Mining TV Twitter Networks for Adaptive Content Navigation and Community Awareness

Hugo Hromic, Andrea Barraza-Urbina, Conor Hayes, and Neal Cantle

Abstract In this work, we explore the potential of mining social media streams for extracting collective knowledge related to television (TV) programming. We propose to provide near real-time online community awareness around a set of current TV shows being broadcasted, as well as an adaptive content navigation experience purely based on automatically updated social relatedness, to enhance the online television services and support the characteristics of modern television viewers. As an example, we take the case study put forth by Raidió Teilifís Éireann (RTÉ), the national public television and radio provider of Ireland. RTÉ also streams its content online and has a strong presence in the Twitter micro-blogging service. To support end-users in exploring the RTÉ catalog and understand what is happening in Twitter related to its programming, we developed the RTÉ XPLORER prototype. By using Adaptive Content Navigation and Community Awareness services, we aim for users to find interesting content faster and participate in an enhanced/richer social experience around their favorite shows. The main contributions of this paper include an analysis of captured Twitter data, an in-detail description of the RTÉ XPLORER and the presentation of a functional prototype that showcases our system within the existing RTÉ streaming service.

1 Introduction

The way we consume television (TV) today has changed. Our TV viewing experience has shifted more towards the use of social media platforms. In this work, we

Hugo Hromic · Andrea Barraza-Urbina · Conor Hayes
Insight Centre for Data Analytics, National University of Ireland Galway, Ireland,
e-mail: {firstname.lastname}@insight-centre.org

Neal Cantle
Raidió Teilifís Éireann (RTÉ) Donnybrook Dublin, Ireland, e-mail: neal.cantle@rte.ie

aim to exploit social media as a source of collective knowledge in order to enhance online television services and support the characteristics of modern TV viewers.

Specifically, we focus on the case study put forth by Raidió Teilifís Éireann (RTÉ), the national public provider of television and radio in Ireland. The company makes its TV content available online through the *RTÉ Player* [1], which is Ireland's most popular Video-On-Demand (VOD) Service [5]. The RTÉ catalog has content for all types of audiences, including national and international programming. To help overcome possible information overload, adaptive content navigation and recommendations are desirable features for finding interesting content more efficiently.

The most challenging restriction found in this case study is the unavailability of personal viewer data. Particularly, in order to offer services tailored to the unique characteristics of users, we face the following challenges: (a) lack of personal preference data such as ratings, (b) lack of access to historical user session information, (c) a dynamic catalog with limited life span of TV programs, and (d) the audience users would be considered anonymous. These restrictions keep us from using conventional information adaptation and recommendation approaches, however we identify a key opportunity in using social interactions immersed within social media as a valuable resource that can be exploited to better understand how TV content is related and consumed according to online user activity.

Additionally, to further boost user engagement with RTÉ content, our solution proposes services that describe the social media context of a TV program. We believe that modern TV viewers are also socially curious, thus we focus on solutions that deliver social awareness by offering a structured view of what is going on in social media related to TV content. Furthermore, we provide quick mechanisms for users to engage with communities in social media that they could feel the most connected with, and in this manner, encourage users to participate in online discussions about more diverse RTÉ programming.

In previous work [5] we defined the RTÉ case study and proposed exploratory solution paths towards offering recommendation services. In this work, we present a significant improvement and progress over our initial steps. We not only carry out in-depth analyses of TV social media data, but we also propose a concrete set of solutions focused on Adaptive Content Navigation and Community Awareness services. Furthermore, we also present the RTÉ XPLORER, a prototype system that leverages social media data to enhance the experience of the RTÉ Player audience.

This paper is structured as follows: we first present the background and related work in Sect. 2. Next, we present our approach for an enhanced TV viewing experience in Sect. 3 and a description of our working RTÉ XPLORER prototype in Sect. 4, together with discussion and results. Finally we provide conclusions and potential interesting future directions in Sect. 5.

[1] http://www.rte.ie/player/

2 Background and Related Work

Information overload is a very clear problem for TV consumers. Program catalogs are vast and users can spend a significant amount of time browsing channels to ultimately end watching only from the top-ten most popular channels [26]. Solutions to help users discover faster and more efficiently content that suits their unique tastes, and in turn can keep them engaged and as loyal customers.

Recommender Systems for TV Services. In face of information overload, Recommender Systems (RS) have emerged as tools to help users discover interesting products by means of relevant proactive suggestions. From a business point of view, RS help expose under-explored sections of the program catalog. As a result, RS help reduce churn and generate higher revenues by capitalizing on long-tail shows [26].

There is great interest on RS for Television. An example is *TV Genius* [26], which proposes the construction of a Bayes network defining the likelihood that a viewer could enjoy a TV program. In [21], the use of an item-based RS approach is described, able to suggest less popular items under a high traffic load. In [13, 3, 1], the integration of collaborative and content-based techniques is proposed. However, conventional RS approaches heavily rely on the existence of a user profile; fed by both, implicit and explicit user ratings. Given the unique requirements of our case study, our approach is presented as a solution that does not rely on user profiles, but instead uses online social media as a source of collective knowledge.

We believe that watching television can lead to a number of social experiences. Particularly, it facilitates social interaction by giving common ground for even strangers to establish a conversation [19]. With the benefits of new communication mediums, users have re-purposed social media networks to share about TV with friends and even total strangers [27]. Thus, it has become increasingly important to understand the new ways people communicate about TV programs to support solutions that respond to the characteristics of modern TV viewers.

Twitter: The Grapevine for TV Fans. Television has found an important marketing ally in Twitter. Twitter is a social network micro-blogging service where users post about their status in short real-time messages called Tweets [14]. Given its open nature, brevity of messages and existence of social relations, Twitter can be also seen as a social awareness stream [18].

In [16], Twitter is characterized as a valuable information spreading medium, by virtue of functionalities such as non-symmetric *following* and *retweets* (re-broadcasts of Tweets), both fundamental for open, extensive and fast information diffusion. Kaplan and Haenlein [15] explain how Twitter is a type of awareness system: "*different Tweets sent out over time can paint a very accurate picture of a person's activities, just like the distinct dots in a pointillist painting can collectively create the beautiful images of a Vincent van Gogh, John Roy, or Chuck Close*". The authors define ambient awareness as "*being updated about even the most trivial matters in other peoples' lives*". In this manner, Twitter is as an ideal platform to satisfy social curiosity for those who want to be socially aware but not necessarily socially

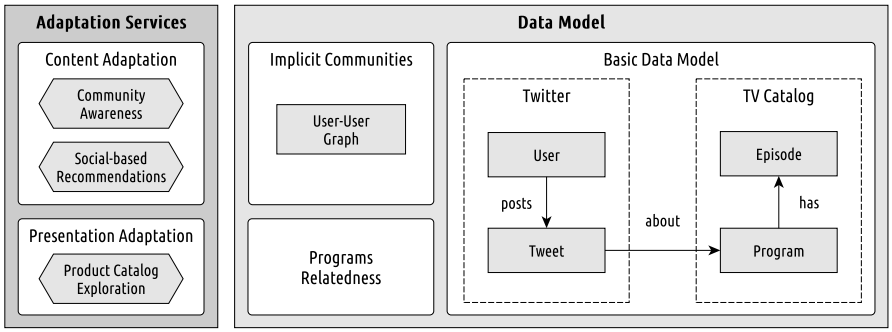

Fig. 1: Conceptual Architecture of our Solution Approach

active. Posting users can contextualize their Tweets by using textual tags (called *hashtags*), thus linking messages to topics or events. These tags make it easy for other users to later read/share about particular streams of Tweets by following/using specific hashtags of interest.

In this paper, we study Twitter activity related to TV content and its value as a source of information to characterize it. We aim to define how two programs are related socially, i.e. via tweeting, retweeting, user mentioning and hashtags usage.

3 Towards an Enhanced Online TV Experience using Twitter

Our main contribution is the design of a solution aimed towards enriching a basic Video-On-Demand (VOD) Streaming Service (i.e. with simple search capabilities) by adding features focused on *Adaptive Content Navigation* and *Community Awareness*. The overall goal is to offer end-users services that support them in exploring the product catalog and help them grasp what is happening in social media related to the broadcaster's programming. In this manner, users can find interesting programs faster and be encouraged to participate in relevant social media communities.

The core components that structure our solution are depicted in the architecture shown in Fig. 1, and are organized in two main logical sections:

(a) *Data Model*: offers the data foundation to enable services. It is supported on a Basic Data Model that represents data from both Twitter and the product catalog of the VOD TV service. The model assumes that all Tweets are associated to one or more programs from the catalog. A concrete instance of this model is discussed in Sect. 4.1 for the RTÉ use case.

(b) *Adaptation Services*: offers tailored services for *Content* and *Presentation* adaptation. On the one hand, Presentation adaption refers to customizing the method used to *display content*, relevant to users' contextual characteristics. Specifically, the Product Catalog Exploration component tailors the navigational structure of content, given up-to-date user and social media data. On the other hand,

Content adaptation processes and adjusts content to offer users a *customized version of services* according to their contextual characteristics. For this, we propose services related to Community Awareness and Recommendations.

In the following Sections we offer an in-depth description of the *Adaptive Content Navigation* and *Community Awareness* components.

3.1 Community Awareness

In this Section, we describe our approach towards providing Community Awareness services for online television catalog systems.

3.1.1 Data Model for Implicit Communities

There is no defined notion of user communities in Twitter. For this reason, in this Section we first establish what a community is in Twitter, under the context of our system. Next, we explain our community detection approach.

What is a Community in Twitter? Humans are social by nature and tend to form tied groups while interacting. Depending on different contexts, those groups can be called clusters, modules or communities. In social media networks, implicit or explicit community structures are known to emerge from user interactions [25]. Defining the concept of a community in a network is not trivial, however the literature often follows this definition: "*a group of nodes more densely connected to each other than to nodes outside the group*" [25, 14, 23]. In Twitter, the same definition of *community* is often adopted [14]. However, we argue that given the unique characteristics of the platform, communities are necessarily *implicit*, and a wider non-personal sense must be adopted. It has been established that Twitter has differentiating characteristics compared to other social media platforms [16], e.g. weak user relationships due to an undefined concept of friendship, non-committed reciprocity, restricted Tweet length, and fast-pacing. Therefore, communities are not as clearly outlined in Twitter as in other social platforms like Facebook. If Twitter users are not connected by close relationships, how could they be possibly connected then?

Engeström discusses the notion of "*object-centered sociality*" [10], arguing that social networks require an intermediate object (of any nature) that connects people together to truly become social. We then argue that there could be underlying communities defined by Twitter users that are connected through common interest objects and not necessarily through explicit social relationships. Users could feel curious if other users agree with them, seek validation of their own ideas, expand and complement them, or look for opinions without the need of forming strong or explicit friendships. In this manner, users can possibly engage in different implicit communities that comment over different social objects. Furthermore, users can act passively over community content, e.g. retweeting, or actively posting comments.

We aim to identify those potential implicit user communities that could have emerged from common interests about TV programs, i.e. our instance of *social objects*. Next, we detail our community detection approach.

Detecting Communities in Twitter. Explicit relations between users can be established given followers, however this information is not enough to determine groups of users that share specific interests. There could be a number of unclear reasons why a user is following another, hence considering following information can introduce noise to our study. We need to clearly identify communities specifically around a set of Television programs.

We want to detect implicit communities that can potentially be created by users while interacting together in Twitter unambiguously in relation to our target TV shows. For this we consider the following *user interactions*: user mentions, replies and retweets. We represent those activities into a graph of interactions that model how users (nodes) relate to each other by a number of interactions (weighted edges). An example of this proposed *User-User graph* can be seen in Fig. 2. This simple network can become complex as the interactions between users increase in time and can be used as the input for a network-based community detection algorithm [12].

The core notion behind our model is to put users together in communities according to how tightly connected they interact among each other compared to users external to the group. We chose the OSLOM (Order Statistics Local Optimization Method) algorithm for user networks [17] as the detection algorithm for mining implicit communities from our User-User graph. OSLOM is capable of finding community structures based on optimizing the statistical significance of clusters with respect to random groups, taking into account edge directions, weights, community hierarchies (i.e. sub-communities) and overlapping (i.e. users in multiple communities at the same time). OSLOM can connect groups of users in communities, but cannot directly associate individual Tweets from member users of the mined communities. To address this, our model assumes that those communities are mined from relatively short data windows, e.g. one hour periods, and that Twitter users highly unlikely will post random content within this short time. We can then assign to each found user in each found community their latest published Tweets that fall within the same time window used for mining the communities.

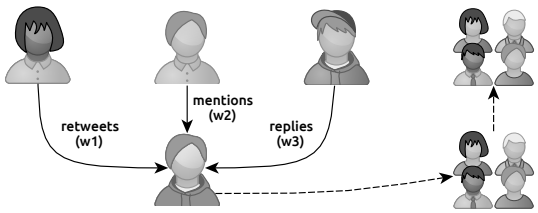

Fig. 2 User-User Twitter interactions graph, users (nodes) relate to each other by a number of distinct interactions (weighted edges)

3.1.2 Services for Community Awareness

We have argued in Sect. 2 that users have a social curiosity for updates on what others think about the TV content they are watching. We propose *Community Awareness* services to offer users with a structured view on the activities of Twitter communities that discuss their favorite shows, and furthermore tools to aid in the interaction with these communities.

In particular, we have detailed in Sect. 3.1.1 that implicit communities discussing TV content can be identified in Twitter. In order to promote user engagement with these communities, we propose to offer widgets customized to the social media characteristics of the programs. These widgets present the most interesting characteristics of the communities associated with the programs, and furthermore, allow users to interact with social media directly. The following are the proposed widgets for Community Awareness services:

(a) *Interesting Conversations*: present potentially interesting conversations (or discussions) to the user that are happening in Twitter associated to the program they are watching. By making users aware of these conversations, the user can better understand other viewer's opinions and also more efficiently engage in discussions with them if she wishes. Incentivizing conversation could lead to the generation of more viewer feedback around programs in Twitter, and hence, generate more feedback for our solution approach. Towards this purpose, we show users the most recent Tweets in a community about the program, the most retweeted Tweets in the community and also related Tweets, i.e. Tweets that were found in the community but that are not necessarily about the same program the user is currently watching.
(b) *Relevant Hashtags*: suggest the most relevant hashtags to use when tweeting about a given program. The aim is to encourage the use of the program's official hashtags and also of other relevant hashtags found that co-occur with the program. Making users aware of these hashtags can stimulate their usage and help our service retrieve more information from Twitter.
(c) *Top Users*: identify the most influential users from Twitter that are commenting on a given program. These users could be interesting users to follow in Twitter. A user is influential if his/her activity contributes to the overall spread of information in the network, e.g. Tweets from this user reach potentially multiple other communities/regions of the network. We identify top users according to their PageRank [20] centrality within the community sub-networks we mine.
(d) *Relevant Media*: showcase the most popular media components that are being shared in Twitter. Media components could include images and links to other web resources. Popularity is determined by the frequency the media item is mentioned within the implicit community.
(e) *Live Tweets*: displays the live Tweets related to a program to better allow users to interact with Twitter while using our services. Real-time interaction with social media is aligned with promoting the second screen phenomena.

To summarize, Community Awareness widgets are designed to help users quickly understand what is happening in social media about their preferred programs, and ideally give them tools to easily share their opinions.

3.2 Adaptive Content Navigation

In this Section, we present our Adaptive Content Navigation solution based on knowledge obtained from live social media activity and Recommendation services.

3.2.1 Data Model for Program Relatedness

We believe that implicit in Twitter data we can find information on the relatedness of TV programs. For this, we adopt the notion that relatedness is a broader concept than similarity [7]. Two items can be related without necessarily being similar, e.g. a car and a wheel. In our case, we would like to determine if two programs are related by virtue of their online social connections without necessarily having to be similar in their content features.

In the context of Twitter, and as a heuristic, we define that two programs are related proportionally to the amount of times they are found together in different social settings. In particular, we define that *two programs are related* if they satisfy any of the following types of co-occurrence:

(a) *Intra-Tweet*: both are mentioned in the same Tweet,
(b) *Inter-Tweet*: both are mentioned by the same user in separate Tweets, and
(c) *Community-based*: both are mentioned in the same user community. Communities are discussed in Sect. 3.1.1.

After the above, we define two Twitter social settings for relatedness: *Tweet-based* (for Inter-Tweet and Intra-Tweet) and *Community-based*.

3.2.2 Services for Social-based Recommendations

The *Social-based Recommendations* component uses knowledge acquired from social media about program relatedness (Sect. 3.2.1), to offer a ranked list of N programs, ordered in terms of relevance in relation to a given input program. A program is relevant if it could be interesting for the user considering his/her current context. Context is restricted to what the user is currently viewing and what is happening in social media only, as in our system users are anonymous and we do not have past browsing history records beyond the current session.

This component uses an Item-Based Recommendation approach, founded on the concept of Conditional Probability-Based Similarity as formulated in [9]. We define $P(j|i)$ as the conditional probability of program j being relevant to the current

context, given that program i has already been determined as relevant. Because we do not have user explicit input, we must rely on the assumption that if the user is currently watching a program then it is relevant to the user's current context. As a result, given the current program the user is viewing, the recommendation service delivers the top-n programs that have the highest probability of being relevant to the current program being watched. Therefore, the challenge is on how to define $P(j|i)$. The formulation of $P(j|i)$ as defined by [9] is in Eqn. 1.

$$P(j|i) = \frac{frequency(i,j)}{frequency(j)} \tag{1}$$

In Sect. 3.2.1, we have argued that Twitter can offer information on the relatedness of different TV programs based on three co-occurrence situations: *Inter-Tweet*, *Intra-Tweet* and *Community-based*. If programs co-occur in any of those situations, we assume they are to an extent related or relevant to each other. Thus, from each setting we can derive a different approximation to the conditional probability distribution function $P(j|i)$. Table 1 defines $frequency(i,j)$ and $frequency(i)$ in Eqn. 1 for each defined type of co-occurrence case. Furthermore, we can deal with the *popularity bias* by separately scaling the output with a value that depends on $P(j)$ [9]. For this, we multiply each $P_k(j|i)$ by $-log_2 P_k(j)$, inspired from the widely used inverse-document frequency (IDF) factor.

Table 1: Probability distributions $P_k(j|i)$ for our proposed co-occurrence types

Co-occurrence Type		$frequency(i,j)$	$frequency(i)$
Intra-Tweet	$P_1(j\|i)$	Number of Tweets that mention both programs i and j together.	Number of Tweets associated to program i.
Inter-Tweet	$P_2(j\|i)$	Number of times the same user, in separate Tweets, mentions both programs i and j.	Number of Tweets associated to program i.
Community-based	$P_3(j\|i)$	Number of communities associated to both programs i and j.	Number of communities associated to program i.

Lastly, to determine the final value for $P(j|i)$, we linearly combine the different sources of evidence using different weights w_k as in Eqn. 2, where K is the number of co-occurrence types we proposed, i.e. sources of evidence.

$$P(j|i) = \sum_{k=1}^{K} w_k \cdot P_k(j|i) \cdot -log_2 P_k(j) \tag{2}$$

In our model, the weights are defined by business rules and empirical intuitions. For example, if evidence from *Inter-Tweet* co-occurrence (w_1) should be the most influential over the final probability, then weight values have to comply with the following restrictions: $w_1 > w_2$ and $w_1 > w_3$.

Extensions to the formulation of $P(j|i)$ can be further explored given different heuristics. For example, the shared content or semantic features of programs, or even on user explicit feedback (when it can be obtained) can be used. As a consequence, diverse information sources could be considered, such as linked open data or even other social media platforms. Proposals towards these approaches are suggested in [5], but still leave ground for further future work.

3.2.3 Services for Adaptive Content Navigation

The goal of the *Product Catalog Exploration* component is to support users in the task of browsing available TV content, and ideally find in less time interesting programs to watch. This component showcases different customized views of the catalog relevant to user and social media features.

First, we define a set of views that will support product catalog exploration. Each view is designed to highlight a set of programs focused on a different perspective of the catalog, for example:

(a) *Because You Watched*: programs related to those the user recently watched, extracted from the Social-based Recommendations component.
(b) *Most Discussed*: shows that are actively discussed in social media, e.g. Twitter.
(c) *Most Discussed New Programs*: new content that is starting to become popular in social media. Introduces users to programs that could become popular soon.
(d) *National Exclusive*: highlights exclusive programming locally produced by the TV broadcaster to help the audience identify faster those types of programs.
(e) *One Time Runners*: programs that are running once.
(f) *Last Time to See*: programs that are expiring soon.
(g) *Most Recent*: showcase new or recently added programs to the catalog.
(h) *Coming Up*: programs that will be added in the near future.

The above views are organized according to the user's latest current browsing history and to what is currently happening in social media at any given moment. More specifically, we identify the content features of programs that the user has recently watched to prioritize the display of the proposed views in a way that matches the user's most recent tastes. Also, the popularity of programs in social media influences this ordering. Items that the user would most probably like are then placed first considering the implicit feedback and popularity found within social media.

4 Discussion and Results

The RTÉ XPLORER prototype is meant to be a tangible representation of how the proposed services in Sect. 3 could be integrated into the RTÉ Player service. In this Section, we discuss and analyze the models and services we proposed previously in the context of the RTÉ use case and its dataset.

4.1 Data Collection Strategy for the RTÉ Basic Data Model

In this Section, we outline our proposed data collection strategy for the RTÉ use case and offer a general description for our captured social dataset. The complete technical report containing detailed analyses can be found in [6].

We collected data from (a) the online catalog of the RTÉ Player, and (b) live dynamic content shared in Twitter relevant to current RTÉ programming. On the one hand, the official RTÉ Player website is our principal data source for the RTÉ product catalog. From this site we capture information on available programs and episodes using a custom Web Crawler. Information describing programs include (but is not limited to): name, description, season number, episode number, category, expiry date and original broadcast date.

On the other hand, RTÉ has a strong presence in social media, mainly on platforms such as YouTube, Twitter and Facebook. We are particularly interested in Twitter as it is an important source of interconnected TV-related content as argued in Sect. 2. We collected data from Twitter using their real-time Streaming API, and Tweets were captured using a curated list of non-overlapping terms and users that are each related to an RTÉ program. Thus, we are able to annotate captured Tweets back to the associated programs we initially intended to listen for.

Descriptive Statistics of Collected Data. Our captured dataset consists of data collected from the period of **July 17, 2015** to **February 17, 2016** (seven months). Our Web Crawler downloaded catalog data from the RTÉ Player for a total of 151 programs. For all those, we have manually built a curated list of 83 hashtags, 91 keywords and 121 users to listen to in Twitter, that do not overlap across programs.

After applying the basic spam detection technique described in [11], we captured a total of **10,596,397 Tweets** authored by **2,131,195 users** related to **138 programs**. From those shows, 77 are national to Ireland and 61 are of international programming. All the 138 programs are distributed in the following categories: Drama (34), Children (24), Factual (23), Entertainment (22), Lifestyle (16), News and Sports (15), Comedy (2), and Religious and Language (2). We found that a share of 80% of the Tweets belong to the top-five programs, leading to a long tail shape of posting activity as expected. More in-depth details can be found in [6].

4.2 Analysis of Implicit Communities in the RTÉ Data Model

In this Section, we study if we are able to automatically discover implicit communities in Twitter related to TV programs, as discussed in Sect. 3.1.

Methodology. We study discussion structures in our Twitter dataset in order to further view the potential of finding implicit communities in Twitter. For this, we consider the concept of conversational reciprocity for activities being returned [23], e.g. if a user mentions a user in a Tweet and that same user replies back. We then borrow

the *Reciprocity* and *Popularity* measures from [8] to examine conversational behaviors in Twitter networks such as the User-User graph shown in Fig. 2. Reciprocity is defined as the average percentage (%) of bidirectional users that have replied or retweeted to each other at least once. Likewise, Popularity is defined as the ratio of the neighboring users that have replied to a user's Tweets at least once.

Results. We computed the Reciprocity and Popularity measures over our dataset and found that from 1,680,015 users that replied or retweeted at least once, a vast majority of 1,667,542 (99.3%) had a Reciprocity of less than 10%. In the case of Popularity, we found that 961,190 users posted Tweets that were not retweets and, again, a big portion of 801,312 (83.4%) had less than 10% of their neighbors replying back. This represents a poor direct conversational behavior in relation to our listened TV programs for our dataset.

However, if we focus on the users that did have high Reciprocity and Popularity, we find that there is a significant bias towards more than 90% in both measures. This suggests that users seem to be polarized on engaging or not in conversations, and when they do, there is an interesting potential for placing them into communities. Moreover, even though users are not directly communicating with each other, they are still participating by using common hashtags and retweeting content in a kind of larger meta-group or implicit community [27]. In conclusion, despite Twitter being not originally designed for organized discussions [23], forms of those are seemingly still possible that further lead to interesting implicit communities.

4.3 Analysis of Program Relatedness in the RTÉ Data Model

In this Section, we offer an in-depth analysis of Twitter being used as a source of collective knowledge for observing RTÉ programs relatedness, considering the three social settings proposed in Sect. 3.2.

Methodology. For each proposed co-occurrence type we build an $N \times N$ matrix M, where N is the number of shows we are listening to in Twitter. Each cell in this matrix $(M_{i,j} : i, j \leq N)$ defines the number of times a particular pair of programs co-occurred depending on a social setting. Thus, we built three matrices: *Inter-Tweet*, *Intra-Tweet* and *Community-based*. For each, we study its sparsity to draw insights over the amount of captured co-occurrence information. Sparsity is defined in Eqn. 3 as the proportion of zero-valued cells compared to the total number of cells.

$$sparsity(M) = \frac{\sum_{i,j \leq N} empty(i,j)}{N^2},\ empty(i,j) = \begin{cases} 1, & M_{i,j} = 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Therefore, to determine the amount of information that can be captured about programs relatedness in Twitter, we study the sparsity of our co-occurrence matrices. Finally, we also draw conclusions by combining all our proposed social settings.

Results. We first computed the sparsity of both matrices for the Tweet-based co-occurrences on a weekly basis over the seven months dataset, accumulating the data from one week to the next. The results are shown in Fig. 3. It can be observed that, as expected, sparsity decreases steadily over time, reaching a minimum of 0.74 for the Inter-Tweet co-occurrence. The Intra-Tweet scheme has a much higher sparsity, most probably due to the difficulty for users to post about two or more programs in the same Tweet given its 140-characters restriction.

Furthermore, we found a strong separation between national and international programs, i.e. there is a low connection across those shows being produced locally and globally. However, we still found weak connections bridging the two islands. From another perspective, we identified a high cohesion between program categories. Users tweet simultaneously about two or more programs more often if these are of the same category. Interestingly, strong co-occurrence was also found in accordance to the topics of the programs. For example, we observed that Twitter users posted closely about cooking shows but clearly separated from other shows in the same Lifestyle category. The same could be seen for Children programs. For example, we found a strong link between three programs: two of them about news broadcasted using sign language and a third about weather news. All those associations are found purely and automatically as an effect of the social collective in Twitter. For the case of Drama (the most popular category in the dataset), those programs created a big cohesive cluster but did not show any particular sub-categorization. This suggests that users tend to discuss more generally about Drama programs, indicating a strong global fan-base for multiple popular Drama shows.

We now study the relatedness of programs under the Community-based co-occurrence setting. We take communities detected at regular intervals of one hour and build the co-occurrence matrix for each time frame. Sparsity values over time can be seen also in Fig. 3. The community-based sparsity lays in-between the Intra-Tweet and Inter-Tweet curves. We found a positive effect of using communities as a relatedness connector in the form that it acted as a noise filter for connections

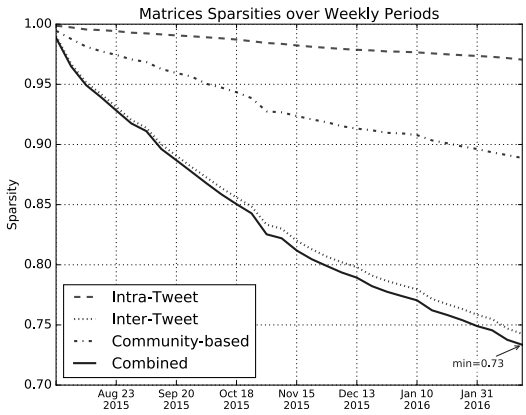

Fig. 3 Accumulated co-occurrence matrix sparsities computed over weekly periods. Sparsity decreases steadily over time in all three settings, indicating that more programs in the catalog are constantly connected together by Twitter interactions

(a) Main View (b) Video View (c) Example Community

Fig. 4: User Interface views of the RTÉ XPLORER prototype. The live working app can be found in http://rtexplorer.insight-centre.org/

between programs. The resulting matrix, despite being sparser than for the Inter-Tweet case, was able to capture most of the same connections with only 24% less coverage, mostly due to the inherent filtering by OSLOM of shows that have a very low tweeting activity. The same types of associations discussed previously were also observed for the community-based social setting.

Finally, we present results from combining our three proposed co-occurrence settings. For this, we combined their three matrices by adding their frequencies together. The intuition for this merging strategy is that the different matrices provide different perspectives for social relatedness of the programs, and by combining them we can blend the social connections from each into a single matrix with enhanced overall coverage. The sparsity of the resulting matrix is shown in Fig. 3. The combined coverage had an improvement of 1.21% over the best found coverage, resulting in a further lower sparsity of 0.73. With this, we now are obtaining social connections for 138 programs in total, 78 national and 60 international. In conclusion, even though complete coverage of the product catalog is not achieved, Twitter is able to provide us with relatedness information for 91% of the listened programs.

4.4 The RTÉ XPLORER*: Functional Prototype*

In this Section, we briefly describe the working functional prototype developed to showcase our RTÉ XPLORER solution. Our prototype provides two main user interface views: (a) the *Exploration View* (Fig. 4a): a landing page where the user can explore different programs ranked according to contextual features, i.e. the user's viewing history and live social media activity, and (b) the *Video View* (Fig. 4b): which offers all information for a chosen video item, e.g. an episode of a program. In addition to the embedded video, this view also shows communities associated to the program (Fig. 4c), latest Tweets and a set of other related videos.

In more detail, the Exploration View offers different types of program carousels to allow a user to explore the product catalog in a more efficient way. This design is inspired by display optimization techniques found in Netflix [2]. Using carousels, diverse content can be offered both horizontally and vertically, increasing the chances for users finding interesting content faster. Currently, the Exploration View has three core sections: *Just for You*, *Today in Social Media* and *More National Content*, but it can be further enhanced to consider more.

When users select an item from the Exploration View, they are taken to the item's dedicated Video View. Here, not only basic information about the particular video is shown, but also its related programs and social media context. At the bottom of this same view, the user can access a feed with the latest Tweets that are related to the program. To the right hand side, the user is offered a widget containing all the implicit communities discovered about the show. This communities panel includes elements for each live user community found: frequent hashtags, community Tweets, most retweeted Tweets, other related Tweets and the top users.

Finally, those carousels where the user is presented with related videos (e.g. Just for You carousels) include a magnifying glass icon to provide explanations on how a pair of programs were associated together. This option displays the underlying co-occurrence information for all three social settings described in Sect. 3.2, including the top-20 co-occurring Tweets, users and discovered communities.

User Experience Study. To investigate the usability and grasp a real-world assessment of the RTÉ XPLORER prototype, we conducted a user survey with a standard System Usability Scale (SUS) at its core [4]. We designed our survey with three parts: (a) general demographics about the user and questions on his/her preferences for using social media while watching TV content, (b) a SUS questionnaire, and (c) a final remarks section for users to leave free-text comments.

After running the survey online, a total of 46 responses were collected, with 43.5% of users being female and 56.5% male. The most common age group (65.2%) was for ages between 26 and 35. We ran the survey on a global scale and 69.6% of the answers were from Ireland residents. However, clear information about RTÉ and its Irish nature was provided as context for international users.

Initial interesting results come from the first section of the survey. A significant 82.6% of users expressed a strong preference for online streaming services instead of more traditional TV broadcasting, highlighting the niche for efficient catalog exploration solutions. On the other hand, 58.6% expressed having curiosity to seek for external opinions in online social media, while 86.9% said they rarely post their own opinions on this medium. This result suggests that there is potential opportunity on bridging those two user groups by means of online community awareness services.

Finally, regarding the results of the SUS-based part of the survey, 63% of the responses indicated that they *would like to use the* RTÉ XPLORER *frequently*. Also, 71.7% of users thought the prototype was easy to use and 87% thought that most people would learn to use it very quickly. We computed the overall SUS score to be **72.2**. If we only consider those users residing in Ireland, hence giving weight to people more familiar with RTÉ, the score is **73.1**. The general consensus states

that scores above 68 are considered to be good, however the system could be further improved to be ideal [22].

5 Conclusions and Future Work

In this paper, we presented a well-defined set of solutions which aim to enrich the experience of a basic Video-On-Demand (VOD) Streaming Service (i.e. with simple search capabilities). Our solutions were designed considering the challenges and restrictions depicted by the RTÉ case study [6]. The most significant limitations are the lack of explicit user feedback (such as ratings) and user session data. This keeps us from maintaining a traditional user profile and also from applying traditional recommendation approaches. As a solution, we found a key opportunity in using Twitter as a source of collective knowledge to determine how TV programs are related to each other; and in turn, use this knowledge to provide users with customized services tailored to their social media context while watching online TV services.

The main contributions presented in this paper include: (a) the portrayal of our solution approach towards an enhanced TV experience, which includes services to support content navigation, recommendations and community awareness, (b) an in--depth analysis of Twitter network data in the context of RTÉ, and (c) the development and implementation of the RTÉ XPLORER functional prototype, designed to offer a tangible representation of how the proposed services can be integrated into the official live RTÉ Player service.

We believe that the RTÉ scenario is a representative case study to explore the potential use of micro-blogging network data to enhance Information Adaptation services. In this paper, we have proposed a set of novel approaches that can be used for any online TV Player. Furthermore, our study methodology can be used with other similarly captured Twitter datasets.

Finally, we emphasize that the richness of this case study allows for plenty of room for future work. From possible directions we highlight the following: assess alternative methods to discover implicit communities using algorithms such as RankClus [24], expand our Twitter data analyses, e.g. study more in detail what are the features that contribute to the growth and decline of communities. In reference to our recommendation approach, we could further add data sources that can offer evidence on program relatedness, e.g. using linked open data.

Acknowledgements This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289 (Insight).

References

1. Ali, K., Van Stam, W.: TiVo: making show recommendations using a distributed collaborative filtering architecture. In: Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 394–401. ACM (2004)
2. Amatriain, X., Basilico, J.: Recommender Systems in Industry: A Netflix Case Study. In: Recommender Systems Handbook, pp. 385–419. Springer (2015)
3. Bambini, R., Cremonesi, P., Turrin, R.: A recommender system for an IPTV service provider: a real large-scale production environment. In: Recommender systems handbook, pp. 299–331. Springer (2011)
4. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the System Usability Scale. Intl. Journal of Human–Computer Interaction **24**(6), 574–594 (2008)
5. Barraza-Urbina, A., Hromic, H., Heitmann, B., Hulpus, I., Hayes, C.: Using Social Media Data for Online Television Recommendation Services at RTÉ Ireland. In: 2nd Workshop on Recommendation Systems for Television and Online Video (RecSysTV) (2015)
6. Barraza-Urbina, A., Hromic, H., Heitmann, B., Yañez, A., Tamatam, H., Hayes, C.: Using Social Media for Online Television Adaptation Services at RTÉ Ireland. Tech. rep., Insight Centre for Data Analytics, National University of Ireland, Galway (2016)
7. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of lexical semantic relatedness. Computational Linguistics **32**(1), 13–47 (2006)
8. Chan, J., Hayes, C., Daly, E.M.: Decomposing Discussion Forums and Boards Using User Roles. ICWSM **10**, 215–218 (2010)
9. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS) **22**(1), 143–177 (2004)
10. Engeström, J.: Why some social network services work and others don't—Or: the case for object-centered sociality (2005). URL http://www.zengestrom.com/blog/2005/04/why-some-social-network-services-work-and-others-dont-or-the-case-for-object-centered-sociality.html
11. Hromic, H., Karnstedt, M., Wang, M., Hogan, A., Belák, V., Hayes, C.: Event Panning in a Stream of Big Data. In: LWA Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML) (2012)
12. Hromic, H., Prangnawarat, N., Hulpuş, I., Karnstedt, M., Hayes, C.: Graph-based methods for clustering topics of interest in twitter. In: International Conference on Web Engineering, pp. 701–704. Springer (2015)
13. Hsu, S.H., Wen, M.H., Lin, H.C., Lee, C.C., Lee, C.H.: AIMED-A personalized TV recommendation system. In: European Conference on Interactive Television, pp. 166–174. Springer (2007)
14. Java, A., Song, X., Finin, T., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56–65. ACM (2007)
15. Kaplan, A.M., Haenlein, M.: The early bird catches the news: Nine things you should know about micro-blogging. Business Horizons **54**(2), 105–113 (2011)
16. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World wide web, pp. 591–600. ACM (2010)
17. Lancichinetti, A., Radicchi, F., Ramasco, J.J., Fortunato, S.: Finding Statistically Significant Communities in Networks. PloS one **6**(4), e18,961 (2011)
18. Naaman, M., Boase, J., Lai, C.H.: Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work, pp. 189–192. ACM (2010)
19. Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., Amento, B.: CollaboraTV: making television viewing social again. In: Proceedings of the 1st international conference on Designing interactive user experiences for TV and video, pp. 85–94. ACM (2008)
20. Page, L., Brin, S., Motwani, R., Winograd, T.: The PageRank citation ranking: bringing order to the web (1999)

21. Pripužić, K., Žarko, I.P., Podobnik, V., Lovrek, I., Čavka, M., Petković, I., Stulić, P., Gojčeta, M.: Building an IPTV VoD recommender system: An experience report. In: Telecommunications (ConTEL), 2013 12th International Conference on, pp. 155–162. IEEE (2013)
22. Sauro, J.: Measuring usability with the system usability scale (sus) (2011)
23. Shi, X., Tseng, B., Adamic, L.: Looking at the blogosphere topology through different lenses. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007), vol. 1001, p. 48109 (2007)
24. Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., Wu, T.: RankClus: integrating clustering with ranking for heterogeneous information network analysis. In: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, pp. 565–576. ACM (2009)
25. Tang, L., Liu, H.: Community detection and mining in social media. Synthesis Lectures on Data Mining and Knowledge Discovery **2**(1), 1–137 (2010)
26. TV Genius: An Integrated Approach to TV & VOD Recommendations (2011)
27. Wohn, D.Y., Na, E.K.: Tweeting about TV: Sharing television viewing experiences via social media message streams. First Monday **16**(3) (2011)