Cad2graph: Automated Extraction of Spatial Graphs from Architectural Drawings

Pratik Maitra[1], Masahiro Kiji[1], Talal Riaz[2], Philip M. Polgreen[1], Alberto M. Segre[1], Sriram V. Pemmaraju[1], and Bijaya Adhikari[1](✉)

[1] University of Iowa, Iowa City, USA
{pratik-maitra,masahiro-kiji,philip-polgreen,alberto-segre, sriram-pemmaraju,bijaya-adhikari}@uiowa.edu
[2] Yelp, San Francisco, USA

Abstract. A significant obstacle to spatial epidemiology in healthcare facilities is the absence of computationally amenable maps of the underlying space. Spatial data for built spaces are typically stored in computer aided design (CAD) architectural files which are difficult to parse, query, and combine with other data sources. To alleviate this difficulty, we design a tool, CAD2GRAPH, which automatically extracts spatial maps from CAD files. To ensure that the spatial map is easily amenable to computation, we represent it as a graph whose vertices represent spatial units of a uniform size and whose edges represent obstacle-free, walkable paths of uniform length connecting adjacent pairs of spatial units. CAD2GRAPH extracts key information such as walls, doors, and room labels from the CAD file and through a series of geometric transformations, extracts a spatial graph.

Keywords: spatial graphs · graph extraction · architectural drawings · epidemiology · healthcare associated infections

1 Introduction

Spatial epidemiology at the scale of healthcare facilities is critical for modelling and combating healthcare associated infections (HAIs). Some example include spatio-temporal clustering of *Clostridioides Difficile* infections (CDI) in hospitals [8], characterizing spatial distribution of healthcare professionals (HCPs) [4,5], optimizing microbial swabbing for disease surveillance [1], and non pharmaceutical interventions to combat CDI and *Methicillin-resistant Staphylococcus Aureus* (MRSA) [3,7]. A major obstacle in spatial epidemiology at the healthcare facility level is the lack of spatial maps of the architectural layout of the facilities. While many healthcare facilities have spatial data, it is often stored as computer aided design (CAD) files. It is non-trivial to analyze these together with other datasets often required for spatial analysis such as healthcare professionals mobility, patient transfers between rooms, and patient-room-doctors interactions [2,6]. On the other hand, if the data present in CAD files

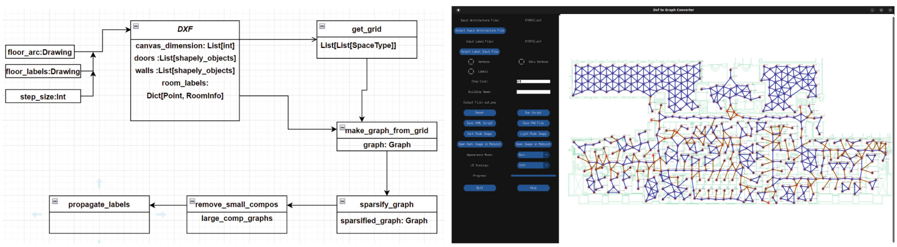
G. De Francisci Morales et al. (Eds.): ECML PKDD 2023, LNAI 14175, pp. 315–319, 2023.
https://doi.org/10.1007/978-3-031-43430-3_22

could be extracted as a spatial graph, it could easily be stored in the same database as other data and be analyzed together. In prior work [5,6], we have used hand-crafted spatial graphs. Generating hand-crafted spatial graphs for the entire University of Iowa Hospitals and Clinics took many months of work by 4-5 undergraduate students, 2-3 masters students, and 3 faculty members. This is a significant effort that not all healthcare facilities can afford.

To address the issues mentioned above, here we develop and demonstrate CAD2GRAPH, a novel tool to automatically generate a spatial graph representing the physical space within a hospital given an input CAD file. CAD2GRAPH carefully reads the outline of the architectural drawing and extracts spatial graph via a series of geometric transformations. Our target audience include data mining researchers who are applying their work towards the understanding and mitigation of HAIs and epidemiologists who are seeking to apply data mining techniques for clinical applications.

2 System Overview

The input to CAD2GRAPH is a CAD file representing a specific floor in a specific building. We first extract the external layout of the floor and structure of the walls and doors. We then construct a two dimensional grid with a pre-defined spacing and overlay the grid on the structure with walls and doors. We then assign a label to each grid node based on whether the given grid node is within a polygon of walls. We then repeat the same process and label the door nodes. We then add edges between the grid nodes in eight directions. Finally, we sparsify the grid and extract spatial graph. To this tool, we added a graphical user interface (in Python). The overview of the system and GUI are presented in Fig. 1.



(a) (b)

Fig. 1. (a) Overview of CAD2GRAPH. (b) The interface of the tool implemented in Python. The left panel consists of interactive elements and the right panel visualizes generated graph on top of the architectural layout.

The system presented here automatically extracts spatial graph $G_L(L, E, W, X)$ from a given CAD file. The graph is defined between the locations L within healthcare facilities including patient rooms, hallways, and so on. Each edge $e(l_1, l_2) \in E$ between two locations l_1 and l_2 indicates that they are in close proximity. The corresponding edge weight depends on whether l_1 and l_2 are within the same closed space or are connected via doors, stairs, and elevators. We provide a high-level summary of the steps involved in CAD2GRAPH next.

1. Canvas construction. We read the CAD file and extract the architectural layout and room labels, positions of walls and doors, and the dimension of the outer most walls. We then construct a 2-d canvass and assign (x, y) co-ordinates to each label read from the CAD file.

2. Grid extraction. We then construct an evenly spaced 2-d grid on the generated canvas. The number of rows and columns on the grid is determined by the size of the canvas and a user-specified parameter ρ. We then assign numeric labels to each point on the grid. Points on walls and doors are labelled 1 and 2 respectively. Others are labelled 0.

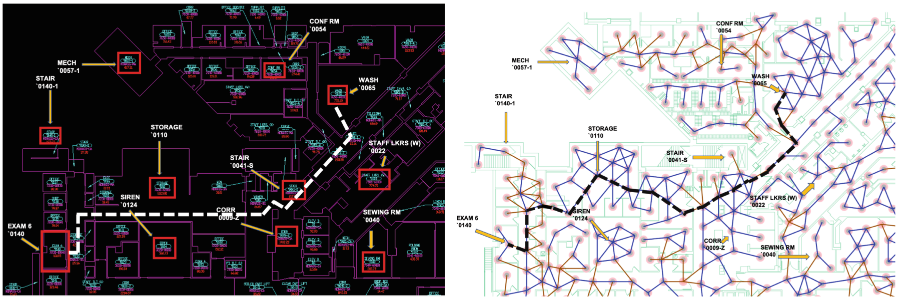
3. Graph extraction from the grid. The next step involves creating a spatial graph $G'(L', E', W', X')$ from the grid defined above. First we go over the labels extracted in step 1 and assign them as nodes L' (note: each room has a single label in the underlying CAD graph). We then add edges E' between the newly added nodes L'. Since the nodes were extracted from the grid, they too are organized in a 2-d space. We connect nodes in horizontal, vertical, and diagonal directions and assign weights depending on whether an edge crosses a door.

4. Graph sparsification. $G'(L', E', W', F')$ could be very dense for small values of ρ. This would imply that even a small room could have multiple nodes inside it, which is not ideal. Therefore, we sparsify $G'(L', E', W', F')$ to obtain a sparse spatial graph $G(L, E, W, F)$ using K-nearest neighbor search [9] and finally we remove small disconnected components. We then add edges between disjoint connected components while ensuring that the newly added edges are between the nodes which are geographically close. Note that only very few edges are added in the post processing step.

3 Demonstration

We run CAD2GRAPH on CAD files obtained from the University of Iowa Hospitals and Clinics (UIHC). Here we present a subsection of the visualization of a CAD file for a floor in the Roy Carver building[1] for demonstration. Figure 2 (a) visualizes the input CAD files. The red rectangles represent a subset of labelled rooms. Figure 2 (b) shows spatial graph extracted by CAD2GRAPH on top of the architectural layout. Here, we are only showing some of the labels in a subsection of the floor for legibility; notice that CAD2GRAPH is able to assign the labels to the correct nodes. As observed, the stairs, storage rooms, mechanical rooms, and staff's rooms are all assigned in the right place. Next, we observe that the cross door edges (in brown) and non-cross door edges (in blue) have been correctly identified: none of the blue edges cross any doors and all brown edges cross a door. Finally, we see a reasonable number of nodes within each open spaces,

[1] https://www.facilities.uiowa.edu/building/0359.



(a) (b)

Fig. 2. (a) Visualization of a subset of the CAD file showing one of the floors of the Roy Carver building in the University of Iowa Healthcare and Clinics.(b) Spatial graph extracted by CAD2GRAPH from the CAD file shown on the left.

only one node in small rooms, and the hallways are represented by single chain of blue edges. These observations are consistent with our design goal.

The dashed white line in Fig. 2 (a) shows obstacle free walkable path from the room EXAM 6 to the room WASH 0065. The dashed black line in Fig. 2 (b) is drawn over the edges along the shortest paths between the two rooms. As observed in the figure, the spatial graph extracted by CAD2GRAPH is actually able to infer edges which correspond to meaningful obstacle-free walkable paths between physical spaces. For additional validation, we first computed euclidean distances between all pairs of rooms in the same floor as above. We then computed shortest hop distance on extracted spatial graph between the same pairs of rooms. The Pearson's correlation between the two distances was 0.83, further validating that the spatial graphs extracted by CAD2GRAPH do capture the underlying architectural space well. A short demonstration video is available online[2].

4 Conclusion

In this paper, we presented CAD2GRAPH, an automated approach to extracting spatial graphs from CAD files. CAD2GRAPH carefully constructs a sparse graph from the architectural information in the input CAD file. We demonstrated a subsection of spatial graph generated from a CAD file obtained from University of Iowa Hospitals and Clinics. Additional demos along with out source code are publicly available. Our results show that the generated graphs are meaningful. These graphs can be stored in relational databases along with other datasets obtained from hospital operations and can be easily leveraged for spatial analysis of epidemics within healthcare facilitates.

[2] https://www.dropbox.com/s/9j6q1l5q11q2uuq/Pr_Final.mp4?dl=0.

References

1. Adhikari, B., Lewis, B., Vullikanti, A., Jiménez, J.M., Prakash, B.A.: Fast and near-optimal monitoring for healthcare acquired infection outbreaks. PLoS Comput. Biol. **15**(9), e1007284 (2019)
2. Cruz-Correia, R., et al.: Integration of hospital data using agent technologies-a case study. Artif. Intell. Commun. **18**(3), 191–200 (2005)
3. Curtis, D.E., Hlady, C.S., Kanade, G., Pemmaraju, S.V., Polgreen, P.M., Segre, A.M.: Healthcare worker contact networks and the prevention of hospital-acquired infections. PLoS ONE **8**(12), e79906 (2013)
4. Curtis, D.E., Hlady, C.S., Pemmaraju, S.V., Polgreen, P.M., Segre, A.M.: Modeling and estimating the spatial distribution of healthcare workers. In: Proceedings of the 1st ACM International Health Informatics Symposium, pp. 287–296 (2010)
5. Hasan, D.H., et al.: Modeling and evaluation of clustering patient care into bubbles. In: 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), pp. 73–82. IEEE (2021)
6. Jang, H., Pai, S., Adhikari, B., Pemmaraju, S.V.: Risk-aware temporal cascade reconstruction to detect asymptomatic cases: for the CDC mind healthcare network. In: 2021 IEEE International Conference on Data Mining (ICDM), pp. 240–249. IEEE (2021)
7. Monsalve, M.N., Pemmaraju, S.V., Thomas, G.W., Herman, T., Segre, A.M., Polgreen, P.M.: Do peer effects improve hand hygiene adherence among healthcare workers? Infect. Control Hosp. Epidemiol. **35**(10), 1277–1285 (2014)
8. Pai, S., Polgreen, P.M., Segre, A.M., Sewell, D.K., Pemmaraju, S.V., et al.: Spatiotemporal clustering of in-hospital clostridioides difficile infection. Infect. Control Hosp. Epidemiol. **41**(4), 418–424 (2020)
9. Peterson, L.E.: K-nearest neighbor. Scholarpedia **4**(2), 1883 (2009)