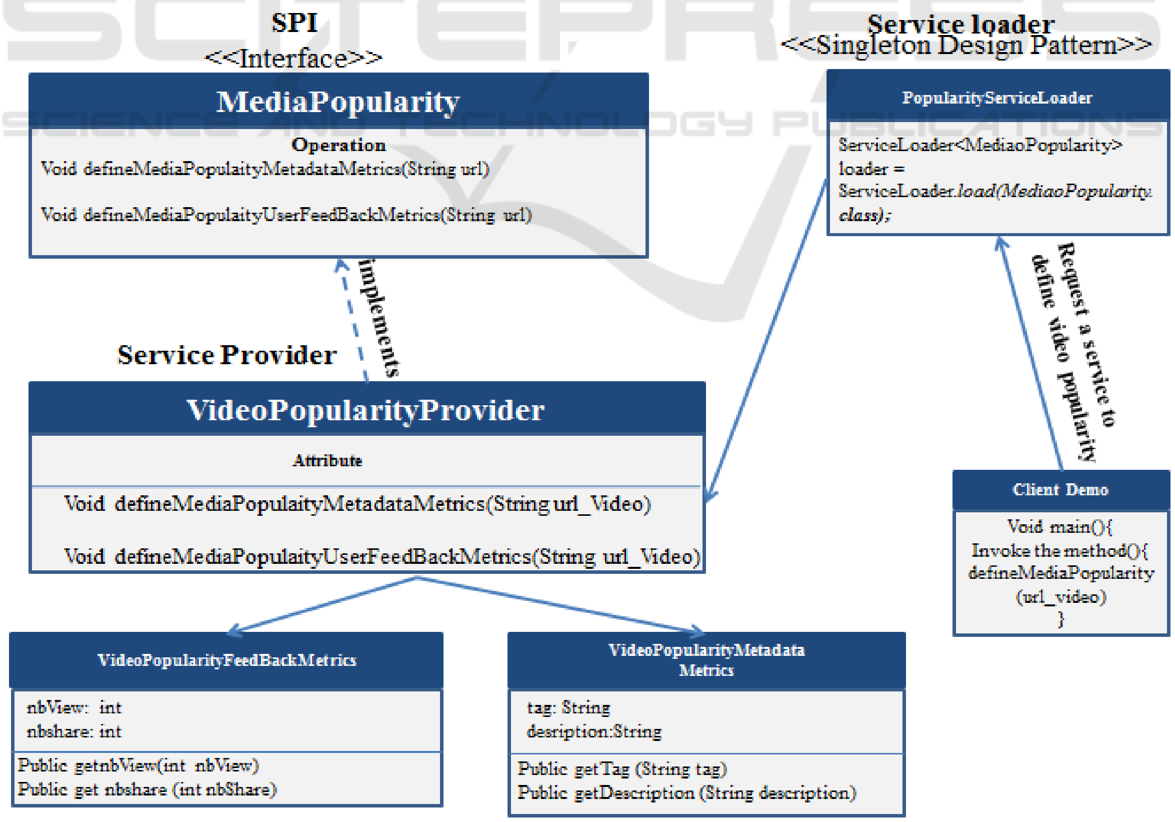

Figure 3: Excerpt from the SPI modeling in relation to Video Popularity.

offered by the media and user popularity interfaces. These services store the social entities URLs and their related popularity information. It is worth also to mention that the proposed solution implements a service loader class introduced by the class PopularityServiceLoader that follows the Singleton design pattern and works as a template for the relationships and interactions between classes and ensuring that only a single instance of a class is ever created.

Figure 3 presents the model of the media popularity SPI implemented for the video popularity provider class and it shows the interaction between the client and the SPI using the popularity service loader class. As they are categorized in the previous hierarchy, the popularity metrics related to each social entity are introduced by a set of classes. The figure also includes two classes related to video popularity metrics which are: VideoPopularityMetadaMetrics and VideoPopularityFeedBackMetrics.

Figure 4 is an excerpt from the whole implemented model. It focuses on the case of video entity but it is worth to mention that the definition of other media entities popularity (i.e. text and photo) implements a user popularity interface previously introduced.

The SPI consumer extends the popularity interfaces and implements its services to instantiate his own popularity according to the application needs and the availability of information. The architecture of the SPI consumption is described in Figure 4. The Client application identifies a task related to a specific social entity (e.g. predict video popularity).

He identifies the target social media websites from which he defines his popularity metrics (e.g. YouTube videos and Facebook videos).The client implements the services relative to the target entity popularity. So, the invocation of the specific services (e.g. in video popularity interface) and the instantiation of popularity is based on the metrics extracted from the target social media. The developed SPI is available on GitHub under the link https://github.com/SebeiHiba/SocEntPopularitySPI.

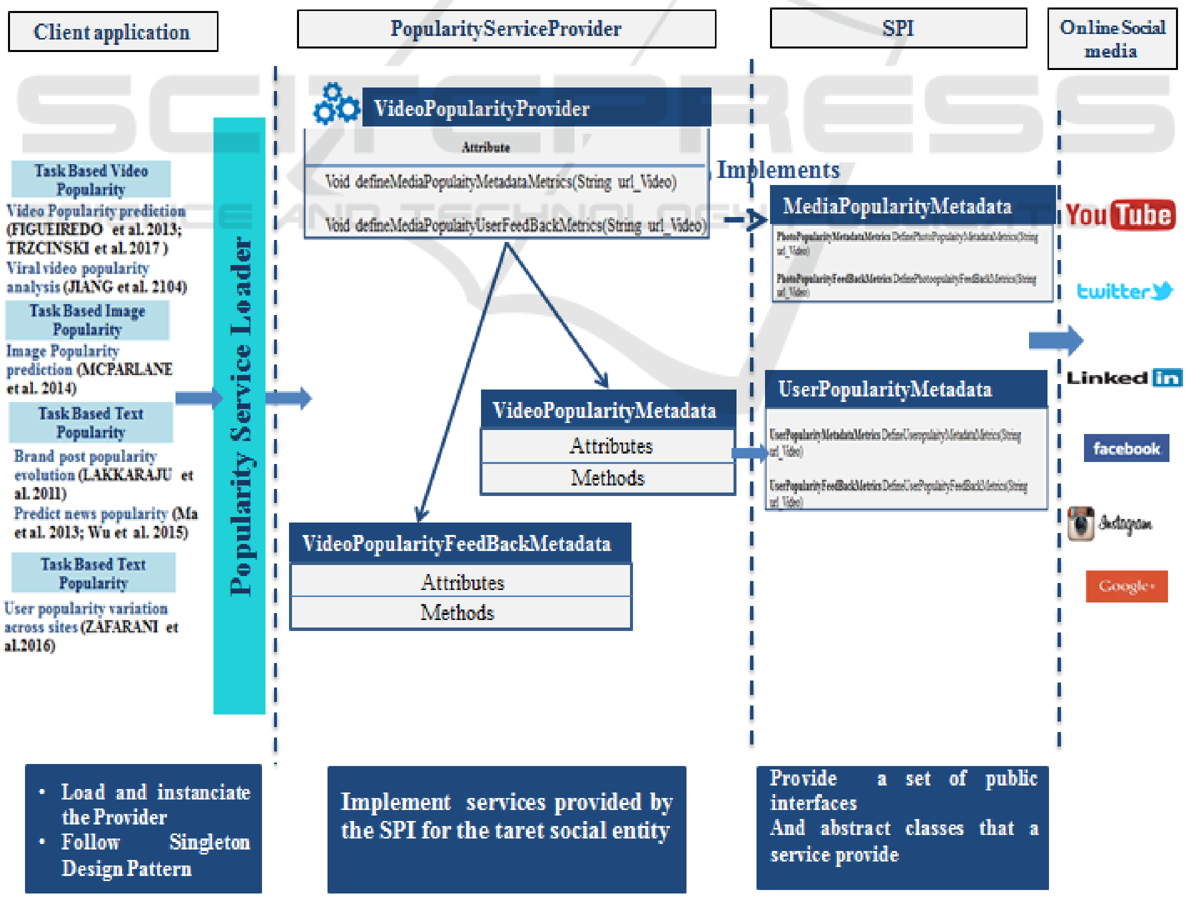

Figure 4: integration of the proposed SPI in the applications based on the analysis of social entities popularity.

The details that are not clear in Figure 4 can be viewed in the code from the previous link.

5 CONCLUSION

In this paper, we dealt with the problem related to the variety of metrics of the quantification of the popularity of social entities (text, video, photo and user) studied across several online social media websites which are Facebook, Twitter, YouTube, Google+ and Flickr. This variety is clear during the investigation of the various studies established to analyse the popularity of the social entities as well as during the extraction of data related to social entities using the various APIs provided by social networking websites as Twitter search API and Facebook Graph API. Our proposal to create a normalized view of these metrics divides it into two main categories: media (i.e. text, photo and video) popularity metrics and user popularity metrics extracted from profiles and pages that present the user' self-presentation. In each one of these categories, the metrics are factorized according to the ones adopted in the related works of popularity analysis also according to the analysis of the extracted data from social networking websites. In addition, the normalized metrics are presented in a hierarchical model to highlight the different factorization levels. Moreover, the normalized view is materialized via in an implemented SPI used as a unified contract between users to express social entities popularity independently of different online social media. The SPI, available for researchers, provides a set of basic services that can be extended to define social entities popularity.

This work can be improved in future by moving it to another level of abstraction through the integration of Resource Description Framework (RDF) to model the different popularity metrics.

REFERENCES

Cappallo, S., Mensink, T. & Snoek, C.G.M., 2015. Latent Factors of Visual Popularity Prediction. In A. G. Hauptmann et al., eds. *ICMR*. ACM, pp. 195–202.

Chatzopoulou, G., Sheng, C. & Faloutsos, M., 2010. A first step towards understanding popularity in YouTube. In *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*. IEEE, pp. 1–6.

Couronné, T., Stoica, A. & Beuscart, J.-S., 2010. Online Social Network Popularity Evolution: An Additive Mixture Model. In N. Memon & R. Alhajj, eds. *ASONAM*. IEEE Computer Society, pp. 346–350.

Figueiredo, F., 2013. On the prediction of popularity of trends and hits for user generated videos. In S. Leonardi et al., eds. *WSDM*. ACM, pp. 741–746.

Gao, S., Ma, J. & Chen, Z., 2014. Popularity Prediction in Microblogging Network. In L. Chen et al., eds. *APWeb*. Lecture Notes in Computer Science. Springer, pp. 379–390.

Gelli, F. et al., 2015. Image Popularity Prediction in Social Media Using Sentiment and Context Features. In X. Zhou et al., eds. *ACM Multimedia*. ACM, pp. 907–910.

Hong, L., Dan, O. & Davison, B.D., 2011. Predicting popular messages in Twitter. In S. Srinivasan et al., eds. *WWW (Companion Volume)*. ACM, pp. 57–58.

Jiang, L. et al., 2014. Viral Video Style: A Closer Look at Viral Videos on YouTube. In M. S. Kankanhalli et al., eds. *ICMR*. ACM, p. 193.

Khosla, A., Das Sarma, A. & Hamid, R., 2014. What makes an image popular? In *Proceedings of the 23rd international conference on World wide web*. International World Wide Web Conferences Steering Committee, pp. 867–876.

Kwak, H. et al., 2010. What is Twitter, a social network or a news media? In M. Rappa et al., eds. *WWW*. ACM, pp. 591–600.

Lakkaraju, H. & Ajmera, J., 2011. Attention prediction on social media brand pages. In C. Macdonald, I. Ounis, & I. Ruthven, eds. *CIKM*. ACM, pp. 2157–2160.

Lerman, K. & Hogg, T., 2010. Using a model of social dynamics to predict popularity of news. In M. Rappa et al., eds. *WWW*. ACM, pp. 621–630.

Li, C.-T. et al., 2016. Exploiting concept drift to predict popularity of social multimedia in microblogs. *Inf. Sci.*, 339, pp.310–331.

Ma, Z., Sun, A. & Cong, G., 2013. On predicting the popularity of newly emerging hashtags in Twitter. *JASIST*, 64(7), pp.1399–1410.

McParlane, P.J., Moshfeghi, Y. & Jose, J.M., 2014. Nobody comes here anymore, it"s too crowded'; Predicting Image Popularity on Flickr. In M. S. Kankanhalli et al., eds. *ICMR*. ACM, p. 385.

Quan, H. et al., 2012. A connectivity-based popularity prediction approach for social networks. In *ICC*. IEEE, pp. 2098–2102.

Szabo, G. & Huberman, B.A., 2010. Predicting the popularity of online content. *Communications of the ACM*, 53(8), pp.80–88.

Trzcinski, T. & Rokita, P., 2017. Predicting popularity of online videos using support vector regression. *IEEE Transactions on Multimedia*.

Wu, B. & Shen, H., 2015. Analyzing and predicting news popularity on Twitter. *Int J. Information Management*, 35(6), pp.702–711.

Zafarani, R. & Liu, H., 2016. Users joining multiple sites: Friendship and popularity variations across sites. *Information Fusion*, 28, pp.83–89. 16.