Artificial Intelligence in Medicine 65 (2015) 113–130

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Synthesis of a high resolution social contact network for Delhi with application to pandemic planning

Huadong Xia, Kalyani Nagaraj, Jiangzhuo Chen, Madhav V. Marathe*

Network Dynamics & Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, 1880 Pratt Drive, Blacksburg, VA 24060, United States

ARTICLE INFO

Keywords:
Networked-epidemiology
Synthetic population
Imputation techniques
Iterative proportional fitting
Labeled graphlets
Pandemic response
Targeted-layered containment

ABSTRACT

Objective: We aim to understand quantitatively how targeted-layered containment (TLC) strategies contain an influenza pandemic in a populous urban area such as Delhi, India using networked epidemiology.
Methods: A key contribution of our work is a methodology for the synthesis of a realistic individual-based social contact network for Delhi using a wide variety of open source and commercial data. New techniques were developed to infer daily activities for individuals using aggregate data published in transportation science literature in combination with human development surveys and targeted local surveys. The resulting social contact network is the first such network constructed for any urban region of India. This time varying, spatially explicit network has over 13 million people and more than 200 million people–people contacts. The network has several interesting similarities and differences when compared with similar networks of US cities. Additionally, we use a high performance agent-based modeling environment to study how an influenza-like illness would spread over Delhi. We also analyze well understood pharmaceutical and non-pharmaceutical containment strategies, or a combination thereof (also known as TLCs), to control a pandemic outbreak.
Results: (i) TLC strategies produce the mildest and most delayed epidemic out-break than any of the individual interventions; (ii) the epidemic dynamics of Delhi appear to be strongly influenced by the activity patterns and the demographic structure of its local residents; and (iii) a high resolution social contact network helps in analyzing effective public health policies.
Conclusion: A high resolution synthetic network is constructed based on surveyed data. It captures the underlying contact structure of a certain population and can be used to quantitatively analyze public health policy effectiveness. To the best of our knowledge, this study is the first of its kind in the Indian sub-continent.

1. Introduction

Today's densely populated urban regions facilitate a swift transmission of infectious airborne diseases [1]. Additionally, urban contact networks in regions like India and China are witnessing rapid growth. Delhi, officially recognized as the National Capital Territory (NCT) of India, is predicted to rise in population from 16.7 million in 2011 to 22.5 million in 2021 primarily due to a high rate of in-migration [2]. In Beijing, the population rose from 12.9 million in 2000 to 18.8 million in 2010 [3]. The ever increasing density of these urban regions can further increase the risk of an unmitigated pandemic. Consequently, public health authorities around the world have focused on developing effective policies to control the spread of diseases. The close coordination among the authorities, use of data driven computational models, and timely interventions have helped in controlling a number of recent outbreaks. Pharmaceutical as well as social distancing based interventions have proved themselves effective in this regard.

An epidemic diffuses through both dimensions of space and time [4]. Work on travel and mobility analysis [5–8] has prepared us for better observations in this regard. *Networked computational epidemiology* is the use of computer models to understand the spatio-temporal diffusion of disease through populations using a synthetic yet realistic representation of the underlying social contact network [9]. The basic approach is now widely accepted in the epidemiology community [10–12]. Researchers agree that a better understanding of social contact network characteristics can provide novel insights into the disease dynamics and intervention strategies for effective epidemic planning.

A methodology to synthesize realistic social contact networks already exists for United States cities. Contact networks for United States cities are generated by following a three-step process. (i)

* Corresponding author. Tel.: +1 540 808 3292; fax: +1 540 231 2891.
E-mail address: mmarathe@vbi.vt.edu (M.V. Marathe).

http://dx.doi.org/10.1016/j.artmed.2015.06.003
0933-3657/

A baseline population is synthesized based on sociodemographic statistics and microsample data from the United States Census. (ii) Mobility patterns from a nationwide household survey and land use data in the form of work, retail, recreational, school, and college locations are used to infer the spatio-temporal mobility patterns of individuals. (iii) User specified interaction criteria is used to estimate region-specific contact networks. The structure of the resulting social networks, calibrated to the above data, has been shown to influence the outcome of disease outbreaks in our simulated epidemic models [9,13].

Since the synthetic network should provide a realistic representation of the contact network specific to that region, the process to generate the contact network utilizes region-specific data. The United States synthetic population captures details of household structure by utilizing the 5% Public Use Microdata Sample for each Public Use Microdata Area that is modeled. The United States National Household Travel Survey (NHTS) [14] captures the interdependence of people's activities in the same household across all surveyed households in the United States. Data with a similar degree of detail is not available for many other regions (including Delhi, India), making it impossible to replicate the United States network generation process for regions outside the United States.

1.1. Summary of contributions

Building on our earlier work, we construct a synthetic social contact network for Delhi. To overcome data limitations for Delhi, we developed several new methods, many of which are generic enough to be easily applied to the synthesis of networks for urban regions in other developing countries. To the best of our knowledge, this is the first such synthetic representation of a contact network for an urban region in South Asia. Using a variety of data sources, a synthetic contact structure with detailed demographic information for each person, a minute-by-minute schedule of their activities, and the locations where these activities take place is generated by a combination of simulation and data fusion techniques. This yields a *dynamic social contact network* represented by a labeled bipartite graph G_{PL}, where P is the set of synthetic individuals and L is the set of locations. If a person $p \in P$ visits a location $\ell \in L$, there is an edge $(p, \ell, label)$ between them, where *label* here is a record of the type of activity of the visit and its start and end times. The synthetic social contact network is: (i) spatially explicit – home locations, work locations, business locations, educational institutions, government institutions, and other places of interest are explicitly represented; (ii) time varying – individuals carry out daily activities for a normative day by potentially visiting several locations, and in turn interacting with other individuals visiting the same locations during the same time period, and (iii) labeled – both individuals and locations carry a range of attributes described in the subsequent sections. Note that it is *impossible* to build such a network by simply collecting field data. The use of generative models to build such networks is a unique feature of this work.

We then use high-performance agent-based simulations to study the spread of an influenza-like illness over the synthetic social contact network of Delhi. We study the efficacy of various intervention strategies, including pharmaceutical and non-pharmaceutical interventions. We rank these strategies by order of their efficacy and discuss how the outcome of the simulated intervention experiments compares with those reported for other cities in the world. Finally, we carry out a detailed sensitivity analysis to assess the robustness of our conclusions.

1.2. Significance

The methodology and results presented in this paper extend our earlier work in a number of ways. First, we employ novel data sources and data integration methods. The methods we developed to synthesize urban scale social contact networks were based on data sources that were easily available. Many of these data sources are not easy to obtain for other countries. For example, the way we model people's activity sequences based on aggregated statistics is new. Furthermore, the basic method is generic and thus is applicable in other countries with similar limitations. Second, our work yields the first synthetic social contact network for Delhi. Very few if any such networks have been synthesized for urban areas in developing countries. Although the focus of the present paper is public health epidemiology, the social contact network synthesized can be used in a number of other applications; e.g. evacuation planning or urban transportation planning [15]. New methods are also presented to analyze massive social contact networks. For example, graphlets describing the demographic structure of people–people contacts are employed to analyze people's interaction patterns – recently these methods were used in the context of analyzing mobility patterns [16,17]. Finally, our results confirm that a targeted layer containment (TLC) strategy is effective in controlling an influenza epidemic in Delhi. TLC strategies have been proposed and extensively analyzed in [18]. As expected, vaccinations are effective, but their effectiveness depends on compliance and vaccine efficacy. Influenza vaccines continue to yield mixed results as has been discussed in epidemiology literature. Thus TLC as a combination of social distancing and pharmaceutical interventions can be seen as a natural and implementable alternative.

The remainder of the paper is organized as follows. Section 2 presents the new methodology to generate the detailed network and an overview of the data used for this methodology. In Section 3, we analyze structural properties of this network and compare effectiveness of different public health interventions in the H1N1 epidemic in Delhi using simulations. We assess the robustness of our epidemiological findings to synthesized social contact network. This is further discussed in Section 4.1. Section 4.2 contains a detailed discussion of graph structural properties of the resulting network and their comparisons with those of the coarse network. Section 4.3 summarizes our efforts in validating the synthesized networks. Finally, Section 5 concludes with remarks on future research.

1.3. Related work

Traditionally, mathematical and computational modeling of epidemics has focused on aggregate models using coupled rate equations [19]. In this approach, a population is divided into compartments according to an individual's health state (e.g., susceptible, exposed, infected, or recovered) and his/her demographic group. The evolution of the infectious disease is then characterized by ordinary differential equations. For analytical tractability, these models assume homogeneous mixing, which limits their use for spatially sensitive processes.

In recent years, high-resolution individual-based computational models have been developed to support the planning, control, and response to epidemics. These models support networked epidemiology, that is the study of epidemic processes over explicit social contact networks. Research in this area can be divided into three distinct subareas.

Work in the first subarea aims to develop analytical techniques and computer simulations over classes of progressively sophisticated random graphs [20,21]. These models relax the mean field assumption to some extent, but still use the inherent symmetries in random graphs to analytically compute important epidemic quantities of interest. The primary goal of these techniques is to obtain closed form analytical results.

The second subarea aims to develop individual-based models using important statistics associated with a given region. Two