CS:4980 Homework 3

Akash Choudhuri

April 2023

2 Late Days Used

1 Submodularity

1.1

To prove that the coverage function (denoted by f) is a monotone submodular function, we denote f by $f(S_i) = |f(S_1 \cup S_2 \cup ... \cup S_i)|$. If we look at all sets A $\subseteq$ B and for all elements i $\notin$ B, $f(A \cup \{i\}) - f(A) = [S_i/S_A]$ and $f(B \cup \{i\}) - f(B) = [S_i/S_B]$. We can easily see that $[S_i/S_B]$ will be lesser than $[S_i/S_A]$ due to the fact that $A \subseteq B$. So, in simpler terms, if B is a larger set than A, if we add a single element which was not in B to B, the fraction of increase in the cardinality of the resultant set will be much lesser than the fraction of increase in cardinality of the resultant set after adding the same single element to A.

To prove monotonicity, if we just take 2 sets S_i and S_j where $i \geq j$, we can see that $f(S_i) = |S_1 \cup S_2 \cup ...S_i| \geq f(S_j) = |S_1 \cup S_2 \cup ... \cup S_j|$. So, we can say that the coverage function is monotonous.

1.2

From Claim 1 in Lecture 14 in class, we can use the inequality $f(S \cup T) - f(S) \leq \sum_{e \in T}(f(S + e) - f(S))$. Now, as f is monotone, we can use the inequality that $f(S \cup T) \geq f(T)$. As we already have an upper bound on $f(S \cup T) - f(S)$, by using the monotone property of f, we can subsequently build $f(T) - f(S) \leq f(S \cup T) - f(S) \leq \sum_{e \in T}(f(S + e) - f(S))$. So, this proves the statement.

1.3

For h to be a submodular function, we need to prove that for any two sets X and Y such that X $\subseteq$ Y and any element e not in Y, $h(X \cup \{e\}) - h(X) \geq h(Y \cup \{e\}) - h(Y)$. Using the definition of h, we can say that $h(X \cup \{e\}) = min(f(X \cup \{e\}), f(V/(X \cup \{e\})))$ and $h(Y \cup e) = min(f(Y \cup \{e\}), f(V/(Y \cup \{e\})))$. Since $X \subseteq Y$, we have $Y \cup \{e\} \subseteq V/(X \cup \{e\})$ and as f is monotone, $f(Y \cup \{e\}) \leq f(V/(X \cup \{e\}))$ and $f(X \cup \{e\}) \leq f(Y \cup \{e\})$.

So, the above statement to prove simplifies to proving $min(f(X \cup \{e\}), f(V/(X \cup \{e\}))) - min(f(X), f(V\ X)) \geq min(f(Y \cup \{e\}), f(V/(Y \cup \{e\}))) - min(f(Y), f(V/Y))$. We can break it into 4 cases:

- Case 1: If $f(X) \leq f(Y)$ and $f(X \cup \{e\}) \leq f(Y \cup \{e\})$, we can use the previously shown result of $X \subseteq Y$ and extend it by saying $Y \cup \{e\} \subseteq (V/X)$. So, as f is monotone, we can get $f(Y \cup \{e\}) \leq f(V/X)$. Thus, the LHS of the above proof is $f(X \cup \{e\}) - f(X)$ and the RHS by following the claim we just made will be $f(Y \cup \{e\}) - f(Y)$. Now, we use the definition of submodularity of function f, we can assert that $f(X \cup \{e\}) - f(X) \geq f(Y \cup \{e\}) - f(Y)$.
- Case 2: If $f(X) \leq f(Y)$ and $f(X \cup \{e\}) > f(Y \cup \{e\})$, in which case the LHS will evaluate to 0 and the RHS will be $f(Y \cup \{e\}) - f(Y)$ and by just using the definition of submodular function, we will always have $0 \leq f(Y \cup \{e\}) - f(Y)$.
- Case 3: If $f(X) > f(Y)$ and $f(X \cup \{e\}) \leq f(Y \cup \{e\})$, in which case the RHS evaluates to 0 and the LHS will be $f(X \cup \{e\}) - f(X)$ and by using the definition of submodularity of f, we will always have $f(X \cup \{e\}) - f(X) \geq 0$.
- Case 4: If $f(X) > f(Y)$ and $f(X \cup \{e\}) > f(Y \cup \{e\})$, in which case the LHS of the proof becomes $f(V/(X \cup \{e\})) - f(X)$ and using the previously shown result of $X \subseteq Y$, we can use the monotone property of f to get $B \cup \{e\} \subseteq V/(X \cup \{e\})$ and thus evaluate the RHS of the proof to finally get the submodular inequality $f(Y \cup \{e\}) \leq f(V/(X \cup \{e\}))$.

This proves that h is submodular.

1.4

As we know that f is monotone submodular, for a pair of subsets A and B with $A \subseteq B$ and for a quantity e not in B, we will get $f(A + e) - f(A) \geq f(B + e) - f(B)$. As g is a concave function we will take the slope of the function g at intervals A, A+e, B and B+e, we will get $\frac{g(f(A+e)) - g(f(A))}{f(a+e) - f(A)} > \frac{g(f(B+e)) - g(f(B))}{f(B+e) - f(B)}$. Now, as the LHS denominator $\geq$ RHS denominator by the definition of submodularity of f, we can assert that the LHS numerator $\geq$ RHS numerator. So, with this, we get $g(f(A+e)) - g(f(A)) \geq g(f(B+e)) - g(f(B))$, which proves that g is submodular by the definition of submodular functions.

2 Sensors for detection is Network Model

2.1

The average fraction of of S,I vs t curve is given in Figure 1. The average number of daily infected people is given in Figure 2.

The average time when the infections is maximum is 32.

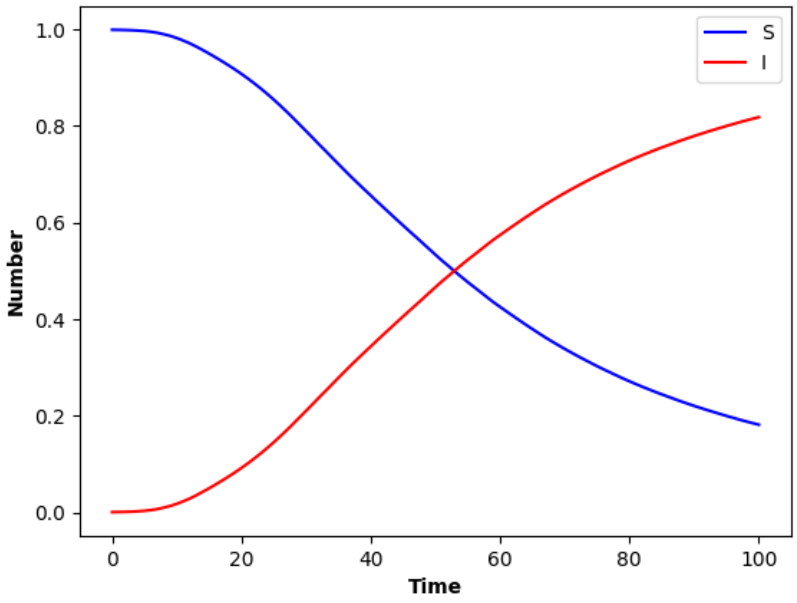

Figure 1: Variation of S,I vs t

2.2

The Average fraction of sensors that are infected with time plot for RANDOM is given in Figure 3 while for the same strategy, the average number of daily infected people is given in Figure 4. The average time when the infections is maximum is 33.

The Average fraction of sensors that are infected with time plot for FRIENDS is given in Figure 5 while for the same strategy, the average number of daily infected people is given in Figure 6. The average time when the infections is maximum is 28.

The Average fraction of sensors that are infected with time plot for CENTRAL is given in Figure 7 while for the same strategy, the average number of daily infected people is given in Figure 8. The average time when the infections is maximum is 9.

2.3

The peak time for RANDOM is 33. The peak time for FRIENDS is 28. The peak time for CENTRAL is 9.

The lead time for RANDOM is 0. The lead time for FRIENDS is 4. The lead time for CENTRAL is 23.

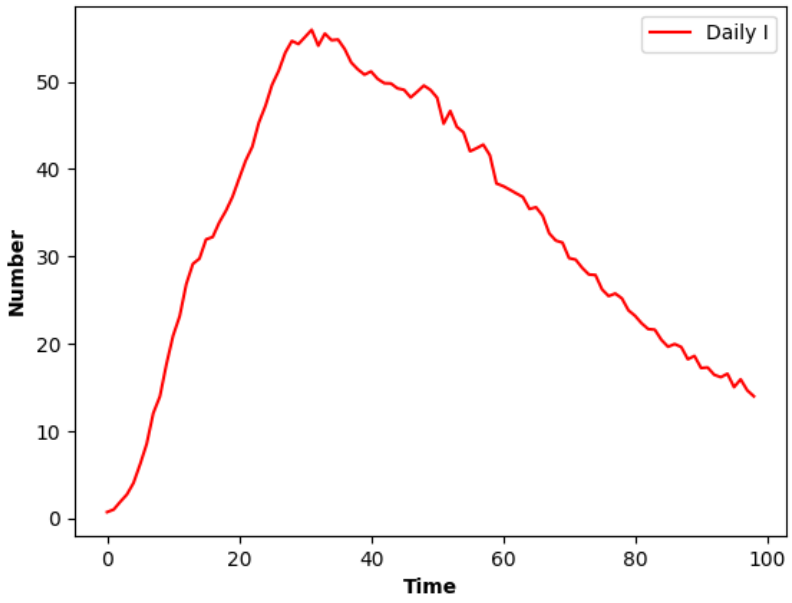

Figure 2: Variation of daily infected people

2.4

The lead time for RANDOM is the highest while the lead time for CENTRAL is lowest. This is because choosing random sensors does not help in detecting the peak in the number of infected people in the population. FRIENDS does a little better because, by choosing a random friend of a random person, we will be effectively be trying to approximate some of the most important people (high degree). On the other hand, with CENTRAL, we are choosing the nodes with highest degrees as sensors, which will act as great sensors about understanding when the infections will peak.

2.5

For k=50 The Average fraction of sensors that are infected with time plot for RANDOM is given in Figure 9 while for the same strategy, the average number of daily infected people is given in Figure 10. The average time when the infections is maximum is 28.

The Average fraction of sensors that are infected with time plot for FRIENDS is given in Figure 11 while for the same strategy, the average number of daily infected people is given in Figure 12. The average time when the infections is maximum is 30.

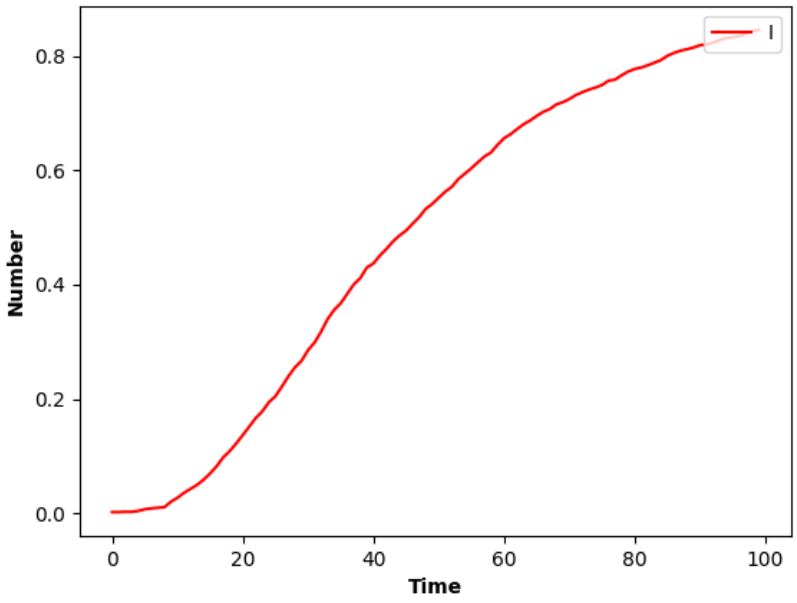

Figure 3: Variation of I vs t for RANDOM

The Average fraction of sensors that are infected with time plot for CENTRAL is given in Figure 13 while for the same strategy, the average number of daily infected people is given in Figure 14. The average time when the infections is maximum is 9.

For k=500 The Average fraction of sensors that are infected with time plot for RANDOM is given in Figure 15 while for the same strategy, the average number of daily infected people is given in Figure 16. The average time when the infections is maximum is 39.

The Average fraction of sensors that are infected with time plot for FRIENDS is given in Figure 17 while for the same strategy, the average number of daily infected people is given in Figure 18. The average time when the infections is maximum is 26.

The Average fraction of sensors that are infected with time plot for CENTRAL is given in Figure 19 while for the same strategy, the average number of daily infected people is given in Figure 20. The average time when the infections is maximum is 24.

We see that for k=50, lead time for RANDOM increases by 5 while lead time for FRIENDS decreases by 2 and the lead time for CENTRAL remains the same. On the other hand, for k=500, lead time for RANDOM decreases by 6 while lead time for FRIENDS increases by 2 and the lead time for CENTRAL

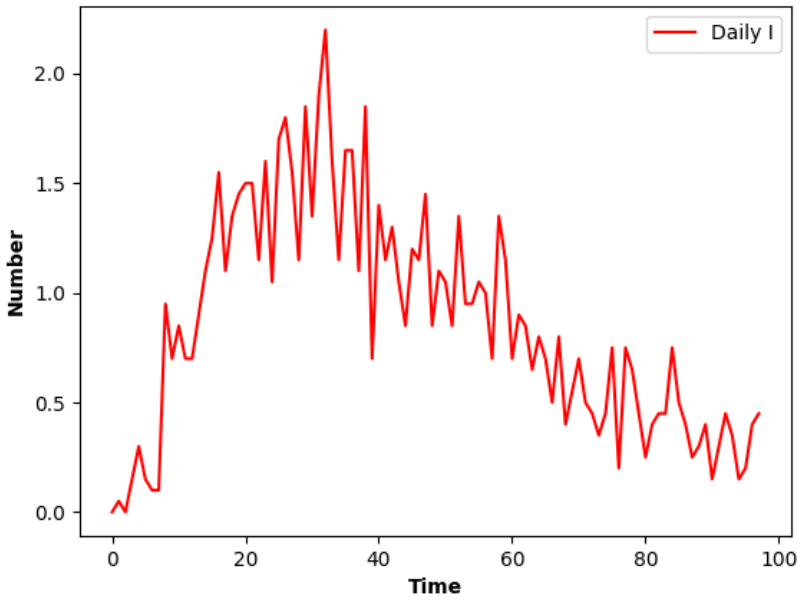

Figure 4: Variation of daily infected people for RANDOM

decreases by 15.

3 Flu Surveillance using Google Symptoms Data

3.1

The Weekly Unweighted ILI of Iowa over 2018-19 flu season is given in Figure 21.

3.2

The single plot showing the trends of all the symptoms over 2018-19 seasons with x-axis showing the Epiweeks and y-axis the symptom trend values is shown in Figure 22.

3.3

The PCC values for Iowa was in Table 1

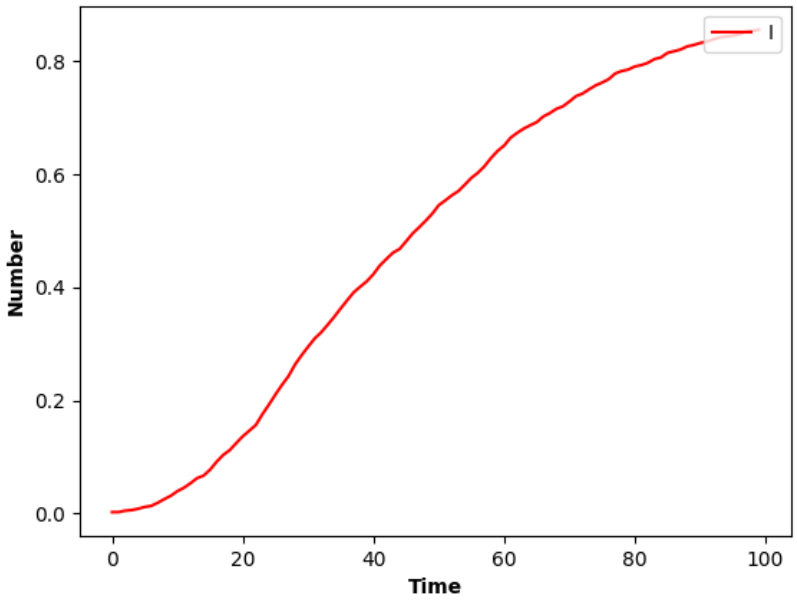

Figure 5: Variation of I vs t for FRIENDS

3.4

The PCC values for California was in Table 2

The PCC values for Texas was in Table 3

The PCC values for New York was in Table 4

The PCC values for Alaska was in Table 5

The PCC values for Georgia was in Table 6

The PCC values for Mississippi was in Table 7.

The primary reason for the variation in PCCs across the states is due to the fact that there are different strain of the virus that is dominant in different parts of USA. Also, the difference in the weather of the states also contribute to the diversity of spread of particular strains of the disease in particular states.

3.5

Based on the PCC value of the most correlated symptom, the Lead Time values are given in Table 8.

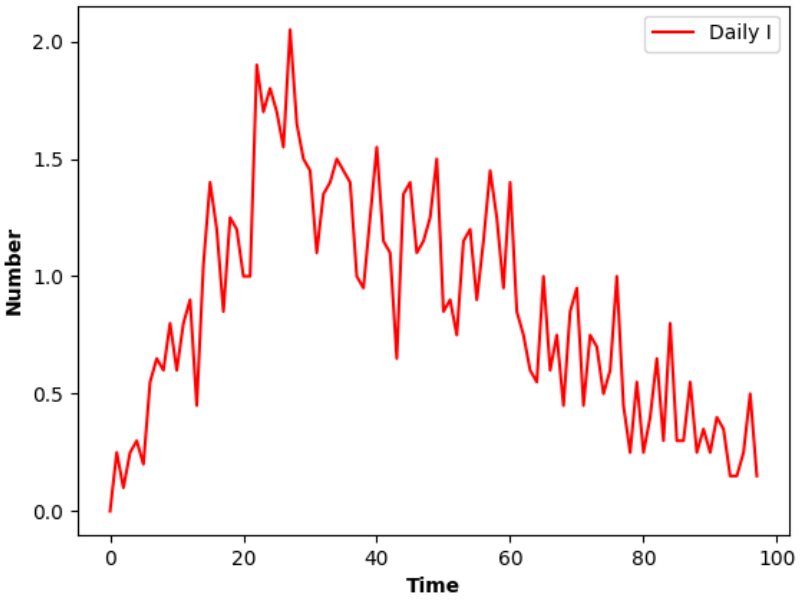

Figure 6: Variation of daily infected people for FRIENDS

Variable	PCC
Fever	0.9636
Low-grade Fever	0.9153
Cough	0.7674
Sore Throat	0.7050
Headache	0.7717
Fatigue	0.5114
Muscle Weakness	0.3002

Table 1: PCC values of different symptoms with ILI for Iowa

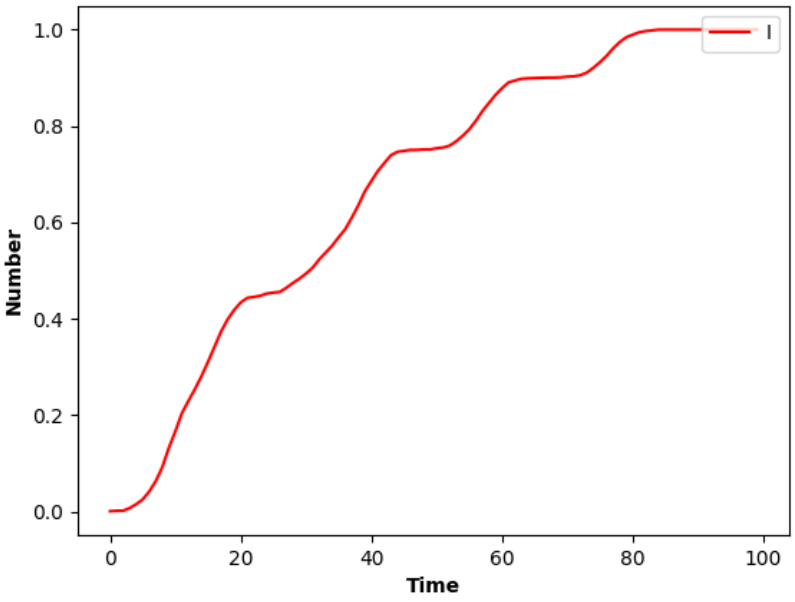

Figure 7: Variation of I vs t for CENTRAL

Variable	PCC
Fever	0.8371
Low-grade Fever	0.8590
Cough	0.9527
Sore Throat	0.8745
Headache	0.5796
Fatigue	0.0011
Muscle Weakness	-0.0585

Table 2: PCC values of different symptoms with ILI for California

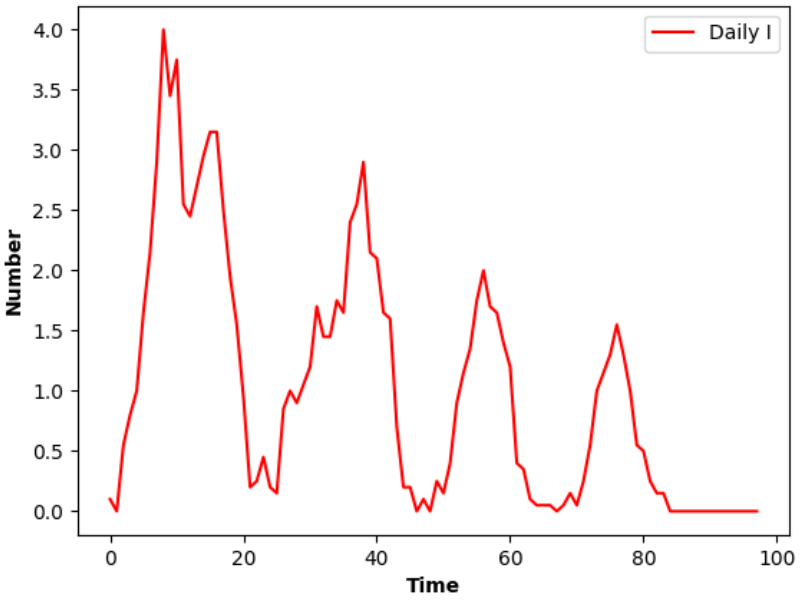

Figure 8: Variation of daily infected people for CENTRAL

Variable	PCC
Fever	0.9366
Low-grade Fever	0.9662
Cough	0.7394
Sore Throat	0.7292
Headache	0.6168
Fatigue	0.4687
Muscle Weakness	0.0188

Table 3: PCC values of different symptoms with ILI for Texas

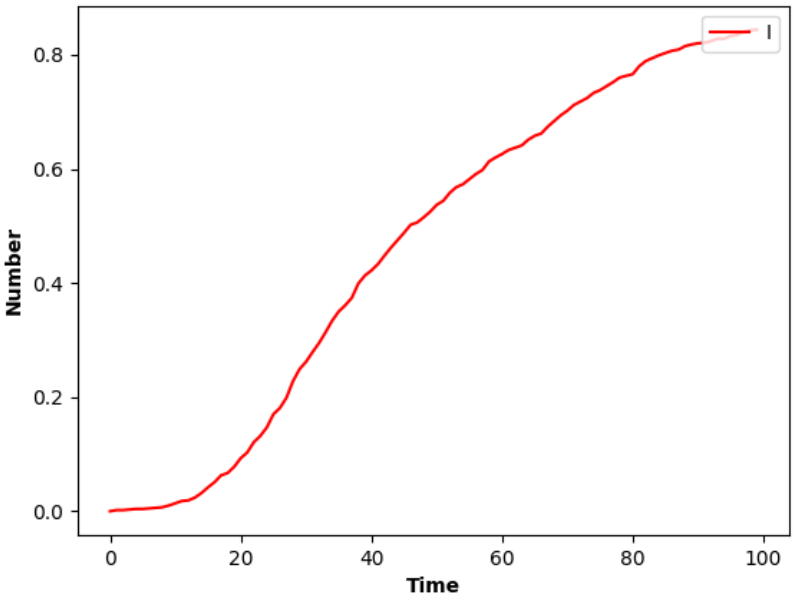

Figure 9: Variation of I vs t for RANDOM k=50

Variable	PCC
Fever	0.9401
Low-grade Fever	0.9630
Cough	0.6631
Sore Throat	0.6768
Headache	0.6631
Fatigue	0.4572
Muscle Weakness	0.1206

Table 4: PCC values of different symptoms with ILI for New York

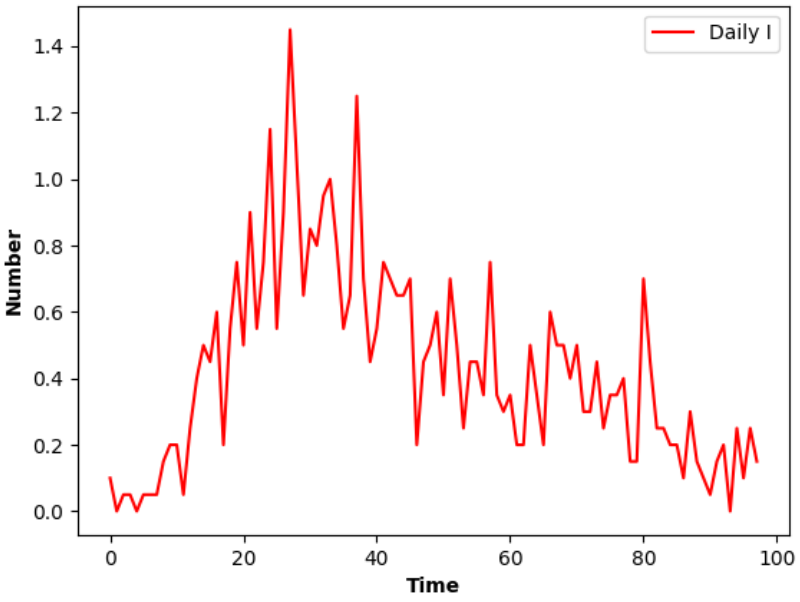

Figure 10: Variation of daily infected people for RANDOM k=50

Variable	PCC
Fever	0.9629
Low-grade Fever	0.6763
Cough	0.9639
Sore Throat	0.8417
Headache	0.5681
Fatigue	0.5255
Muscle Weakness	0.5019

Table 5: PCC values of different symptoms with ILI for Alaska

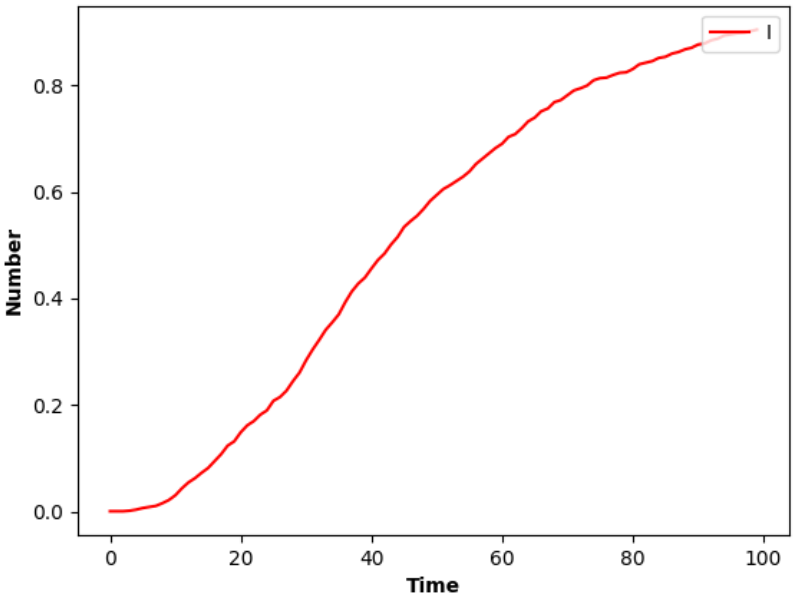

Figure 11: Variation of I vs t for FRIENDS k=50

Variable	PCC
Fever	0.9385
Low-grade Fever	0.8291
Cough	0.9512
Sore Throat	0.8895
Headache	0.3091
Fatigue	-0.2839
Muscle Weakness	-0.5721

Table 6: PCC values of different symptoms with ILI for Georgia

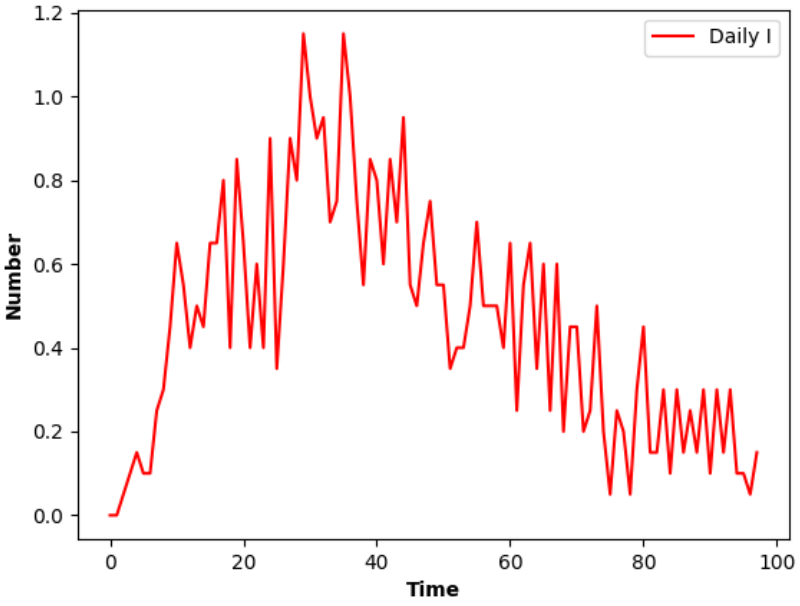

Figure 12: Variation of daily infected people for FRIENDS k=50

Variable	PCC
Fever	0.9080
Low-grade Fever	0.8506
Cough	0.8387
Sore Throat	0.8324
Headache	0.5214
Fatigue	0.4989
Muscle Weakness	0.0509

Table 7: PCC values of different symptoms with ILI for Mississippi

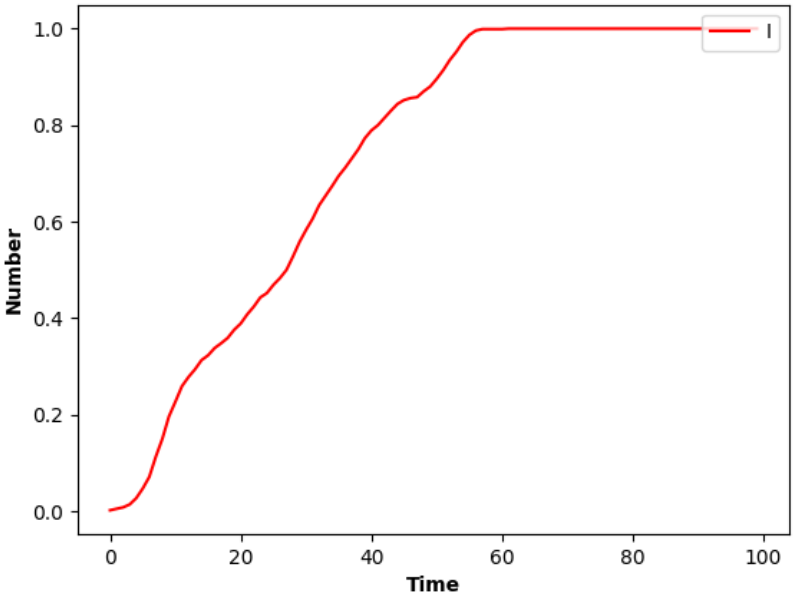

Figure 13: Variation of I vs t for CENTRAL k=50

State	Symptom	Lead Time
Iowa	Fever	1
California	Cough	1
Texas	Low Grade Fever	18
New York	Low Grade Fever	32
Alaska	Cough	1
Georgia	Cough Fever	1
Mississippi	Fever	32

Table 8: Lead Times of different States

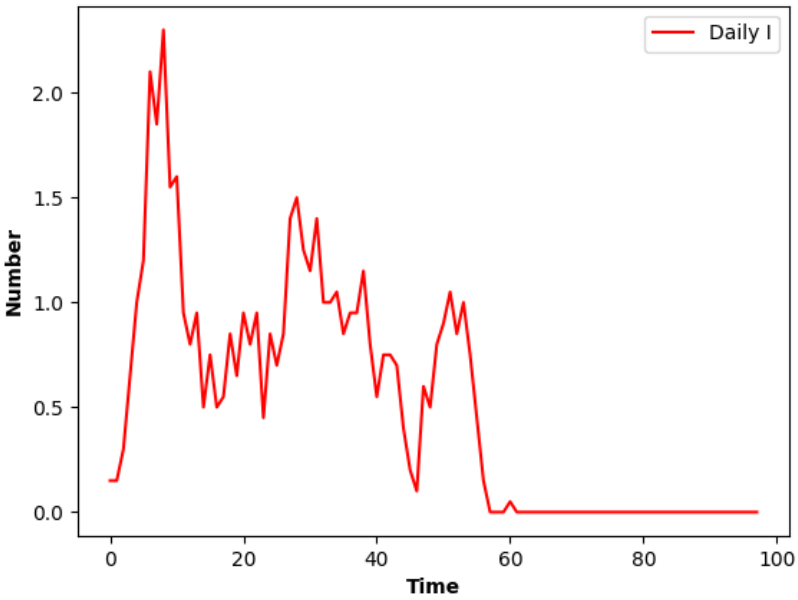

Figure 14: Variation of daily infected people for CENTRAL k=50

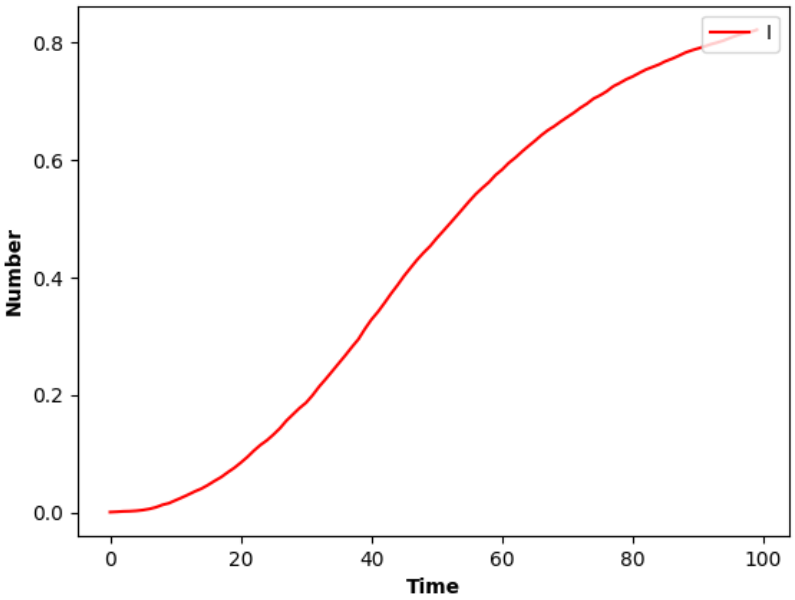

Figure 15: Variation of I vs t for RANDOM k=500

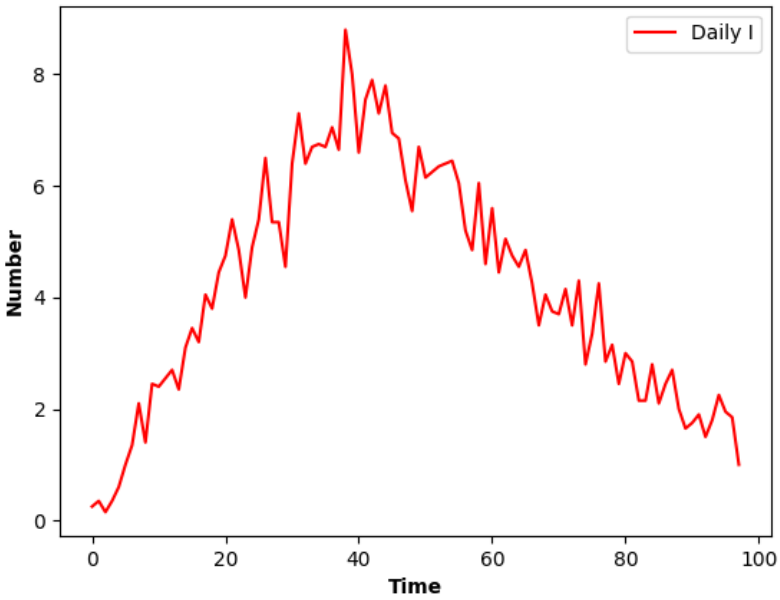

Figure 16: Variation of daily infected people for RANDOM k=500

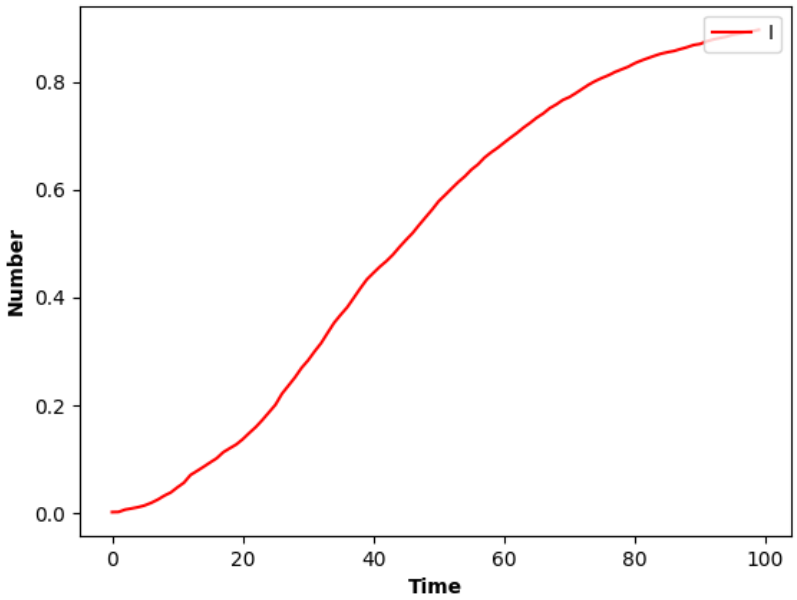

Figure 17: Variation of I vs t for FRIENDS k=500

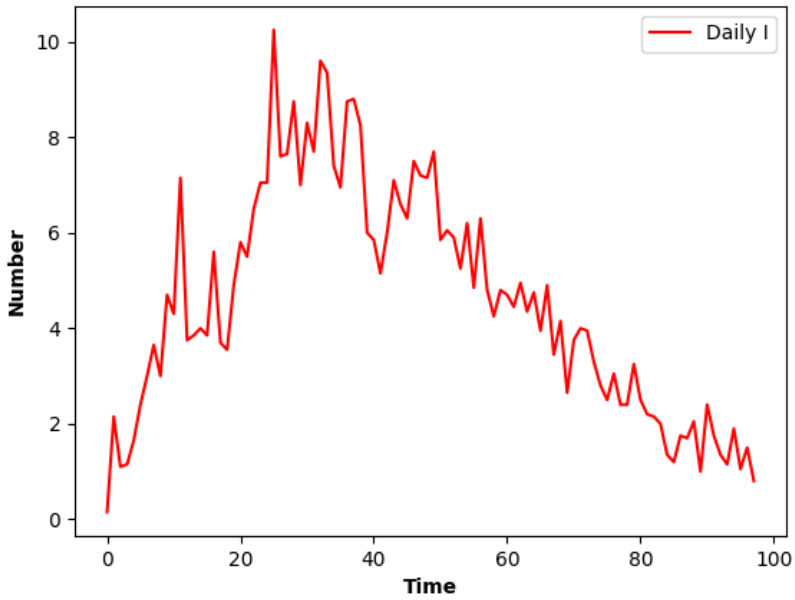

Figure 18: Variation of daily infected people for FRIENDS k=500

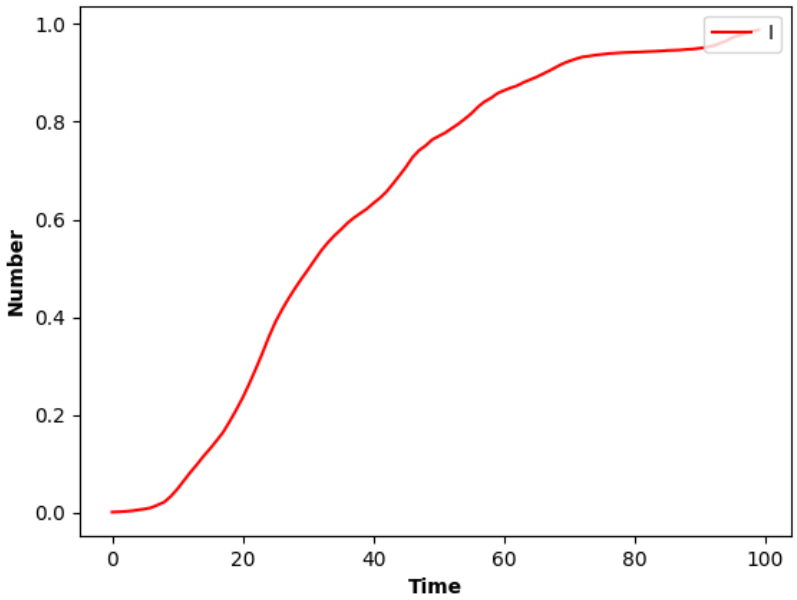

Figure 19: Variation of I vs t for CENTRAL k=500

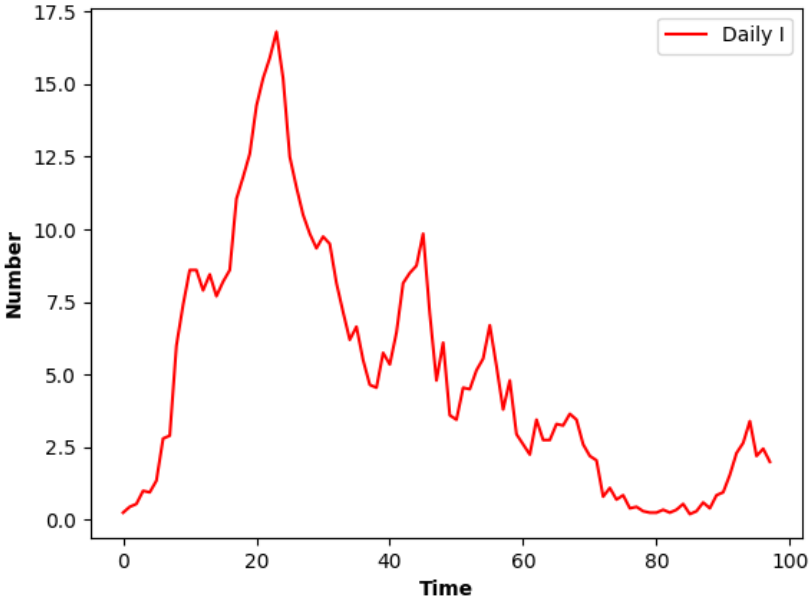

Figure 20: Variation of daily infected people for CENTRAL k=500

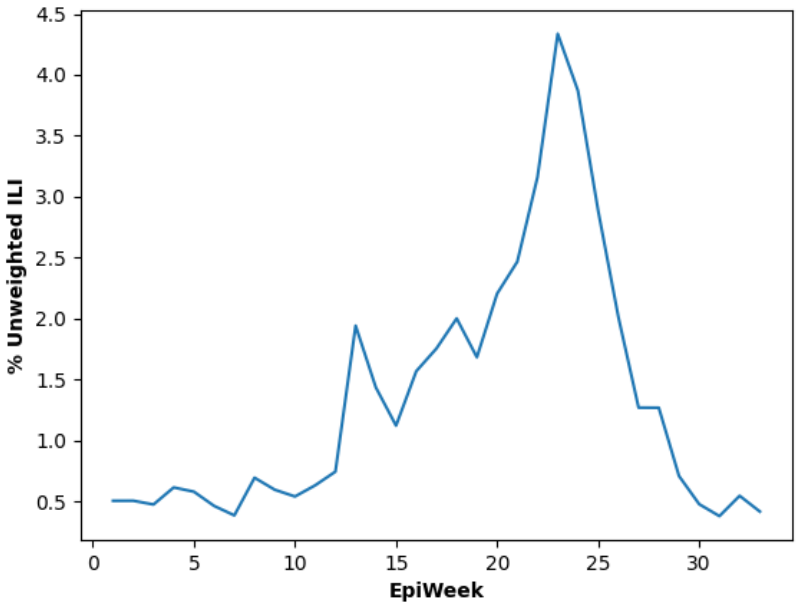

Figure 21: Unweighted ILI for Iowa for 2018-19 Flu Season

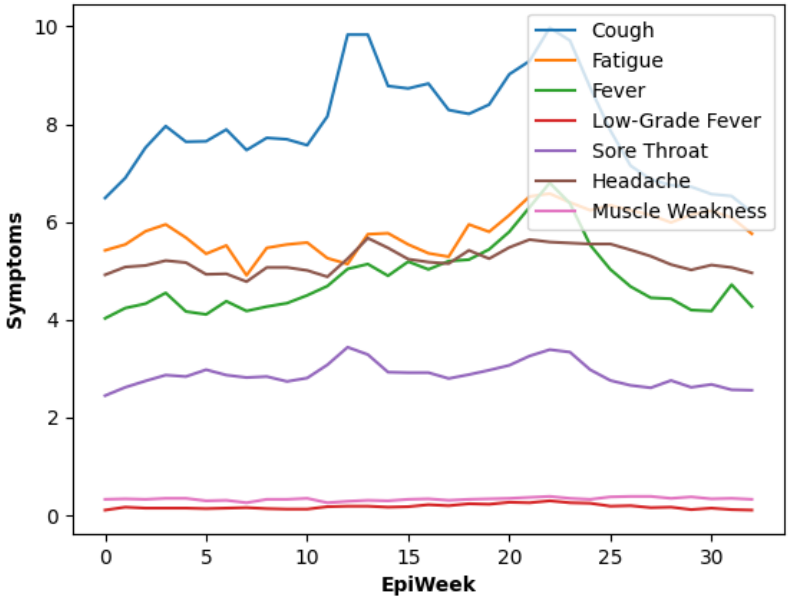

Figure 22: Symptoms plot for Iowa

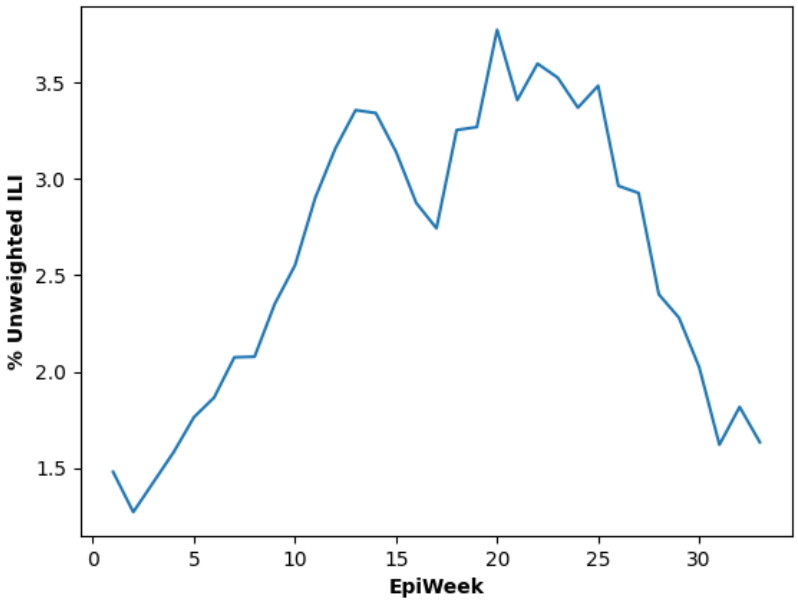

Figure 23: Unweighted ILI for California for 2018-19 Flu Season

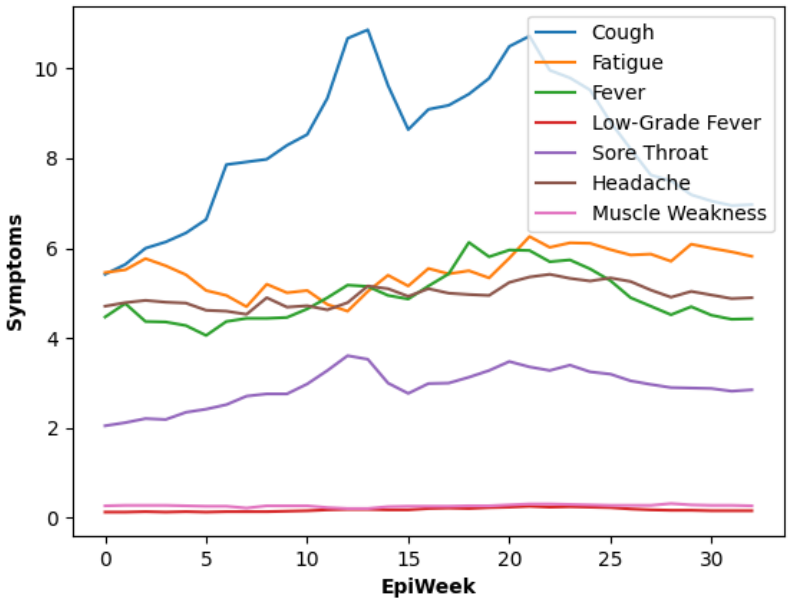

Figure 24: Symptoms plot for California

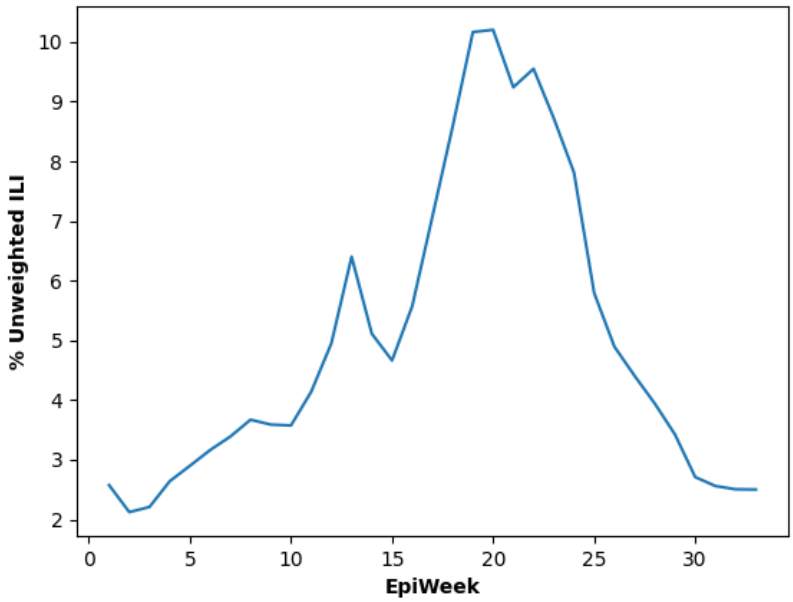

Figure 25: Unweighted ILI for Texas for 2018-19 Flu Season

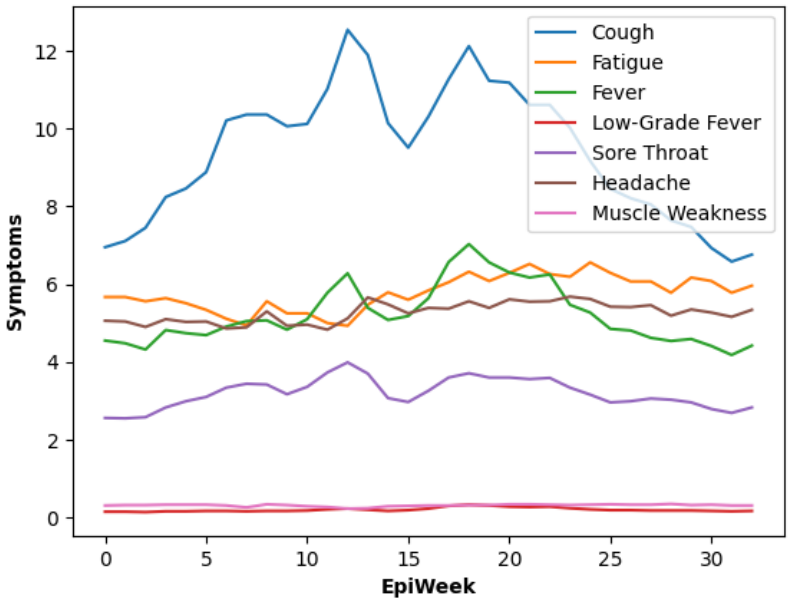

Figure 26: Symptoms plot for Texas

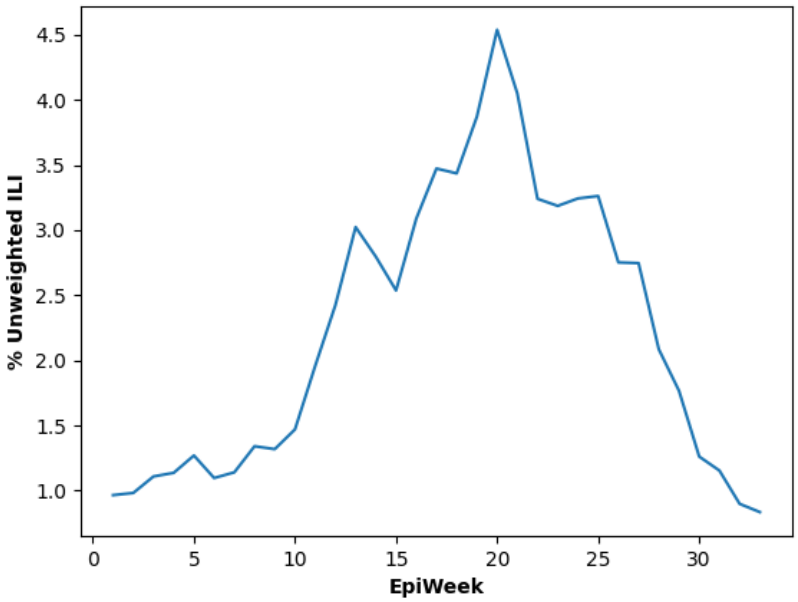

Figure 27: Unweighted ILI for New York for 2018-19 Flu Season

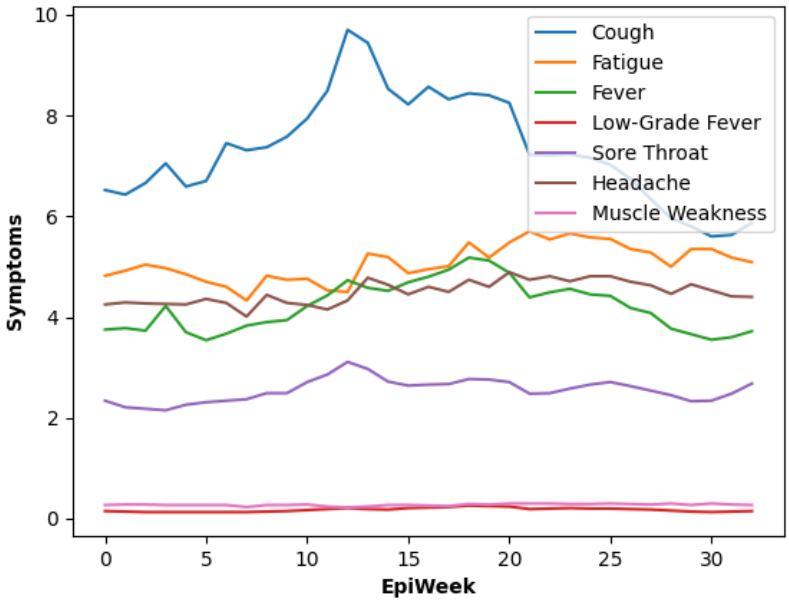

Figure 28: Symptoms plot for New York

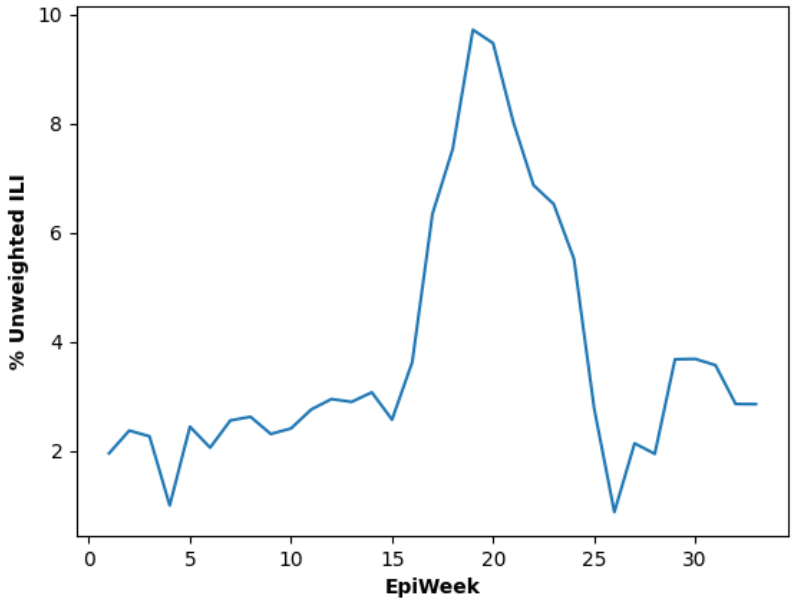

Figure 29: Unweighted ILI for Alaska for 2018-19 Flu Season

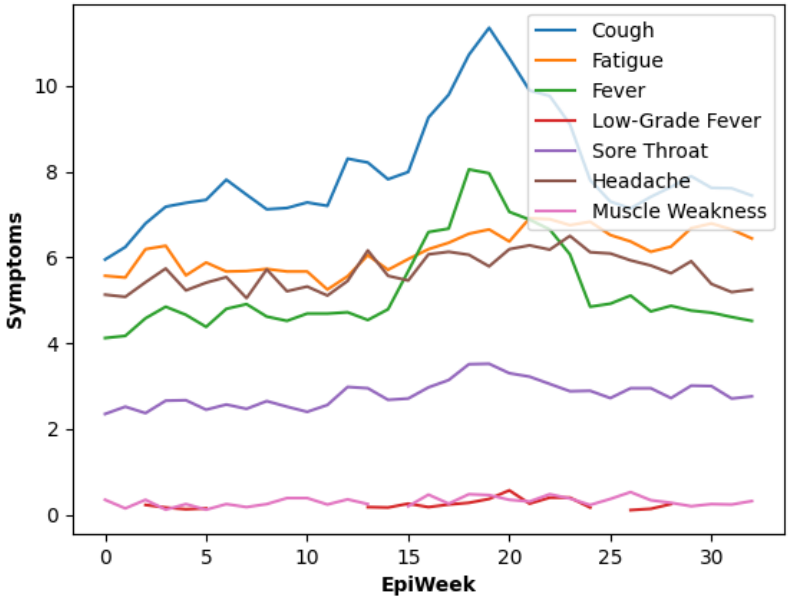

Figure 30: Symptoms plot for Alaska

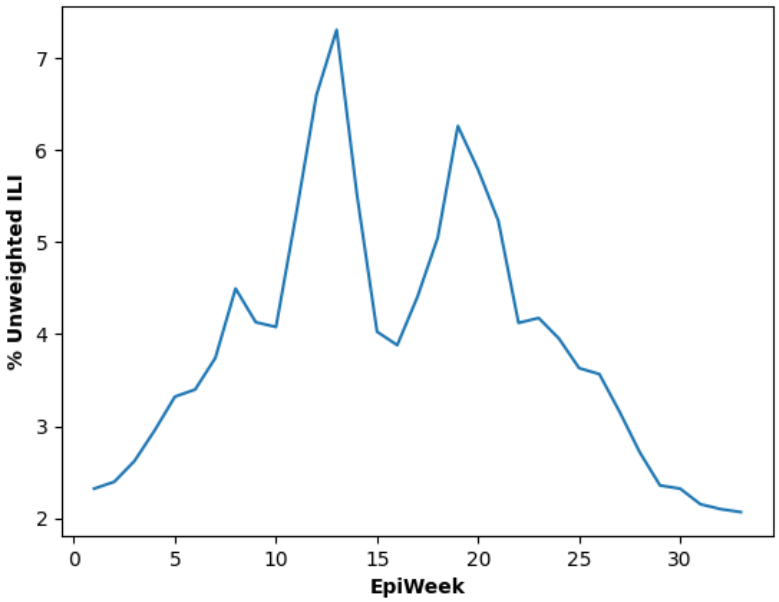

Figure 31: Unweighted ILI for Georgia for 2018-19 Flu Season

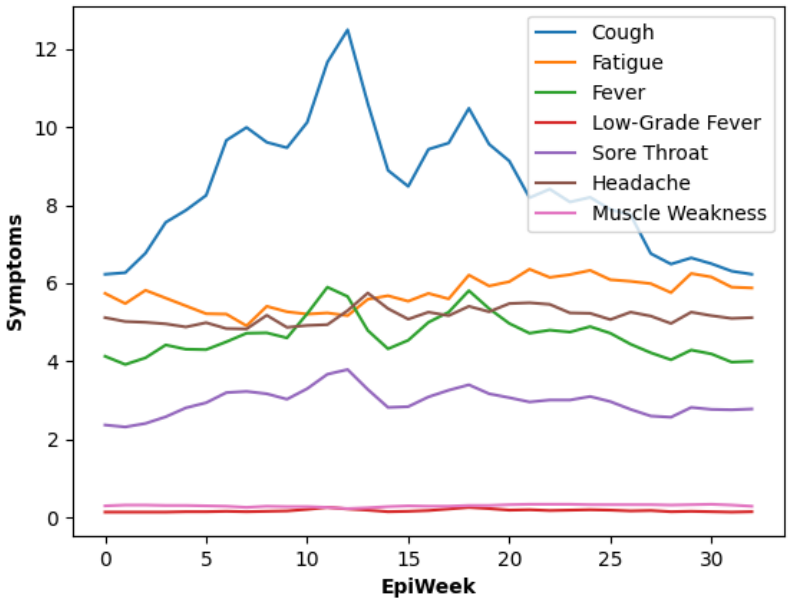

Figure 32: Symptoms plot for Georgia

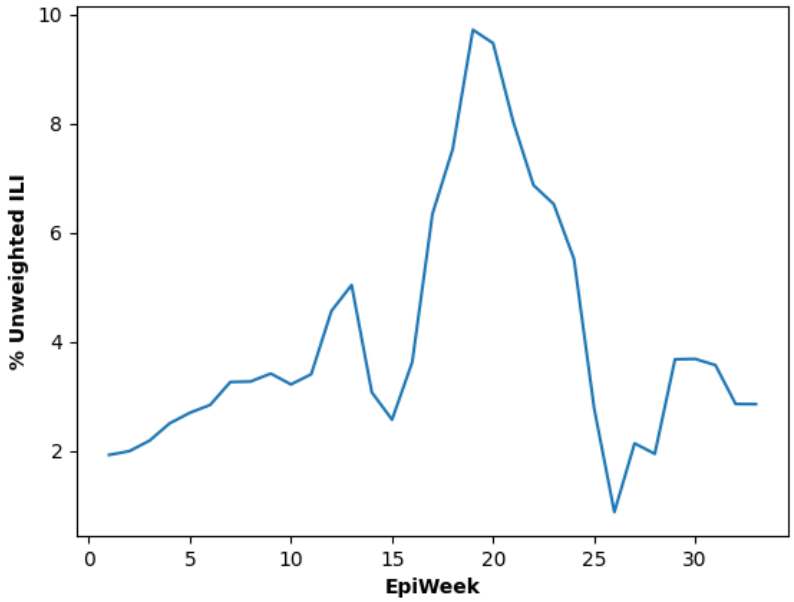

Figure 33: Unweighted ILI for Mississippi for 2018-19 Flu Season

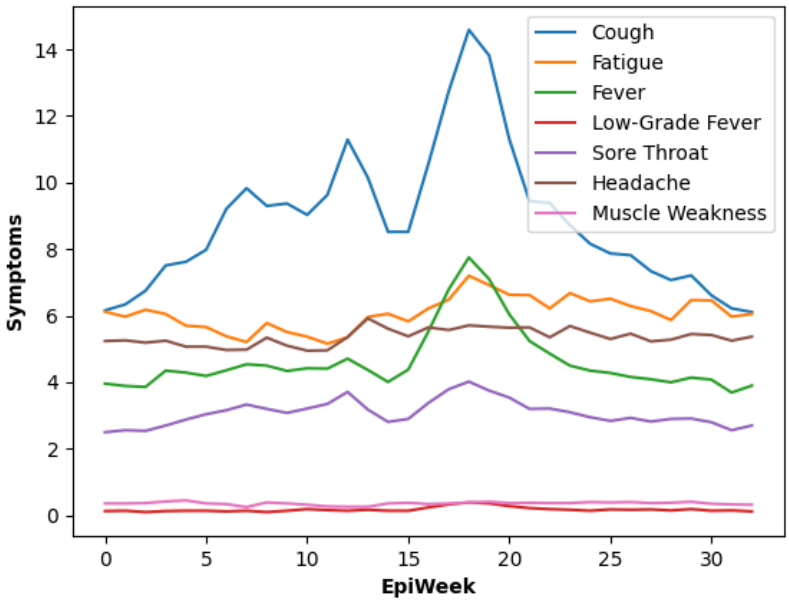

Figure 34: Symptoms plot for Mississippi