A Framework for Brand Reputation Mining and Visualisation

Ayesh Alshukri, Frans Coenen, Yang Li, Andrew Redfern and Prudence W.H. Wong

Abstract Online brand reputation is of increasing significance to many organisations and institutes around the globe. As the usage of the www continues to increase it has become the most commonly used platform for users and customers of services and products to discuss their views and experiences. The nature of this www discussion can significantly influence the perception and hence the success of a brand. Brand Reputation Mining (BRM) is a process to help brand owners to know what is being said about their brand online. This paper proposes a BRM framework to provide support for enterprises wishing to conduct brand reputation management. The proposed framework can be generically applied to collect, process and display the reputation of different brands. A key feature is the visualisation facilities included to allow the display of the results of reputation mining activities. The framework is fully described and illustrated using a case study. The concepts expressed in this paper have been incorporated into the "LittleBirdy" brand reputation management product commercially available from Hit Search Ltd.

1 Introduction

Brand reputation has always been an important issue with respect to many organisations (both commercial and non-commercial), particularly in the context of consumer facing organisations. Recommendation and "word-of-mouth" are an important factor in how organisations are perceived and can have a substantial influence

Ayesh Alshukri and Andrew Redfern. e-mail: {AAlshukri, ARedfern}@hitsearchlimited.com
Hit Search, Liverpool Innovation Park, L7 9NG.

Ayesh Alshukri, Frans Coenen, Yang Li and Prudence W.H. Wong. e-mail: {a.alshukri, coenen, y.li48, pwong}@liverpool.ac.uk. Dept of Computer Science, University of Liverpool, L69 3BX.

This work was supported by the UK Technology Strategy Board Knowledge Transfer Partnerships programme (Grant No. KTP008896)

on the success, or otherwise, of an organisation; it is easy to tarnish a brand if negative social activities are associated with it [9]. With the advent of the www, and the prolific use of social media, it has become very important for organisations to manage their online reputation. A growing percentage of users maintain that blogs and online fora are credible ways of finding out about products and services [9]. Organisations that wish to preserve their reputation are therefore interested in knowing "what is being said about them" on social media. However, the increasing predominance of Consumer Generated Content (CGC) on the www (examples include blogs, news forms, message boards and web pages/sites), makes it virtually impossible for organisations to manually monitor the reputation of their brand based on human effort alone [11]. One solution is to automate the process.

This paper proposes a framework for conducting effective Brand Reputation Mining (BRM). In this context BRM is concerned with the identification of "mentions" expressed across electronic (social) media with respect to a particular company, institution or organisation, and determination of the sentiment of such mentions. The information gathered from such a BRM activity can be used by organisations to: (i) compare their performance against competitors, (ii) assess specific marketing strategies and (iii) gauge how a particular product or service is received in the market place. The successful conduct of BRM entails three broad challenges: (i) the identification and collection of mentions on www media (for example with respect to social networks, blogs and news sites); (ii) the application of data mining techniques to the gathered information in order to determine the sentiment associated with opinions expressed in the form of mentions or to group mentions according to the views expressed; and (iii) the presentation of the data mining results obtained in a manner that allows it to be acted upon. In the case of the third challenge visualisation techniques are seen to be the most appropriate solution. However, such visualisation also presents a significant challenge concerned with how best to present the data in a meaningful manner.

The proposed BRM framework can be generically applied to collect, process and display the reputation of brands belonging to organisations. The framework includes mechanisms for the collection of information from www sources such as social media, news and blogs. The framework also includes mechanisms for mining the collected data, this includes: (i) the application of sentiment and opinion mining techniques and (ii) the discovery of topical groupings (using hierarchical clustering). The most significant element of the framework is a collection of visualisation options whereby the BRM outcomes can be displayed so that users can effectively digest the collected information and obtain a view concerning their brand in the context of online media. The framework has been incorporated into a commercial brand reputation management system called LittleBirdy (www.littlebirdy.buzz).

The remainder of this paper is organised as follows. In Section 2 some related work is presented. Section 3 gives a formal description of the BRM problem, while Section 4 describes the proposed BRM framework. Section 5 then presents a case study and an evaluation of the framework. Some conclusions are presented in Section 6.

2 Related Work

Social media is providing a new form of communication platform which is both unregulated and unofficial. By definition social media content is outside of the direct control of organisations [3]. However, social media activity can also provide substantial opportunities for organisations. For example the ability for organisations to be speedily aware of, and able to identify, disgruntled customers according to activity on social media is seen to be particularly beneficial in that organisations can quickly implement some action in an attempt to safeguard their product [2]. Another example is the swift identification of the use of copyrighted material on www fora such as YouTube (for video sharing), Flikr (for photograph sharing) and SlideShare (for presentation sharing), amongst many others [2]. BRM is therefore seen as an important aspect of the modern commercial world. It is thus not surprising that there has been substantial recent work directed at BRM. Examples can be found in [4, 11, 9, 8, 10, 5].

The main focus of the work presented in Morinaga [4] is to determine the reputation of a company/brand by focusing on mining online opinions concerning their products. By using collected texts it is assumed that factual information about a product is not required, concentrating on opinion only so as to focus on the experience of a product that individuals are writing about. The main distinguishing factor between Morinaga's work and the BRM framework presented in this paper is the data mining approach. Morinaga does use sentiment analysis, but there is no topic discovery element as included in the BRM framework (which uses a hierarchical clustering approach for topic discovery).

In the work by Ziegler et al [11] a system is described that uses RSS (Rich Site Summary) feeds to collect news related to large corporations. The feeds are categorised using taxonomies from the DMOZ open directory project www.dmoz.org. The BRM framework presented in this paper focuses on current trends in social media and blogs amongst other sources of information. This differs from Ziegler's work where the focus of the reputation mining effort is directed at large corporations and only uses news feeds as the information source (no social media). It is argued that restricting BRM activity to news feeds may not be the most appropriate method of collecting brand mentions, particularly if we wish to focus on user opinions.

In the work by Spangler [9] a system called COBRA (COrporate Brand and Reputation Analysis) is proposed in order to provide corporate brand monitoring and alerting. The main focus of the COBRA system is to identify product categories, topics, issues, and brands to be monitored. It does this by focusing on broad keyword based queries related to a brand in order to retrieve "sufficient" data. This approach can be very wasteful in terms of bandwidth usage and data storage. Distinction with the BRM framework proposed is that the framework's focus is on brand mentions, consequently the collected data is significantly less "noisy" (little irrelevant data is collected), hence the data mining applied at a later stage is more effective.

3 Formal Description

This section presents a formal description of the BRM problem. A *mention* of a particular brand comprises a body of text collected from an online social media site of some form. We use the notation m_i to indicate a particular mention. A data set comprising n mentions, collected over a period of time, is denoted using the notation $M = \{m_1, m_2, \cdots, m_n\}$. The content of each mention m_i is represented using a feature vector F_i (how this is generated will become clear later in this paper) founded on a feature-space model [7]. Each dimension in the feature space represents some attribute a_i whereby each attribute can take two or more values (if it could take only one value it would be a constant and therefore not of interest with respect to the BRM). We indicate the set of values for an attribute a_i using the notation $a_i.V = \{v_1, v_2, \cdots\}$, we indicate a particular value v_j associated with a particular attribute a_i using the notation $a_i.v_j$. Thus we have a global set A of attribute-value pairs. Thus each F_i is a subset of A ($F_i \subset A$). The complete data set M is therefore represented by a set F of feature vectors such that $F = \{F_1, F_2, \cdots F_n\}$ (note that there is a one to one correspondence between M and F). The first challenge is thus to translate M into F.

Once F has been established the next stage is to apply some sentiment mining to the content of F. For each vector encoded mention in F we wish to attach a sentiment value. A coarse sentiment scoring was conducted using the label set $\{positive, negative, neutral\}$. Thus for each mention $m_i \in M$ there exists a corresponding sentiment label $s_i \in S$ (where $S = \{positive, negative, neutral\}$). The most appropriate sentiment label is derived using a classification algorithm which relies upon the features in F to assign a label to each mention. Details of the classification algorithm is presented in section 4.2.1.

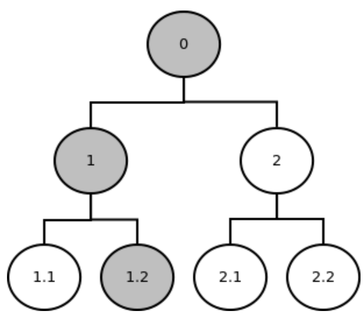

Fig. 1 Example of a BRM hierarchical clustering

The next stage in the proposed BRM framework (see below) is to describe the identified mentions in terms of a hierarchical structure using a hierarchical clustering algorithm that uses the similarity of the features from F for each mention. Thus for each mention $m_i \in M$ there exists a “path” in the hierarchy $P = \{c_1, c_2, \cdots c_x\}$, where x denotes the number of levels in the hierarchy and c_i denotes the cluster ID of a cluster belonging to level i. An example of a hierarchical clustering is shown in Figure 1. The example shows a set of hierarchical clusters at various levels. Level

1 represents the complete collection of mentions (so the set M) and is indicated by cluster 0. Level 2 shows the complete collection of mentions segmented into two groups, cluster 1 and cluster 2. Level 3 shows that clusters 1 and 2 have been further segmented into two sub-groups each indicated by the notation 1.1, 1.2, 2.1 and 2.2 respectively. Thus a mention that is a member of cluster 1.1 is also a member of cluster 1 and of course cluster 0 (as the latter contains all mentions). Figure 1 also shows a path (highlighted) associated with a particular mention, $P = \{0, 1, 1.2\}$. Further details concerning the hierarchical clustering algorithm is presented in Section 4.2.2. It should be noted that this structure is independent of the sentiment score assigned to each mention, thus each cluster in the hierarchy will contain mentions that feature different sentiment labels.

4 Brand Reputation Mining Framework

This section presents details of the BRM framework which has been developed in order to address the challenges of BRM (as presented in Section 1 above). The framework comprises three stages: (i) data collection, (ii) data mining and (iii) visualisation. Each of these is discussed in detail in the following three sub-sections.

4.1 Data Collection

Two main approaches are used for data collection. The first is a server side RSS reader and the second a social media related Application Programming Interface (API's). Both approaches were utilised in order to collect and gather mentions at regular time intervals. For the first relevant RSS feeds need to be identified with respect to the brand in question. Examples include news website RSS feeds and industry specific feeds which are updated on a regular basis. Social media platforms offer access, via an API, so that relevant mentions can be obtained based on a submitted query. The query in this case needs to include terms related to the brand in question, for example brand name or related products and services.

The data is then processed and added to the collection of mentions M (at the start $M = \emptyset$) in such a way that each $m_i \in M$ has associated with it details of: (i) the source it was gathered from and (ii) the time it was gathered. Gathering data from the web inevitable leads to duplication and noise. For this reason a parsing function was applied to each chunk of collected data in order to act as a filter to "clean" the received data before storage. This function first confirmed that the mentions in the currently collected data "reference" the brand in question, if not these mentions were removed. The parsing function then searched for and removed duplicate mentions in the data. Finally it checked that the remaining mentions weren't already included in M, if so these mentions were also removed. The remaining mentions in the collected data were then added to M. The physical storage was provided using a scalable cloud database. The collection process continues in this manner with new data being collected at regular intervals, mentions that have expired (are older than some user specified time span) are removed from M so as to prevent M from containing out of date information and/or scaling to a size where it can no longer be processed in almost real time. Once a reasonably sized collection M has been obtained it can be processed, analysed and viewed at any time.

4.2 Data Mining

The objectives of the data mining stage were: (i) to conduct sentiment analysis and (ii) the grouping of mentions into related topics. To achieve these objectives two data mining techniques were used. The first technique comprised a sentiment classifier using the class label set $\{positive, negative, neutral\}$. The second technique was a hierarchical clustering algorithm which was used to identify groupings of related topics in a hierarchy structure. This idea was to segment the data into related topics on a level by level basis. Grouping the mentions into topics allows for top level analysis to be done. For example identifying what the currently most popular topics related to a particular brand are and what the comparative preponderance of these topics is. Further details of these two processes are given below in the following two sub-sections.

4.2.1 Sentiment Analysis

The sentiment analysis method used was based on an off the shelf, Naive Bayesian, sentiment classifier called uclassify (www.uclassify.com). The data set used to train the classifier was 40,000 Amazon product reviews from 25 different product genres [1]. This sentiment classifier was selected because: (i) it was directly related to the application of BRM, (ii) preliminary testing indicated that the recorded accuracy was acceptable compared with other sentiment classifiers considered, and (iii) the training data uses the same critical language style as that of someone commenting on a brand or service. The uclassify classifier requires input in the standard feature vector format, this a set F of the form described above. The classifier returns a value between 0 and 1, where 1 indicates a positive sentiment and 0 a negative sentiment.

Table 1 shows the pseudo code for allocating a sentiment label s, taken from the set $S = \{positive, negative, neutral\}$, to a mention m. The input is a feature vector F describing mention m (as defined above). The output is a sentiment label s to be associated with the mention. The sentiment mining algorithm first determines the polarity value associated with the mention m (line 3), the second part of the algorithm then determines the sentiment class label to be assigned to the mention using the polarity value p and a threshold T. If $p \geq 1 - T$ (line 4) then $s = positive$ (line 5). If $p \leq T$ (line 7) then $s = negative$ (line 8). Otherwise $s = neutral$ (line 10). It was found experimentally that $T = 0.2$ was the most appropriate value to be

```
Algorithm Sentiment_Analysis (F)
T = 0.2;
s = null;
p = Bayesian_Classifier(m);
if p >= 1 − T then
    s = positive; {positive sentiment}
else
    if p <= T then
        s = negative; {negative sentiment}
    else
        s = neutral; {neutral sentiment}
    end if
end if
return s;
```

Table 1 Pseudo code for the sentiment analysis algorithm.

used in the context of the work presented in this paper. Although a sentiment class label set of size three was used it is clear from the above that a larger number of class labels (depending on the application domain and end user requirements) could equally well have been adopted.

4.2.2 Hierarchical Clustering

The purpose of the hierarchical clustering, as noted above, is to group mentions into a hierarchical format such that each mention belongs to one cluster in each level of the hierarchy. This idea is to reduce the potentially large number of mentions that may have been identified by "compartmentalising" them into smaller groups which can be readily understood and further analysed. In other words the purpose of the clustering is to identify topic hierarchies within the collection M. Once topic clusters have been identified statistical analysis can be conducted with respect to each cluster (topic or sub-topic). To obtain the desired hierarchical clustering a divisive (top-down) hierarchical clustering algorithm was used, the alternative would be a conglomerative (bottom-up) hierarchical clustering. Divisive hierarchical clustering operates in a breadth first manner by repeatedly dividing the candidates at each level and each branch of the growing hierarchy into k clusters. In effect it can be thought of as an iterative k means process.

The pseudo code for this procedure is shown in Table 2. The input is the set of mentions M and the number of clusters k desired at each level. It is a recursive process in that the *divisive_cluster* procedure is called with respect to each discovered branch. The process continues until the number of mentions considered at some leaf node in the hierarchy is less than k (line 1) or until as a sufficiently cohesive clustering configuration was reached (line 5). This was measured in terms of the Silhouette Coefficient [6]. The Silhouette Coefficient of a cluster configuration is a measure of both the cohesiveness and separation of a configuration. It is a number between -1.0 and $+1.0$. The nearer the coefficient is to +1.0 the better the cluster configuration. Experiments were conducted using a variety of values for k and it was found that $k = 3$ produced the most effective results with respect to BRM.

```
Procedure divisive_cluster (M, k)
if (|M| < k) then
    exit
end if
C = set of k clusters {c_1, c_2, ..., c_k}
if (Silhouette_Coefficient(C) > σ) then
    exit
end if
for all c_i ∈ C do
    M' = subset of M in c_i
    divisive_cluster (M', k)
end for
```

Table 2 Pseudo code for the hierarchical clustering procedure.

Once the hierarchical clustering had been completed each cluster was allocated a label designed to describe its content. This was done by extracting the words that appeared most frequently in the feature vectors representing the mentions contained in a given cluster and identifying the words whose occurrence count was above some threshold or if there were more than three frequently occurring words selecting the top 3. The identified words thus constituted the cluster label.

4.3 Visualisation

The following section presents details on the visualisation stage of the BRM framework. The objective of the visualisation stage is to effectively communicate the brand reputation patterns and new knowledge that has been determined in the previous data mining stage (Section 4.2). The approach used to achieve this was to create a sequence of charts and diagrams, in a dashboard style interface, which made the information easily digestible by the end users of the system in a visually appealing manner. The dashboard interface comprised several different visualisations, these individual visualisations can be categorised as follows: (i) Wheel Charts, (ii) Treemaps and (iii) Circle Packings. Each category of visualisation was developed in order to highlight a certain aspect of the data mining effort. The complete list of data mining visualisations is given in Table 3. Each of the visualisations is described in more details below.

The **Standard Wheel Chart** shows all clusters (with labels) in one visualisation. The Inner core of the wheel shows all the data; while each ring, moving to the outer edge, shows the sub clusters (identified by the hierarchical clustering) of the previous ring. This chart can be used to instantly see a comparison of the number of mentions in each of clusters.

The **Interactive Wheel Chart** displays data in a similar way to the wheel chart explained above, but in this case adds a level of interactivity. When a user selects

Table 3 A list and description of the visualisation techniques provided with the BRM framework

Chart Name	Representation	Function	Tool or API
		Wheel Charts	
Standard	Multi-Level Pie Charts	Displays hierarchical cluster relationships and proportions.	Fusion charts
Interactive	Multi-Level Pie Charts	Created using D3 Library with some added interactive user exploration features	D3 Javascript Library
Advanced	Multi-Level Pie Charts	Adds an extra layer of information related to the proportion of sentiments in each category.	D3 Javascript Library
		Treemap	
Treemap	Display hierarchical (tree-structured) data as a set of nested rectangles	Shows a visual representation of a data tree, clusters and sub-clusters. This chart can be used to explore various levels of the hierarchy.	Google Chart API
		Circle Packing	
Standard	A hierarchical layout using recursive circle packing	Displays each category in a circle and shows the hierarchical relationships.	D3 Javascript Library
Advanced	A hierarchical layout using recursive circle packing	Adds an extra layer of information in the form of the proportion of sentiments in each category with several pie charts.	D3 Javascript Library
Timeline	A hierarchical layout using recursive circle packing and multiple data sets.	Displays the changes in different data sets using animation; data sets may be from many different successive time periods.	D3 Javascript Library

a cluster the chart is transformed so that this cluster becomes the “root node” in the centre of the chart. This allows for fast exploration of the hierarchical clustering outcomes. Large volumes of data can be explored by navigating clusters and sub-clusters in this manner.

The **Advanced Wheel Chart** displays the hierarchical clusters within each level as segments of a wheel. The cluster size of each sentiment category is displayed as a variation of the main colour for each cluster. Adding the sentiment information as an extra layer of data, means this chart can be used as an alternative interpretation to the wheel charts above while still maintaining the advantages of viewing large data sets effectively.

The **Treemap Chart** shows blocks of varying sizes and colours. Each block represents a cluster (with its label). The size of a block is a reflection of the number of mentions within the cluster. Each block has a colour, from a variable range, which is used to indicate the overall sentiment category of a cluster. The chart also features an interactive element whereby a desired hierarchical level can be selected so that the user can drill down further into the data. The treemap chart allows the user to quickly make comparisons between clusters and sentiments at a single glance.

The **Standard Circle Packing Chart** is constructed by representing the clusters within a level of the hierarchy as a pie chart. An algorithm is then used to recursively display all clusters and sub-clusters in one chart.

The **Advanced Circle Packing Chart** is constructed based on the standard circle packing chart but in this case adding an extra layer of information which shows the sentiment categories and their relative cluster sizes.

The **Timeline Circle Packing Chart** is a variation of the advanced circle packing chart described above, but with an animation element. This chart can display mentions at various points in time. This chart can also display changes to cluster sizes and sentiment categories as the chart cycles over various points.

5 Case Study

The most appropriate mechanism for gaining an appreciation of the operation of the BRM framework is by considering an example. This section therefore presents a case study which uses the presented BRM framework as applied to a specific "real world" brand reputation problem. Section 5.1 describes the scenario used and how the data set was derived. Section 5.2 presents the resulting visualisations produced by the BRM framework with respect to the brand used in this case study.

5.1 Data set

The particular data set used as a focus for the case study is the popular TV talent show "Britain's Got Talent" (BGT); other countries have a similar show. The show provides a forum for members of the public to audition, a panel of judges select the best acts which then go on to the next stage where each week a number of contestants get voted off; this continues until there is one winner. This particular brand was selected as it is an example of a popular brand that has a significant presence on social media where "debates" are conducted concerning the antics of the contestants, presenters and the panel of judges. The collated BGT data set comprised 14,280 Twitter mentions, collected from April to June 2013.

It should be noted that the BGT brand used in this work has a strong correlation with other household brands (in the context of the proposed BRM framework). Thus the brand name, and associated hash tags, are used on social media platforms to identify discussions about the BGT brand in the same manner that any other brand might be discussed. The opinions of the judges and contestants of the BGT brand discussed on social media can be considered to be analogous to products or services offered with respect to more traditional companies/brands. The case study is therefore designed to demonstrate that customer satisfaction can be derived from the analysis of such media using the proposed BRM framework.

5.2 Results

This sub-section describes the outputs from using the BRM framework with respect to the BGT brand. The first part of this sub-section presents the results from the hierarchical clustering algorithm. The second part concentrates on the visualisation of the combined hierarchical clustering and sentiment analysis.

The hierarchical clustering algorithm was applied to the BGT data set. The resulting output is shown in Figure 2. From the figure it can be seen that the clustering algorithm identified some clear groupings regarding the BGT dataset. In particular it can be seen that the BGT brand has a clear set of top level sub topics related to the judges, contestants, presenters and the TV show itself. From the sub-clusters it can be seen that the TV show topic contains sub-topics regarding dissatisfaction with the scheduling overlap of BGT and another popular TV talent show ("The Voice"). With respect to the generic application of the BRM framework, this could be a particular brand with clusters showing the most popular talking points reflected on social media. This could also highlight issues with competitor's products or services.

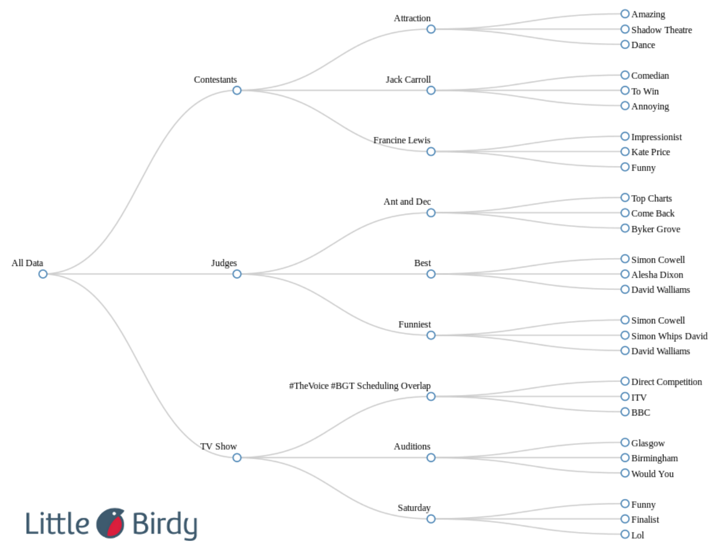

Fig. 2 Illustration of the hierarchical clusters within the BGT data set.

The **Advanced Wheel Chart** is shown in Figure 3. The chart shows all the hierarchical clusters and the cluster sizes with respect to each sentiment. The cluster with the most fragmented sub-clusters can be seen very easily (clusters at the top of the chart in Figure 3). The "TV show" and "Contestant" clusters have smaller sized sub-clusters representing more in-depth topics, for instance certain locations for auditions, and so on. The larger clusters represent topics of much broader interest. These cluster topics can be related to a more general aspect of the BGT brand.

The interactive element to this chart can also aid with the exploration of the information presented. This allows the user to drill down into certain topics of interest that may have had a previously unknown association with the brand. An example in the BGT case is that of the presenters topping the music charts with a re-released song. The sentiment categories are displayed in a pop-up box when the user hovers over a cluster.

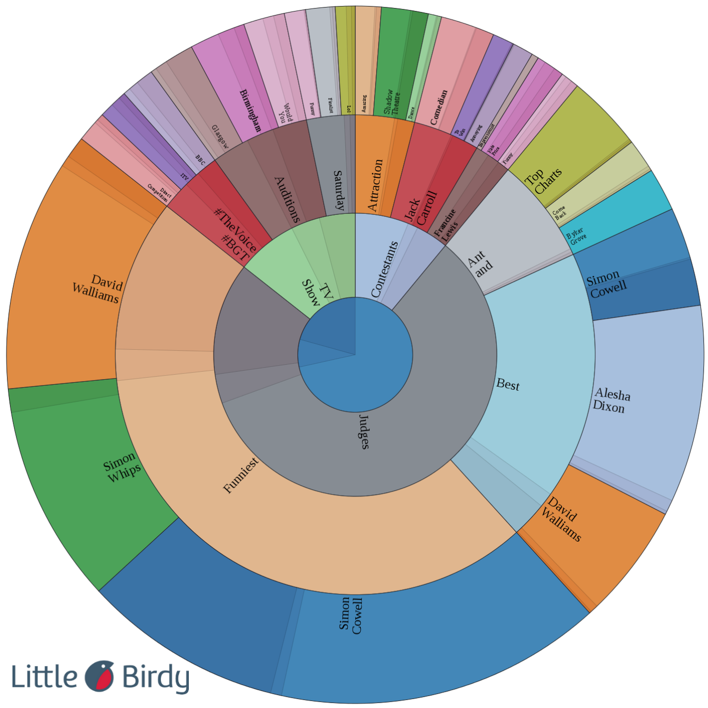

Fig. 3 Advanced Wheel Chart for the BGT data set.

The **Treemap Chart** is shown in Figure 4. The chart shows blocks of varying sizes and colours. The size is a reflection of the number of mentions, colour is used to indicate the overall sentiment category of a cluster, in this case red is negative while blue is positive. This chart very clearly highlights the sentiment of the most popular clusters within the hierarchy. For example it can be seen that the "Simon Cowell" cluster (top right) has an overwhelming negative sentiment while the "Byker Grove" (a no longer running TV show in which the presenters were featured) cluster (bottom left) has mostly positive sentiment. The Treemap chart allows the user to quickly make a comparison of the clusters and their associated sentiment at a single glance. Thus it can be seen that the popularity of the judges is indicated by the size and colour of the blocks, this is shown in the case of judge "Simon Cowell" (negative sentiment) and "Byker Grove" (positive sentiment).

The **Circle Packing Chart** is shown in Figure 5. The chart represents the clusters within a level of the hierarchy as a pie chart. The darkness of the colour of the pie chart reflects the sentiment categories, from darkest to lightest representing positive to negative. This chart includes an animation element. The animation can effectively display data at various points in time. This chart will display changes to cluster sizes

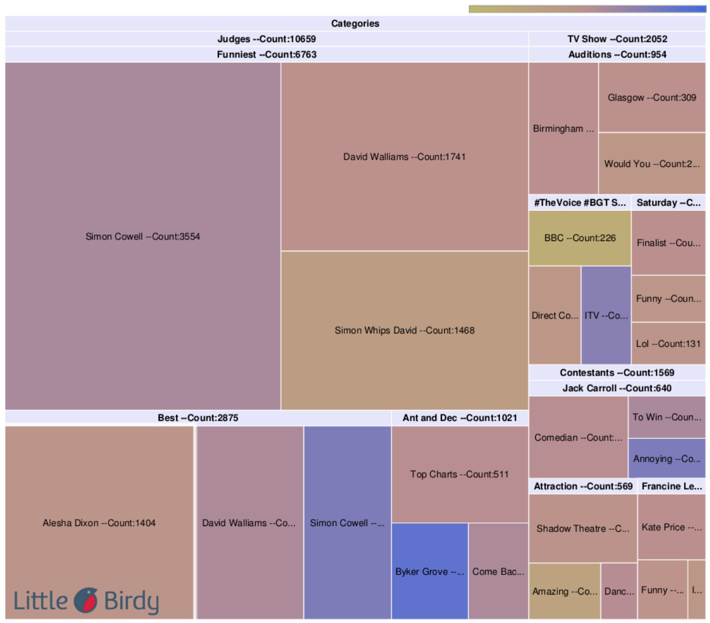

Fig. 4 Treemap Chart for the BGT data set.

and sentiment categories as the chart is "cycled" through time. It can thus be used to display changes in cluster size and sentiment of the various cluster topics over (say) a number of weeks. In terms of more generic brand reputation mining this can be considered as a mechanism for displaying change in brand perception over time.

6 Conclusion

This paper has presented the BRM framework to address the Brand Reputation Mining (BRM) problem. The BRM framework encompasses the identification and collection of mentions of a particular brand as expressed on social media and the wider www. The framework incorporates a clustering mechanism to identify a cluster hierarchy within a collection of mentions concerning some brand and mechanisms to determine the sentiment associated with individual clusters (elements or attributes of the brand in question). The final stage of the proposed framework consists of a set of visualisation tools to display the results of the clustering and sentiment analysis. The evaluation comprised a case study to demonstrate the effectiveness of the proposed BRM framework in the context of brand reputation management. The case study demonstrated that the framework can effectively be used to mine the online reputation of brands. Analysis of the visualisation demonstrated previously unknown patterns and provided for an insight into the brand that would have previously been lost or very difficult to discover via manual analysis.

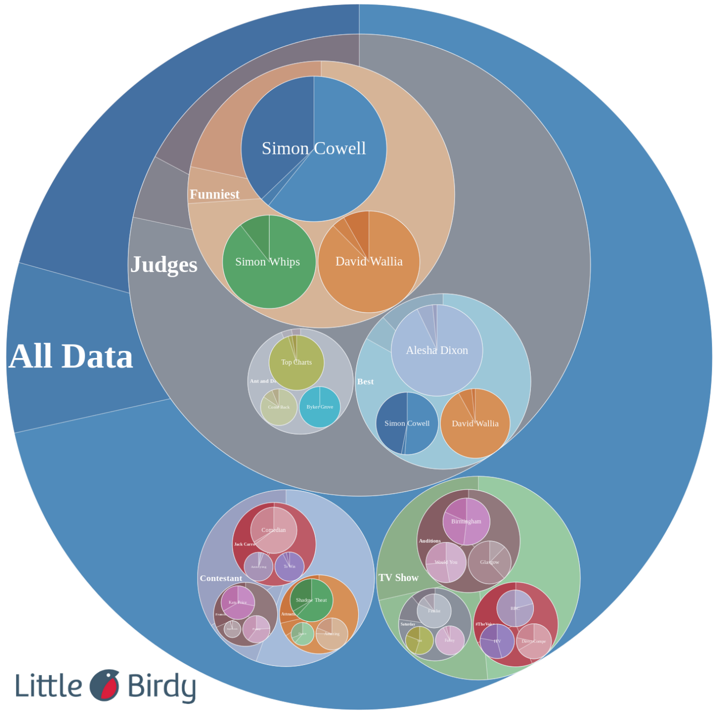

Fig. 5 Circle Packing Chart for the BGT data set.

References

1. J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *ACL*, 2007.
2. A. M. Kaplan and M. Haenlein. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1):59–68, January 2010.
3. W. G. Mangold and D. J. Faulds. Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4):357–365, July 2009.
4. S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In *The eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 341–349. ACM New York, NY, USA, 2002.
5. M. Morzy. New Algorithms for Mining the Reputation of Participants of Online Auctions. In *First International Workshop, WINE 2005*, volume 52, pages 112–121, Hong Kong, China, 2005. Springer Berlin Heidelberg.
6. P. J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 1987.
7. G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, November 1975.
8. F. Skopik, H. Truong, and S. Dustdar. Trust and reputation mining in professional virtual communities. *Web Engineering*, pages 1–15, 2009.
9. S. Spangler, Y. Chen, and L. Proctor. COBRA Mining web for corporate brand and reputation analysis. *Web Intelligence and Agent Systems*, pages 246—-254, 2009.
10. J. Yi and W. Niblack. Sentiment mining in WebFountain. In *Proceedings. 21st International Conference on Data Engineering, ICDE 2005*, pages 1073 – 1083, 2005.
11. C. N. Ziegler and M. Skubacz. Towards automated reputation and brand monitoring on the web. *Web Intelligence, 2006. WI 2006. IEEE*, 2006.