CSE 8803 EPI: Data Science for Epidemiology, Fall 2022

Lecturer: B. Aditya Prakash
Scribe: Rachel Calder, Zun Wang

Sep 15, 2022
Lecture # : 6 and 7

Lecture 6

1 SI1I2S Model

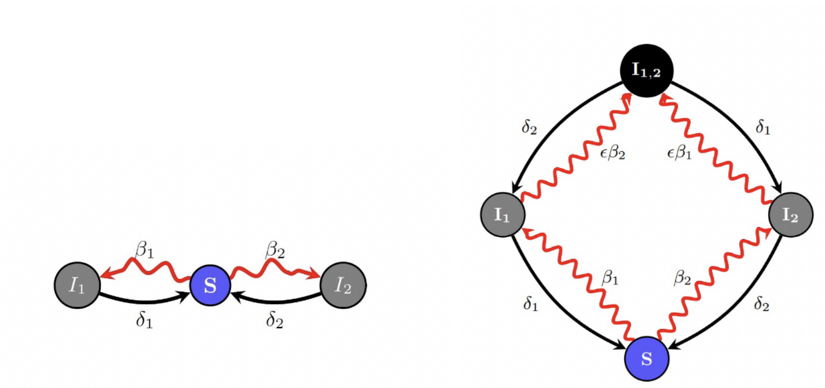

Figure 1: SI1I2S Model and S$I_{1|2}$IS Model

A SI1I2S model can be used to can be used to model competing infections for a single susceptible node where only one of the infections will be picked to infect the node, generating mutual immunity. A S$I_{1|2}$IS Model involves the possibility of various interaction factors, ϵ, for two viruses: Full mutual immunity: $\epsilon = 0$, partial mutual immunity (competition): $\epsilon < 0$, and Cooperation: $\epsilon > 0$.

Remember, for a single virus: $\lambda * \beta/\delta < 1$

What happens when two viruses are competing? This can be calculated by taking the footprint of the steady state of virus 1 and dividing it by the footprint of the steady state of the virus 2 over time. In this instance, over time, it can be observed that the winning or stronger virus will end up being the solely infecting the population and the weaker virus will die out. When there are two weak viruses, they will both die out.

Lecture 7

2 Lecture Summary

In this lecture, we discuss how to think of the SIR type models as networks to understand the spread of disease. We discuss the use of these networks for contact tracing as well as how networks can be built off of different data sources.

3 Common Network Models

In a network model, a circle represents a person, and square represents a location and a line represents a connecting edge. There are four common network models as seen in Fig 3. A person to person network involves unweighted connections. In a person to location network, people are interacting indirectly through a location. Locations can represent different zipcodes or even rooms within a building, like a hospital. A oneway population hcw can include directed edges. In the case of a travel city to city network, nodes can represent cities with weighted edges representing the volumn of travel between them.

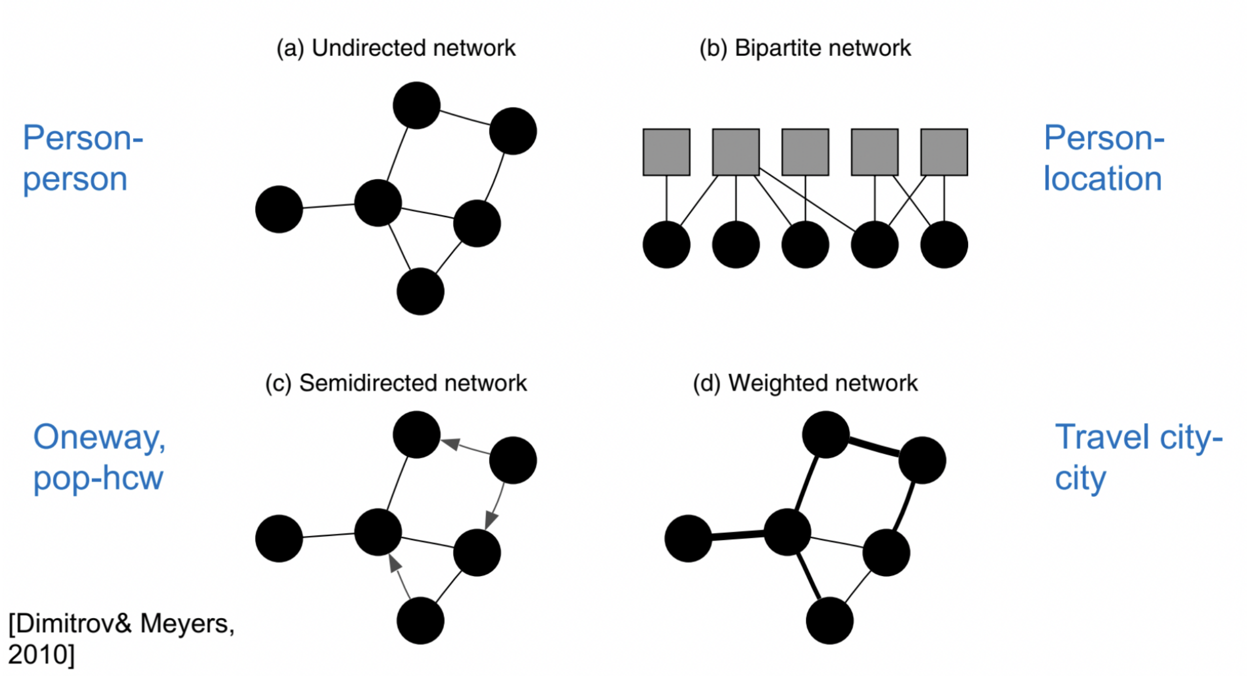

Figure 2: Common Network Models

4 Building a Network Model via Mobility Data

When building a network, one needs define the nodes involved in the model as well as what a contact edge would represent. For example, for respiratory diseases, and edge would represent close proximity. For sexually transmitted diseases, and edge would represent sexual contact, needle sharing, etc. A distribution of epidemiological contacts can be collected through mobility data[4]. Mobility can be acquired through a combination of surveys and trace data. Mobility data would ideally be easy to collect and provide substantial information into transmission. This would involve a useful frequency of collection. Each collection method has unique challenges to consider Fig 3.

Methods	Advantages	Disadvantages
Survey & direct	Multi purposed use; fewer biases; can capture multiple correlations	can be expensive to collect data observations;
Wi-Fi localization	Accuracy; Energy usage 50% GPS	Providing access point is expensive
GPS localization	High spatial precision: 5m; Can distinguish between transportation modes	High battery (energy) usage; expensive; sampling biases; No (low quality) signal in indoor environment
Cellular network localization (passive) (Call Data Records);	Automatically generated;	Sparse in time; Lower spatial resolution (175m); Needs more filtering; sampling biases; Proprietary
Cellular network localization (active)	More accuracy than passive localization; Less expensive than previous methods	More costly than passive form; sampling biases; Proprietary and thus not publicly available

Figure 3: Advantages and disadvantages of mobility data [3].

4.1 Considerations of Mobility Data

When collecting mobility data, one must also consider what could happen if the data is being used with malicious intent. Often times, extra precautions are made to mask the individuals personally identifiable information. Also, when scaling a model, one should assess that the model spans socio-economic categories.

4.2 Examples of Systems using Mobility Data

4.2.1 RFID tags and Localization

One great example is the use of RFID (Radio Frequency Identification) tags paired with localization data. Recall from previous lecture, the MIT Reality dataset investigated the capability of smartphones to track human interaction in a certain community, and in this case around the MIT Media Laboratory [2][5]. They discovered that the decision of identifying one another as friend is significantly correlated with spending time after work/at weekend, in other words same localization information in certain time period. They also claimed that there are periodicity in one person's behavior and the interaction between people can in a way be predicted.
Another example locates in a high school, where wireless sensor motes were distributed to students, faculty, and staff. Using the localization data, they build a social network with 762868 CPIs (close proximity interactions) at a maximal distance of 3 meters across 788 individuals. They did 100 simulations on each of the 788 individuals with SEIR model imposed over the network and found that the secondary infections and R0 are in agreement with school absenteeism data during the experiment period.

4.2.2 COVID-19 examples

There are numerous examples used for COVID-19:

- Maps and directions in Apple
- Location history in Google
- High resolution imagery in Facebook
- POI access in Safegraph
- Mobile phone data for Cubeiq
- etc ...

5 First Principle Approach for Constructing Social Contact Networks

When building a contact network based on how agents are acting in the model, the first principle approach emphasizes some core aspects of one individual's information that need to be addressed:

- **Who**: Demographics of individuals
- **What**: Sequence of their activities
- **When**: Time of their activities happened
- **Where**: Places/locations of their activities
- **Why**: the Reasons for their activities

These are aspects that can change along the time as disease spread or other interventions go on. Noticeably, the challenge here is that usually there is no one dataset that will give all aspects of data one is looking for when building in this kind of agent-based model. Instead, one need to synthesize multiple datasets and domain knowledge to cover all the aspects needed for the first principle approach. After successfully aggregate all the information, one can use the network to model behavioral changes, such as hypothesizing the absence of certain conditions to observe the changes accordingly.

5.1 Aggregation data from various sources

One application example is shown in Figure 4. This is a good visualization of synthesizing multiple data streams into a social contact network. The "Who" data is collected from the census data of different areas combined with social media. Then they figure out the synthetic population (ex: who are in the same household) and their "When", "What", and "Where" through different sources. Aggregating all these information build the synthetic social contact network.

Figure 4: Synthetic Social Contact Network

5.2 Example: COVID-19 in MA

Since behaviors change after interventions, in order to accurately model the behavioral changes of people, one can split the mobility in terms of layers. As shown in Figure 5 [1], this is a mobility study in the Massachusetts. They split the population into children and adult to visualize their fraction vs. location accordingly. They investigate their movement throughout the day and separate the data in terms of location layers: school layer, workplace and community layer, and household layer. During COVID-19, many schools switched to remote and the network should change accordingly by removing the school layer to accurately reflect the mobility.

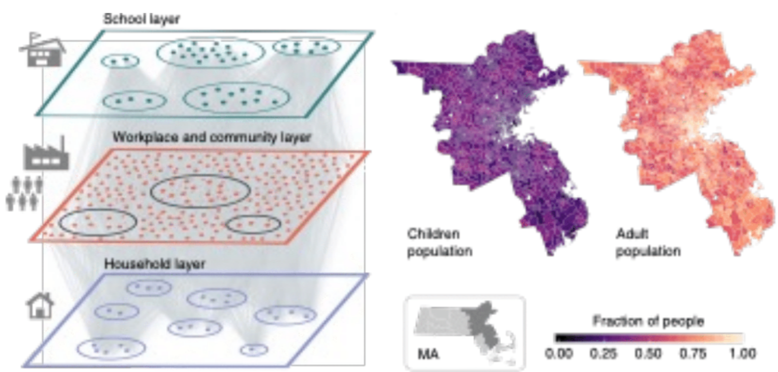

Figure 5: Layers of Mobility in MA

5.3 Multi-source data: Copenhagen Networks

Another example of a multi-source data aggregation is the Copenhagen Networks Study.They collected data from various sources uploaded by users and obtained from 3rd party servers. Users' data include WiFi access, Bluetooth scans, location estimates, etc. 3rd party servers like Facebook can provide friend list, likes, tags, etc, or like university administration can provide course grades. These various sources were aggregated into a single network and researchers can access the API to do investigation. They provided a temporal aggregation of the Bluetooth network as shown in Figure 6, showing how people are connected and how the structure changes along a small period of time, and the granularity here is very important for observation. The network is in use at GT for COVID-19 purpose, include phone-base proximity alerting and Infrastructure-based (WiFi) social interaction.

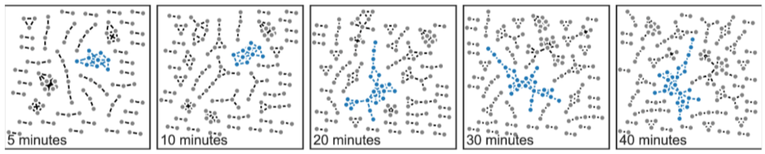

Figure 6: Copenhagen networks: temporal aggregation of Bluetooth network

5.4 WiFi as a coarse location sensor

Just as mentioned in the previous example, passive sensor stream can be extremely powerful when collecting data for analysis. Specifically, WiFi can act as a coarse location sensor, the regular authentication to the campus network from devices is the indication of location of the user. GT has over 7000 access points across 250 buildings, so these location information can be very accurate. For application in COVID-19, the WiFi information can provide details such as which students have close contact in a room for certain period to estimate risks.

5.5 WiMob application

Instead of route data that cover many places, WiFi Mobility data targets fewer spaces and can be more specific. Usual practice include remote classes/localized closures. One project done in Fall 2019 investigate the relationship between WiFi Mobility Network versus Enrollment. They visualize the enrollment and WiFi mobility in Figure 7. In the first week, there are plenty of enrollment and people come to classrooms, while in the 10th week there are fewer enrollment but fewer people come to classrooms. At the end of the semester, the network becomes very dense: enrollment and WiFi mobility paired very well because people come to take exams. In this case, since as the time goes, the enrollment does not change much but the attendance changes a lot during the semester, and in this case enrollment overestimates the efficacy of remote instruction.

5.6 Dynamic COVID Model

One project focuses on building a dynamic COVID SEIR model. They used the dynamic collocation network as the underlying contact network. By capturing the asymptomatic transmissions and isolating symptomatic individuals and considering the external infections from the surrounding neighborhood, they calculate the confirmed cases real time.

5.7 Multi-network and Co-evolution

In the real world, many changes in the behaviors can only be explained when incorporating multiple networks. The diffusion of behaviors can be affected by different interventions in neighbor-networks, such as intervention/policies on social information networks that leads to diffusion of public information and disease dynamics on social contact networks that

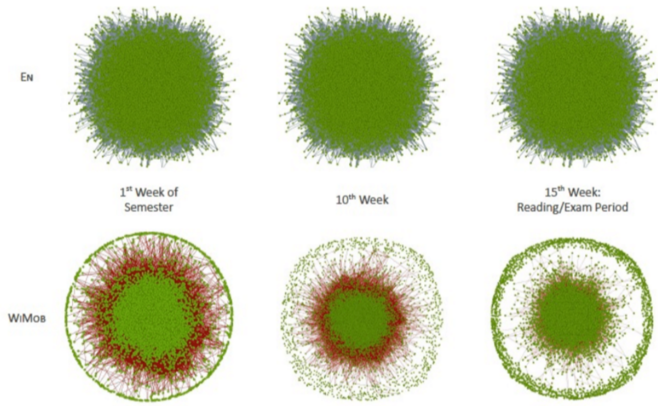

Figure 7: WiMob data vs. Enrollment

leads to epidemics changes. In this way, we can think of how misinformation can have huge impact on people behavior.

5.8 If data is plenty

Many of the times there are plenty of data, and some question cannot be answered just with collecting more data, such as on social.web cascades. But one can infer the underlying propagation network from set of observed cascades such as using Machine Learning methods or incorporating more generally surveillance information.

References

[1] Aleta. Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. In *Nature Human Behavior*, pages 964–971, 2020.

[2] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.

[3] M. Marathe and A. Vullikanti. Computational epiemiology. In *Computer Science*, 2014.

[4] T. J. Misa. Communities of computing: Computer science and society in the acm. In *Communities of Computing*, page 422, 2016.

[5] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.