

Social Glass: A Platform for Urban Analytics and Decision-making Through Heterogeneous Social Data

Stefano Bocconi, Alessandro Bozzon, Achilleas Psyllidis, Christiaan Titos Bolivar,
Geert-Jan Houben
Delft University of Technology
Delft, The Netherlands
{s.bocconi, a.bozzon, a.psyllidis, c.titosolivar, g.j.p.m.houben}@tudelft.nl

ABSTRACT

This demo presents Social Glass, a novel web-based platform that supports the analysis, valorisation, integration, and visualisation of large-scale and heterogeneous urban data in the domains of city planning and decision-making. The platform systematically combines publicly available social datasets from municipalities together with social media streams (e.g. Twitter, Instagram and Foursquare) and resources from knowledge repositories. It further enables the mapping of demographic information, human movement patterns, place popularity, traffic conditions, as well as citizens' and visitors' opinions and preferences with regard to specific venues in the city. Social Glass will be demonstrated through several real-world case studies, that exemplify the framework's conceptual properties, and its potential value as a solution for urban analytics and city-scale event monitoring and assessment.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/ Hypermedia; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – search process, query formulation

Keywords

Urban Analytics, User Analysis, Social Data

1. INTRODUCTION

The growing urban populations pose broad challenges to contemporary cities that span across domains as diverse as transportation, energy, and environment. As a consequence, there is an increasing demand for methods and tools that extend the current approaches towards urban systems. In this direction, the growing availability of data stemming from interconnected devices and social media platforms provides emerging opportunities for supporting urban planning and decision-making processes. *Social Data*, i.e. data generated

by users in social context like online social networks, recently attracted a lot of attention as a potential source of knowledge to better understand urban environments, citizens' activities, and (environmental) phenomena. Many examples hinted to a successful usage of social data (microblog posts, images, etc.) to gain knowledge about urban areas (a neighbourhood, a city, or an urbanised region) [2][3].

The analysis and integration of social data in planning support systems have potential to provide meaningful insights about urban dynamics that would have otherwise been impossible to gain, simply by using static governmental records. Before adoption and consideration by scientists and decision-makers, social data must be subject to meticulous study that qualifies their ability to represent or describe the targeted reality. Typical questions are: 1) Do data possess enough (temporal and spatial) resolution to represent the studied phenomenon? 2) Are there biases (of cultural or technological nature) that might influence the conclusions that could be drawn by such data? 3) How does the world depicted by social data compare with common or consolidate knowledge (e.g. from census or municipal databases)?

In this demo we introduce **Social Glass**, a novel web-based platform that supports the analysis, valorisation, integration, and visualisation of large-scale and heterogeneous urban data in the domains of city planning and decision-making. We show how publicly available social datasets from municipalities, social media streams (e.g. Twitter, Instagram and Foursquare), and resources from knowledge repositories can be systematically combined and visually explored to gather insights about urban phenomena. **Social Glass** enables the mapping of demographic information, human movement patterns, place popularity, traffic conditions, as well as citizens' and visitors' opinions and preferences with regard to specific venues in the city. A crowdsourcing module allows a pro-active interaction with individuals and groups (drawn from social media or human computation platforms) for data creation, cleansing, and interpretation purposes.

The main contribution of our work is a urban analytics platform that blends state-of-the-art methods and tools for social media content analysis, user modelling, semantic data integration, crowdsourcing, and visual data exploration. **Social Glass** has been deployed for the monitoring of several city-scale events, including the 2014 edition of the "Milan Design Week", "Como Summer Holiday Seasons", and the "Amsterdam Light Festival".

Additional information about the **Social Glass** platform is available at <http://www.social-glass.org>.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). IW3C2 reserves the right to provide a hyperlink to the author's site if the Material is used in electronic media.
WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
<http://dx.doi.org/10.1145/2740908.2742826>.

2. THE SOCIAL GLASS PLATFORM

The **Social Glass** platform sits on top of a multi-layered architecture, depicted in Figure 1. The demonstration leverages on content retrieved from *social media* (e.g. Twitter, Instagram, FourSquare), mobile phone data, and spatial statistics and demographics. The platform is however able to work with a broad range of data sources, including human-related sensor data such as dynamic sensor streams (e.g. environmental, traffic).

The *Ingestion and Analysis* tier is responsible for the acquisition, cleansing, and analysis of social data, possibly with the explicit involvement of users. The *Fusion* tier caters for integration interoperability issues across different data sources and usage domains. Finally, the *Exploration and Visualisation* offers user interfaces for data exploration, comparison, and urban analytics.

2.1 Ingestion and Analysis

The diversity, quantity, and speed of social data sources pose several scalability challenges. **Social Glass** addresses such challenges with an architectural design optimised for modularity, loose coupling and scalability. Each functionality is performed by a single module, and each module focuses on a specific functionality; a module can independently and redundantly be deployed, so as to enable dynamic allocation and distribution of computation and storage resources. The communication between modules is implemented by means of message queues, and producers can post messages without knowing what module will consume them at the other end of the pipeline. Thanks to this modular, loosely coupled architecture, the **Social Glass** back-end can accommodate new data sources or analysis components, with relatively low effort.

The *Ingestion and Analysis* tier is roughly organised in three sub-systems, respectively devoted to *Social Media Analysis*, *Crowdsourcing*, and *Sensor Analysis*. This demonstration showcases the functionalities of the former two sub-systems, described in the following sections.

2.1.1 Social Media Analysis Sub-System

The Social Media Analysis (SMA) subsystem is devoted to the generation of information about a given urban environment from the content created by users of popular social media platforms like Twitter, Instagram, and FourSquare. To this end the SMA subsystem includes the following analysis modules, all devoted to the generation of insights about city users, points of interest, and their spatial and temporal relationships.

Point of Interest Mapper. This module maps geo-localized microposts to urban Points of Interest (POI). In the current implementation, this is achieved through the “check-in intent” Foursquare API¹, which selects the most popular venues among the ones that are in the proximity of the user post. The API returns a list of venues and we select the first venue within a 40-meter radius. Thanks to the POI mapper, it is possible to enable the spatio-temporal analysis of an urban area according to the functional categorisation (e.g. restaurant, museum, transportation) of a location.

Semantic Analyser. This module processes the content of micro posts to extract name entities (such as a person’s

name, or the name of a location) from its textual content, and connects them to entities in DBPedia². In this way it is possible to associate social media users, spatial areas, and/or temporal intervals with the entities mentioned in the related posts, so as to create topics profiles. Entity extraction relies on DBPedia Spotlight, which features annotation functionalities in several languages.

Demographic Profiler. Given social media users observed in an urban area of interest, the module estimates their *gender* and *age*. This is achieved by means of a multi-modal decision tree classifier: starting from the profile picture and the name of a user, the module combines the output of a state-of-the-art face detection and analysis component **Face++**; with the output of a dictionary-based gender recognition module **Genderize**, which consumes the home location of a user to disambiguate country-dependent names (e.g. “Andrea” is a male or female name in different countries).

Home Locator. This module estimates the home location of users thanks to a recursive grid search, based on the geo-location of their microposts. The search finds the actual place where the user posts most often, and uses that as an approximation of the user’s home location. This is achieved by clustering the adjacent areas where the user posts the most, and subdividing those further until a square of approximately 20 meters. Once identifying the coordinates of the estimated home location, we use a reverse geocoding service (**Geo-names**) to identify the *city* and *country* of origin of the user.

User Role Identification. Functional to urban analytics purposes is the characterisation of the *role* a user plays in a city. This module classifies users in relation to their home city, by the estimated city and country of residence with the currently analysed urban area. A user can belong to one of the following classes: *Resident*, if the city of the users home location is the same as the city under study; *Commuters*, if the city of the users home location is different than the city under study, but it is still in the same country; and *Foreign Tourist*, if the user’s home location is in a different country.

Path Extraction. The module constructs paths by examining all geo-located posts of a user in a fixed time period. Paths are formed by concatenating the coordinates of subsequent posts. Each point in the path is mapped to a known venue using the Venue Mapper module. To gather patterns in user paths, we use the PrefixSpan (Prefix-projected Sequential Pattern mining) algorithm. This algorithm enumerates frequent sequential patterns from a set of sequences, where frequent means that a sequence occurs in no less than N sequences, with N a threshold value.

2.1.2 Crowdsourcing Sub-System

To cater for issues related to data sparseness, veracity, and sense-making, **Social Glass** includes a crowdsourcing sub-system, working on top of the CroKnow³ platform. The module can operate in two modes: 1) *Social Sensing* mode, which has the ability to contact social media users in order to request services such as on-demand data creation, cleansing and linkage; 2) *Human Computation*, to engage with anonymous crowd from human computation platforms such as CrowdFlower and Amazon Mechanical Turk. Examples

¹<https://developer.foursquare.com/docs/venues/search>

²<http://dbpedia.org/>

³<http://www.wis.ewi.tudelft.nl/CroKnow>

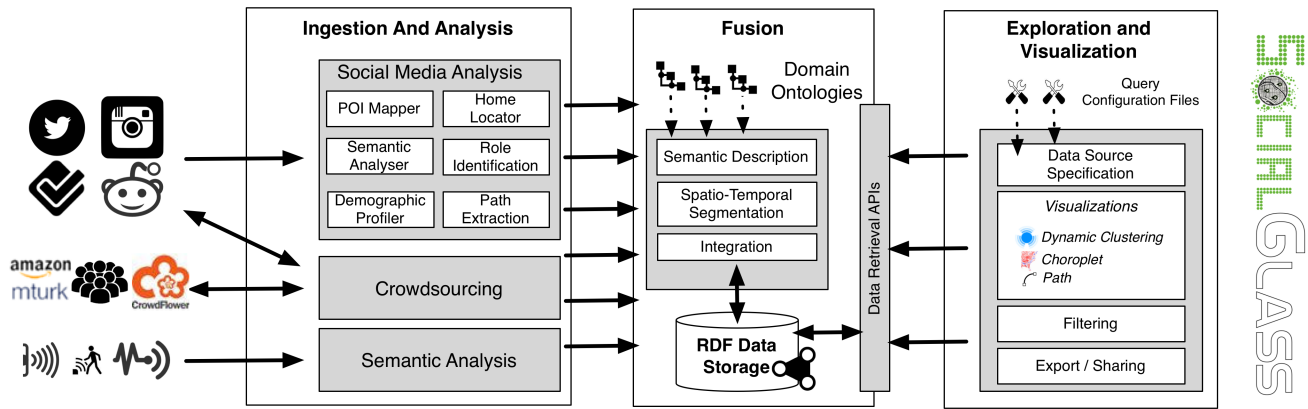


Figure 1: Architecture of the Social Glass platform

of operations enabled by the crowdsourcing module are: on-demand sensing of urban or environmental (e.g. rain, temperature) phenomena; disambiguation of textual and visual content w.r.t. sentiment or tone; verification of relatedness and appropriateness of images w.r.t. an observed event, and a targeted user group.

In this demonstration we will showcase the *Social Sensing* mode. When receiving the identifiers of people to be potentially engaged, the module takes care of sending them welcome and reminder messages to motivate them to opt-in in the service. If a user to be engaged has not received any message yet, the module invites them through the references social media (e.g. Twitter) to follow its account. If the user is already a follower (i.e. she already expressed interest in participating in the crowdsourcing activity), then the module can send her a direct message containing a link to the data operation to be performed.

Using their Twitter account, users can access a personal dashboard showing a summary of their planned or completed activities. The crowdsourcing module can also be used as a simple communication tool with a reference crowd. For instance, Figure 2d shows an example of dashboard configured to show venues for the Milano Design Week scenario [1]. There, crowdsourcing has been used to gather feedback about the perception of usefulness and interest in the recommended venues, so to improve the performance of the recommender.

2.2 Fusion

The goal is to structure and encode social data gathered from the ingestion tier. When combining data from complex urban environments it is easy to encounter syntactic and semantic discrepancies, mainly due to spatial, temporal and/or thematic diversities.

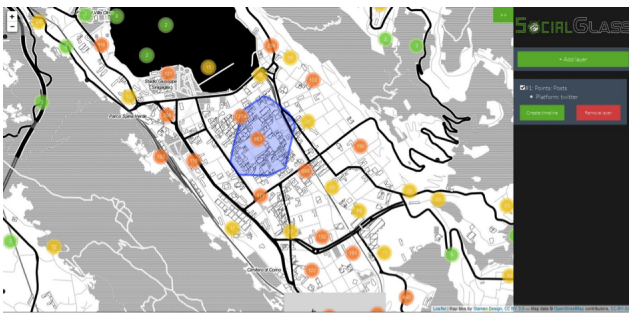
A coherent data representation simplifies their usage, and unlocks their combined value. This is achieved by formally representing urban systems, the relations among them and the respective data sources with several semantically enriched knowledge models, based on ontologies [4]. This particular aspect of the framework harnesses the potential of combined data silos, allowing for a more complete image of specific urban situations. By exploiting such knowledge representation models, the system enables data discovery and reuse for the provision of new services and applications.

2.3 Exploration and Visualisation

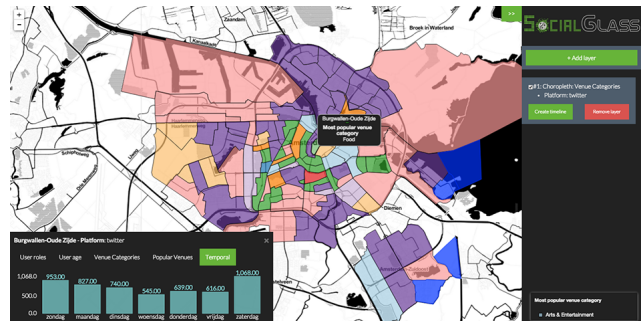
The *Exploration and Visualisation* tier uses the APIs of the *Fusion* tier to provide user interfaces for urban analytics. Figure 2 depicts some examples of data visualisations showcased by the demo. Interactive data exploration tools allow users to drill into the gathered data, to extract insights about the spatio-temporal analysis of the observed urban environment. To ease exploration, the user interface allows the creation of an arbitrary number of *layers* that can be created, deactivated, and organised in an arbitrary order. Each layer visualises a partition of the retrieved data, and is characterised as follows.

Source. A layer insists on a source that is typically originated by enriched, but siloed, social media (e.g. Twitter, Instagram), sensor, and statistical data. Thanks to the *Fusion* tier, it is also possible to specify custom queries spanning multiple data sources. To provide a uniform spatial and temporal segmentation of heterogeneous data, we exploit the notion of *city Pixel*, i.e. a geographical partition of the urban space (defined geometrically – e.g. square of 250x250 meters – or administratively – e.g. districts or neighbourhoods), temporally split into time slots.

Visualisation. *Social Glass* currently offers the following set of map-based visualisations: *a) Clustered Points.* As depicted in Figure 2a, this visualisation displays geo-located objects, e.g. micro posts, point of interest, sensor data. Objects are dynamically clustered according to their spatial proximity to a common location. Zooming modifies the cluster granularity, thus giving an overview of the number of objects in the area. *b) Choropleth.* The visualisation builds upon the urban space partitions of the underlying data source. For each partition, it can show: 1) a colour representing a category; or 2) a colour shade, in proportion to the measured variable. Figure 2b shows an example of the social media activity intensity in the different administrative district of Amsterdam. *c) Path* (Figure 2c) shows arcs between locations, typically POIs, that denote a paths taken by users to traverse the city. The thickness and colour of the arc show the popularity of that path among the considered set of users. By clicking on an area or a path, a popup appears in the lower left part of the screen. It contains several information about the area clicked on, such as: 1) the distribution of users according to their role, age and gender; 2) the popularity of point of interest and venue cat-



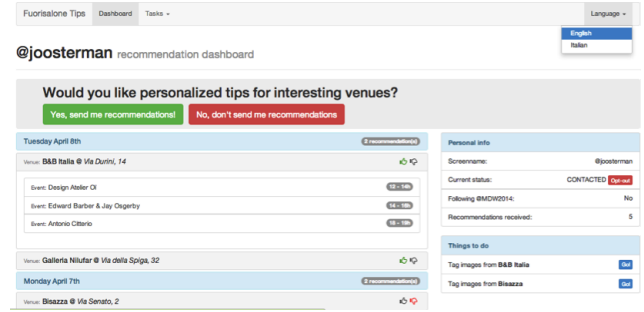
(a) Clustered visualisation of geo-located microposts in the *Como* scenario.



(b) Choropleth visualisation of social media activity intensity in the *Amsterdam* scenario.



(c) Visualisation of tourists' paths in the *Amsterdam* scenario.



(d) The personal dashboard in the *Milan* scenario

Figure 2: Examples of data visualisations and crowdsourcing dashboard of the *Social Glass* platform.

egories; 3) temporal distribution of micro posts across the week days; 4) semantic profile of the users that were posting from the area; and 5) other static data taken from public data sources (e.g. crime rate).

Filter. Once retrieved, data can be further filtered according to a number of conditions, including: *a) Time Span:* when temporally annotated, only data produced within a given temporal window of interest are shown; *b) Point of Interest Category:* when linked to POIs, only data produced within a given temporal window of interest are shown; *c) Activity Intensity:* for layers based on social media or sensor data it is possible to set maximal and minimal threshold of activity; *d) User type:* for layers related to users, it is possible to filter data according to the particular user type (or role) linked to them.

Export and Sharing. Users are also provided with functionalities for exporting the current result set in various formats (CSV, JSON, RDF) and save the current exploration status for sharing or further reuse.

3. DEMO HIGHLIGHTS

The attendees will be given the opportunity to have a hands-on experience with the *Social Glass* platform. They will be allowed to interact with several data sources, and explore data with the available palette of visualisations and filtering operations.

The demonstration will focus on 4 scenarios: *a)* “Milano Design Week”, which features Twitter data collected for one week over the Milan metropolitan area; *b)* “Como Summer Holiday Seasons”, which combines Twitter and mobile phone data gathered during a period of 4 months in the city of Como; *c)* “Amsterdam Light Festival”, where Twitter

and Instagram data are integrated with municipality data sources, and statistics collected from the event’s organisers; and *d)* WWW 2015, an instance where data from the Florence area are ingested and analysed in near real-time.

The demonstration will also highlight the challenges posed by social data in urban analytics. Drawing from previous [1] and novel experiments, we will show how, contrary to the current understanding, social media data are *not always sufficient* for deriving a detailed and *accurate* view of activities within an urban area.

Acknowledgements. This research is partially funded by the EIT ICT Labs, and by the Amsterdam Institute for Advanced Metropolitan Solutions.

4. REFERENCES

- [1] M. Balduini, A. Bozzon, E. D. Valle, Y. Huang, and G. Houben. Recommending venues using continuous predictive social media analytics. *IEEE Internet Computing*, 18(5):28–35, 2014.
- [2] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. In *5th International AAAI Conference on Weblogs and Social Media (ICWSM’11)*, pages 81–88. AAAI, 2011.
- [3] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.
- [4] A. Psyllidis and N. Bilorla. Ontopolis: A semantic participatory platform for performance assessment and augmentation of urban environments. In *International Conference on Intelligent Environments 2104*, pages 140–147, June 2014.