Synthetic Populations for Epidemic Modeling

Abhijin Adiga, Richard Beckman, Keith Bisset, Jiangzhuo Chen, Youngyun Chungbaek, Stephen Eubank, Sandeep Gupta, Maleq Khan, Chris J. Kuhlman, Eric Lofgren, Bryan Lewis, Achla Marathe, Madhav V. Marathe, Henning S. Mortveit[1], Eric Nordberg, Caitlin Rivers, Paula Stretz, Samarth Swarup, Amanda Wilson and Dawen Xie — *Network Dynamics and Simulation Science Laboratory, Virginia Tech*

Infectious disease outbreaks can emerge without warning and spread rapidly through a population. They place a huge cost upon society, both directly through disease and mortality, and indirectly through impacts upon the food supply, the economy, and so on. The WHO estimates that, in lower income countries, almost one-third of all deaths are due to infectious diseases.

In recent years, we have seen several examples of epidemics that quickly spread across large populations, such as the ongoing Ebola outbreak, the swine flu epidemic of 2009, and the SARS epidemic of 2003, in addition to more localized (but still large) epidemics such as the Cholera outbreak in Haiti in 2010 and seasonal flu and malaria outbreaks in various parts of the world.

Epidemics typically spread through contact between infectious and susceptible persons. Therefore it is important to understand the social contact structure in a population in order the characterize or forecast the spread of an epidemic. Mathematical modeling of epidemics has contributed to epidemiology for a long time. These models typically use systems of ordinary differential equations to analyze the rate of progression of the disease through a population.

More recently, these models have been extended to the network setting, leading to insights about how network topology affects epidemic characteristics. At the same time, with increasing availability of data and computing power, sophisticated computational methods for epidemic modeling have emerged.

These large-scale simulation methods rely on approximating or synthesizing the social contact network for a large population. The social contact network is the network of physical, potentially disease-carrying, contacts between people in a region, such as through collocation at home, work, or school locations. In addition to the topology of the network, information about demographics and activity types, represented as node and edge labels, can be used to derive forecasts of which segments of the population are most likely to be affected by the epidemic.

These networks cannot be obtained through direct surveys for a couple of reasons. First, since the entire population needs to be represented, the scale of the problem is beyond the scope of typical surveys. Second, most people simply do not know all the people they come into contact with on a typical day because social contacts can occur in a broad range of activities.

Therefore, computational models rely on sophisticated techniques to synthesize a representation of the social contact network by integrating data from multiple sources [1]. The resulting networks are approximations of the actual social contact network.

We present a novel synthetic information resource to study the spread of infectious diseases in three West African countries – Guinea, Liberia and Sierra Leone. It is motivated by the current outbreak of Ebola in West Africa that is unprecedented in its scope, size and extent. The synthetic information is synthesized by integrating a number of open and commercially available data feeds using statistical and machine learning techniques.

This unique resource comprises of spatially explicit synthetic populations, daily activities and contact networks for each of the three countries. The resulting data sets are large and content rich. The synthetic resource can be used to develop detailed agent based models for the Ebola (and other infectious diseases) in these three countries.

A key requirement for synthesizing these data sets is timeliness. Our intent is to provide these data sets to the scientific community to support real-time computational epidemiology as it pertains to Ebola response. Towards this end, we have used data sets that are available right now. Further work is being done to improve these information resources as additional data becomes available. We have also placed these datasets on the web; see `http://www.vbi.vt.edu/ndssl/ebola`, free for anyone to download and use – this step has been

[1]* Corresponding author: Henning S. Mortveit; email: henning.mortveit@vt.edu; Telephone: +1 540 231-5327; Fax: +1 540 231 2891

taken keeping in mind unprecedented situation on the ground and with the aim of allowing modelers across the world to develop new approaches for Ebola response. We remark that their use of synthetic populations is not limited to epidemics, but to many systems with humans in the loop.

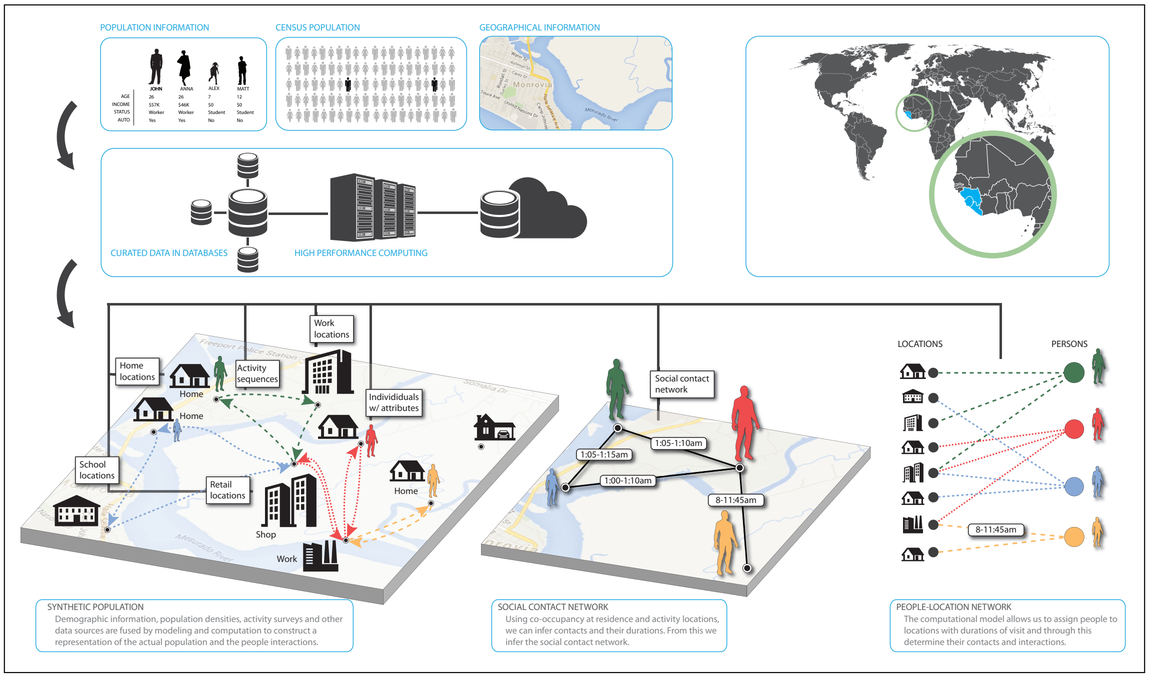

Figure 1: Elements involved in the construction of a synthetic population with contacts.

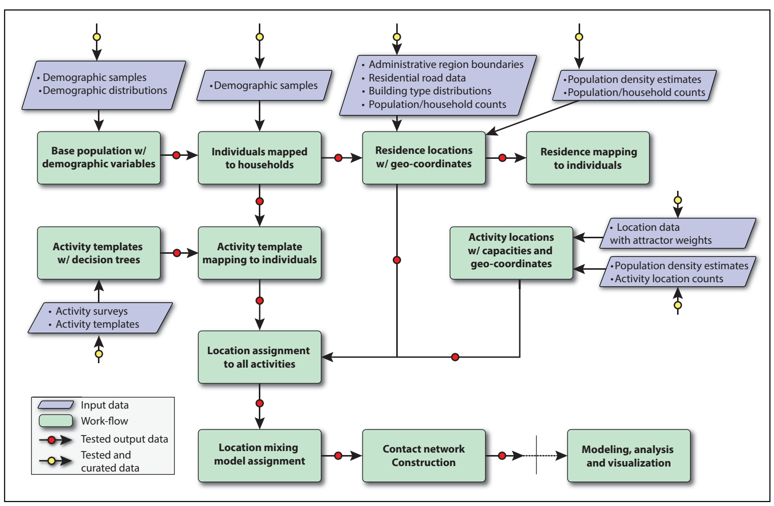

Figure 2: Overview of computational design for constructing a synthetic population with contacts.

References

[1] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, *Modelling disease outbreaks in realistic urban social networks*, Nature, 429 (2004), pp. 180–184.