www.elsevier.com/locate/dcm

#Living/Minimum Wage: Influential Citizen Talk in Twitter

Nuria Lorenzo-Dus, Matteo di Cristofaro

PII: S2211-6958(16)30008-3
DOI: http://dx.doi.org/10.1016/j.dcm.2016.02.003
Reference: DCM108

To appear in: *Discourse, Context and Media*

Received date: 25 November 2015
Revised date: 14 February 2016
Accepted date: 20 February 2016

Cite this article as: Nuria Lorenzo-Dus and Matteo di Cristofaro #Living/Minimum Wage: Influential Citizen Talk in Twitter, *Discourse, Context and Media,* http://dx.doi.org/10.1016/j.dcm.2016.02.003

#Living/Minimum Wage: Influential Citizen Talk in Twitter

Manuscript authors:

Prof Nuria Lorenzo-Dus, English Department, Swansea University, UK

Dr Matteo Di Cristofaro, Linguistics and English Department, Lancaster University, UK

Corresponding author details:

Prof Nuria Lorenzo-Dus

English Department

Swansea University

Singleton Park Campus

Swansea SA2 8PP

South Glamorgan, Wales, UK

Tel: 00 44 1792 204892

n.lorenzo-dus@swansea.ac.uk

Abstract: There has been a major cultural shift away from 'passive' consumption to more active production of digital texts by citizens. Yet, this does not mean that we all participate in digital media in the same ways and for the same reasons. Nor does it mean that we all have the same level of access to digital networks. This article seeks to contribute to a better understanding of the diversity and fluidity of citizen participation in digital environments by examining the discourse style of a particular group of digital users, namely citizens whose contributions become crowdsourced to prominence in microblogging. We refer to this form of citizen participation as 'influential', in as much as the discourse of these citizens attracts inordinate levels of attention and can trigger social contagion. We conduct a Corpus-Assisted Discourse Study of a corpus of tweets posted by a group of citizens who emerge as 'influential' within a Twitter debate about the minimum / living wage.

Our analysis reveals that their discourse style is characterised by (i) limited content originality but a high participation rate; (ii) a continuum of thematic engagement; (iii) high levels of emotionality; and (iv) a preference towards stance-taking acts that convey full confidence in one's views.

Keywords: social media influence, corpus-assisted discourse studies, Twitter, the living wage, stance.

1. Introduction

Digital communication constitutes the backbone of everyday life in many societies, 'always on' (Baron 2008) having become the default mode of social engagement for many of us. As digital citizens, we participate in social life in more and more varied ways than even just a decade ago. Several hybrid terms have been coined – such as 'produser' and 'co-creator' (Bruns 2007) – that articulate citizens' 'increased production prowess' (Van Dijck 2009:42) across digital environments. The notion of 'participatory culture' (Jenkins et al 2009; Jenkins 2014) captures a major cultural shift away from 'passive' consumption to more active production of digital texts by citizens.

Yet, living in a participatory culture does not mean that we all participate in digital media in the same ways and for the same reasons (Goode et al 2010). Nor does it mean that we all have the same level of access to digital networks. This article seeks to contribute to a better understanding of the diversity and fluidity of citizen participation in digital environments by examining the discourse style of a particular group of digital users: citizens whose contributions become crowdsourced to prominence in microblogging. These citizens not only attract inordinate levels of attention from others, including high-profile institutions, but can also trigger social contagion (Cha et al 2010). Throughout the article, we refer to them as 'influential citizens': they are neither celebrities nor official representatives of powerful institutions; their tweets get massively propagated (they may go viral) and acted upon (e.g. retweeted) the most. We examine their discourse through a case study of a concrete practice (debating) in relation to a particular social issue (the living / minimum wage) on Twitter.

2. Citizen Participation and Influence in Twitter

Social media are a key player in the current cultural shift away from citizens' passive consumption of, and towards active involvement in, the production of digital texts. This shift is seen to have resulted in the establishment of a 'participatory culture' (Jenkins et al 2009:xi), which is characterised by

> relatively low barriers to artistic expression and civic engagement, strong support for creating and sharing creations, and some type of informal mentorship whereby experienced participants pass along knowledge to novices. In a participatory culture, members also believe their contributions matter and feel some degree of social connection with one another (at the least, members care about others' opinions of what they have created).

The notion of participatory culture has been critiqued for its overly optimistic overtones of enhanced media and citizen empowerment (see, e.g., Hay and Couldry 2011). Yet, right from the outset, Jenkins et al (2009) acknowledged three key challenges to it, namely the participatory gap (linked to the digital divide that still exists across and within many societies), the need for transparency regarding means and forms of participation, and the ethics of participation. Importantly, too, the notion of participatory culture predates the internet. Within the Social Sciences, concepts such as the 'revalorisation of lay knowledge' in the media (Livingstone and Lunt 1994), the 'demotic turn' in broadcasting (Turner 2010) and the 'ordinarisation' of television (Bonner 2003) document a progressive but marked increase, from approximately the 1980s, in citizen participation across 'traditional' media.[1] Science and Technology scholars have also highlighted the increased value assigned to citizen participation in social life – a so-called 'third wave of science studies' considers 'the argument for citizen participation on expertise grounds to have been won at least in principle' and is now interested in better understanding the processes and outcomes of such participation (Evans and Plows 2007: 828). In order to do so, it is widely accepted that we need to move beyond lay-expert or producer-consumer binaries and to focus instead on understanding citizen participation as comprising multiple facets and being dependent upon locally-performed identities (Van Dijck 2009, Thornborrow 2015). This is especially so in digital platforms such as Twitter, in which much communication revolves around citizens

[1] See also Jenkins' (1992) work on television's participatory culture.

sharing their knowledge and views and evaluating the knowledge and views of others within large virtual communities (Zhang et al 2010, Zappavigna 2013).

Launched in 2006, Twitter is a text-based microblogging service where users can send messages (tweets) of up to 140 characters. Twitter users can place a hashtag symbol (#) before a single character, a word or an up-to-140-character sentence (without spaces) that thus becomes the topic around which further tweets are grouped. By aggregating tweets in this way, hashtags contribute to the three main functions of Twitter, namely news reporting of events as they happen, continuous discussion of events deemed to be newsworthy, and commentary on current events from the users' personal viewpoints (Bruns and Burgess 2012). Commentary relates most closely to the 'ambient' properties of Twitter (Hermida 2010; Bruns and Burgess 2012; Zappavigna 2013), whereby this microblogging platform serves as an always on, indirect communication medium between users. The non-reciprocal nature of Twitter networks means that hasthag-facilitated ambient affiliation can be 'asymmetrical and need not involve dialogic exchanges.' (Page 2012: 184).

Example (1), taken from the corpus used in this study, illustrates the ambient affiliation function of hashtags and other Twitter conventions:

(1) RT @OccupyAustin: Join the #FightFor15! #FastFoodGlobal

Day of Action for Living Wages! THU 11:30AM

In (1), the names of two events ('Fight for 15' and 'Fast Food Global') are used as hashtags and treated as hyperlinks by the Twitter service: by clicking on them, one is directed to Twitter pages that list all the tweets containing them, effectively enabling Twitter users to access 'with just one click' a virtual community around those hashtags. This makes hashtags like the ones in (1) useful mechanisms for accessing – and potentially influencing – '*ad hoc* communities without the need to establish mutual follower / followee relationships with any members of those communities' (Bruns and Burgess 2012:3). Example (1) also includes two other Twitter conventions: '@' and 'RT'. The symbol '@' precedes usernames to convert them into hyperlinks and performs a range of mainly addressivity-related functions (see e.g. Honeycutt and Herring 2009). 'RT' (Re-Tweet) is a tweet that is forwarded to one's Twitter followers, but in which original attribution is retained. RTs play a key part in mediating follower/ followee relations, including validating others' views and gaining followers (boyd et al 2010). A further Twitter convention, not used in (1) but frequent in our corpus, is 'via', which enables users to forward tweets that preserve original attribution but admits changes to

original content.

These Twitter conventions are thus far from mere technical affordances of the Twitter service. They also fulfil important participation structuring, agenda framing, community forming and opinion articulation functions (see, e.g., Hansen et al 2011; Bastos et al 2013; Puschmann 2015), often through crowdsourcing practices. Crowdsourcing designates a participative practice in which 'an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task.' (Estellés-Arolas and González-Ladrón-de-Guevara 2012:192). The term conjures up an image of egalitarian digital participation that does not live up to reality. The call may be open, and the task may be voluntarily undertaken by many. However, the likelihood of one's contribution to the task standing out, as it were, from the crowd – let alone to influence the task's outcome – is contingent upon a range of factors. Citizen participation in social media is, after all, not only varied but also unevenly distributed (Van Dicjk 2009; Van Dijck and Nieborg 2009; Goode et al 2010; Page 2012). Hierarchies of participation operate across social media and, whilst fluid, they are determined in part by differences in discursive style amongst users and user groups (Weller et al 2013). For instance, celebrities, corporations and 'ordinary' users are known to select and deploy hashtags differently when trying to 'command the potential attention of an audience within the linguistic marketplace of Twitter'. Whereas ordinary users favour the construction of affiliated over individuated self-identities, celebrity figures and corporate accounts tend to, respectively, 'project their identity as engaged with their audience and to endorse the values of their followers' (Page 2012:198). [2]

The issue of how Twitter users seek to command attention from other users has generated considerable interest within social network science studies, too. Findings repeatedly show that open web systems develop in ways whereby small groups of users – estimated at between 10% and 20% of all users - attract inordinate levels of attention and can exercise social influence, including triggering 'social contagion' (Cha et al 2010). This minority group is variously described in the literature as 'leaders' (Sonnerbichler 2010), 'experts' (Weber et

[2] Page (2012) appositely borrows the metaphor of the 'linguistic marketplace' from Bourdieau (1977) to describe self-branding practices in social media genres, whereby those genres' users deploy different linguistic resources in order to promote their visibility.

al 2007), 'emergent elites' (Paracharissi and Oliveira 2012; Meraz and Papacharissi 2013), 'discussion catalysts' (Himelboim et al 2009), and 'superparticipants' (Graham and Wright 2013). In our work, and drawing upon extant studies of influence on Twitter, we use the term 'influential citizens'.

Although influence is a notoriously difficult concept to define and measure, there are two broad academic views on it. One considers influence to reside within a small group of individuals who have exceptional persuasion skills (e.g. Gladwell 2002). The other challenges the idea that influence can be the possession of a few, arguing instead that anyone can be influential (what Watts (2007) calls 'accidential influentials') and that influence is therefore a matter of there being a critical mass of easily influenced people (e.g. Domingos and Richardson 2001). Empirical evidence is inconclusive but, within the context of Twitter, broadly supports the former view. A number of studies have identified various user practices that are conducive to influence. These include limiting one's tweets to a single topic / hashtag and keeping high levels of personal involvement (Cha et al 2010), maintaining high levels of activity (Romero et al 2011) and posting tweets that express a negative mood and a sense of community (Quercia et al 2011).

The above studies are largely based on social media analytics and mathematical modelling methods. They therefore provide a useful starting point for more detailed examination of influential citizens' discourse style, which is the main aim of this article. To our knowledge, only two studies to date have examined in part the *discourse* of influential Twitter citizens: Papacharissi and Oliveria (2012) and Meraz and Papacharissi (2013). We say in part because, although these studies combine social media analytics, content and 'discourse-based' methods, the latter does not entail micro-level examination of the discourse features of the corpus. Both studies draw upon a corpus of 1.5 million tweets relating to the #egypt hashtag during a one month period in 2011, pre and post resignation of former Egyptian President Hosni Mubarak, and reveal that tweets produced by influential citizens were emotive, made frequent use of personal stories, and displayed 'heightened conversationality' (2013).

3. Methodology

3.1 Data:

Our study draws upon two purpose-built corpora collected over a ten-day period between 26th May and 4th June 2014: (i) a reference corpus of 1.6 million tweets and (ii) an analytic corpus of 38,400 tweets from twelve hashtags that were thematically linked to the living / minimum wage debate. The living wage is broadly defined as the minimum income deemed necessary to maintain a safe, decent standard of living (Alderman and Greenhouse 2014). The actual amount varies across communities but includes more needs than those within the minimum wage, which is set by law. Employers may choose to pay the living wage on a voluntary basis. There is a vigorous debate regarding the pros and cons of a living versus a minimum wage, which tends to resurface whenever wage and / or poverty policies are announced by governments. The debate is both social and discursive. Research has competently focussed on the former, concluding amongst other things that 'the greatest successes in securing the living wage have been made through bottom-up processes of organising and campaigning.' (Lawton and Pennycook 2013:10). Although social media epitomises such processes, it remains under-examined to date and justifies our decision to select a Twitter debate on this issue as our case study.

3.2 Framework and Procedure

Our work adopts a Corpus-Assisted Discourse Studies (CADS) framework, which has proven felicitous for understanding the main discourses around social topics ranging from immigration to social benefits in print and social media (e.g. Baker et al 2013; Zappavigna 2013; Baker and McEnery 2015). As its name indicates, CADS works at the interface of Corpus Linguistics methods and Discourse Studies theories and analytic concepts. It tends to follow a 'serendipitous journey' (Partington 2009: 286) so that research is empirically designed in order to test pre-assumptions through corpus-based analyses of actual texts. In our case, research into Twitter influence and citizen participation informs the analyses of our corpora. These analyses are treated as an initial '"map" … pinpointing areas of interest for a subsequent close analysis' (Baker et al 2008: 28; see also articles in Baker (ed.) 2015). As per the CADS approach, our study is premised on the belief that quantitative and qualitative discourse research methods can be fruitfully integrated. This is reflected in the methodological procedures we adopted in order to build and analyse our corpus, which we next describe.

3.2.1 Corpus building

3.2.1.1 Reference corpus: The reference corpus was collected via the Twitter API facility – through the use of Twitter4j (https://dev.twitter.com/overview/api) – and used to pinpoint the discourse peculiarities of the analytic corpus. The decision to build our own reference corpus, rather than resort to a general language corpus, such as the British National Corpus, was motivated by (i) the belief that general language corpora are not adequate for comparison purposes given the idiosyncrasy of Twitter language, including its 140 character per tweet limitation (see also Baker and McEnery 2015); and (ii) the desire to collect a synchronic corpus, rather than rely on general Twitter corpora, gaining access to which can also be quite complex given Twitter's Terms of Service (https://twitter.com/tos).

3.2.1.2 Analytic corpus: Three steps were followed in order to build our analytic corpus. Firstly, we manually identified three terms related to the living / minimum wage debate, namely poverty, living wage, and minimum wage. We used these terms as 'thematic seeds' in the commercially available Twitter analytics tools Twitonomy (https://www.twitonomy.com/) in order to derive a list of frequent hashtags for the living / minimum wage debate, namely: #austerity; #compassion; #cuts; #homeless; #justice; #livingwage; #lowwage; #minimumwage; #poor; #poverty; #raisethewage; and #recession.[3] These hashtags were treated as 'discourse topics' (Baker and McEnery 2015), that is, as categories that reflected and constructed the discourse around the living / minimum wage debate in our corpus. Choosing to tweet about this debate via one or another hashtag, for example #livingwage instead of #lowwage, contributes to construct a tweeter's stance towards the debate: to 'evaluate' a 'stance object' (the living / minimum wage debate) in a specific way and, in so doing, to 'position' oneself , and 'align' oneself with others, including through the use of other linked hashtags (Du Bois 2007).

Secondly, we used Klout (https://klout.com/home) in order to identify the most influential user for each of the twelve hashtags, manually identifying and discarding celebrity and

[3] Only tweets that used the exact form of the chosen search term as a hashtag were collected – grammatical (e.g. #theminimumwage) and misspelt (#miniumwage) variants and occurrences of the search terms not marked as hashtags were excluded. This was because neither Twitter nor Twitter analytic tools allow for automatic identification and collection of hashtag variants. Hashtags identified initially through Twitonomy but manually discarded on account of their non-thematic relevance included: #2ch, #logsoku, #gop, #iceland, #uniteblue, and #renewui.

institutional Twitter accounts. Klout is one of the most widely used pieces of software for ranking social media influence (see Anger and Kittl 2011; Quercia et al. 2011; Edwards et al. 2013). Although the algorithm used for calculating Klout influence scores is not publicly available, it is known to include the following variables: following count; followers' count; number of retweets, list memberships, number of spam or dead accounts following, influence levels of those who retweet a given user, unique mentions, and ratio of reactions generated by that user compared to the amount of content shared by him / her. Information about each of these variables is collected by Klout and compared to data points from eight different social media networks: Facebook, Instagram, Google+, Bing, Linkedin, Foursquare, Klout and Twitter (https://klout.com/corp/score). Klout scores range from 1 to 100, with higher scores indicating higher influence levels (Edwards et al. 2013).

Thirdly, we collected a sample of the, at that point in time, most recent 3,200 tweets for each of our influential users within their relevant hashtag. The 3,200 figure is the maximum number of tweets per user that can be collected via Twitonomy. This constitutes one of two limitations in our corpus building method, for our analytic corpus does not include all the tweets potentially contributing to each of the users' Klout scores. It is worth pointing out, however, that Twitter historical data is sold either by Twitter or through companies that started collecting tweets when Twitter was launched in 2006. Getting unlimited access to this data is extremely expensive.[4] A second limitation concerns the fact that, although the tweets we collected came from hashtags that had been created up to six years beforehand (see Table 1), we were unable to examine hashtag influence flows across time. Again, it is worth pointing out that there are commercial services that provide information about a hashtag's first use on Twitter (e.g. http://ctrlq.org/first/) but they do not provide sufficient details regarding how it evolves across an extended period of time, thus precluding temporal analysis of hashtag development.

Table 1 provides an overview of our analytic corpus. It includes an illustrative example per hashtag / user, the date when the hashtag was created, and the total number of tweets, followers and words per hashtag / user and Klout score at the data collection point.

[please insert Table 1 here]

[4] See, e.g., https://gnip.com/historical/ and http://sifter.texifter.com/.

Although we did not limit our corpus geographically or by gender, as shown in Table 1, the influential citizens that we identified through the three-step process we have outlined came mainly from the United States of America and were male. In order both to preserve user anonymity and to facilitate reader comprehension for this paper, in Table 1 and from this point onwards, the influential citizens' hashtags are also used as their user names, placing the latter in italics (e.g. discourse topic #austerity → user name *austerity*). In doing this, we echo Baker and McEnery's (2015:247) point that, despite the absence of a 'common consensus around "best practice" … it is important that ethical concerns do not prohibit that [social media] research from being carried out … and that researchers are able to show their data to readers, in order to demonstrate transparency.'

3.2.2 Corpus analysis:

Once collected, we converted all the tweets in our analytic and reference corpus into .txt file format. For the analytic sub-corpora this entailed creating a single .txt file per influential citizen. For the reference corpus, and given the corpus size capability of the semantic tagger to be used at a later stage, we randomly selected c. 95,000 tweets and saved them into a single .txt file. We next cleaned all the .txt files in readiness for analysis, deleting from each tweet: date of posting; names of other users addressed through @ (by deleting the structure @[WORD]); the *RT* term; the "http://" and "https://" in front of web-links; the # character; internal links to Twitter pages; illegal non-ASCII characters; and ID numbers included by Twitonomy.

Next, we used the semantic and the Part Of Speech (POS) taggers available in Wmatrix (Rayson 2009) to identify word frequencies, keywords and key semantic domains for the twelve analytic sub-corpora and the reference corpus. We used two Wmatrix measures to calculate and rank the strength – or keyness – of keywords and semantic domains: Log-Likelihood (LL) and LogRatio. The former, which is at present considered the standard measure, is a statistical significance measure, i.e. 'it tells us how much evidence we have for a difference between two corpora' (Hardie 2014). The latter is an effect-size measure that therefore 'represent[s] how big the difference between two corpora is for a particular

keyword' or semantic domain. [5] The statistical significance of a given semantic domain is represented in Wmatrix in two different ways. One is by means of a table that lists, in descending significance order of frequency, the semantic domain labels and provides details regarding their LL and LogRatio values (see Tables 2 and 3 in Section 4.2). The other is a series of word clouds where the label of each semantic domain is represented in text. The size of each semantic domain label is dependent on its LL score: the higher the semantic domain's LL value, the larger the font size in which it is textually represented in the word cloud (see Figures 1 and 2 in Section 4.2).

Grouping keywords into semantic domains was helpful in being able to identify the 'aboutness' of the discourse of these influential citizens. However, it did not reveal other aspects of their discourse that may play a relevant role in their having been crowdsourced to prominence. Because of this we also conducted cluster, collocation and Key-Word-In-Context (KWIC) concordance analyses for each of the twelve influential citizen corpora using Wmatrix and AntConc.[6] This enabled us to identify and better understand patterns in the use of emotionality, impoliteness, modality and so forth, as we discuss in the remainder of this article.

4. The discourse of influential citizens around the living / minimum wage debate in Twitter

Overall, the discourse of influential citizens in our corpus was characterised by: limited originality but high participation rate (4.1), varying degrees of thematic engagement (4.2), high levels of emotionality (4.3); and high levels of self-confidence (4.4). Although non-mutually exclusive, in what follows we discuss each of these practices independently for clarity of presentation.

4.1. Limited originality by high participation rate

Our twelve influential citizens made very frequent use of the Twitter convention 'via', that is, they often relayed others' tweets and / or imported web content via hyperlinks, preserving

[5] In practical terms, as noted by Hardie (2014), the relationship between the two measures is such that 'every extra point of Log Ratio score represents a doubling in size of the difference between the two corpora, for the keyword under consideration' (emphasis in the original); the same holds true for the keyness of semantic domains.

[6] Antconc: A freeware corpus analysis toolkit for concordancing and text analysis (http://www.laurenceanthony.net/software/antconc/)

original attribution but also making some changes to the original content. This was especially the case for *recession* and *justice.* The strongest keyword for both users was 'via' and the second strongest lexical item for *recession* was 'abc' which, upon close inspection, corresponded to the US news channel 'ABC'. This news channel was indeed used as the main web source from which *recession* imported content when discussing the living / minimum wage in financial terms within his hashtag. The second strongest keyword for *justice* was 't4us', which designates a group of tweeters who commit to retweeting each other's posts (https://tagdef.com/t4us).

In as much as research has shown the prevalence of re-posting and importing practices in microblogging (see, e.g. boyd et al 2010; Cha et al 2010; Puschmann 2015), this result is perhaps not surprising. However, it seems counter-intuitive that a discourse practice that relies on other-authored content and hence shows limited individuality or 'uniqueness' may characterise the discourse of influential citizens on Twitter. Three possible reasons for this are worth considering. Firstly, relaying / importing content is only one of a set of salient practices by influential citizens in our study. The other practices relied on the performance of individual stance-taking acts (see Sections 4.2 – 4.4). Secondly, and especially in the context of social policy campaigning, relaying / importing practices can be used to broker information strategically so as to give rise to alternative social agendas (see e.g. Bastos et al 2013; Penney and Dadas 2014). As such, these practices may position users as influential figures with the 'authority' to select 'important' content to be relayed to others.[7]

Thirdly, all twelve users displayed high participation rate scores, that is to say, they tweeted very frequently within their respective hashtags. Participation rate has been found to act as a crucial predictor of perceived expertise across numerous offline inter-personal / inter-group communication contexts (Littlepage et al 1995). In digital environments, and in the absence of physical cues, a high participation rate is also known to contribute to establish one's social presence, that is, one's belonging to a virtual community. Furthermore, there is a positive correlation between the ability to establish one's social presence and an increase in social influence (Walvoord et al 2008). And that is what we found in our study: 'doing being frequent tweeters' was one of the means by which the twelve influential citizens in our study established their social presence. They kept very active profiles within the Twitter hashtag community in which they participated and ultimately became influential.

[7] We are grateful to one of the anonymous reviewers of this paper for suggesting this interpretation.

4.2 A Continuum of Thematic Engagement

Social media theory postulates that influential social media users need to maintain a high level of thematic engagement, that is, they need to post content that is thematically relevant to the issue/s in which the virtual communities that have crowdsourced them to prominence engage (Cha et al 2010). Our analysis of keywords and key semantic domains does not show a direct positive correlation between thematic coherence and influence scores. Instead, the results of our KWIC concordance analyses revealed a 'thematic engagement continuum' across the twelve influential citizens whose tweets we examined. The tweets posted by seven of these users, namely *livingwage, minimumwage, raisethewage, recession, austery, homeless* and *cuts,* predominantly drew upon semantic domains that were strongly / moderately relevant to civic participation issues, including issues directly relevant to their hashtag. Other users, namely *lowwage, poverty, poor, compassion* and *justice,* predominantly posted tweets within a mixed bag of semantic domains that were only loosely or hardly relevant to civic participation issues, including issues that were unrelated to their hashtag. These semantic domains ranged from 'Sports' (*lowwage*), 'Living creatures' (*lowwage*) and 'Geographical names' (*poverty*) to 'Speech acts' (*poor*), 'Religion and the supernatural' (*compassion*) and 'Food' (*justice*).

As an illustration, let us consider the top ten key semantic domains for *livingwage* (Table 2; Figure 1) and *lowwage* (Table 3; Figure 2). [8]

Table 2: Top 10 semantic key domains for *livingwage*

	Semantic domain	LL	LogRatio
1	Politics	1429.74	4.17
2	Law and order	452.13	2.84
3	Government	339.31	2.77
4	Ethical	154.15	3.28
5	In power	137.82	1.42
6	Belonging to a group	130.88	1.53
7	Money and pay	106.07	2.07
8	Substances and materials generally	104.26	2.23

[8] A table containing the strongest key semantic domains for each user vis-à-vis our reference corpus is provided in the Appendix.

9	Grammatical bin[9]	86.81	0.17
10	Personal names	69.86	0.48

Figure 1 – Key Semantic Domains word cloud for *livingwage*

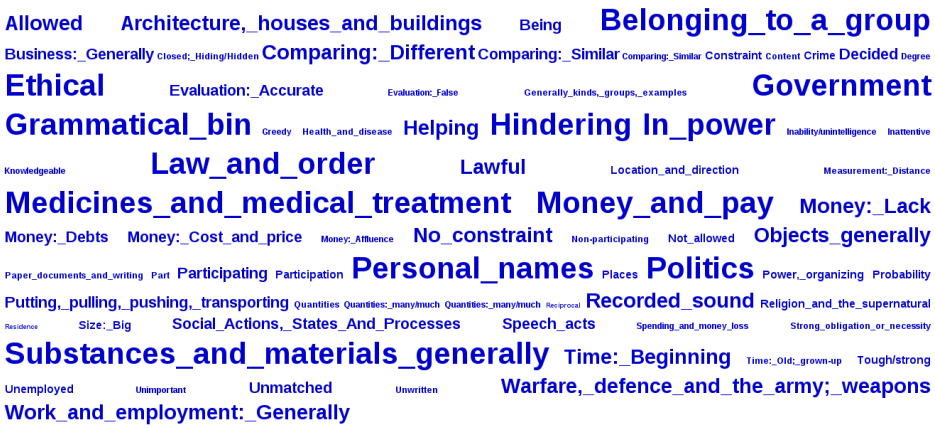

Table 3: Top 10 semantic key domains for *lowwage*

	Semantic domain	LL	LogRatio
1	Living creatures: animals, birds, etc.	167.64	1.47
2	Grammatical bin	108.9	0.19
3	Sports	106.04	1.16
4	Personal names	78.18	0.52
5	Games	45.38	1.89
6	No obligation or necessity	44.58	1.54
7	Warfare, defence and the army; weapons	43.31	1.16
8	Government	34.02	1.23
9	Speech	33.92	0.53
10	Politics	33.58	1.2

Figure 2: Key Semantic Domains Word Cloud for *lowwage*

[9] This is defined in WMatrix as a domain comprising: 'Prepositions/adverbs/conjunctions, etc.' (Archer, Wilson, Rayson 2002:36)

As shown in Table 2 / Figure 1, the strongest ten semantic domains in the tweets by *livingwage* pointed to civic engagement topics, with 'Politics', 'Law and order', 'Government' being the strongest three. Note, too, that these domains were extremely more frequent in her tweets than in the reference corpus, as indicated by their LogRatio values. The LogRatio value for *livingwage*'s strongest semantic domain ('Politics'), for instance, indicates that the frequency of use of that semantic domain was over eight times higher than in the reference corpus. The examples below (2a-2c) are illustrative of tweets by *livingwage* within, respectively, the 'Politics', 'Law and order' and 'Government' key semantic domains:

(2a): Where in the world can you deny citizens access to healthcare and living wages, then get them to vote for you?

(2b): NCGA Mor Protesters arrested at MoralMonday rally focusing on workersRights. MUST WATCH Protesters urge NC legislators on living wage

(2c): Gov. Deval Patrick signs measure that will give Massachussets the nation's highest state minimum wage: $ 11 an hour

In contrast, as shown in Table 3 / Figure 2, the strongest key semantic domains in the tweets by *lowwage* bore no thematic relevance to either his hashtag (#lowwage) or any issues relating to civic engagement in the public political sphere. The strongest semantic domain for this influential citizen, 'Living Creatures', included sports teams named after animals, such as 'Chicago Bulls'. Similarly, the fourth top domain, 'Personal names', mainly contained names of athletes and television celebrities with no known connection to the living / minimum wage debate. The first and fourth domain, thus, mainly supported the third top domain: 'Sports'. Together, they placed this user towards the end pole of the non-thematic engagement continuum, as Examples (3a-3c) illustrate:

(3a): Got cockteased with playing golf today. Now all I'm left with is not plying golf. Going to fill the void with basketball. (Semantic domain: 'Games')

(3b): if the bulls keep their picks, they take payne and Shabazz (or gay harries). (Semantic domains: 'Living Creatures'; 'Personal Names')

(3c): Kids cheering against the US are the worst type of kids. PEOPLE DIED SO YOU COULD ROOT FOR USA SOCCER. AND VOTE. (Semantic domain: 'Sports').

4.3 Emotionality

The performance of emotionality emerged as a salient practice in our analytic corpus, and it was discursively realised through orthographic and / or lexical means. Orthographic means were used by all twelve influential citizens and it is probably a 'generic feature' of much social media communication, rather than a discursive marker of influence per se. A number of media commentators, for example, have pointed out, indeed bitterly complained about, the over-use of exclamation marks in digital communication, coining terms such as 'serial exclamation pointers'[10] to describe online users who draw extensively (excessively, in their view) on orthography to express emotions and also to convey opinions, present facts and so forth. Academic research has moved from treating these orthographic means as markers of 'excitability' in the speech of women, a phrase that unfortunately connotes emotional randomness, to acknowledging their non-genderness and multifunctionality. Waseleski (2006), for instance, identifies three broad functions for them in digital communication, namely their being markers of: friendliness (expressions of cordiality and thanking), factuality (statements, regardless of their truth value) and, in only c. 10% of her data, excitability (sarcasm, flaming and effusive thanks). Although all three functions were present in our corpus, exclamation marks and capitalisation were primarily used as markers of excitability (principally flaming[11]), as in (4), and factuality, as in (5).

(4)

I still keep wondering WHERE ARE ALL THESE FUCKING JOBS get the jobs first then put the pressure on you BLOODY!!

(5)

GOP is COUNTING on us NOT turning out to VOTE in HUGE numbers. THIS YEAR, we MUST prove them WRONG!

As for the lexical means used to perform emotionality, these were salient in the discourse of the influential citizens positioned towards the positive end pole of the thematic engagement

[10] See, for example: https://www.bostonglobe.com/lifestyle/style/2012/04/25/how-mail-and-texting-have-driven-people-overuse-exclamation-points-confessions-serial-exclamation-pointer/bSKe7sq0TEZLHcq1bq5A7M/story.html; and http://blog.hubspot.com/marketing/exclamation-point-flowchart.

[11] The term flaming is used here as in Waseleski's (2006) work, which includes a continuum from the expression of annoyance to the verbalisation of personal insults.

continuum in our corpus. These citizens used emotional lexis when tweeting messages that contained 'calls for action' (Example (6a)) and / or when expressing other-directed 'negative' emotions, principally disappointment, anger and frustration (Examples (6a) – (6b)):

(6a)

Step down. You're not up to the job. Your ministry sucks. We couldn't think any less of you.

(6b)

We are still suffering because of the idiotic antics of the gop! anyone who blames PBO is delusional! We must stand up and fight

Examples (6a) – (6b) illustrate a further pattern in the corpus, revealed through KWIC concordance analyses of other-directed emotional lexis: the co-occurrence of 'negative' emotionality and social group polarisation. Negative emotionality was often performed through us-versus-them discourse structures and, as such, functioned simultaneously as a disaffiliation mechanism and a means to establish and / or reinforce communities of like-minded Twitter users. In (6a), for instance, second person singular and first person plural pronouns were explicitly contrasted through effective use of punctuation within the last three sentences, each of which containing an instance of impoliteness. In the case of 'You're not up to the job' and 'Your ministry sucks', the impoliteness strategy of 'explicitly associating the other with a negative aspect' (Culpeper 2011) entailed a derogatory evaluation that was nevertheless verbalised through evidential modality (see 4.4 for a discussion of the use of evidential modality in the corpus). This set the grounds upon which, in the last sentence of the tweet, the us-versus-them dichotomy was justified. The dichotomy was also expressed through an impoliteness strategy, on this occasion, the 'condescend, scorn, ridicule' strategy (Culpeper 2011) in (6b). This served both to belittle the impoliteness target and to bind together an explicit, though referentially imprecise, 'we' group. It was this in-group binding aspect within the performance of negative emotionality that was salient within the most thematically engaged users in our study.

The finding confirms previous research on negativity and emotionality being markers of influence on Twitter (Quercia et al 2011). The result may be partly explained by drawing upon the field of Cognitive Psychology, where research has shown that individuals in a negative mood employ distinct cognitive processing styles that enable them to produce ideas that become influential (e.g. Forgas 2001). Additionally, within the field of Computer-

Mediated Communication, research has provided some empirical evidence for the existence of what is known as 'the online disinhibition effect' (Suler 2004). This argues that anonymity – or at least lack of public visibility – can promote increased levels of self-disclosure and trust (benign disinhibition) but also increased aggression (toxic disinhibition). Both benign and toxic disinhibition are discursively realised through 'emotion talk', with toxic disinhibiiton being mainly performed through impoliteness and other forms of verbal aggression.

4.4 Confidence

KWIC concordance analyses revealed that the twelve influential citizens in our study regularly used evidential modality in order to perform evaluative (stance-taking) acts. This is illustrated in Examples (7a) – (7b), where these citizens' assessments of two issues (homelessness and taxation, respectively) are presented as indisputable, indeed as a 'reality check in the case of *homeless*, and hence asserted, rather than merely evaluated, through language: [12]

(7a)

Reality check: Starving the homeless won't end homelessness (*homeless*)

(7b)

Getting rid of the carbon tax is economic vandalism (*austerity*)

Frequently, too, influential citizens in the corpus evaluated issues and the actions of other users / social actors through deontic modality, with a clear emphasis on 'collective duty'. They often worded these evaluations through directives, as (7c) and (7d) illustrate:

(7c)

Make no mistake this budget is all about taking from students (*cuts*)

(7d)

[12] The modality-related terms 'epistemic' and 'evidential' are used here in the sense of de Haan (2005), who does not see evidentiality as a sub-type of epistemic modality but as separate on the basis that epistemic modality is about evaluating evidence through language, whereas evidential modality is about asserting that evidence through language.

Protect voting rights here and now *(livingwage)*

Our twelve users may have derived part of their influential citizen 'status' from the assuredness and confidence with which they phrased their tweets. Indeed, there is evidence from empirical research in the field of psychology that expressions of confidence operate as an expertise proxy across many contexts, especially in communicative contexts involving large groups (Zarnoth and Sniezek, 1997). This tendency to link confidence with expertise and influence is known as 'confidence heuristic' (e.g. Price and Stone 2004) and is explained in terms of correctness (confident assertions are believed to be more likely to be correct than tentative statements, Keren and Teigen 2001) and social benefits accruing through following someone who comes across as communicatively confident (Zarnoth and Sniezek 1997).

5. Conclusions

The aim of our study was to characterise the discourse practices of citizens who, through crowdsourcing, become influential in the context of a particular debate on Twitter. Through quantitative and qualitative analyses we have identified the regular performance of four such practices, namely (i) their use of Twitter conventions that somewhat limit content originality but are coupled with high participation rates; (ii) a continuum of thematic engagement; (iii) high levels of emotionality; and (iv) a preference towards stance-taking acts that convey full confidence in one's views.

State-of-the art software employed within the burgeoning field of social media analytics uses participation rate as one of the factors whereby to identify and rank influence in social networking environments. We used one such software product (Klout) in order to initially identify influential citizens in our data. However, our results revealed that the statistical and algorithmical methods used for Klout (and generally other similar software) provided only limited insight into the ways through which influence was sought and achieved in our Twitter corpus (and, likely, other social media). This is because other, discursive factors also play a key role, i.e., (ii) – (iv) above. Of these, the presence of a thematic engagement continuum across the twelve influential citizen corpora is of particular note in the broader context of what seems to be required to be seen as an expert in 'the Twitterverse'.

Deliberation in microblogging (and other social media) contexts has been accused of failing to provide actual opportunities for pluralistic discussion, let alone meaningful civic engagement (Hill and Hughes 1999, Ranerup 2003). This owes to its being thought to consist

primarily of a minority of highly vocal individuals 'depositing' a random set of self-centred views online without any real interest in further discussion. Our results support this view only in part. Granted, the discourse of some of the twelve influential citizens examined was only loosely thematically relevant to the content of the hashtag within which the citizen in question had become influential. In a couple of cases no thematic engagement with the respective topic could be ascertained through the corpus-assisted discourse tools we used, including manual reading of numerous KWIC concordance lines and collocations. Nevertheless, these constituted the negative end pole in a continuum of thematic engagement which, on the whole, provided a digital platform for the wide framing of civic participation issues, rather than a narrow-focussed deliberative arena. Simply put, providing 'the bigger picture' emerged as a valued discourse practice when it came to gaining influence on Twitter, even if some of the details contained therein were not directly relevant to the issues being debated.

How influential citizens perform this wide framing discursively matters, too. In this regard, and although we acknowledge that influence is also determined by non-discursive factors such as topic and technical infrastructure (e.g. Romero et al 2011), our findings point tentatively to a discourse style for influential citizens in Twitter where higher levels of thematic engagement co-occur with higher negative emotionality and frequent expressions of self-assuredness and confidence. The verbal release of negative emotions in online settings has been described in terms of incivility: social network sites have been accused of generating heated, impolite discussion, with anonymity being regarded as a likely reason for such behaviour. Again, our data supports this only in part. The tweets from the twelve influential citizens in our data did include frequent face-threatening acts, performed through emotionally phrased impoliteness strategies, and often relied on group polarisation discourse strategies. However, given the users' influential 'status', lack of public visibility is unlikely to have played a major role. Instead, the aim of their emotionally and confidently worded tweets seemed to reside in displaying divergent perspectives around the living / minimum wage, and in the constant negotiation of those perspectives. We see this as contributing to the consolidation of a participatory culture that may facilitate socio-political deliberation outside 'authoritative, concentrated sources of collective political intelligence such as Parliament and Congress' (Chadwick 2009:5). What is more, negative feelings were not the only kind of emotionality saliently performed in the data. Calls for action that sought to get other tweeter users involved as citizens in society were also performed through 'emotion talk'. The

discourse of these influential citizens, it seems, has the potential to shift the authority balance between political institutions and representatives on the one hand, and voters, activists and citizens on the other, making the latter a more visible and potentially potent force to reckon with.

References

Alderman, Liz; Greenhouse, Steven. "Fast food in Denmark serves something atypical: Living wages", New York Times, October 27, 2014. http://www.nytimes.com/2014/10/28/business/international/living-wages-served-in-denmark-fast-food-restaurants.html?_r=00

Anger, Isabel, and Christian Kittl. "Measuring influence on Twitter." In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, edited by Stefanie Lindstaedt and Michael Granitzer. Graz: ACM, 2011.

Archer, Dawn, Andrew Wilson, Paul Rayson. *Introduction to the USAS category system*, 2002 http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf

Baron, Naomi. *Always On: Language in an Online and Mobile World*. Oxford: Oxford University Press, 2008.

Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michał Krzyżanowski, Tony McEnery, and Ruth Wodak. "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press." *Discourse & Society* 19, no. 3 (2008): 273-306.

Baker, Paul, Costas Gabrielatos, and Tony McEnery. *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge: Cambridge University Press, 2013.

Baker, Paul, and Tony McEnery. "Who benefits when discourse gets democratised?: analysing a Twitter corpus around the British Benefits Street debate." In *Corpora and Discourse Studies: Integrating Discourse and Corpora* edited by Paul Baker and Tony McEnery, 244-265. London: Palgrave Mcmillan, 2015.

Bastos, Marco Toledo, Rafael Luis Galdini Raimundo, and Rodrigo Travitzki. "Gatekeeping Twitter: message diffusion in political hashtags."*Media, Culture & Society* 35, no. 2 (2013): 260-270

Bonner, Frances. *Ordinary television. Analysing popular TV*, Brisbane: University of Queensland Press, 2003.

Boyd, Danah, Scott Golder, and Gilad Lotan. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on System Sciences* edited by, 1-10. Koloa: IEEE, 2010.

Bruns, Axel. *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. Bern: Peter Lang, 2007.

Bruns, Axel, and Jean Burgess. "Researching news discussion on Twitter: New methodologies."*Journalism Studies* 13, no. 5-6 (2012): 801-814.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and P. Krishna Gummadi. 'Measuring User Influence in Twitter: The Million Follower Fallacy.' In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.* edited by Tumasjan, A., T. O. Sprenger, P. G. Sandner, and I. M. Welpe, 10-17. Menlo Park: AAAI Press, 2010.

Chadwick, Andrew. "Web 2.0: New Challenges for the Study of E-Democracy in an Era of Informational Exuberance." *I/S: A Journal of Law and Policy for the Information Society, no. 5 (2009):* 9-41.

Culpeper, Jonathan. *Impoliteness. Using language to cause offence*. Cambridge: Cambridge University Press, 2011.

De Haan, Ferdinand. 'Encoding speaker perspective: Evidentials.' *Linguistic Diversity and Language Theories* 72 (2005): 379-397.

Domingos, Pedro and Matt Richardson. "Mining the network value of customers". In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57-66. San Francisco: KDD, 2001

Du Bois, John W. "The stance triangle." *Stancetaking in discourse: Subjectivity, evaluation, interaction* 164 (2007): 139-182.

Edwards, Chad, Patric R. Spence, Christina J. Gentile, America Edwards, and Autumn Edwards. "How much Klout do you have… A test of system generated cues on source credibility." *Computers in Human Behavior* 29, no. 5 (2013): A12-A16.

Estellés-Arolas, Enrique, and Fernando González-Ladrón-de-Guevara. "Towards an integrated crowdsourcing definition." *Journal of Information Science* 38, no. 2 (2012): 189-200.

Evans, Robert, and Alexandra Plows. "Listening without prejudice? Re-discovering the value of the disinterested citizen." *Social Studies of Science* 37, no. 6 (2007): 827-853.

Fairclough, Norman. *Media discourse,* London: Arnold, 1995.

Forgas, Joseph P. *Feeling and thinking: The role of affect in social cognition*. Cambridge: Cambridge University Press, 2001.

Gladwell, Malcolm. *The tipping point: How little things can make a big difference*. New York: Little, Brown, 2006.

Goode, Luke. "Cultural citizenship online: the Internet and digital culture." *Citizenship Studies* 14, no. 5 (2010): 527-542.

Hay, James, and Nick Couldry. "Rethinking convergence/culture: an introduction." *Cultural Studies* 25, no. 4-5 (2011): 473-486.

Hardie, Andrew. http://cass.lancs.ac.uk/?p=1133 . 2014

Hermida, Alfred. "From TV to Twitter: How ambient news became ambient journalism." *Media/Culture Journal* 13, no. 2 (2010).

Hill, Kevin A., and John E. Hughes. *Cyberpolitics: Citizen activism in the age of the Internet*. Lanham: Rowman & Littlefield Publishers, 1999.

Himelboim, Itai, Eric Gleave, and Marc Smith. "Discussion catalysts in online political discussions: Content importers and conversation starters." *Journal of Computer Mediated Communication* 14, no. 4 (2009): 771-789.

Jenkins, Henry. *Textual Poachers: Television Fans and Participatory Culture*. Cambridge, MA: MIT Press, 1992

Jenkins, Henry, Ravi Purushotma, Margaret Weigel, Katie Clinton, and Alice J. Robison. *Confronting the challenges of participatory culture: Media education for the 21st century*. Cambridge, MA: MIT Press, 2009.

Jenkins, Henry. "Rethinking 'rethinking convergence/culture." *Cultural Studies* 28, no. 2 (2014): 267-297.

Keren, Gideon, and Karl Halvor Teigen. "The probability—outcome correspondence principle: A dispositional view of the interpretation of probability statements." *Memory & cognition* 29, no. 7 (2001): 1010-1021.

Lawton, Kayte, and Matthew Pennycook. *Beyond the Bottom Line.* London: IPPR Resolution Foundation, 2013. http://www.ippr.org/files/images/media/files/publication/2013/01/beyond-bottom-line_living-wage_Jan2013_10162.pdf

Littlepage, Glenn E., Greg W. Schmidt, Eric W. Whisler, and Alan G. Frost. "An input-process-output analysis of influence and performance in problem-solving groups." *Journal of Personality and Social Psychology* 69, no. 5 (1995): 877

Livingstone, Sonia, and P. K. Lunt. *Talk on Television: The critical reception of audience discussion programs*. London: Routledge, 1994.

Liu, Yabing, Chloe Kliman-Silver, and Alan Mislove. "The tweets they are a-changin': Evolution of twitter users and behavior." ' In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 15-24. Menlo Park: AAAI Press, 2014.

Meraz, Sharon, and Zizi Papacharissi. "Networked gatekeeping and networked framing on# Egypt." *The international journal of press/politics* (2013): 1-29

Page, Ruth. "The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags." *Discourse & Communication* 6, no. 2 (2012): 181-201.

Papacharissi, Zizi, and Maria de Fatima Oliveira. "Affective news and networked publics: The rhythms of news storytelling on #Egypt." *Journal of Communication* 62, no. 2 (2012): 266-282.

Partington, Alan. "Evaluating Evaluation and Some Concluding Thoughts on CADS." In *Corpus-assisted discourse studies on the Iraq conflict: Wording the war* edited by John Morley and Paul Bayley, 261-303. London and New York: Routledge, 2009.

Penney, Joel, and Caroline Dadas. "(Re) Tweeting in the service of protest: Digital composition and circulation in the Occupy Wall Street movement." *New Media & Society* (2013): 74-90.

Penrod, Steven, and Brian Cutler. "Witness confidence and witness accuracy: Assessing their forensic relation" *Psychology, Public Policy, and Law* 1, no. 4 (1995): 817-845.

Price, Paul C., and Eric R. Stone. "Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic." *Journal of Behavioral Decision Making* 17, no. 1 (2004): 39-57.

Puschmann, Cornelius. "The form and function of quoting in digital media." *Discourse, Context & Media* 7 (2015): 28-36.

Quercia, Daniele, Jonathan Ellis, Licia Capra, and Jon Crowcroft. "In the mood for being influential on twitter." In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 307-314. Piscataway: IEEE, 2011.

Ranerup, Agneta. "Tools for Consumer Democracy?: The Role of Cybermediaries in Quasi-Markets." *E-service Journal* 2, no. 1 (2003): 90-113.

Rayson, Paul. "Wmatrix: a web-based corpus processing environment." (2009). http://ucrel.lancs.ac.uk/wmatrix/

Romero, Daniel M., Brendan Meeder, and Jon Kleinberg. "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter." In *Proceedings of the 20th international conference on World wide web*, 695-704. New York: ACM, 2011.

Suler, John. "The online disinhibition effect." *Cyberpsychology & Behavior* 7, no. 3 (2004): 321-326.

Thornborrow, Joanna. *The discourse of public participation media.* London: Routledge. 2015.

Tolson, Andrew. *Media talk*. Edinburgh: Edinburgh University Press, 2005.

Turner, Graeme. *Ordinary people and the media. The demotic turn.* London: Sage, 2010.

Van Dijck, José. "Users like you? Theorizing agency in user-generated content."" *Media, Culture, and Society* 31, no. 1 (2009): 41-58.

Van Dijck, José, and David Nieborg. "Wikinomics and its discontents: a critical analysis of Web 2.0 business manifestos." *New Media & Society* 11, no. 5 (2009): 855-874.

Walvoord, Ashley AG, Elizabeth R. Redden, Linda R. Elliott, and Michael D. Coovert. "Empowering followers in virtual teams: Guiding principles from theory and practice." *Computers in Human Behavior* 24, no. 5 (2008): 1884-1906.

Waseleski, Carol. "Gender and the Use of Exclamation Points in Computer-Mediated Communication: An Analysis of Exclamations Posted to Two Electronic Discussion Lists." *Journal of Computer-Mediated Communication* 11, no. 4 (2006): 1012-1024.

Watts, Duncan J., and Peter Sheridan Dodds. "Influentials, networks, and public opinion formation." *Journal of Consumer Research* 34, no. 4 (2007): 441-458.

Zappavigna, Michele. *Ambient affiliation: A linguistic perspective on Twitter*. London: Sage, 2013.

Zarnoth, Paul, and Janet A. Sniezek. "The social influence of confidence in group decision making." *Journal of Experimental Social Psychology* 33, no. 4 (1997): 345-366.

Zhang, Yixiand, Yulin Fang, Kwok-Kee Wei and Huaping Chen. "Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities." *International Journal of Information Management* 30, no. 5 (2010): 425-436.

Table 1: Analytic Corpus (Details at data collection point. Users' country of origin and gender are as indicated by the users. Date refers to the account creation date.)

User	Illustrative tweet containing the # for the influential user	Date	No. tweets / tokens	Follower count	Klout score
austerity (Australia / female)	Solidarity to a million-plus public sector workers on strike in UK today over #austerity cuts. ausunions	07-02-2012	85,411/ 45,272	2,928	62.24524
compassion (US / female)	If a Dharma practioner does nothing to benefit all sentient beings they are walking on ONE LEG. Dharma is Wisdom AND #compassion!	17-09-2012	131,799/ 43,422	134,833	71.09747
cuts (Tasmania / male)	Budget2014 #cuts inequality Looking to cut money in defence? Start with the Army History Unit's study into stained glass ausd…	02-09-2012	36,484/ 41,996	3,383	63.76673
homeless (Canada / male)	A start: housing for	10-03-2009	28,153/ 42,795	5,264	63.72059

	'some' #homeless - the solution: housing for all homeless				
justice (US / female)	Teaching My Children #Justice in an Un-Just World via	21-06-2007	92,494/ 39,697	21,714	74.48165
livingwage (US / female)	Back from MoralMonday. This child of UAW members expects a livingwage and equal pay for all Americans!	08-01-2011	100,709/ 46,162	7,214	63.32992
lowwage (US / male)	FastFoodStrike and $15/hour to stand up for #lowwage workers	05-01-2011	7,300/ 41,676	169	33.62881
minimumwage (US / male)	Well, HQ is in Seattle w/ their $15 #minimumwage - Your cup of Starbucks about to get even more expensive	09-12-2008	36,306/ 45,413	19,764	69.26137
poor (US / male)	kkk was formed during reconstruction. they were #poor whites, no trade skills, who were jealous of blacks flourishing while they struggled	19-08-2010	134,705/ 32,407	3,574	68.23324
poverty (US / male)	Gender inequality	17-04-2009	19,521/ 39,831	732	64.8066

	haunts many women well into retirement, leaving many living below #poverty				
raisethewage (US / male)	Add your name if you think working full-time for $14,500 a year isn't enough: #raisethewage	30-09-2008	35,479/ 43,267	6,191	68.26928
recession (US / male)	Obamacare could cause DOUBLE-DIP #recession: MakeDCListen Repeal tcot	30-07-2012	17,914/ 38,414	1,456	50.48324

Appendix: Key Semantic Domains

AUSTERITY

	Semantic domain	LL	LogRatio
1	Government	1097.99	3.95
2	Politics	493.24	3.07
3	Geographical names	436.4	1.32
4	Personal names	384.32	1.01
5	Grammatical bin	339.41	0.32
6	Money and pay	335.11	3.05
7	Law and order	310.31	2.52
8	Sailing, swimming, etc.	182.24	2.62
9	In power	169.59	1.54
10	Speech acts	168.5	1.04

COMPASSION

	Semantic domain	LL	LogRatio
1	Speech acts	1083.33	2.14
2	Religion and the supernatural	780.58	2.3
3	Other proper names	492.73	1.48
4	Helping	374.42	2.18
5	Emotional Actions, States And Processes General	236.74	2.87
6	Entire; maximum	231.89	1.41
7	Psychological Actions, States And Processes	229.58	4.41
8	Grammatical bin	228.33	0.27
9	Life and living things	225.72	5.14
10	Sad	171.78	2.02

CUTS

	Semantic domain	LL	LogRatio
1	Government	829.43	3.68
2	Grammatical	535.66	0.41
3	Politics	402.51	2.92
4	Money and pay	385.2	3.22
5	Money generally	161.36	1.64
6	Belonging to a group	159.28	1.69
7	Speech: Communicative	143.23	1.06
8	Deciding	140.34	-8.53
9	Speech acts	121.79	0.92
10	Money: Cost and price	108.51	2.62

HOMELESS

	Semantic domain	LL	LogRatio
1	Non-resident	2531.22	9.34
2	Architecture, houses and buildings	2074.25	4.74
3	Geographical names	865.57	1.73
4	Residence	863.78	3.45
5	Cheap	677.09	4.9
6	Politics	589.6	3.25
7	Money and pay	578.44	3.6
8	Places	469.22	2.6
9	Government	350.38	2.81
10	Belonging to a group	252.03	1.97

JUSTICE

	Semantic domain	LL	LogRatio
1	Food	973.76	2.41
2	Cheap	655.16	5.84
3	The Media: Books	632.63	3.9
4	Success	425.15	2.44
5	Business: Selling	347.08	2.28
6	Investigate,examine, test, search	300.9	2.31
7	No obligation or necessity	144.08	2.37
8	Giving	141.42	1.56
9	Cleaning and personal care	127.97	2.2
10	Degree: Approximators	126.75	2.43

LIVINGWAGE

	Semantic domain	LL	LogRatio
1	Politics	1429.74	4.17
2	Law and order	452.13	2.84
3	Government	339.31	2.77
4	Ethical	154.15	3.28
5	In power	137.82	1.42
6	Belonging to a group	130.88	1.53
7	Money and pay	106.07	2.07
8	Substances and materials generally	104.26	2.23
9	Grammatical bin	86.81	0.17
10	Personal names	69.86	0.48

LOWWAGE

	Semantic domain	LL	LogRatio
1	Living creatures: animals, birds, etc.	167.64	1.47
2	Grammatical bin	108.9	0.19
3	Sports	106.04	1.16
4	Personal names	78.18	0.52
5	Games	45.38	1.89
6	No obligation or necessity	44.58	1.54
7	Warfare, defence and the army; weapons	43.31	1.16
8	Government	34.02	1.23
9	Speech	33.92	0.53
10	Politics	33.58	1.2

MINIMUMWAGE

	Semantic domain	LL	LogRatio
1	Geographical	501.19	1.4
2	Warfare, defence and the army; weapons	457.38	2.7
3	Crime	371.33	3.32
4	Grammatical bin	239.29	0.28
5	Politics	237.55	2.42
6	Government	207.42	2.36
7	Measurement: Volume	171	4.64
8	Personal names	151.12	0.68
9	Places	148.72	1.73
10	Law and order	145.67	1.94

POOR

	Semantic domain	LL	LogRatio
1	People: Female	302.91	1.95
2	Speech acts	200	1.21
3	Recorded sound	189.28	3.12
4	Pronouns	182.93	0.33
5	Discourse Bin	163.87	0.89
6	Anatomy and phisiology	157.91	0.98
7	Comparing: Similar	154.81	2.46
8	Evaluation: Bad	152.9	1.76
9	Emotional Actions, States and Processes General	149.19	2.61
10	Kin	114.84	1.15

POVERTY

	Semantic domain	LL	LogRatio
1	Politics	294.4	2.67
2	Government	289.3	2.7
3	Grammatical bin	221.59	0.28
4	Geographical names	221.19	1.04
5	Healthy	123.92	3.3
6	In power	118.28	1.39
7	Sports	116.05	1.22
8	Belonging to a group	102.93	1.45
9	Happy	91.59	0.94
10	Personal names	84.74	0.54

RAISETHEWAGE

	Semantic domain	LL	LogRatio
1	Politics	1744.94	4.44
2	Government	443.33	3.05
3	Warfare, defence and the army; weapons	397.67	2.6
4	Law and order	320.92	2.57
5	Personal names	294.94	0.92
6	Money and pay	273.39	2.88
7	In power	225.24	1.74
8	Grammatical bin	175.15	0.24
9	Belonging to a group	138.07	1.59
10	Speech acts	122.93	0.92

RECESSION

	Semantic domain	LL	LogRatio
1	Medicines and medical treatment	629.26	3.36
2	Law and order	553.11	3.13
3	Geographical names	507.48	1.46
4	Warfare, defence and the army; weapons	487.84	2.84
5	Politics	330.76	2.79
6	Government	295.15	2.74
7	Grammatical bin	239.99	0.29
8	Personal bin	203.07	0.81
9	Dead	202.08	2.07
10	Crime	191.53	2.79