A Taxonomy for Uncertainty in Static Geo-Located Networks

Tatiana von Landesberger and Sebastian Bremm

TU Darmstadt

Abstract. Static graphs are analysed in various domains such as social science, finance, or communications. Often, these networks are geo-located, i.e., the nodes/edges have a geographic location. For example, in social networks, the connected people have a home location.
Experts strive to gain knowledge about the network structure and its geo-location. They often use visualization for this purpose.
Static geo-located graphs may include uncertainty. The uncertainty can affect both graph's nodes and edges. For example, uncertainty may be with regard to the existence of links between nodes, or the node locations.
For the visualization of geo-located graphs incorporating uncertainty, it is important to define the types of uncertainty that may appear in networks and to distinguish the various degrees of uncertainty. Although various typologies of uncertainty of geo-located data have been proposed, graph component is largely missing. Therefore, we extend current taxonomies with graph aspects.

Keywords: Uncertainty, Geo-Location, Graphs, Typology

1 Introduction

Geo-located graphs (or networks) are composed of nodes and edges, which have geographic location. For example, a geo-located social network is composed of persons connected through their friendships. It includes information on person's home location.

Understanding geo-located graphs (or networks) is a necessary task in various application domains. For example, social scientists wish to analyze friendships or team connections for gaining knowledge about social structures and their geographic composition.

The analyzed graph datasets are often subject to uncertainty. The uncertainty may affect both nodes and edges of the network as well as their geographic location. For example, it may be uncertain whether two persons are friends, and/or it may be unclear where a person exactly lives. The uncertainty may stem from inexact data, or may be a result of modeling algorithms.

Domain experts often analyse networks with the help of their graphical representation [2, 7]. It may reveal interesting data patterns, such as exact differences between two globally similar communities. Including uncertainty in this analysis is necessary [1]. In case of graphs, many uncertain relationships need to be assessed in detail in order to draw robust conclusions about the graph structure and the location distribution of entities in the network.

Visualization of static geo-located graphs including uncertainty requires a sound theoretical basis. It needs a typology of uncertainty in graphs for developing suitable visualization methods. Although there are comprehensive characterizations of uncertainty in the visualization area, e.g., [3–6], they cannot be directly applied to geo-located graphs with uncertainty. They lack several important characteristics such as the uncertainties in links, or the influence of edge existence on graph structure (see Section 2).

In this position paper, we propose an extended taxonomy of uncertainty specially for geo-located graph data (see Section 3). In this way, we propose an answer to the question "Which types of uncertainty exist in geo-located graphs?".

2 Related work

In visualization area, a comprehensive characterization of uncertainty types for various kinds of data such as general, geo-located or time-dependent has been presented [2–6]. In this paper, we focus on static geo-located network data.

The taxonomy of geo-located data uncertainty by Thomson [6] and extended by McEachren [4, 5] defines the following types of uncertainty: accuracy/error (difference between observation and reality), precision (exactness of data values), completeness (existence of missing data), consistency (agreement of interrelated data components), timeliness (certainty about currency of data), credibility (trustfulness of data sources), subjectivity (involvement of human interpretation or judgment in data construction) and interrelatedness (relationship of data sources/authenticity). Precision can be further distinguished into several subtypes, such as bounded uncertainty (e.g., distance between 20km and 25km), statistical uncertainty (e.g., the distance is normal distributed with mean 25km and variance of 3km), or abstract (distance of ca. 50km, or > 50km) [2].

Although these types can be applied also to graph data, they lack several important aspects, such as the interplay of link and node existence, or the influence of edge existence on graph structure.

3 Advanced Typology of Geo-Located Graph Uncertainty

We present an amended uncertainty characterization for static geo-located graphs.

We discuss three main aspects: uncertainty in nodes, uncertainty in edges and the interplay of these uncertainties in graphs. The uncertainty of geographic location is part of the three aspects, as we see geographic location a special attribute of a node or of an edge.

Node uncertainty: See Figure 1left.

- *Location*: Uncertainty about node's location. For example, uncertain person's location can be in form of an imprecise geographic location data such as the person located near Frankfurt. Alternatively, it can be in form of an information only on high level geographic of structures: The person is located in Germany, but we are unsure where exactly (in which town and in which street). Or, the information may be missing completely.
- *Existence*: Uncertainty whether the node exists. For example, missing persons in social networks. This can result from incomplete information, e.g., not all friends are stored in Facebook.
- *Ambiguity/Duplicity*: Uncertainty, whether two nodes are possibly one and the same node. For example, when extracting networks from articles, person duplicates can occur due to variance in name spelling.
- *Attributes' values*: Uncertainty about attribute values of a node. For example, imprecise person's age, or missing occupation.
- Grouping: Uncertainty about the belonging to a certain group (e.g., uncertainty about to which community a person belongs to).

Edge uncertainty: See Figure 1right.

- *Location*: Uncertainty about exact edge location. As an example, this can be found in transportation networks, where the exact train or person's route between two stations in not known at all, is known only to a certain degree (i.e., due to low geographic resolution), or only several location inner-points are known (i.e., several landmarks between two stations).
- *Existence*: Uncertainty about the existence of a connection. For example, there can be an uncertainty whether the road is already open to public, whether the train goes on Sunday, or whether two persons are friends.
- *Attributes' values*: Uncertainty about attribute values of an edge. For example, there can be uncertainty about the duration of a link between two stations.
- Ambiguity of edge's end nodes (a person has a friend in a city, but it is unclear who it is exactly).
- *Direction*: Is edge directed? In which direction? This can happen for a new street, when it is unclear whether it is a one way street.

Interplay of node and edge uncertainty within graphs:

- *Interdependence of node-edge uncertainty*: Node and edge uncertainties are interrelated.
  - *Existence:* A (non-)existence of a node implies (non-)existence or redirection of its adjacent edges (e.g., if it is unknown that a person is the network, its contacts are also uncertain).
  - *Location:* Uncertainty about the location of end node implies uncertainty about the location of edge's end.
  - *Inconsistency between nodes and edges:* The information in the edges and nodes may be inconsistent. For example, the end location of an edge may not match the information on the location of its end node.

  - *Grouping:* The uncertainty about node grouping implies uncertainty about edges between groups. For example, if a node may belong to two groups, also the edges connecting this node need to lead to both groups. This results in ambiguous group edges.
  - *Ambiguity:* Ambiguous nodes lead to ambiguous edges. If a node may have a duplicate, then the edges connecting the two duplicates need to be the same, these edges are then ambiguous.
- *Graph structure uncertainty*: Node and edge uncertainties have implications on graph structure and graph analysis (see Fig. 2).
  - *Existence:* (Non-)existence of an edge has implications on graph structure. For example, non-existence of a friendship may split a network into two separate sub-networks.
  - *Edge direction:* Edge direction determines existence of cycles and network motifs. Uncertainty about edge direction implies uncertainties about structural network properties.
  - *Edge attributes:* The uncertainty about the exact values of edge attributes may have implications on network paths. For example, uncertainty about traveling time between two geographic points may imply uncertainty about the shortest route between places on the map.
  - Edge ambiguity: Ambiguous edges may lead to ambiguous (shortest) paths in a graph, it may also imply possible existence of cycles or non-existence of unique paths, which is an important indication of graph structures such as trees or directed acyclic graphs.

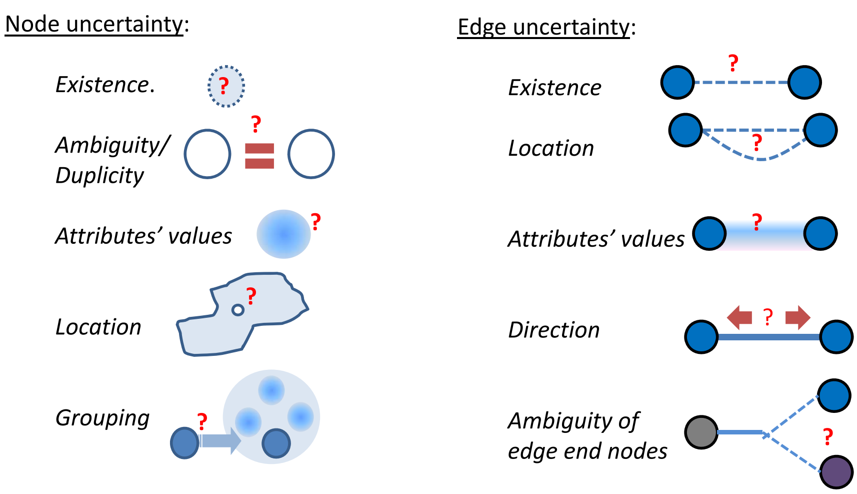

Fig. 1. Overview of node and edge uncertainties.

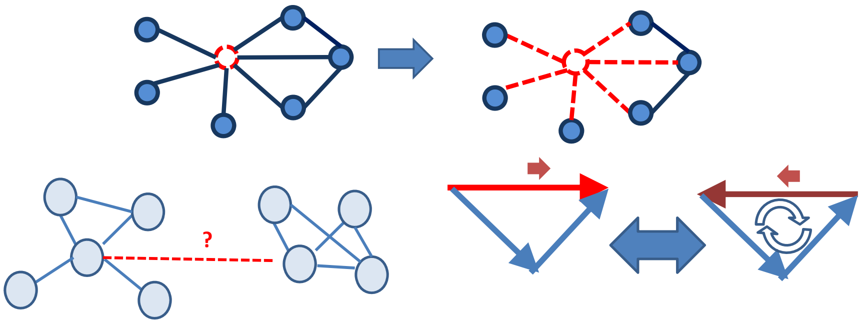

Fig. 2. Interplay of node and edge uncertainty within graphs. Top:Interdependence of node-edge existence. Bottom: Examples of the infuence of node and edge uncertainty on graph structure.

This typology relates to the taxonomy of Thomson [6]. For example, edge existence uncertainty may result from outdated information (e.g., old data vintage), imprecise geographic location value (home in Frankfurt area), or missing information (e.g., incomplete data source in social networks).

4 Conclusions and Future Work

We presented an extended typology of uncertainty in geo-located networks. This typology is important for classifying the types of uncertainty in the data. These types have impact on the visualization.

In the future, we would like to focus on the impact of geo-located graph uncertainties on the visualization and on dynamic geo-located graphs.

References

1. Boukhelifa, N., Duke, D.J.: Uncertainty visualization: why might it fail? In: CHI'09 Extended Abstracts on Human Factors in Computing Systems. pp. 4051–4056. ACM (2009)
2. Brodlie, K., Osorio, R.A., Lopes, A.: A review of uncertainty in data visualization. In: Expanding the Frontiers of Visual Analytics and Visualization, pp. 81–109. Springer (2012)
3. Gschwandtner, T., Gärtner, J., Aigner, W., Miksch, S.: A taxonomy of dirty time-oriented data. In: Multidisciplinary Research and Practice for Information Systems, pp. 58–72. Springer (2012)
4. MacEachren, A.M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., Hetzler, E.: Visualizing geospatial information uncertainty: What we know and what we need to know. Cartography and Geographic Information Science 32(3), 139–160 (2005)

5. MacEachren, A.M., Roth, R.E., O'Brien, J., Li, B., Swingley, D., Gahegan, M.: Visual semiotics & uncertainty visualization: An empirical study. Visualization and Computer Graphics, IEEE Transactions on 18(12), 2496–2505 (2012)
6. Thomson, J., Hetzler, E., MacEachren, A., Gahegan, M., Pavel, M.: A typology for visualizing uncertainty. In: Electronic Imaging 2005. pp. 146–157. International Society for Optics and Photonics (2005)
7. Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J.J., Fekete, J.D., Fellner, D.W.: Visual analysis of large graphs: State-of-the-art and future research challenges. In: Computer graphics forum. vol. 30, pp. 1719–1749. Wiley Online Library (2011)