Community detection and analysis with trajectory history of taxi drivers

Carles Sierra
Artificial Intelligence Research Institute, IIIA-CSIC, Bellaterra, Catalonia
pcchair@ijcai-17.org

Abstract

The *IJCAI–17 Proceedings* will be printed from electronic manuscripts submitted by the authors. The electronic manuscript will also be included in the online version of the proceedings. This paper provides the style instructions.

1 Introduction

It is natural that people find and form a community to share information. They get benefits from others, and they also contribute to the community. Active communities communicate with members often and achieve their goals. However, some others consist of people who have the similar preference but just flock together. This results from homophily, love of the same. In this research, we aim to quantify interaction in a community and to detect active communities.In taxi industry, drivers share information such as locations where they can easily pick up passengers without long queueing time. It means that there is community structure in taxi driver's network. However, it is difficult to find communities without drivers' contact list, and the list does not necessarily mean real communities.

The advanced information technology collects all data from anywhere and at any time. In the case of the taxi industry, vehicles' information such as position and speed is regularly recorded, which constitutes trajectory history. Whenever a driver picks up and drop off a passenger, servers store locations, fare, duration, etc.. This information makes us overcome the challenge mentioned before.

The massive trajectory data generated by moving objects such as people, animal and vehicles gives us opportunities to understand behavior of objects and to utility that knowledge. Zheng [2015] summarizes various trajectory data mining research and, especially, introduces several trajectory pattern mining research. They are aim to find a group of objects that shows move together or similar patterns. One of applications is recommendation systems. Some researches define the similarity between users with trajectory history [Xiao *et al.*, 2014; Liu and Wang, 2016]. Xiao et al. [2014] focuses on semantic location, not just physical locations, and manipulate an information retrieval method, TF-IDF (term frequency-inverse document frequency). Liu and Wang [2016] tries to get more accurate similarity score. Not only consider locations, but this research also include factors such as temporal duration, spatial proximity to other objects and movement velocity. The research related to recommendation systems can also find a group of people and groups. However our interest is active communities. A community is a group of people but a group does not necessarily means a community in terms of homophily and influence between members.

Some researches have more interests in influence between people. Shimizu [2008] propose a Viral Diffusion Model to capture information flows in a community. This research use a mailing list data and measures the impact of users' comment by comparing similarity in subsequent comments. Song et al. [2010] also focuses on communication between users. It quantifies relationship through Weighted Harmonic Rule which includes the intensity of the conversation. Arel et al. [2009] explicitly distinguishes peer influence and homophily effect. With sampling and statistics approach, this research quantifies peer influence in product adoption decisions.

Many community detection researches mainly focus on how to partition a given graph into subgraphs. Plantié and Crampes [2013] introduces many community detection algorithm and says that modularity has been used by many researchers to measure the performance of a community detection algorithm. This is because high values in modularity means there are many intra-community edges but few edges between different communities. Blondel et al. [2008] directly uses modularity with a heuristic approach, which iteratively merges communities if the mergence gives a better modularity value. Steinhaeuser and Chawla [2008] and Expert et al. [2011] also use modularity in algorithms, but they modifies it depending on problems they focuses. Chen et al. [**?**] models a community detection problem using game theory. In this research, agents can choose communities, a gain function is modularity and Nash equilibriums represents community structures.

In this research, we quantify drivers' influence to other drivers with a linear regression model, build weighted directed graphs and applies the well-known partition algorithm to detect communities. Furthermore, we compare some statistics of communities and validate our framework by another regression model that estimates locations where members in the same community share. A challenge in this work is that

we do not have prior information about drivers' contact list. Also, to find active communities, we need to exclude homophily effect in a systematic way. The remainder of this research is organized as follows. In Section 2, we introduce background about taxi industry, dataset and some assumptions. In Section 3, our framework is explained in detail. In Section 4, we show our results and validate them in another perspective. Lastly, we summarize our research and conclude with discussions.

1.1 Length of Papers

Each accepted full paper is allocated six pages in the conference proceedings, excluded references. References can take up to one page. Up to two additional pages may be purchased at a price to be announced per page for any accepted paper. However, all *submissions* must be a maximum of six pages, plus at most one for references, in length.

1.2 Word Processing Software

As detailed below, IJCAI has prepared and made available a set of LaTeX macros and a Microsoft Word template for use in formatting your paper. If you are using some other word processing software (such as WordPerfect, etc.), please follow the format instructions given below and ensure that your final paper looks as much like this sample as possible.

2 Style and Format

LaTeX and Word style files that implement these instructions can be retrieved electronically. (See Appendix A for instructions on how to obtain these files.)

2.1 Layout

Print manuscripts two columns to a page, in the manner in which these instructions are printed. The exact dimensions for pages are:

- left and right margins: .75″
- column width: 3.375″
- gap between columns: .25″
- top margin—first page: 1.375″
- top margin—other pages: .75″
- bottom margin: 1.25″
- column height—first page: 6.625″
- column height—other pages: 9″

All measurements assume an 8-1/2″ × 11″ page size. For A4-size paper, use the given top and left margins, column width, height, and gap, and modify the bottom and right margins as necessary.

2.2 Format of Electronic Manuscript

For the production of the electronic manuscript, you must use Adobe's *Portable Document Format* (PDF). A PDF file can be generated, for instance, on Unix systems using `ps2pdf` or on Windows systems using Adobe's Distiller. There is also a website with free software and conversion services: `http://www.ps2pdf.com/`. For reasons of uniformity, use of Adobe's *Times Roman* font is strongly suggested. In LaTeX2e, this is accomplished by putting

```
\usepackage{times}
```

in the preamble.[1]

Additionally, it is of utmost importance to specify the American **letter** format (corresponding to 8-1/2″ × 11″) when formatting the paper. When working with `dvips`, for instance, one should specify `-t letter`.

2.3 Title and Author Information

Center the title on the entire width of the page in a 14-point bold font. Below it, center the author name(s) in a 12-point bold font, and then center the address(es) in a 12-point regular font. Credit to a sponsoring agency can appear on the first page as a footnote.

Blind Review

In order to make blind reviewing possible, authors must omit their names and affiliations when submitting the paper for review. In place of names and affiliations, provide a list of content areas. When referring to one's own work, use the third person rather than the first person. For example, say, "Previously, Gottlob [1992] has shown that…", rather than, "In our previous work [Gottlob, 1992], we have shown that…" Try to avoid including any information in the body of the paper or references that would identify the authors or their institutions. Such information can be added to the final camera-ready version for publication.

2.4 Abstract

Place the abstract at the beginning of the first column 3″ from the top of the page, unless that does not leave enough room for the title and author information. Use a slightly smaller width than in the body of the paper. Head the abstract with "Abstract" centered above the body of the abstract in a 12-point bold font. The body of the abstract should be in the same font as the body of the paper.

The abstract should be a concise, one-paragraph summary describing the general thesis and conclusion of your paper. A reader should be able to learn the purpose of the paper and the reason for its importance from the abstract. The abstract should be no more than 200 words long.

2.5 Text

The main body of the text immediately follows the abstract. Use 10-point type in a clear, readable font with 1-point leading (10 on 11).

Indent when starting a new paragraph, except after major headings.

2.6 Headings and Sections

When necessary, headings should be used to separate major sections of your paper. (These instructions use many headings to demonstrate their appearance; your paper should have fewer headings.)

[1] You may want also to use the package `latexsym`, which defines all symbols known from the old LaTeX version.

Section Headings
Print section headings in 12-point bold type in the style shown in these instructions. Leave a blank space of approximately 10 points above and 4 points below section headings. Number sections with arabic numerals.

Subsection Headings
Print subsection headings in 11-point bold type. Leave a blank space of approximately 8 points above and 3 points below subsection headings. Number subsections with the section number and the subsection number (in arabic numerals) separated by a period.

Subsubsection Headings
Print subsubsection headings in 10-point bold type. Leave a blank space of approximately 6 points above subsubsection headings. Do not number subsubsections.

Special Sections
You may include an unnumbered acknowledgments section, including acknowledgments of help from colleagues, financial support, and permission to publish.

Any appendices directly follow the text and look like sections, except that they are numbered with capital letters instead of arabic numerals.

The references section is headed "References," printed in the same style as a section heading but without a number. A sample list of references is given at the end of these instructions. Use a consistent format for references, such as that provided by BibTEX. The reference list should not include unpublished work.

2.7 Citations

Citations within the text should include the author's last name and the year of publication, for example [Gottlob, 1992]. Append lowercase letters to the year in cases of ambiguity. Treat multiple authors as in the following examples: [Abelson *et al.*, 1985] or [Baumgartner *et al.*, 2001] (for more than two authors) and [Brachman and Schmolze, 1985] (for two authors). If the author portion of a citation is obvious, omit it, e.g., Nebel [2000]. Collapse multiple citations as follows: [Gottlob *et al.*, 2002; Levesque, 1984a].

2.8 Footnotes

Place footnotes at the bottom of the page in a 9-point font. Refer to them with superscript numbers.[2] Separate them from the text by a short line.[3] Avoid footnotes as much as possible; they interrupt the flow of the text.

3 Illustrations

Place all illustrations (figures, drawings, tables, and photographs) throughout the paper at the places where they are first discussed, rather than at the end of the paper. If placed at the bottom or top of a page, illustrations may run across both columns.

Illustrations must be rendered electronically or scanned and placed directly in your document. All illustrations should be in black and white, as color illustrations may cause problems. Line weights should be 1/2-point or thicker. Avoid screens and superimposing type on patterns as these effects may not reproduce well.

Number illustrations sequentially. Use references of the following form: Figure 1, Table 2, etc. Place illustration numbers and captions under illustrations. Leave a margin of 1/4-inch around the area covered by the illustration and caption. Use 9-point type for captions, labels, and other text in illustrations.

[2]This is how your footnotes should appear.

[3]Note the line separating these footnotes from the text.

Acknowledgments

The preparation of these instructions and the LaTeX and BibTeX files that implement them was supported by Schlumberger Palo Alto Research, AT&T Bell Laboratories, and Morgan Kaufmann Publishers. Preparation of the Microsoft Word file was supported by IJCAI. An early version of this document was created by Shirley Jowell and Peter F. Patel-Schneider. It was subsequently modified by Jennifer Ballentine and Thomas Dean, Bernhard Nebel, and Daniel Pagenstecher. These instructions are the same as the ones for IJCAI–05, prepared by Kurt Steinkraus, Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Lab.

A LaTeX and Word Style Files

The LaTeX and Word style files are available on the IJCAI–17 website, `http://www.ijcai-17.org/`. These style files implement the formatting instructions in this document.

The LaTeX files are `ijcai17.sty` and `ijcai17.tex`, and the BibTeX files are `named.bst` and `ijcai17.bib`. The LaTeX style file is for version 2e of LaTeX, and the BibTeX style file is for version 0.99c of BibTeX (*not* version 0.98i). The `ijcai17.sty` file is the same as the `ijcai07.sty` file used for IJCAI–07.

The Microsoft Word style file consists of a single file, `ijcai17.doc`. This template is the same as the one used for IJCAI–07.

These Microsoft Word and LaTeX files contain the source of the present document and may serve as a formatting sample.

Further information on using these styles for the preparation of papers for IJCAI–17 can be obtained by contacting `pcchair@ijcai-17.org`.

References

[Abelson *et al.*, 1985] Harold Abelson, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs*. MIT Press, Cambridge, Massachusetts, 1985.

[Aral *et al.*, 2009] Sinan Aral, Lev Muchnik, and Arun Sundararajan. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. In *Proceedings of the National Academy of Sciences*, pages 21544–21549, December 2009.

[Baumgartner *et al.*, 2001] Robert Baumgartner, Georg Gottlob, and Sergio Flesca. Visual information extraction with Lixto. In *Proceedings of the 27th International Conference*

on Very Large Databases, pages 119–128, Rome, Italy, September 2001. Morgan Kaufmann.

[Blondel *et al.*, 2008] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10008, 2008.

[Brachman and Schmolze, 1985] Ronald J. Brachman and James G. Schmolze. An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9(2):171–216, April–June 1985.

[Expert *et al.*, 2011] Paul Expert, Tim S. Evans, Vincent D. Blondel, and Renaud Lambiotte. Uncovering space-independent communities in spatial networks. In *Proceedings of the National Academy of Sciences*, pages 7663–7668, January 2011.

[Gottlob *et al.*, 2002] Georg Gottlob, Nicola Leone, and Francesco Scarcello. Hypertree decompositions and tractable queries. *Journal of Computer and System Sciences*, 64(3):579–627, May 2002.

[Gottlob, 1992] Georg Gottlob. Complexity results for nonmonotonic logics. *Journal of Logic and Computation*, 2(3):397–425, June 1992.

[Levesque, 1984a] Hector J. Levesque. Foundations of a functional approach to knowledge representation. *Artificial Intelligence*, 23(2):155–212, July 1984.

[Levesque, 1984b] Hector J. Levesque. A logic of implicit and explicit belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, pages 198–202, Austin, Texas, August 1984. American Association for Artificial Intelligence.

[Liu and Wang, 2016] Siyuan Liu and Shuhui Wang. Trajectory community discovery and recommendation by multi-source diffusion modeling. *IEEE Transactions on Knowledge and Data Engineering*, PP(99):1–1, 2016.

[Nebel, 2000] Bernhard Nebel. On the compilability and expressive power of propositional planning formalisms. *Journal of Artificial Intelligence Research*, 12:271–315, 2000.

[Plantié and Crampes, 2013] Michel Plantié and Michel Crampes. *Survey on Social Community Detection*. Springer, London, 2013.

[Shimizu, 2008] Hiroyuki Shimizu. Implicit relation analysis for mining social network of community. In *Proceedings of the IADIS International Conference on WWW/Interne*, Freiburg, Germany, October 2008.

[Song *et al.*, 2010] Meesun Song, Wonkyu Lee, and Junghwan Kim. Extraction and visualization of implicit social relations on social networking services. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, pages 1425–1430, Atlanta, Georgia, 2010. AAAI Press.

[Steinhaeuser and Chawla, 2008] Karsten Steinhaeuser and Nitesh V. Chawla. *Community Detection in a Large Real-World Social Network*. Springer, Boston, MA, 2008.

[Xiao *et al.*, 2014] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, 5(1):3–19, 2014.

[Zheng, 2015] Yu Zheng. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.*, 6(3):29:1–29:41, May 2015.