EXPLORING CARTOGRAPHIC DESIGN IN SOCIAL-NETWORK MAP MASHUPS

FIELD K.(1), O'BRIEN J.(2), CARTWRIGHT W.(3)
(1) Esri Inc, REDLANDS, UNITED STATES ; (2) Risk Frontiers, Macquarie University, NORTH RYDE, AUSTRALIA ; (3) Geospatial Science, RMIT University, MELBOURNE, AUSTRALIA

ABSTRACT

Online mapping landscapes have never been as rich or diverse yet this challenges cartography as new map makers present their work often eschewing long-established cartographic conventions. Blogs, micro-blogs and online forums are used to communicate ever more and maps are becoming an important part of that online world. The way that this information is mapped is often poor and based on little more than placing a geometric symbol at an imprecise geo-located point. Symbols often overlap, are poorly scaled and conflict with both each other and the map background. In this paper, we explore why many map mashups of social network data fail to communicate. We then explore what value can be extracted from the social network site Twitter that might be represented spatially and how the data can be used purposefully. We develop a range of maps using the ArcGIS javascript API in concert with Open Streetmap data and Arc Server map services to develop web maps for effective communication. We deal with the cartographic issues caused by mapping tweets at coincident points, representing the temporal dimension of the data and developing suitable designs to more effectively visualize Twitter feeds.

1. INTRODUCTION

The purpose of design is to clarify and to simplify and the power of design lies in its nuance which must be planned and skilfully achieved (Baer and Vacarra, 2008) yet we are surrounded by a world of poor web maps brought about largely by the democratisation of cartography and the massive increase in consumer-grade tools for the creation and consumption of maps by non-experts. This is particularly true of so-called map mashups (Field, 2008). One of Edward Tufte's oft-quoted statements informs us that "clutter is a failure of design, not an attribute of information" (Tufte, 1990) yet many mashups suffer information overload. With increasingly complex datasets, environments and interactions comes a need for clean, effective design, particularly in multimedia cartography displayed online (Cartwright et al. 2007).

The growth in social networking via blogs, micro-blogs and online forums has given rise to online maps defining our place and providing spatial context for our activities. People reveal locations for their identity and activities (Gibin et al. 2008, Field 2009a), the natural extension of which is the desire to represent them in map form yet the cartography is often lacking which renders the information one-dimensional (Field and O'Brien 2010a, O'Brien and Field in press).

This paper explores the design of mashups of the micro-blog postings (tweets) created by users of the social networking tool Twitter. Examples of Twitter maps are critically evaluated to demonstrate some of the cartographic design issues that need tackling. We then demonstrate how bringing together an understanding of online map services and web development with cartography and meaningful design can yield more effective Twitter maps. This work contributes to the development of effective map design for the Geoweb more generally (Haklay et al. 2008, Graham 2009) and reinforces the ideas that visual communication is more effective when deliberate aesthetic and cognitive choices are made to support the intended goal (Bertin 1983).

2. THE LACK OF CARTOGRAPHIC DESIGN IN MASHUPS

Today, a large proportion of web maps are authored by users leveraging web services as a result of advances in Web 2.0 and online mapping environments (Haklay et al. 2008, Graham 2009). The generation of map mashups using data from several sources results in a tendency for maps to become overtly cluttered which can reduce their effectiveness as communication devices (Rosenholtz et al. 2005, Fairbairn 2006). Mashups are relatively simple to construct and publish online but the combination of the data type, screen display, variable viewing scale, live data feeds and uncertainty in metadata presents challenges for revealing patterns cartographically.

With open data initiatives, more people are accessing more data and richer data and the drive to publish some newly available dataset online in map form as rapidly as possible seems to have become a goal in itself. This rapid approach to publishing mashups means sophisticated thought being applied to working out what the data means or what patterns might be revealed is often missing. When Wikileaks published the names and postcode of members of the British National Party in the UK (Wikileaks) the data was
swiftly mashed up with a Google base map service. The resulting map was a cartographic disaster that portrayed thousands of overlapping push-pins that revealed very little (Figure 1).

Figure 1. British National Party membership list mashup (TechCrunch Europe, © 2008 Google, Map Data © 2008 Tele Atlas)

The data itself only located a person to a postcode area, not a point, so the use of a push-pin was an inappropriate mechanism to accurately portray the data. It inferred a much more accurate position than was inherent in the data, simply revealed areas of larger population and failed to communicate anything meaningful over a generic base map. Such mashups tend to be conceptually simple, being little more than one layer of information being displayed in tandem with another. Preliminary data analysis should look at how different dimensions are ordered, how they cluster or show similarities, what outliers exist and so on in order to design a map that reveals patterns. Frye (2009) presents a good review of some of the problems with mashups and suggestions for making them more legible through better configuration of different layers of information.

3. EVALUATING THE DESIGN LIMITATIONS OF TWITTER MASHUPS

Twitter has become a hugely popular online social networking tool that allows users to post tweets of 140 characters in a similar fashion to SMS messaging but using an internet browser or mobile app. In essence, Twitter is a Web-based Internet Relay Chat (IRC) client and users create Twitter profiles that subsequently allow them to post tweets, follow others and also be followed. Twitter is based around a simple concept that asks users to post tweets to update their profile with “What’s happening?”. Since 2009 the Twitter API has allowed other web services and applications to leverage the various information streams it provides yet the spatial aspect of the data has received little attention (Field 2009a).

Embedding certain information within a tweet has allowed others to leverage both the tweet location and some attribution information in different ways. Using #hashtags allows people to embed key words into a tweet which can be collated and subsequently explored or visualized collectively where numerous people are tweeting about a similar subject; a principle that underpins the notion of the ‘Semantic Web’ to support online interaction and to augment established forms of collaboration with deeper interaction and consensus (Berners-Lee, 2001). Non-map examples of the visualisation of Twitter data illustrate how an overview of the interconnectedness amongst a collection of tweets can be harnessed (e.g. Clark: Twitarcs, Schmidt: Social Collider).

The use of embedded information in a spatial context is best exhibited by the uksnow map. As snow began to fall in February 2009 a #uksnow tag took hold on Twitter providing real time crowdsourced reports of snowfall where the tweet had "#uksnow", a UK postcode (first part only) and a subjective score out of 10 indicating how hard it was snowing. Figure 2 illustrates the UKsnow Twitter mashup that uses a push-pin as a locational device in tandem with Google Maps base data. The symbols reflect the snow intensity (using increasingly saturated reds), based on the subjective scores. Overlapping symbols hide a great amount of detail and rather that the reader inferring that a cluster of symbols is equivalent to a large snowfall, it simply reflects a large number of tweets being sent from that vicinity. The use of a standard Google base map creates unnecessary clutter – a problem with this style of mashup in general. The areas
with no markers do not necessarily reflect areas of no snow but areas of no tweets; a fact further illustrated by the urban bias in the information. The quality of the human sensors is debatable since no two people will assess snowfall intensity in the same way. Even when mapped, the spatial resolution is poor since an outward postcode will locate all point symbols to the centre of what is in fact a relatively large area (see also Marsh 2009, 2010).

Figure 2. UKsnow Twitter Map (Darbyshire, © 2009 Google, Map Data © 2009 Tele Atlas)

Field and O'Brien (2010a) review a range of mashups that make some use of the spatial component of Twitter and summarise a number of other problems that have characterised early attempts to visualize tweets effectively. Locations are often randomly assigned making the spatial component misleading (e.g. GeoTweeters). Variables such as relative spatial distance or frequency of communication are rarely reflected through design (e.g. UMmapper). Very little attempt is made to organise information using themes, topics, a numeric measure or even place (e.g. ArcGIS Twitter extension by Hussein). Temporal data is rarely visualized and, if presented at all, is shown in list form alongside a map (e.g. UKsnow Twitter Map by Darbyshire, #uksnow Tweets by Marsh, #uksnow Map 2.0 by Marsh). Where individuals have not set a location in their profile and do not use native GPS support on mobile devices, their location is set at 0o latitude, 0o longitude which clearly puts them off the African coast in the Atlantic Ocean. In all maps there is no attempt to explore ways of representing tweets from coincident or similar locations. Most maps use account level location which is not dependable (as it's impossible to verify whether a user is at their manually defined location). Finally, many tweets form part of conversational timelines but temporal dimensions are not reconciled.

Field and O'Brien (2010a) review a range of mashups that make some use of the spatial component of Twitter and summarise a number of other problems that have characterised early attempts to visualize tweets effectively. Locations are often randomly assigned making the spatial component misleading (e.g. GeoTweeters). Variables such as relative spatial distance or frequency of communication are rarely reflected through design (e.g. UMmapper). Very little attempt is made to organise information using themes, topics, a numeric measure or even place (e.g. ArcGIS Twitter extension by Hussein). Temporal data is rarely visualized and, if presented at all, is shown in list form alongside a map (e.g. UKsnow Twitter Map by Darbyshire, #uksnow Tweets by Marsh, #uksnow Map 2.0 by Marsh). Where individuals have not set a location in their profile and do not use native GPS support on mobile devices, their location is set at 0o latitude, 0o longitude which clearly puts them off the African coast in the Atlantic Ocean. In all maps there is no attempt to explore ways of representing tweets from coincident or similar locations. Most maps use account level location which is not dependable (as it's impossible to verify whether a user is at their manually defined location). Finally, many tweets form part of conversational timelines but temporal dimensions are not reconciled.

Harvesting Twitter data has become a popular pursuit for those wishing to create mashups See for instance Twittervision (Troy), Flutter (Bradshaw), Twittearth (Digitas), London Twittercloud (DigitalUrban), London Weekend Twitter Map and Tweetography (Urbanick), NodeXL (Smith), Mentionmap (Asterisq) and Invisible Cities (Schmidt). All of these mashups suffer from a fundamental problem which has been succinctly described by Steinitz (2010) who states that "much of it is visualisation but not communication...there is little point to much of it and it’s not particularly 'useful'. It’s made worse because there is a preoccupation with making things go round and round and up and down…worse if it’s accompanied by music".

There are useful spatial contexts that can be derived from tweets but there remain problems with the organisation and visualization of the mapped outputs. For instance, there is very little attempt to order information semantically by using themes, topics, common values, a numeric measure or even place (other than by proximity). Neither is there an attempt to address the problem of how to represent tweets sent from coincident or similar locations cartographically or the temporal dimension. A problem when using live feeds of data such as Twitter is that the position of the overlaid data is an unknown parameter in the map's design and construction. The base map and thematic detail cannot easily be designed to work together spatially or temporally.

4. DEVELOPING A FIT FOR PURPOSE TWITTER MAP WITH PURPOSE
The project described here was driven by our own need for an online Twitter map to support student learning on fieldwork. The fostering of a collaborative approach to data gathering is only partially served by traditional fieldwork (Drummond et al. 2006) and limitations exist in terms of what different groups of students achieve. For instance, groups tend to operate independently from each other and gather data in different ways (e.g. classifying features using different object types, capture resolutions, attribution and detail) which creates problems when data is combined.

As part of a wider project (Field, 2009b; Field and O’Brien, 2010a, 2010b), our approach was to develop a mobile learning collaboratory which students could engage with using their personal technologies. The development of an online map that acted as a place-holder giving spatial context to the online collaboration was an important part of the design to facilitate asynchronous collaboration. A proof-of-concept was originally developed to support student fieldwork in 2009 (Field, 2009b) to explore the value of combining social networking tools with map mashups for learning. Our #malta09 TweetMap made use of the Twitter search API to find the #malta09 hashtag which was being used to denote content related to the 2009 fieldwork activities. The technical design and implementation of the application is described in Field and O’Brien (2010a) and Figure 3 illustrates the prototype #malta09 TweetMap showing some of the student tweets over a standard Google Maps image layer map service.

Although the prototype demonstrated the technical feasibility of the mashup and supported a range of pedagogic goals (Field 2009b, Field and O’Brien 2010a, 2010b, 2010c), it suffered from many of the same drawbacks and criticisms we previously identified on other Twitter mashups.

The Association for Geographic Information (AGI) GeoCommunity 2009 conference (in Stratford-upon-Avon, UK) provided a case study to explore ways of representing spatially coincident tweets since tweets were sent predominantly from the conference venue by Twitter users amongst the 600+ delegates. Our desire was to explore ways of representing the spatial component of the tweets more effectively whilst also reconciling the temporal component to make the conversation timeline more visible. Our AGI #geocom TweetMap ignored the reported location of the attendees and focussed instead on the content of their tweets arranging them in different ways around the conference venue as the spatially coincident point. A range of alternative approaches were designed. The first used a random positional location (Figure 4) with more recent tweets denoted by larger place markers and common words being represented by similar colours (e.g. green identified comments relating to the keynote lecture).
Our second approach used a series of alternative cartographic representations for place markers arranged in a spiral around the venue with most recent tweets largest and on the outside of the spiral. The original tweet is therefore given a spatial context by proximity to the conference venue with location being used innovatively as a visual variable (Bertin 1983, MacEachren 1995), offset from the venue in an ordered, predictable fashion as a golden (logarithmic) spiral (Knopfmachera 2007) to indicate the temporal dimension and positioned chronologically amongst other tweets.

Figure 5a shows the same colour-coded theme used on the randomly distributed marker map so some sense could be made of tweets with related themes based on keywords in the content. Older tweets were towards the centre of the spiral and represented with a smaller marker symbol to promote more recent tweets to the foreground using larger marker symbols. Figure 5b replaced the thematic colour-coding with a blue-red dichromatic colour ramp to show red symbols as ‘current hot topics’. Figure 5c used a uniform pictorial design to reflect the general nature of the tweeted data (i.e. by using a recognisable Twitter logo). Finally, Figure 5d shows the Twitter profile avatar used as a marker symbol. Figures 5b-d all lose the ability for users to identify tweets thematically though user testing revealed this to be of little concern. The avatar version (Figure 5d) proved to be the most popular with 78% of the sample of 124 people surveyed at the conference saying it provided both a greater quantity and richer information (Field and O’Brien 2010a). Users preferred the ability to follow tweets by contributor rather than by thematic topic.
The #geocom TweetMap informed development work for the 2010 Malta fieldcourse so that the spatial and temporal detail were more suitably represented. Further work in the design of the map was undertaken to specifically tackle the figure-ground relationship and to provide a more appropriate base map. In order to have full control over the underlying map service, ArcServer™ and the ArcGIS® API for Javascript was used to construct the map so that our base layer of choice could be served. This allowed us to fully control the appearance of the base map (by designing in ArcGIS®). The use of map services from our own server gave us added value by enabling us to serve historic imagery that students could use in the field as a context for their work. It would be feasible to serve data from the Esri® World Topographic map template or ArcGIS® online using the same process though for the small areas we are interested in on Malta the level of detail was insufficient. Instead, we served Cloudmade styled OpenStreetmap data as a tiled map service.

The 2010 #malta10 TweetMap was used in three ways during the 2010 fieldcourse. Firstly, students were required to use it as a general collaboratory during the fieldcourse which enabled the collection of a chronological field log. Figure 6 illustrates the use of simple marker symbols at the small scale zoom level to locate students over a sparse base map. The marker symbols allow mouseover actions that shows a popup of the tweet.

Secondly, the #malta10 TweetMap was used to support the land use mapping exercise in Mellieha and was developed for to enable exchange of ideas, interpretations of landscapes and the search for common frameworks for data collection. Figure 7 shows a medium scale zoom and since the exercise required some ability to navigate the urban area, roads and building detail was added. Historic imagery could also be served as well as using Google Maps imagery map services so students could explore landscape change visually whilst they were undertaking the exercise. The avatar marker symbols allow mouseover actions that shows a popup of the tweet.
Finally, the #malta10 Tweetmap was used to visualize, in real time, an exercise designed to explore differences in navigation based on a comparison of use between paper maps and GPS (Figure 8). The exercise required a large scale base map that gave some contextual detail (e.g. points of interest) and place/road names and the temporal timeline approach was used to visualize the tweets and log the temporal detail of the order in which students reached the waypoint. The avatar marker symbols allow mouseover actions that shows a popup of the tweet. A visual word tag cloud was rendered around the waypoint as part of a different exercise. This approach allowed us to extract key information from the separate tweets and represent it in an aggregated way to provide some quantitative view of the data contained in the tweets.
5. CONCLUSIONS

The pedagogic aspects and technical implementation of the Tweetmaps are explained elsewhere (Field and O'Brien 2010a, b, c) but here, we illustrate the value of providing a coherent map display that treats the base map service as an important component in tandem with developing techniques to represent overlaid information that are both fit for purpose and fit the purpose. For each of the three exercises that used the #malta10 TweetMap a range of different considerations were dominant in defining what information the map should be designed to show. The design was then a reflection of that purpose.

Simple and ineffective push-pin style mashups need to be replaced by more complex maps that reveal far more information through the processing of data and the representation, cartographically, of characteristics of the data. Tufte (2009) provides a fitting explanation of the general theory that defines our approach here: 'to clarify, add detail’. In this sense we are recognising that in many mashups the information is in chaos yet instead of discarding information we fixed the design (Tufte, 2009).

ACKNOWLEDGEMENTS

This work is part funded by the transforming curriculum delivery through technology JISC e-learning programme Mobilising Remote Student Engagement (MoRSE) project. It is also part-funded by the Geography, Earth and Environmental Sciences (GEES) Subject Centre Learning and Teaching Development fund programme Mobile Decision-making in the Cloud (MobiDIC) project. The authors completed the work whilst working at Kingston University London.

REFERENCES


