Social Media Analytics for YouTube Comments: Potential and Limitations[1]

Mike Thelwall

School of Mathematics and Computing, University of Wolverhampton, UK

The need to elicit public opinion about predefined topics is widespread in the social sciences, government and business. Traditional survey-based methods are being partly replaced by social media data mining but their potential and limitations are poorly understood. This article investigates this issue by introducing and critically evaluating a systematic social media analytics strategy to gain insights about a topic from YouTube. The results of an investigation into sets of dance style videos show that it is possible to identify plausible patterns of subtopic difference, gender and sentiment. The analysis also points to the generic limitations of social media analytics that derive from their fundamentally exploratory multi-method nature.

Keywords: social media analytics; YouTube comments; opinion mining; gender; sentiment; issues

Introduction

Questionnaires, interviews and focus groups are standard social science and market research methods for eliciting opinions from the public, service users or other specific groups. In industry, mining social media for opinions expressed in public seems to be standard practice with commercial social media analytics software suites that include an eclectic mix of data mining tools (e.g., Fan & Gordon, 2014). Within academia, there is also a drive to generate effective methods to mine the social web (Stieglitz, Dang-Xuan, Bruns, & Neuberger, 2014). From a research methods perspective, there is a need to investigate the generic limitations of new methodological approaches.

Many existing social research techniques have been adapted for the web, such as content analysis (Henrich & Holmes, 2011), ethnography (Hine, 2000), surveys (Crick, 2012; Harrison, Wilding, Bowman, Fuller, Nicholls, et al. 2016; Tijdens & Steinmetz, 2016), or network analysis (Ackland & Gibson, 2013; Corley, Cook, Mikler, & Singh, 2010). Some quantitative methods focus on counting web activities, such as tweets, retweets hashtags, keywords, or YouTube video comments, and use these for analyses of the levels of interest in a topic or set of resources (Bruns & Stieglitz, 2012; Sugimoto & Thelwall, 2013), or for a time series analysis of trends in interest (Bruns & Stieglitz, 2013; Stieglitz & Krüger, 2011; Thelwall, 2007). In contrast, the computational approach uses algorithms to generate new insights (Giglietto, Rossi, & Bennato, 2012). Examples include community detection based upon connections between users or resources (Jürgens, 2012), and automatic categorisation (Bouman, Drossaert, & Pieterse, 2012), with sentiment polarity being the most common example (Thelwall, Buckley, & Paltoglou, 2011). Using natural language processing techniques, it may also be possible to extract highly specific information, such as the symptoms of illegal drug

[1] This is an Accepted Manuscript of an article published by Taylor & Francis in *International Journal of Social Research Methodology*.

use at specific dosage levels (Chary, Park, McKenzie, Sun, Manini, & Genes, 2014). Theoretical analyses may also use an analytic approach (e.g., Dynel, 2014).

An alternative strategy, social media analytics (Stieglitz, Dang-Xuan, Bruns, & Neuberger, 2014), combines different methods to generate insights, such as, “sentiment analysis, topic modeling, social network analysis, trend analysis etc.” (Fan & Gordon, 2014). This extensive method triangulation and the ability to study a phenomenon dynamically (Edwards, Housley, Williams, Sloan, & Williams, 2013) partially offset the low sampling validity of social web data by merging different types of information (e.g., comments, likes, hit counts) to look for deeper insights. Although social media analytics software has been developed to enable the user to explore their data in multiple different ways (Burnap, Rana, Williams, Housley, Edwards, et al., 2015), combined methods or at least guiding frameworks are needed to help researchers to select and evaluate appropriate analysis strategies.

YouTube is a logical free source of social web data because it has been world’s second or third most popular website since October 2007 according to Alexa.com[2], and is already exploited by commercial organisations (Fan & Gordon, 2014). Twitter is analysed more in academic publications (e.g., Pak & Paroubek, 2010) but is limited by the presence of spam and bots, and restrictions on free data collection. The multiple purposes for YouTube and its international and inter-generational audiences make it a potentially valuable source of information about the act of watching videos, the issues depicted in them and their uses and gratifications.

YouTube has been investigated for the accuracy of videos about important topics (Briones, Nan, Madden, & Waks, 2012), its communication value (Lewis, Heath, Sornberger, & Arbuthnott, 2012), the content of a small themed set of videos (Jaspal, Turner, & Nerlich, 2014) and for the system itself (Thelwall, Sud, & Vis, 2012). There are also analyses of topics on YouTube that use a single primary method, such as content analysis (e.g., Desai, Shariff, Dhingra, Minhas, Eure, & Kats, 2013; Smith, Fischer, & Yongjian, 2012).

There are a few published general purpose social media analytics strategies as well as discussions of the advantages of combining methods (e.g., Lünich, Rössler, & Hautzer, 2012). The *Vista* method exploits time series visualisations to track changes in sentiment, username mentions and keyword frequency over time for an event on Twitter, allowing explorations of subtopics through deeper queries of the data (Hoeber, Hoeber, El Meseery, Odoh, & Gopi, 2016). A YouTube-specific general purpose method analysed the success of six anti-smoking videos through their manually filtered comments, various metrics (e.g., views, Likes), networks of interactions between commenters, automatically detected comment sentiment and a manual content analysis. These were combined to produce an evidence-based evaluation of the success of the campaign (Chung, 2015).

Finally, and most importantly, social media analytics methods have the following generic problems.

- They are lengthy to describe because they involve multiple methods. This makes them difficult to learn and fully evaluate. It also makes them an awkward object of academic research.

[2] http://web.archive.org/web/20090601000000*/http://www.alexa.com/topsites and http://web.archive.org/web/20050201000000*/http://www.alexa.com/site/ds/top_sites?ts_mode=global&lang=none

- Generic social media analytics method are complex since they involve a combination of automatic processing and human judgements (e.g., to filter out spam) as well as analytical decisions that are specific to any given topic.
- Given a large textual dataset, there are many different methods that could reasonably be used to analyse it. For example, topic modelling is a common approach not used here. There are also many available variations of the core methods – such as the different inter-document distance metrics that could be used within network diagrams. A researcher must therefore make a pragmatic decision to limit the number of analyses carried out and the variations to be explored. This contrasts greatly with simple numerical datasets and defined problems, where there may be a single best analysis (e.g., paired t-test).
- Any method is difficult to effectively evaluate from a general perspective since it may work well for one topic but not another, it has multiple components that would need to be assessed separately, and possibly in different ways. For example, the current article is long despite analysing a single topic, omitting the background information about that topic necessary for the reader to evaluate the findings, and presenting a limited subset of the results.

The current article illustrates the above problems by introducing and critiquing a method for gathering and analysing YouTube comments on a relatively large scale (larger than any previously published academic analyses of a topic on YouTube) to gain quick free insights into an issue, with a focus on gender, sentiment and discussion themes. An application to dance videos is used to discuss the generic potential and limitations of research methods based on social media analytics.

Research questions

This paper introduces the Comment Term Frequency Comparison (CTFC) social media analytics method to investigate YouTube culture around a specific topic. This paper critiques this method with four specific research questions.

- RQ1: Can the CTFC method identify plausible and/or insightful subtopic dimensions of discussions about a topic in YouTube comments?
- RQ2: As above for gender dimensions.
- RQ3: As above for sentiment dimensions.
- RQ4: As above for networks of relationships between topics.

The YouTube *CTFC* method comprises a technique for gathering relevant YouTube comments and techniques for analysing them, all supported by the free Webometric Analyst and Mozdeh software. Like grounded theory and similar qualitative approaches, the primary analysis method is exploratory and can be characterised as a (sophisticated) fishing expedition because its goal is to gain insights into the topic rather than to test hypotheses. It is therefore impossible to assess the efficacy of the method rigorously. Moreover, the method is a combination of multiple different approaches so that it would be difficult to perform a more sophisticated evaluation based upon multiple topics. Nevertheless, there is a need for general purpose exploratory YouTube analysis methods, given the popularity of the site, and these limitations are generic to any such attempt.

The CTFC method

This section gives an overview of the CTFC method. For an extended version with additional technical details and software instructions, see Appendix 3.

CTFC Step 1: Data gathering and filtering

The data gathering step involves creating a list of relevant videos and downloading their comments. The recommended data gathering stages are illustrated here with the dance videos topic.

(1) *Topic definition and delineation*: The scope of the project is defined at the outset. This is designed to guide decisions about which videos are relevant and should be guided by the goals of the research project. *Dance*: the scope of the project is videos with a primary focus on a single dance style, including instructional and performance videos.

(2) *Initial subtopic query set generation*: The topic is split into a set of subtopics and a YouTube query generated to match each one. *Dance*: The subtopics are dance styles and the initial query set was generated from a Wikipedia page on Dance style categories (en.wikipedia.org/wiki/List_of_dance_style_categories) with dance style names recorded as phrase searches.

(3) *Query testing and refinement*: Each query in the set generated in the previous step is tested in YouTube.com to ensure that it generates a high percentage (90% as a guideline figure) of correct matches.

(4) *Video list generation*: Software is used to download a list of videos matching the queries. *Dance*: The list of 36,702 videos with title matches was used.

(5) *Video list checking*: The list of videos matching the queries is checked and false matches removed.

(6) *Comment downloading*: Software is used to download the comments on the matching videos.

(7) *Duplicate commenter removal*: Users are restricted to a maximum of one comment each to protect the word frequency analysis from the actions of prolific individuals.

(8) *Comment pre-processing*: The comments are loaded into the analytics software.

(9) *Language filtering*: Comments not in the chosen language are filtered out. A simple way to achieve this approximately is to exclude all comments that do not contain any terms that are common in the selected language and rare in others (Grefenstette, 1995).

The above steps produce dataset of YouTube comments in the chosen language, together with the identity of the videos commented upon, the commenter and the query used to find the videos.

CTFC Step 2: Time series graph

A time series graph is produced to assess the typical dates of comments as background information.

CTFC Step 3: Subtopic word frequency analysis

The subtopic word frequency analysis seeks issues that are characteristic of each subtopic in the sense of being more discussed by people that comment on the subtopic than on other subtopics of the broad topic. It is important to keep the comparison within the broad topic so that the characteristic issues of subtopics are more fine-grained. A related approach has previously been used with Twitter (Bryden, Funk, & Jansen, 2013). Subtopics could also be analysed using topic modeling (Blei, Ng, & Jordan, 2003) instead of word frequencies. Topic modeling produces clusters of mutually related terms that tend to represent topics. In the current context, topic modelling would identify issues within each individual subtopic but could not then compare them between subtopics and so it is not used here.

The word frequency analysis method is to compare the frequency of terms that match one subtopic with their frequency for the remaining subtopics to find terms that are unusually frequent for the chosen subtopic. The chi-squared statistic is used for this by listing terms that are more frequent for the subproject in descending order of chi-square value.

This method is imperfect for several reasons. It relies upon individual words whereas concepts may be expressed through phrases. If key concepts are expressed using phrases that only include common words (e.g., “to be or not to be”) then the method will not identify any of the words within the phrase and the concept will be missed. This could be resolved by using natural language processing techniques to extract key repeated phrases, but this uses much computing time and adds to the amount of information to be processed. This alternative phrase-based approach may be considered if there is relatively little information or if a deeper analysis is needed and time and resources are available for it. A second problem is that the comments may contain substantial unrelated discussions, such as an off-topic argument between two commenters, and these can result in spurious issues being identified. This can be guarded against by reading a random sample of comments from the subtopic containing the identified keyword. Finally, the terms for each individual subtopic depend on the other subtopics in the set. For example, if there is only one subtopic about swing dancing then the term “swing” is likely to be near the top of its term list, but if there are several swing subtopics, then this would make the term less specific to the subtopic and reduce its chi-squared value. There does not appear to be a solution to this issue and so it must be accepted that the results of a word frequency analysis are not necessarily comprehensive.

The top terms (e.g., the 50 with the highest chi-square value) identified by the basic word frequency analysis are manually examined for relevance and insights into the subtopic. A term’s importance should be interpreted using the fact that it is relatively frequent in subtopic comments compared to the other subtopics. The top terms are likely to include predictable nouns associated with the subtopic, including the subtopic name, but terms with a less obvious association are more interesting. The presence of irrelevant terms that need to be identified manually and removed is expected from any word frequency comparison approach (Thelwall, Prabowo, & Fairclough, 2006).

CTFC Step 4: Gender differences analysis

The gender differences analysis seeks issues that disproportionately originate from male or female commenters. This can point to gender differences in opinions about the videos as well as the aspects of the videos that are discussed, although not their causes. Gender differences may give insights into a topic even if a study does not have a specific focus on gender. For this, terms in the female-authored comments are compared to terms in the male-authored comments (a) overall for the project and (b) for each individual subproject. Although commenter gender information is (no longer) provided by YouTube, it can be inferred from commenters’ names. Commenters must be registered within YouTube and have a channel name (possibly with spaces) and a username (without spaces). Potential first names should be extracted from the first of these and from the second if it uses camel case (e.g., PardeepSingh) and then compared against a dictionary of common first names, organised by predominant gender.

In principle, it may also be interesting to conduct comparative word frequency analyses by age range, nationality, and other demographic variables but no information about these is available from YouTube.

CTFC Step 5: Sentiment analysis

The sentiment analysis seeks issues that elicit particularly strong positive or negative sentiment. These are identified by a word frequency comparison of comments with strong sentiment against the remainder. This is achieved through automatic sentiment strength detection, such as with the program SentiStrength that is designed for the short informal text of social media posts, including YouTube comments (Thelwall, Buckley, & Paltoglou, 2012). Other sentiment analysis programs (e.g., Taboada, Brooke, Tofiloski, Voll, & Stede, 2011) are likely to give broadly similar results. For positive sentiment, the technique is the same as for the basic word frequency analysis except for splitting the comments into two sets, one of which has a positive sentiment strength score of at least 3 (moderate) out of 5, and the second set containing the remaining comments. For negative sentiment strength, the procedure is the same except that the split is based on comments having a negative sentiment strength of at least 3 out of 5. Both analyses are conducted across the whole project or within individual subprojects.

CTFC Step 6: Overview network

The overview network is a network diagram of strength of relationships between subtopics. Although subtopic similarity can be analysed by many different statistical methods, such as clustering, factor analysis or multi-dimensional scaling, the method recommended here is a simple network diagram, with subtopics represented by nodes in the network and lines occurring between nodes when their comments tend to use similar words. This network is therefore a comment term similarity network. A network is preferable to cluster analyses and multidimensional scaling because subtopics on YouTube can be expected to relate to others in multiple different ways, rather than through a few dimensions of difference.

The logical metric to use to compare subtopic term similarity is the standard information retrieval metric of cosine similarity based upon Term Frequency and Inverse Document Frequency (TF-IDF). This measures the virtual angle between two subtopics, with wider angles occurring for pairs of subtopics that tend to use different terms. Here the term frequency is the number of comments within the subtopics that contain a term rather than the total number of terms in all comments within the subtopics (i.e., multiple occurrences of a term within a single comment are ignored). The TF-IDF weighting gives more importance to terms that are common for the two subtopics compared and rare in other subtopics. Other measures of document similarity could also be used (e.g., Huang, 2008).

CTFC Steps 0 and 7: Pilot testing and insight verification

The above analyses are likely to be influenced by spam, off-topic conversations in comments, and off-topic videos. The initial results may therefore not give substantial insights. Two strategies are recommended to maximise the value of the method: pilot testing and insight verification.

A small-scale initial pilot test is run to identify any large scale obvious problems with the initial queries for subtopics. All analyses are attempted in the pilot study to quickly check for likely sources of off-topic information so that the subtopic queries can be adjusted to avoid them in the main test.

The second strategy is to trace and verify the causes of all insights during the main analysis. This means reading comments containing the relevant terms, except in the case of the network diagrams, where this is not possible. In cases where the cause of an apparent insight is spam or off-topic comments, the insight should be ignored. If there are too many problems at this stage then the analysis is repeated with modified queries to eliminate the problems. Thus, for a particularly serious project, the method is repeated until the insights reflect the topic rather than spurious factors.

Results

This section describes the result of applying the CTFC method to the topic of dance in YouTube videos, as described above. See Appendix 1 for background information about the issue.

Step 2: Time series graph

Comments on the videos matching the search spanned an 11-year period, with disproportionately many from recent years (Figure 1). The increase over time is partly due to older videos being deleted and older comments not being retrieved for videos with greater than the per video maximum from the YouTube API.

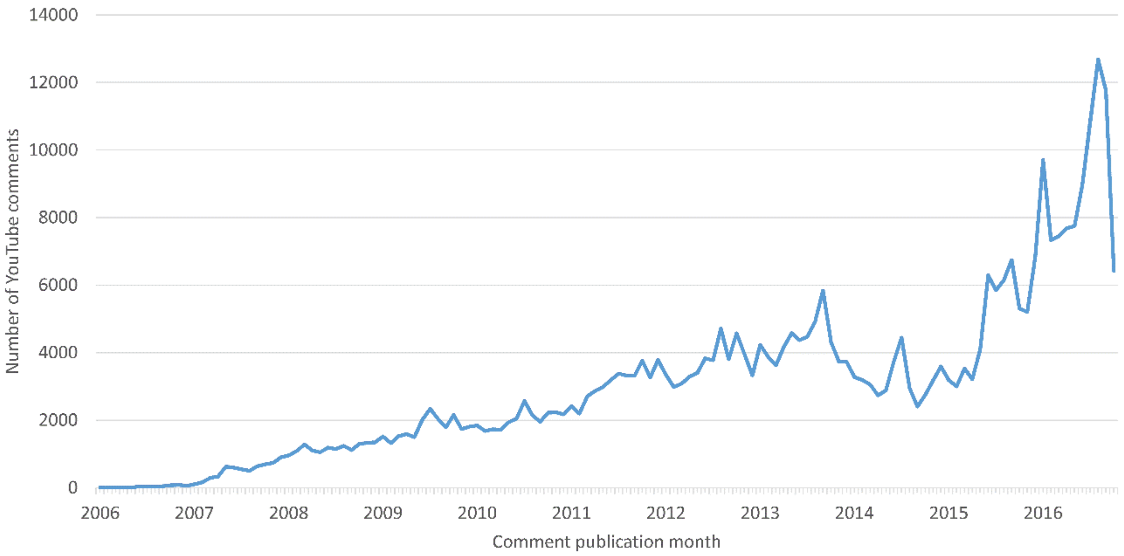

Figure 1. Overall comment frequency per month for the complete dataset.

Step 3: Subtopic word frequency analysis

For each of the 30 dances (i.e., subtopics), the five words most strongly associated with the videos matching the dance style searches were analysed (Table 1). These are terms that occur more frequently in comments for a given dance video set compared to the remaining dance video sets. Although a fuller analysis would cover more terms, analysing the top five for each dance may give insights into the aspects of the dance that are most comment-worthy on YouTube, and hence help to evaluate transparently whether CTFC Step 3 is likely to be useful.

Most of the top terms are dance genre names, dancer names, songs, or dance moves (Table 1), none of which seem likely to give deep insights into an individual dance genre. A partial exception is that in some cases the music-terms point to an above average degree of association between the dance and the music, or a subsumption of the dance within the music (hardcore dancing, electro dance). Terms that mention a country or nationality are not particularly insightful either, except in the case of Bollywood dance, where analysis of the comments containing the terms points to an individual high profile example of the dance, given by a successful Miss America candidate as part of the contest.

Sentiment terms are rarer in Table 1. Contemporary dance is disproportionately seen as being *beautiful*, Hit Dem Folks as being *lit*, with the dancers *killing* it, whereas Zumba is *fun*. In the first two cases the positive reception of the dances are partly language-specific, however. The contrasting language styles of these two cases point to the importance not only of the sentiment expressed but also the term selected to express it. Thus, the other dance styles are either appreciated less or are appreciated with more commonly used terms. Thus, the key sentiment finding is that both dance genres are appreciated with unusual terms. For Zumba, the term fun suggests enjoyment through participation, perhaps set against a context in which people expect not to enjoy exercise.

The context terms help to identify topics that are driven by a primarily non-dance issue, such as those associated with a video game (Ballroom, Disco dancing, Electric boogaloo). In addition, they point to ritual contexts in which a dance may be used (Bachata, Cheer dance, Hardcore dancing) or a dance’s origins (Crip walk).

Overall, then, whilst most of the terms do not give deep insights into dance genres, some may and the terms can also point to data collection problems and individual important contexts for the dance.

Table 1. The top five terms for each dance. These are relatively common in comments for each dance compared to comments for other dances (English comments only, includes non-gender commenters). Terms associated with the dance name, moves, music, and origins are in bold. Pronouns and names of people are in italic. Sentiment terms are underlined. Terms associated with appearance are italic and underlined. Terms associated with lessons are bold and underlined.

Dance genre	**Top 5 terms**	**Comment**
Acro dance	**acro** *Chloe Abby Katrina* **tuck**	Back tuck: move
Bachata dance	**bachata Dominican** *chambelan quince* **ven**	Ven tu: song name
Ballet dance/dancing	**ballet pointe ballerina** *Kaylee* **splits**	On pointe, splits: positions
Ballroom dance	*Squall Rinoa* ff8 fantasy **ballroom**	ff8: Computer game
Belly dance	**belly** *Sadie Shakira Joven Ellen*	
Bollywood dance	**Indian Bollywood India** *Nina* America	Miss America, Nina Davuluri
Breakdancing	*Bridgette* **breakdancing** *Roger* **bboy** *Mr*	
Cheer dance	**cheer** team **cheerleader** tryout **cheerleading**	
Contemporary dance	**contemporary** *Travis Tate* beautiful *Robert*	
Crip Walk	**crip walk blood gang** *Gosu*	
Disco dance	*toby* **disco** *Jessica Sylvester Rayman*	
Disco dancing	map *muselk* server *Alfie* **disco**	Team Fortress 2: Computer game disco-themed custom map
Electric boogaloo	*Jesse* game **boogaloo** space *Karkat*	Dragon Age 2, FTL: Computer games
Electro Dance	**electro** mix *yarus* remix mixes	*Electro dance* is a music style name
Hard dance	*Monstercat* **hardstyle** *Stonebank Ravine* **hard**	*Hard dance* is a music style name
Hardcore dancing	**hardcore** pit **mosh moshing metal**	Moshing, mosh pit: dances; Metal: music style
Hip Hop dance	*Kenneth Matt* Jabbawockeez *bailey dana*	Jabbawockeez: Hip Hop dance crew.
Hit Dem Folks	lit **folks hit** *Meechie* killed	Lit, killed: excellent;
House dance	*momo* **house mix** *Junho Chaeyeon*	“House dance mix” a common phrase
Jazz dance	**jazz** flute **turn pirouette jazz-funk**	“Freedom jazz dance” piece for flute; Pirouette: dance move
Jumpstyle dance	**jumpstyle** *Klaas elan* hardjump tecktonik	Hardjump, tecktonik: dance styles
Kpop dance	**kpop** *Infinite BTS VIXX Miu*	
Line dance	**line wobble cotton** boots **linedance**	Wobble: dance move; Cotton eyed joe: song; Boots: clothing
Lyrical dance	*Jadine* **lyrical** *Nadine Gerald Ade*	
Melbourne Shuffle	**shuffle shuffling melbourne shuffler** lmfao	Lmfao: funny
Popping dance	**popping** *Dytto Hoan Nelson Gucchon*	
Reggaeton dance	**reggaeton** *Maga Yomo Inga* calla	Bella Calla: song
Robot dance	**robot** *Usher mj Michael Jackson*	
Salsa dance	**salsa** *Eugene Claudia* **Cuban** quimbara	Quimbara: song
Tap dance	**tap tapping** cover Irish *Tamera*	Just (Tap) Dance: Song (cover version); Mislabelled Irish dance video;
Zumba dance	**zumba** workout *Vijaya Madelle* fun	Zumba is a dance fitness program

Step 4: Gender differences analysis

The five words most strongly associated with females and males within the comments for each dance style (Table 2) were analysed for insights into gender differences. Again, the purpose of analysing just the top five terms is to evaluate whether Step 4 is likely to be useful.

Females. In the entire corpus, the 10 most female-associated terms are: she, amazing, her, beautiful, cute, omg, belly, ballet, really, workout. These include female pronouns, dance styles, positive sentiment terms and a use of dance (workout).

Positive sentiment in individual dance genres: Positive sentiment terms are a female-associated characteristic, including cute (M:1 topic; F:6 topics), omg (M:0; F:8), amazing (M:0; F:8), liked (M:0; F:2), thanks (M:0; F:3), awesome (M:0; F:1), good (M:0; F:1), beautiful (M:1; F:3), but with some exceptions: great (M:1; F:0), sexy (M:1; F:0), hot (M:1; F:0).

Pronouns: Pronouns are a female-associated characteristic, including she (M:0 topic; F:8 topics), your/you're/u (M:0; F:4), her (M:0; F:3). Pronouns (and thanks: M: 0, F:3) suggest a more direct involvement in the people in the video (she/her) as well as exchanges with other commenters (your/you're/u).

Appearance: Comments on the appearance of the dancers are a female-associated characteristic, including cute (M:1 topic; F:6 topics), shoes (M:0; F:1), hair (M:0; F:1), makeup (M:0; F:2), outfit (M:0; F:1), with the exception of sexy (M:1; F:0), hot (M:1; F:0).

Learning or practicing: Terms associated with learning are female-associated, including learn/learned (M:0; F:1), class/classes/teach (M:0; F:2), workout (M:0; F:2), and perhaps also can/can't (M:1; F:3).

Table 2. Top words (highest chi-square) for female compared to male authored comments for the dance and vice versa. Terms associated with the dance name, moves, music, and origins are in bold. Pronouns and names of people are in italic. Sentiment terms are underlined. Terms associated with appearance are italic and underlined. Terms associated with lessons or workouts are bold and underlined.

Dance subtopic	**Top 5 terms for females**	**Top 5 terms for males**
Acro dance	how **acro** year *Abby u*	original watch must spectacular honest
Bachata dance	main *dress* cute chambelan* did	en *Edwin* salsa **bachata** style
Ballet dance/dancing	*she her* **pointe** when year	god art watched system culture
Ballroom dance	cute omg *makeup* **ballroom classes**	game best second already people
Belly dance	**move** *skinny* thanks *she I'm*	*Joven* sexy *he* instant hot
Bollywood dance	*she* with else *you're* loved	**India** shit at anti-white since
Breakdancing	*Annie* snow **house** *Bratayley* gymnastic	shit game **bboy** bitch *Wolf*
Cheer dance	**cheer** trying out do team	NU** sa ang na lang
Contemporary dance	amazing beautiful *Maddie you're* use	funny bitch front film bich
Crip Walk	liked amazing good xx hell	**walk** c clown gangsta real
Disco dance	**workout** *Jessica* fun challenge day	**mix** lo does who la
Disco dancing	*she* amazing *her* beautiful cute	shit fuck shuffle man fucking
Electric boogaloo	*Karkat* hardy voice act too	better does ship lot *Jesse*
Electro Dance	*shoes* wonderful kept watching walking	don't keep fuck **track DJ**
Hard dance	cute v *makeup hair* beautiful	on **hardstyle** dj remix fucking
Hardcore dancing	*she* girl **hardcore** omg shower	**hardcore mosh** *your* fag if
Hip Hop dance	amazing *Kenneth she Dana* omg	crew *Jabbawockeez* shit name best
Hit Dem Folks	killed cute her *outfit* liked	shit bro kid cop *tre*
House dance	*Junho Momo* ending *Taecyeon* Wooyoung	mix man fuck music name
Jazz dance	**turn** amazing *she* really loved	*he* guy isn't play playing
Jumpstyle dance	can amazing **jumpstyle** omg haha	can man **jumpstyle hardjump** *they*
Kpop dance	really infinite them best vixx	**Asian** girl porn beautiful sex
Line dance	fun make **class** have **teach**	de DJ left dude old
Lyrical dance	*your* well on did have	kc night similar took man
Melbourne Shuffle	omg xx thanks meh awesome	smack alpha twin vid nice
Popping dance	omg *Dytto she* cant amazing	name more *John* **popping** battle
Reggaeton dance	cute am can don't **reggaeton**	cute bro hot *ass* all
Robot dance	*Michael* omg amazing wow *usher*	guy camera judge **popping** up
Salsa dance	*Eugene Claudia judge* really couple	video great name *Anthony* from
Tap dance	much omg been **learn learned**	play deck guy bass band
Zumba dance	fun thanks much **workout** *Jessica*	Google information *his* shed esta

* Chambelan (Spanish) escort **NU: National University (Philippines). Related terms are from bilingual comments.

Males. The 10 most male-associated terms in the whole corpus are: shit, fuck, shuffle, man, fucking, crip, dude, bro, shuffling, hardstyle. These include male common nouns, dance styles and swear words.

Swearing: Strong and moderate swearwords are male-associated, including fuck/fucking (M:4 topics; F:0 topics), and shit (M:5; F:0). The term shit was usually associated with mildly negative sentiment, but also had a few positive uses, such as in the phrase, "[pronoun or name] killed that shit". Fucking was usually employed as a booster term in a negative context and fuck was used in varied negative ways.

Music: Name is a male gendered term (M: 4; F: 0) and typically associated with a comment requesting the song name. This aligns with the higher use of explicitly music-related terms by males, such as play/playing (M:2; F:0), mix/remix (M:3; F:0), and music (M:1; F:0).

Common nouns: Several generic terms for people, especially male(s), were male associated, including people (M:1; F:0), man (M:4; F:0), dude (M:1; F:0), bro (M:2; F:0), and bitch (M:2; F:0), but not girl (M:1; F:1). This suggests a more abstract perspective compared to the use of pronouns by females.

Overall, the gender differences analysis seems to give much more substantial insights into gendered reactions to videos, both overall and for individual dance genres. The strongly gendered comments are unsurprising given the importance of gender for most types of dance.

Step 5: Sentiment analysis

The ten terms associating most strongly with positive sentiment (see online appendix) overall were: **Please; nice; wow; beautiful; loved; job (e.g., nice/great/good job); pretty; hope; perfect; keep (going/up the good work/it up)**. For five dance genres, the term dancer suggests positive comments about the dancers, and for others the terms hilarious/lmao (8 dances) suggest amusement. Other occasional sentiments are respect (4) and inspire/inspired/inspiration (4). Occasional topics include mix/mixes (4) and workouts (4).

For the Melbourne Shuffle, terms in commonly echoed track listings were excluded for sentiment classification of the titles (*Scantraxx Roots - Headhunterz Vs Abject Superstar DJ - Dark Oscillators Smack my derb - Alpha Twins Young Birds - Patrick Bunton*).

The ten terms associating most strongly with negative sentiment overall were: **Shit; fuck; killed; stupid; wtf; hate; idiot; dislike; die; dead**. For several dances terms associated with fear are evident (afraid/scared: 4 dances) suggesting performance worries (except for the Kpop comments). The negative sentiments in some comments were expressions of sadness elicited by a performance, which is implicit praise for it: crying/cry (4), sad (1 - 7 times used in other contexts). Another common term is annoying (7 dances), directed at dancers, commenters, minor parts of the video, and the video production.

The terms choreo/choreography/choreographer (5 dances) caused sentiment classification errors (matching the base negative term stem chore*) and were excluded from the results, as was the term holy, as found in the phrases "holy shit" and "holy crap". The songs *I'll Hurt You - Busta Rhymes ft. Eminem* and *The Quick and the Dead – Rudebrat* caused incorrect sentiment results for Popping dance and their constituent terms were excluded.

Overall, whilst the negative sentiment terms seem to point to individual incidents rather than general themes about the dance genres, the positive terms suggest different ways in which the videos are enjoyed (i.e., gratifications gained from them).

Step 6: Networks

The network (Figure 2; see also Appendix 2) show the existence of several clusters of dances with videos that have attracted similar comments.

Melbourne Shuffle/Hard dance/Jumpstyle dance connecting to *Electro dance/ House dance*: These all have more male than female commenters and are participatory rather than performance dances.
Acro/Ballet/Contemporary/Jazz: The first three of these are classical dance styles jazz dance and all four are theatre-based performance arts. These all have more female than male commenters.
Lyrical/Cheer/Zumba dance: These do not have a natural association except for having more female than male commenters. Lyrical dance was formed from ballet, jazz and contemporary dance and so fits better within a different cluster. They associate because all have some bilingual English/Filipino comments.
Salsa/Bacheta/Reageton: These are Latin dance styles. They have more female than male commenters.

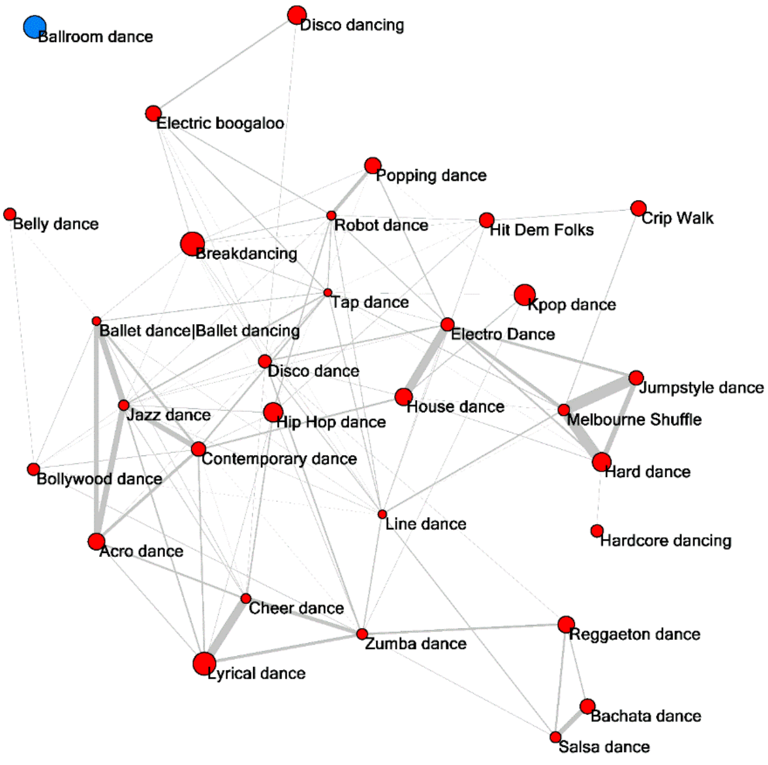

Figure 2. A network of the (cosine) similarity between the terms used in comments posted to each dance topic videos. Only the strongest 25% of all connections are shown. Disconnected nodes are shaded in blue. Thicker lines indicate higher cosine similarity between topics. Node area indicates comment volume.

It is difficult to evaluate the usefulness of the networks since they point to patterns that seem reasonably obvious but also have unexpected gaps.

Discussion and limitations

RQ1: Can the CTFC method identify plausible and/or insightful subtopic dimensions of discussions about a topic in YouTube comments? Whilst the subtopic comparison results seem to be broadly plausible (the discussion around Table 1), none of the terms seem likely to surprise an expert on these dance styles and so it is difficult to ague that the method has been insightful.

RQ2: Can the CTFC method identify plausible and/or insightful gender dimensions of discussions about a topic in YouTube comments? The method was successfully able to identify a range of themes and attitudes that were predominantly from either females or males for individual dances and across the dance topic. For instance, there was a male focus on music for some dances and a female association with positive sentiment. Whilst these seem plausible, none seem obvious and so it seems reasonable to claim that they are insightful.

RQ3: Can the CTFC method identify plausible and/or insightful sentiment dimensions of discussions about a topic in YouTube comments? The sentiment terms gave some plausible insights into why dances were liked, such as for individual dancers or if they were considered to be humorous. This information seems plausible and may also be useful to a researcher of dance genres. It could therefore be claimed to be somewhat insightful.

RQ4: Can the CTFC method identify plausible and/or insightful networks of relationships between topics? The network diagrams were plausible in some of the clusters but with clear anomalies. Network diagrams are difficult to evaluate for accuracy but can be useful a starting point for interviews with experts to trigger discussions and as a starting point for their analysis of structure (Cross, Borgatti, & Parker, 2002). The network diagrams presented here seem to be adequate for this purpose. A limitation of the networks is that it is impossible to see the reason for the strength of the connections in the graphs and they could be due to similarity in language styles, sentiment or topics of discussion, and are partly due to coincidences.

Both the networks and the word frequency analyses point to some subtopics not being a good fit for the dance category. One problem is that the method used to select the dances for analysis apparently selected at least one rare dance (Electric Boogaloo) that attracted many videos and comments only because of its inclusion in a video game. The findings around this dance were therefore not insightful about dance in general or even about this dance itself. The Bollywood dance example is similar in the sense that discussions around the topic were arguably not relevant to dance as much as the wider social and cultural issues triggered by its performance by a Miss America winner. Issues such as these could be circumvented by an additional round of manual filtering to remove Electric Boogaloo and Miss America-related comments. In practice, a decision to conduct this extra round of filtering would depend on the underlying goals of any analysis.

A generic problem for any analysis of YouTube comments is that “the feedback from those who did not post comments is unknown” (Chung, 2015). Away from YouTube, the attitudes of people who do not watch relevant YouTube videos is similarly unknown.

An important limitation is that the results influenced by the mix of types of dance video for each genre. Some were instructional, others were professional or amateur performances, and others seemed to be more about the music than the dance. Whilst it would have been possible to manually filter the videos so that they would only have been about one aspect, such filtering is time consuming and reduces the total number of comments available for analysis.

A more generic limitation is that the value that can be extracted for a topic may vary considerably between topics, with some yielding nothing and others perhaps yielding more than dance. Thus, the current paper illustrates the potential and limitations of the method but does not prove its usefulness for any given other topic. Another generic limitation is that the word frequency methods include a degree of unpredictability because key concepts that do not have unique names may not be picked up. Thus, important subtopic, gender and sentiment issues may have been overlooked because they were typically described with common words and hence did not rank highly in the word lists. Thus, the CTFC method is intrinsically not comprehensive and prone to overlooking important issues. In addition, whilst some comments were about the dances themselves but others were about specific events in videos that might not be relevant to the dance. Discussions about a fight at a salsa dance are an example of this. Highly commented videos that mention the topic frequently in a peripheral way (e.g., Electric Boogaloo; perhaps Miss America for Bollywood dance) are also a problem for detailed insights and point the importance of manual filtering and keeping in mind clear goals for an analysis to aid this filtering.

Finally, the implementation and evaluation of the CTFC method are subjective to the author and it is a human trait so identify patterns where there are none, so the interpretation of the positive aspects of the data may be optimistic.

Conclusions

The CTFC method has described some aspects of the dances analysed and the context in which they are discussed on YouTube. The results highlight gender, sentiment and sub-topic differences between the dances that could serve as a starting point for deeper analyses of the topic, such as through interviews or ethnographies. Although the method is supported by software, manual checking is needed of the initial queries and the term frequency lists produced, and the results seem likely to be influenced by some irrelevant discussion for any topic.

Applications of the CTFC method are most likely to be successful for topics that are extensively discussed on YouTube, especially if the discussions tend to be narrowly focussed on the topic of interest rather than on other issues. The method would therefore be particularly useful for discussing large scale YouTube-specific phenomenon but might also be useful in other contexts to give an initial exploratory analysis of an issue that has not been researched before. In this context, some of the findings might be useful to confirm or deny the researcher’s initial understanding of a topic that has not been researched much before and for which there is not a body of background evidence to rely upon.

> Finally, the results illustrate that social media analytics methods are almost inevitably *exploratory* and hence, even though they are likely to involve quantitative methods, are unlikely to be assessable through traditional hypothesis testing because the null hypothesis would not exist before the analysis. For social media analytics, this echoes a previous recognition that in the age of knowing capitalism sociology needs, “a radical mixture of methods coupled with renewed critical reflection” (Savage, & Burrows, 2007, p. 896; see also: Savage, & Burrows, 2009; Beer & Burrows, 2013) that is likely to be increasingly descriptive. Text-based social media analytics are challenging because they are not basic facts, survey question responses or discussions with researchers; instead they use a range of strategies to make limited and sensible exploratory deductions from collections of public web texts.

References

Ackland, R., & Gibson, R. (2013). Hyperlinks and networked communication: a comparative study of political parties online. International Journal of Social Research Methodology, 16(3), 231-244.

Beer, D., & Burrows, R. (2013). Popular culture, digital archives and the new social life of data. Theory, Culture & Society, 30(4), 47-71.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of machine Learning research, 3(1), 993-1022.

Briones, R., Nan, X., Madden, K., & Waks, L. (2012). When vaccines go viral: an analysis of HPV vaccine coverage on YouTube. Health communication, 27(5), 478-485.

Bouman, M. P., Drossaert, C. H., & Pieterse, M. E. (2012). Mark my words: the design of an innovative methodology to detect and analyze interpersonal health conversations in web and social media. Journal of Technology in Human Services, 30(3-4), 312-326.

Bruns, A., & Stieglitz, S. (2012). Quantitative approaches to comparing communication patterns on Twitter. Journal of Technology in Human Services, 30(3-4), 160-185.

Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. International Journal of Social Research Methodology, 16(2), 91-108.

Bryden, J., Funk, S., & Jansen, V. A. (2013). Word usage mirrors community structure in the online social network Twitter. *EPJ Data Science*, 2(1), 1.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., & Conejero, J. (2015). COSMOS: Towards an integrated and scalable service for analysing social media on demand. International Journal of Parallel, Emergent and Distributed Systems, 30(2), 80-100.

Chary, M., Park, E. H., McKenzie, A., Sun, J., Manini, A. F., & Genes, N. (2014). Signs & symptoms of dextromethorphan exposure from YouTube. PloS ONE, 9(2), e82452.

Cheng, X., Liu, J., & Dale, C. (2013). Understanding the characteristics of internet short video sharing: A YouTube-based measurement study. IEEE Transactions on Multimedia, 15(5), 1184-1194.

Chung, J. E. (2015). Antismoking campaign videos on YouTube and audience response: Application of social media assessment metrics. Computers in Human Behavior, 51(1), 114-121.

Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data mining in influenza mentions on the web and in social media. International Journal of Environmental Research and Public Health, 7, 596-615. doi:10.3390/ijerph7020596

Crick, M. (2012). Social media use in the Bronx: New research and innovations in the study of YouTube's digital neighborhood. Journal of Technology in Human Services, 30(3-4), 262-298.

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. California management review, 44(2), 25-46.

Desai, T., Shariff, A., Dhingra, V., Minhas, D., Eure, M., & Kats, M. (2013). Is content really king? An objective analysis of the public's response to medical videos on YouTube. PloS ONE, 8(12), e82469.

Dynel, M. (2014). Participation framework underlying YouTube interaction. Journal of Pragmatics, 73(1), 37-52.

Edwards, A., Housley, W., Williams, M., Sloan, L., & Williams, M. (2013). Digital social research, social media and the sociological imagination: Surrogacy, augmentation and re-orientation. International Journal of Social Research Methodology, 16(3), 245-260.

Fan, W., & Gordon, M. D. (2014). The power of social media analytics. Communications of the ACM, 57(6), 74-81.

Giglietto, F., Rossi, L., & Bennato, D. (2012). The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. Journal of Technology in Human Services, 30(3-4), 145-159.

Grefenstette, G. (1995). Comparing two language identification schemes, In Proceedings of 3rd International Conference on Statistical Analysis of Textual Data (JADT 1995), Rome, Italy (pp. 263 268).

Harrison, D., Wilding, J., Bowman, A., Fuller, A., Nicholls, S. G., Pound, C. M., & Sampson, M. (2016). Using YouTube to disseminate effective vaccination pain treatment for babies. PloS ONE, 11(10), e0164123.

Henrich, N., & Holmes, B. (2011). What the public was saying about the H1N1 vaccine: perceptions and issues discussed in on-line comments during the 2009 H1N1 pandemic. PloS ONE, 6(4), e18479.

Hine, C. (2000). Virtual ethnography. Oxford, UK: Sage.

Hoeber, O., Hoeber, L., El Meseery, M., Odoh, K., & Gopi, R. (2016). Visual Twitter Analytics (Vista) Temporally changing sentiment and the discovery of emergent themes within sport event tweets. Online Information Review, 40(1), 25-41.

Huang, A. (2008). Similarity measures for text document clustering. In Proceedings of the sixth New Zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand (pp. 49-56).

Jaspal, R., Turner, A., & Nerlich, B. (2014). Fracking on YouTube: Exploring risks, benefits and human values. Environmental Values, 23(5), 501-527.

Jürgens, P. (2012). Communities of communication: making sense of the “social” in social media. Journal of Technology in Human Services, 30(3-4), 186-203.

Lewis, S. P., Heath, N. L., Sornberger, M. J., & Arbuthnott, A. E. (2012). Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. Journal of Adolescent Health, 51(4), 380-385.

Lünich, M., Rössler, P., & Hautzer, L. (2012). Social navigation on the internet: A framework for the analysis of communication processes. Journal of Technology in Human Services, 30(3-4), 232-249.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In LREc 2010. Brussels, BE: European Language Resources Association (pp. 1320-1326).

Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. Sociology, 41(5), 885-899.

Savage, M., & Burrows, R. (2009). Some further reflections on the coming crisis of empirical sociology. Sociology, 43(4), 762-772.

Smith, A. N., Fischer, E., & Yongjian, C. (2012). How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? Journal of Interactive Marketing, 26(2), 102-113.

Stieglitz, S., Dang-Xuan, L., Bruns, A. & Neuberger, C (2014). Social media analytics: An interdisciplinary approach and its implications for information systems. Business & Information Systems Engineering, 6(2), 89-96. doi:10.1007/s12599-014-0315-7

Stieglitz, S., & Krüger, N. (2011). Analysis of sentiments in corporate Twitter communication - A case study on an issue of Toyota. Proceedings of the 22nd Australasian Conference on Information Systems, paper 29. Available: http://aisel.aisnet.org/acis2011/29

Sugimoto, C.R. & Thelwall, M. (2013). Scholars on soap boxes: Science communication and dissemination via TED videos. Journal of the American Society for Information Science and Technology, 64(4), 663-674.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406-418.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173.

Thelwall, M., Prabowo, R. & Fairclough, R. (2006). Are raw RSS feeds suitable for broad issue scanning? A science concern case study. Journal of the American Society for Information Science and Technology, 57(12), 1644-1654.

Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube videos: From Guatemalan rock to el big bang. Journal of the American Society for Information Science and Technology, 63(3), 616-629.

Thelwall, M. (2007). Blog searching: The first general-purpose source of retrospective public opinion in the social sciences? Online Information Review, 31(3), 277-289.

Tijdens, K., & Steinmetz, S. (2016). Is the web a promising tool for data collection in developing countries? An analysis of the sample bias of 10 web and face-to-face surveys from Africa, Asia, and South America. International Journal of Social Research Methodology, 19(4), 461-479.

Appendices

The appendices are online at https://figshare.com/s/2024440df704ed615f62 or doi:10.6084/m9.figshare.5257939.