Classifying and visualizing the social facet of location-based social network data

Thomas Gründemann  
Institute of Cartography  
TU Dresden  
Dresden, Germany  
thomas.gruendemann@tu-dresden.de

Dirk Burghardt  
Institute of Cartography  
TU Dresden  
Dresden, Germany  
dirk.burghardt@tu-dresden.de

Abstract—There has been limited research in the GIScience that takes social aspects into account. This paper develops a classification of the social facet in location-based social network data based on a literature review and an exploratory survey of different location based social networks. The four proposed components of the social facet are: social actor, socio-demographic features, social group and social role. Furthermore, using a sample data set, various options for representing social attributes that are suitable for visual analysis are presented.

Keywords—social, classification, visualization, social network data

I. INTRODUCTION

Location-based social networks (LBSN) offer a huge amount of user-generated data. They provide a variety of opportunities for the research areas of the GIScience to understand geographic processes and spatial relationships.

However, the potential of LBSN data has not yet been fully exploited. Research in the GIScience is often limited to the spatial, temporal and thematic facet of LBSN data by analyzing only the geolocation, timestamp and content of the user contributions. By creating their own profile sites and relationships to other participants, the users disclose information with social characteristics that can be used to better understand the behavior of the users within the social network. Therefore, we must not only focus on user contributions, but also on the social facet of LBSN data that characterizes the users themselves and their connections to other users. Despite the extensive research using LBSN data, little differentiation of data types with social characteristics is made. The contributions of this work are as follows.

- Developing a classification for describing the social facet in LBSN data
- To demonstrate how the classification can be applied in visual analytics.

II. REVIEW OF THE LITERATURE

A. Defining the social facet in LBSN data

The Oxford English Dictionary1 defines notion social in a broader sense as “Needing companionship and therefore best suited to living in communities”. This means that the concept of the social describes the community as a precondition for action. In the context of online LBSN, community (often called as online community) represents a definite number of actors grouped together by different characteristics and interacting with each other. Thus, the users form the basis of every LBSN. [1] state that “The very word ‘social’ associated with media implies that platforms are user centered and that they facilitate communal activities […]” (p. 11).

Social networking services (SNS) lay the foundation for the interaction among the users by providing the infrastructure. Many definitions for the term social network service are also centered around the user. [2] define SNS as web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they share a connection, and view and traverse their list of connections and those made by others within the system. [3] synthesized existing definitions and identified four commonalities of SNS: Social media services are Web 2.0 Internet-based applications were users interact among each other. User-generated content is the basic requirement for the functioning of the social network. Individuals and groups create site-specific user profiles within the SNS to enable network connections and SNS facilitate the development of social networks by connecting a profile with those of other individuals and/or groups.

These definitions illustrate that the users (social actors) and the relations between the users are the essential elements of the social network concept. The social facet can therefore be defined as the part of the social network that characterizes the user on the basis of various social attributes.

B. Classifications of social components in LBSN data

Literature on the social component in LBSN data lacks a comprehensive classification framework. Existing research has tended to focus on particular research questions rather than investigate a broad set of issues; for example [4], [5], [6] and [7] use demographic attributes, such as gender, religion or ethnicity to describe users. Social actors are classified in [8] and [9] using the information in the profiles. [10] and [11] describe users according to their influence in the social network and [12] according to the community they belong.

The classification of the social component is closely linked to the conceptualization of social network data types. [13] deduce user-related data types from the definition of SNS by [2] and derive additional data types by focusing on the goal of the SNS. Based on the taxonomy of social networking data by [14], [15] divides the personal data processed by Facebook into ten user-centric categories. [16] also take the approach of [14] and develop it further into the two main categories: explicit data and implicit data. [17] propose a fine-grained taxonomy of social network data

---

1 https://en.oxforddictionaries.com/definition/social
types and distinguish between the two main categories service provider-related and user-related data.

All approaches aim at developing a general classification of social network data types and differ in the level of granularity. To the best of our knowledge, there exist no classification that focus on social components in social network data types.

III. METHODOLOGY

Because of the complexity of social components in social network data and the limited empirical research, the development of a classification was derived from the literature review and an exploratory survey of different LBSN. The literature review focused on papers developing conceptual frameworks of social network data types and research describing methods for extracting and analyzing social attributes from SNS. This was followed by an in-depth analysis of the five LBSNs Twitter, Facebook, Instagram, Flickr and LinkedIn to extract social network data with social attributes.

IV. FINDINGS

Based on the analysis of existing literature and the five LBSNs in section II it can be deduced that primarily two social network data types contain social attributes: (1) profile data and (2) network data.

Each LBSN offers their members the ability to create a profile. It contains personal data describing the characteristics of the user, e.g. name, age, gender but also interests and preferences. Which profile data is present, depends on whether the LBSN is a platform for general purposes (e.g., Facebook) or for rather specialized ones (e.g. LinkedIn). Name is mandatory on all inspected LBSNs. Birthday and gender are only required on Facebook. LinkedIn forces the user to indicate his country and postal code for networking purposes and his job status. Although all LBSNs offer the user the ability to provide a variety of additional attributes such as basic info, contact info, work and education, the provision of this attributes is optional.

The profile enables to connect with others, because users can only be identified and found based on the personal information contained in the profiles [3]. The connections that form the basis for social interactions between the users, are described by the network data. Network data may be uni- or bidirectional (e.g. friendship or follower) and differ in the strength of a connection (e.g. interaction activity). The main difference between LBSNs is whether the connections are bidirectional (e.g., Facebook, LinkedIn) or unidirectional (e.g., Twitter, Instagram, Flickr). Facebook is the only LBSN that supports both types of connections at the same time.

Extracting social attributes from available LBSN is a difficult task. Not all attributes are explicitly provided by the users and can be extracted directly from the profile entries, such as age or gender. Social attributes can also be derived from one or more user-provided information (e.g. using the user's name to determine the age) or inferred by analyzing the user's behavior. Furthermore, it is possible to derive attributes from the combination of information. For example, if a significant number of contacts live in one city, it can be concluded that the user might live there as well [16].

V. PROPOSED CLASSIFICATION OF SOCIAL COMPONENTS IN LBSN DATA

From the findings of the literature review and the in-depth analysis, four different social components can be determined: (A) social actor, (B) socio-demographic features, (C) social group and (D) social role. The components “social actor” and “socio-demographic features” describe the individual characteristics of the actor (personal identity) and the components “social group” and “social role” describe the actors integration into certain groups (collective identity).

A. Social actor

An actor is a social entity that interacts with other entities. On LBSN the concept of the social actor can refer to various types of entities, mainly individuals, organizations (companies, institutions), and bots. The social actor is only the digital representation of the type of user. The social actor of the type individual is in the physical world a real person. Responsible for the online presence of organization is often the public relations department or the social media manager of the organization. A bot in OSN is a computer algorithm that automatically produces content.

Distinction should be made which intentions or purposes guide the actions of the different user types. Individuals often want to promote themselves, get the latest news, or stay in touch with friends. Organizations use social network services for target-oriented advertising, marketing campaigns and to communicate with customers, vendors, and the public at large [18]. Bots intent to inform other users in the case they provide content from automated sources (e.g. sensors, news feeds) or to alter the behavior of other individuals or organizations by exhibiting human-like behavior [9].

B. Socio-demographic features

The component "Socio-demographic features" contains a variety of attributes that primarily characterize the social actor of the type "individual". The main attributes are summarized in Table 1.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Example of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Female, Male</td>
</tr>
<tr>
<td>Age</td>
<td>Age in years</td>
</tr>
<tr>
<td>Marital status</td>
<td>Married, Single</td>
</tr>
<tr>
<td>Place of origin</td>
<td>Germany, Berlin</td>
</tr>
<tr>
<td>Nationality</td>
<td>German</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Religion</td>
<td>Christian, Muslim</td>
</tr>
<tr>
<td>Education</td>
<td>Primary education, Master</td>
</tr>
<tr>
<td>Occupation</td>
<td>Journalism, Healthcare, Industrial occupations</td>
</tr>
</tbody>
</table>

The extent and type of socio-demographic features vary in the LBSN. They depend on the input fields provided by the SNSs and the willingness of the users to disclose the information. Socio-demographic features play an important role in the social network data.
Fig. 1. Parallel coordinates plot showing the socio-demographic attributes of the female MPs

role for the classification of the users. They enable statements about the user composition on the LBSN and the representativeness of the data for explanations and predictions of social developments and interrelations.

C. Social group

An essential characteristic of LBSNs is the formation of relationships to other users leading to social groups (in the context of social media also referred to as community). The social group consists of users, who interact around a shared purpose, which is of central importance to the functioning of the social group [19]. Given this description, it is clear that social groups can be formed around an infinite number of shared purposes. However, they can be grouped into several categories:

- **private communities**: evolve around leisure activities, hobbies or other non-professional interests
- **professional communities**: formed around shared professional interests
- **commercial communities**: formed around products or companies

In the network structure, the social group is characterized by the fact that users of the group are more densely connected internally than with the rest of the network. Several methods have been developed which take advantage of this feature to detect social groups in LBSN.

D. Social role

In addition to the formation of social groups, the users develop specific roles within the group. The social role in LBSNs is characterized by the position within the social network structure. A rough classification can include:

- **Opinion leaders** ("influentials"): provide new ideas, or solutions to the group members
- **Disseminators**: act as a link between different social groups
- **Common users**: have no special role

The social roles are very important on how information and ideas flow through the social group. Especially opinion leaders play an important role as strategic points in transmitting information. To determine the social role of the users within the network, a number of centrality measures have been proposed (e.g. degree, closeness, betweenness, eigenvector). Table II summarizes the characteristics of these centrality measures.

<table>
<thead>
<tr>
<th>Centrality measure</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>Measures how much an actor is connected to other actors in the network. It is a measure of popularity, since a high value indicates nodes that can directly spread information to many other actors.</td>
</tr>
<tr>
<td>Closeness</td>
<td>Measures the average length of paths from an actor to all other actors in the network. Actors with small length path are considered more important, since they get information faster than those with high length path.</td>
</tr>
<tr>
<td>Betweenness</td>
<td>Measures the extent to which an actor lies on a path between other actors. Nodes with a high value represent important pathways of information between different actors or groups.</td>
</tr>
<tr>
<td>Eigenvector</td>
<td>Measure the related influence. An actor is more important if it is connected to important neighbors. The centrality value of a node corresponds to the sum of the values of its neighbors.</td>
</tr>
</tbody>
</table>
VI. REPRESENTATION OF THE SOCIAL FACET IN VISUAL ANALYTICS

This section briefly describes the representation of the social facet and its components in visual analytics. The goal of visual analytics is to gain insights from large and complex datasets. The approach combines the power of automated data analysis with the human ability to quickly capture patterns visually [20].

As an example, we used the members of the German Bundestag who are active on Twitter – in the following referred to as members of parliament (MPs). This group consists of 509 users (social actors), all of the type “individual”. We fetched the user profiles and the follows/followed by network from Twitter as well as additional socio-demographic attributes from the webpage of the German Bundestag.

Socio-demographic attributes that characterize the social actor are multivariate and can be very extensive. A parallel coordinates plot (PCP) arrays the multiple variables alongside one another and reveals how the social actors show similar or different profiles (Fig. 1). For example, the female MPs show the following pattern: Most of them are married or single, or without stating the marital status, and are mainly found in the left-wing parties (SPD, Bündnis 90/Die Grünen and Die Linke) No pattern can be seen with regard to the number of children and age.

The representation as parallel sets plot (PSP) is suitable for a quantitative visualization of the correlations between the socio-demographic attributes. Furthermore, it is possible to show the composition of groups with the same socio-demographic attributes. The PSP in Fig 2 shows two aspects: The relationship between gender, age and religion (left side) and the relationship between gender, marital status and number of children (right side). It is noticeable that the majority of female MPs aged between 40 and 79 (64%) do declare no religious affiliation. But the majority of younger women (57%) belong to the Christian religion - in the age group 20 – 29 even all female MPs. There is a more nuanced picture among male MPs: at the age between 40 – 59 the majority (56%) belong to the Christian religion and at the age between 20 - 39 and 60 - 79 the majority (58%) declares no religious affiliation. With regard to the marital status and the number of children, the following picture emerges: The largest proportion are married MPs (41%) and most of whom have children (81%). Whereas the majority of unmarried MPs (71%) has no children.

A large part of the field of social network visualization has been associated with graph drawings that are pictorial representations of the structural components of the network, i.e., its social actors and their relations. Therefore, graph drawings are best suited to visualize the components social group and social role. The key premise of a good spatial layout is that actors that are closely related should appear closer in the network visualization, while actors that are not related should appear farther apart. This basic notion of proximity helps analysts to identify social groups.
(community clusters) at a glance. Fig. 3 shows several interesting patterns. On the one hand, the MPs of the individual parties form strongly demarcated groups - recognizable by the color of the nodes, which represents party affiliation. On the other hand, MPs of different parties are more connected when they have politically similar views, as it is the case for the two coalition partners of the current government CDU/CSU and SPD. The differentiation of the parties to the AfD is also well recognizable from the spatial arrangement of the nodes.

Changing the size and color of the nodes are mainly used to visualize the social role of the social actor. The size of the nodes in Fig. 4 indicates the value of the betweenness centrality (Section V). An actor has a high betweenness centrality, if it is part of many shortest paths. They thereby form important bridges of the flow of information between different actors and groups and thus play the social role of a disseminator - e.g. Dr. Michael von Abercron (CDU) and Stefan Liebich (Die Linke).

An example for the visualization of spatial patterns is shown in Fig. 5. It shows the 27 male MPs of the party Bündnis 90/Die Grünen. The MPs are represented by circles and spatially located according to their constituency. The color and size of the circle are used here to visualize demographic characteristics. The color indicates the marital status and the size the number of children of the MP. The distribution can be interpreted as follows: Three quarters of the male members of Bündnis 90/Die Grünen have their constituency in the southwest of Germany. The majority of the MPs in this area who have declared the marital status are married (10), followed by single (4) and one MP in same-sex marriage. 80% of the married MPs have 2 or more children. However, only 2 unmarried MPs have their constituency in the eastern part of Germany. Regional and socio-demographic background play a role in political decisions. The knowledge of these backgrounds can thereby provide important explanations of the decisions of the MPs.

The examples described here are simple applications, but they illustrate the great potential of analysis of the social facet, providing new insights into LBSN data. More advanced visual analytics methods can be found in [21] and [22]. They describe novel methods and interactive graphics to analyze and visualize demographic variables.

Social network data encode a myriad of information and there are technological and cognitive limitations that restrict the amount of information a user is capable to process from looking at a static picture. To overcome these limitations and cope with the increasing size and complexity of the networks, current tools rely on interactivity to foster exploration via operations such as filtering, abstraction, and navigation.

VII. CONCLUSIONS

Despite the growing body of research addressing user-generated data in LBSNs, the social facet that characterizes the user (social actor) on the basis of various social attributes is currently not well understood in GIScience. The lack of a generally accepted classification of social components in the LBSN data underpins the argument.
However, the user is the central building block of LBSNs, since the spatial, temporal and thematic facets are defined or created by the social actor. Without considering the social facet, it is difficult to fully grasp and describe geographic processes and spatial relationships.

To address these shortcomings, we developed a classification of the social facet consisting of the components social actor, socio-demographic features, social group and social role. It builds on a literature review and an exploratory survey of different LBSNs. The classification proposed here is intended as a guideline to identify and extract relevant data with social attributes.

Visual analytics helps to gain insight from the data. Based on an example dataset different options for the visualization of social attributes were presented. Particularly suitable are parallel coordinates and parallel sets plots for representing socio-demographic features and graph drawings for visualizing the components social group and social role.

REFERENCES


