CS4980: Computational Epidemiology

Sriram Pemmaraju and Alberto Maria Segre
Department of Computer Science
The University of Iowa

Spring 2020
https://homepage.cs.uiowa.edu/~sriram/4980/spring20/

Computer Science

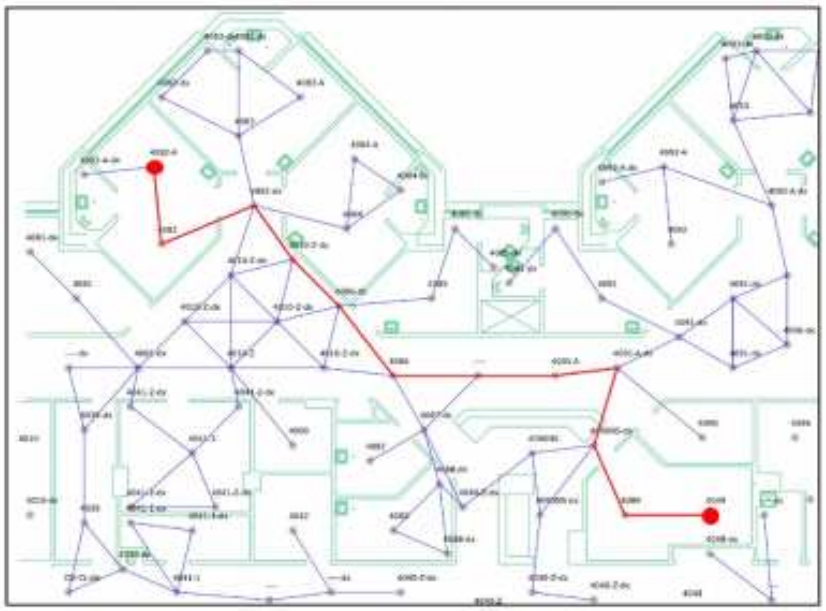
CDI Transmission

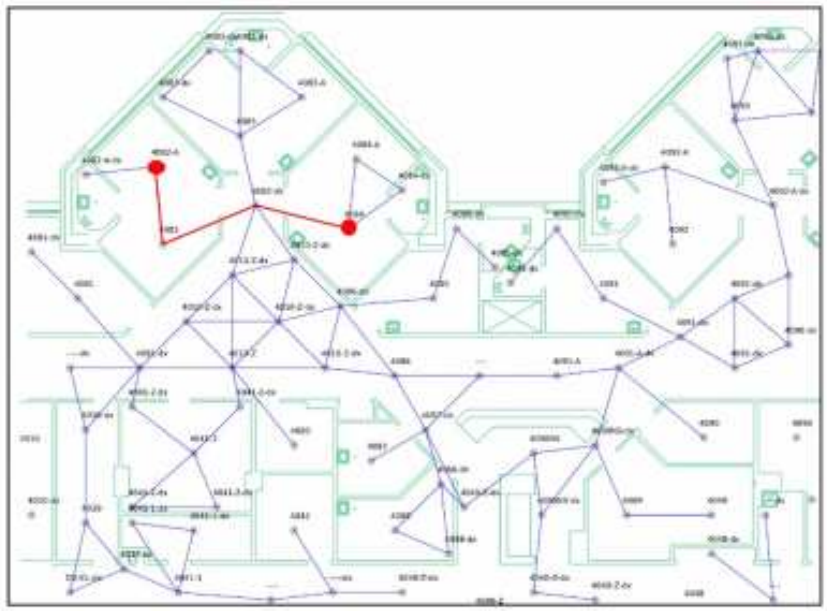
Samore (1999) lists 3 mechanisms for CDI transmission: **direct** (*e.g.*, from HCW hands), **environmental** (*e.g.*, from spores left in the environment) and **endongenous** (*i.e.*, self colonized).

Each of these pathways can be addressed by a different intervention (*e.g.*, better hand hygiene, deep cleaning at discharge, or improved ABX Rx and patient transfer practices).

Effective intervention rely on understanding which of these pathways is in play.

Consider Space and Time for CDI

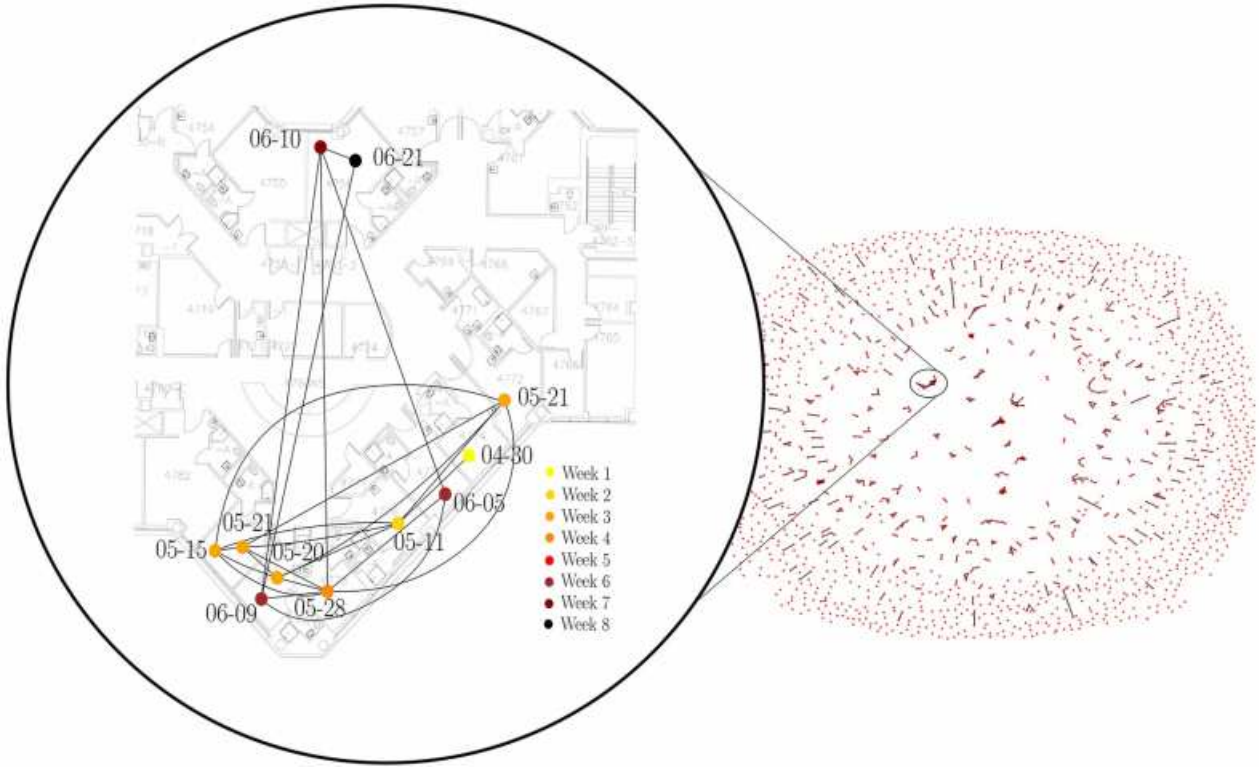




Recall our goal is to see if the observed "clustering" of CDI is accidental or the result of some underlying pathway.

Construct a *case proximity graph* for CDI using various t and d values based on timestamp and UIHC location of positive CDI test result.

The Case Proximity Graph (t=14, d=5)



How can we use such case proximity graphs to "measure" the spatiotemporal relationship between CDI cases?

Deriving a Metric of Spatiotemporal Correlation

What we need is a measure of whether the observed space/time correlation is something that is meaningful or happens by chance. There are several ways to test this condition statistically.

Deriving a Metric of Spatiotemporal Correlation

What we need is a measure of whether the observed space/time correlation is something that is meaningful or happens by chance. There are several ways to test this condition statistically.

The Knox test uses two $C \times C$ matrices, s and t, where C is the number of CDI cases, where s_{ij} is 1 iff cases i and j are within threshold D of each other. Similarly, t_{ij} is 1 iff cases i and j are within threshold T of each other.

Deriving a Metric of Spatiotemporal Correlation

What we need is a measure of whether the observed space/time correlation is something that is meaningful or happens by chance. There are several ways to test this condition statistically.

The Knox test uses two $C \times C$ matrices, s and t, where C is the number of CDI cases, where s_{ij} is 1 iff cases i and j are within threshold D of each other. Similarly, t_{ij} is 1 iff cases i and j are within threshold T of each other.

Summing $s_{ij} \times t_{ij}$ for $i < j$ yields a test statistic that counts how many cases are close enough in space and time.

Testing the Metric of Spatiotemporal Correlation

We next want to determine if the observed measure represents an "unusual" measure of correlation between space and time or if it simply arises by chance.

Testing the Metric of Spatiotemporal Correlation

We next want to determine if the observed measure represents an "unusual" measure of correlation between space and time or if it simply arises by chance.

To measure the "unusualness" of the observed value, we repeatedly randomly permute row/columns of one of the two matrices and compute the Knox metric for each of the permuted cases (this is a Monte Carlo estimation process).

Testing the Metric of Spatiotemporal Correlation

We next want to determine if the observed measure represents an "unusual" measure of correlation between space and time or if it simply arises by chance.

To measure the "unusualness" of the observed value, we repeatedly randomly permute row/columns of one of the two matrices and compute the Knox metric for each of the permuted cases (this is a Monte Carlo estimation process).

This process produces a distribution of Knox metrics where there is no expectation of space/time correlation.

Testing the Metric of Spatiotemporal Correlation

We next want to determine if the observed measure represents an "unusual" measure of correlation between space and time or if it simply arises by chance.

To measure the "unusualness" of the observed value, we repeatedly randomly permute row/columns of one of the two matrices and compute the Knox metric for each of the permuted cases (this is a Monte Carlo estimation process).

This process produces a distribution of Knox metrics where there is no expectation of space/time correlation.

We then compare the observed metric with the distribution.

The Mantel Test

The Knox test has some deficiencies; for one, it is sensitive to D and T thresholds.

The Mantel Test

The Knox test has some deficiencies; for one, it is sensitive to D and T thresholds.

An alternative test is the Mantel test, which is structurally similar to the Knox test, but where the matrices contain actual distance and time differences rather than indicator values.

The Mantel Test

The Knox test has some deficiencies; for one, it is sensitive to D and T thresholds.

An alternative test is the Mantel test, which is structurally similar to the Knox test, but where the matrices contain actual distance and time differences rather than indicator values.

Here, we calculate not the number of co-located indicator variables but the sum of the correlations of the two distances at corresponding matrix locations.

The Mantel Test

The Knox test has some deficiencies; for one, it is sensitive to D and T thresholds.

An alternative test is the Mantel test, which is structurally similar to the Knox test, but where the matrices contain actual distance and time differences rather than indicator values.

Here, we calculate not the number of co-located indicator variables but the sum of the correlations of the two distances at corresponding matrix locations.

The Monte Carlo estimation process is the same as for the Knox test.

The Mantel Test: Details

Because the measures in the two matrices are not directly comparable, we first normalize each matrix by transforming it into a matrix of Z scores (subtract the mean of the matrix from each element and divide the element by the standard deviation).

The Mantel Test: Details

Because the measures in the two matrices are not directly comparable, we first normalize each matrix by transforming it into a matrix of Z scores (subtract the mean of the matrix from each element and divide the element by the standard deviation).

Then, compute Pearson's r statistic over the corresponding normalized matrix elements; this is the cross product over a triangular portion of the matrix.

The Mantel Test: Details

Because the measures in the two matrices are not directly comparable, we first normalize each matrix by transforming it into a matrix of Z scores (subtract the mean of the matrix from each element and divide the element by the standard deviation).

Then, compute Pearson's r statistic over the corresponding normalized matrix elements; this is the cross product over a triangular portion of the matrix.

$-1 \leq r \leq 1$ is a measure of linear correlation between the two values; values of 1 or -1 indicate all values are perfectly aligned on a diagonal.

The Mantel Test: Details

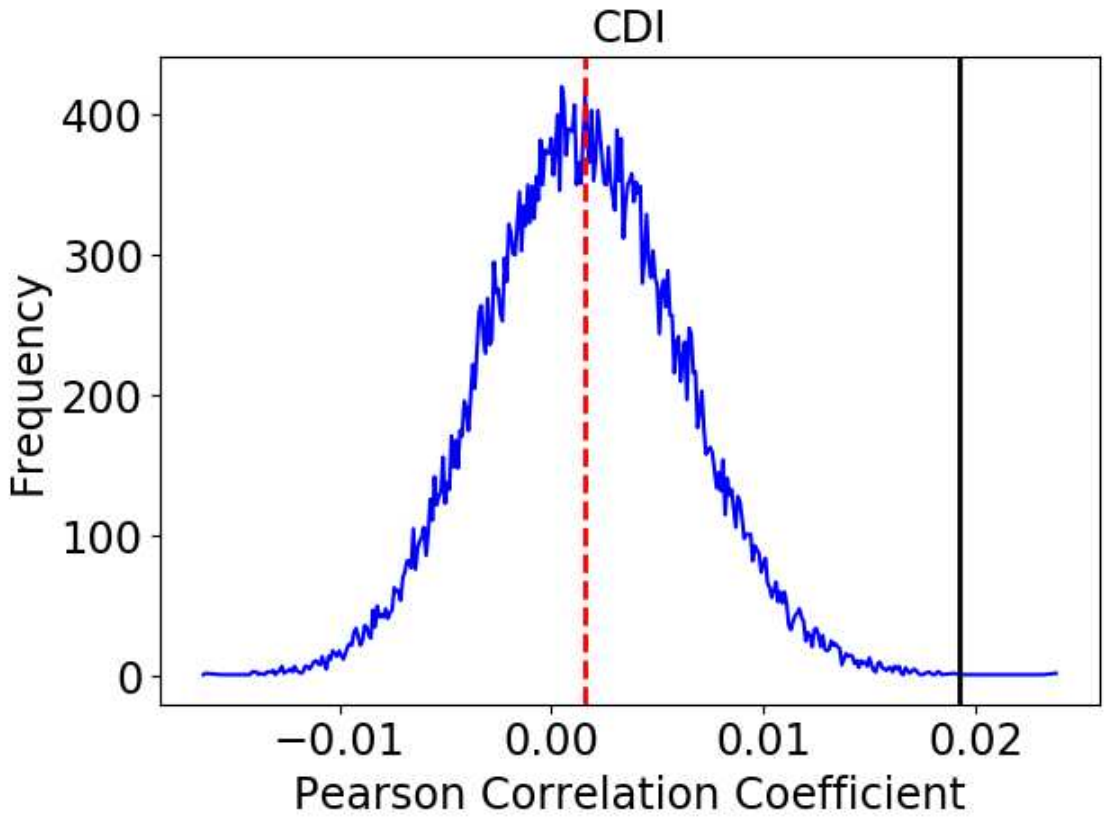
Because the measures in the two matrices are not directly comparable, we first normalize each matrix by transforming it into a matrix of Z scores (subtract the mean of the matrix from each element and divide the element by the standard deviation).

Then, compute Pearson's r statistic over the corresponding normalized matrix elements; this is the cross product over a triangular portion of the matrix.

$-1 \leq r \leq 1$ is a measure of linear correlation between the two values; values of 1 or -1 indicate all values are perfectly aligned on a diagonal.

The permutation test (a form of bootstrapping, where we randomize the correspondance of matrix elements) can be used to derive a p-statistic (count number of times $r_{bootstrap}$ exceeds $r_{observed}$). Confidence intervals can also be derived in a similar fashion.

Result: CDI Clustering



Result of the Mantel test on 20,000 permutations of space/time for CDI clusters; black line is the observed value, dotted red line the experimental mean.

CDI Clustering

Does this really mean that CDI clustering is a function of the bacterial infection? Or is there another explanation?

CDI Clustering

Does this really mean that CDI clustering is a function of the bacterial infection? Or is there another explanation?

What we need is a counterfactual, like John Snow's brewery workers.

CDI Clustering

Does this really mean that CDI clustering is a function of the bacterial infection? Or is there another explanation?

What we need is a counterfactual, like John Snow's brewery workers.

Consider *aspiration pneumonia*, an infection of the lungs that is mechanically induced by aspirating saliva or other substances.

CDI Clustering

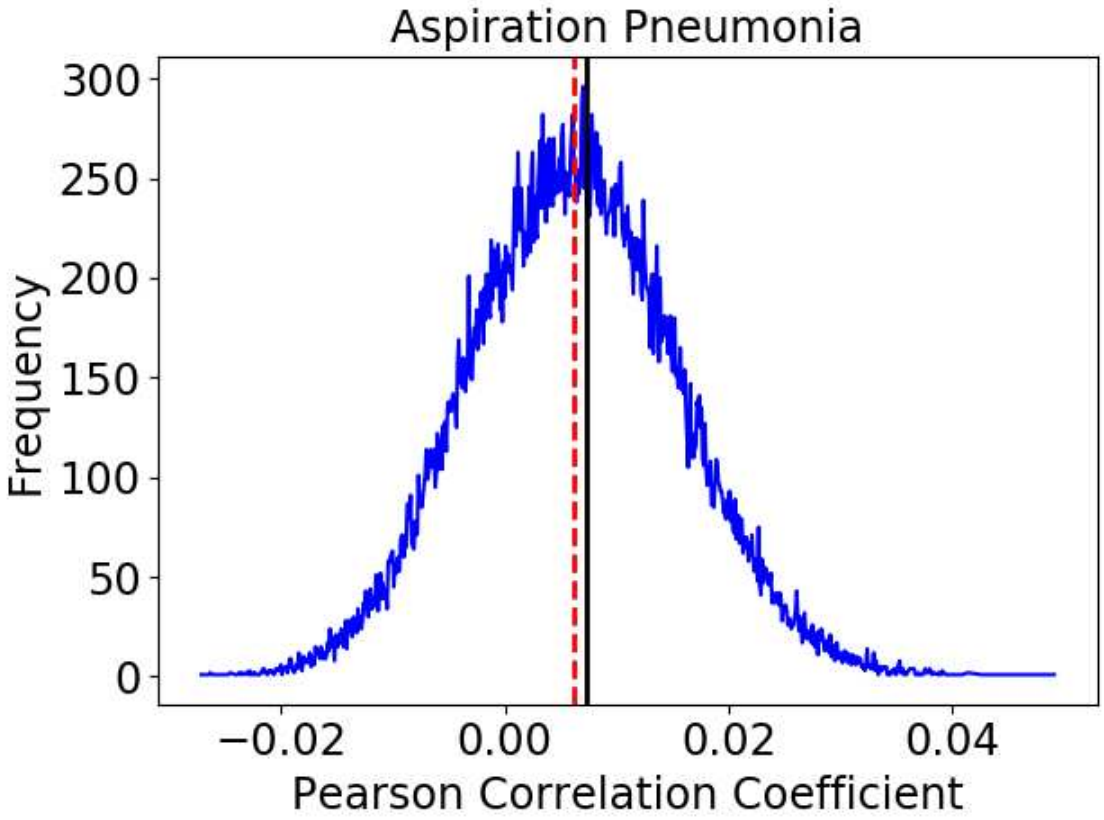
Does this really mean that CDI clustering is a function of the bacterial infection? Or is there another explanation?

What we need is a counterfactual, like John Snow's brewery workers.

Consider *aspiration pneumonia*, an infection of the lungs that is mechanically induced by aspirating saliva or other substances.

We built a case proximity graph for 790 cases of AP from the UIHC data; because AP is not contagious, we do not expect to observe any spatiotemporal correlation between them.

Result: AP clustering



Result of the Mantel test on 20,000 permutations of space/time for AP clusters; black line is the observed value, dotted red line the experimental mean.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Yet no such relationship exists for AP cases.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Yet no such relationship exists for AP cases.

The latter is as expected, since AP is not an infectious disease.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Yet no such relationship exists for AP cases.

The latter is as expected, since AP is not an infectious disease.

The results strongly suggest direct or environmental transmissiom of CDI.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Yet no such relationship exists for AP cases.

The latter is as expected, since AP is not an infectious disease.

The results strongly suggest direct or environmental transmissiom of CDI.

The results argue against endogenous (asymptomatic, self-colonized) transmission of CDI, contradicting Walker (2010), although it does not entirely rule out this pathway.

Results

The Mantel test (similarly, the Knox test) clearly show a spatiotemporal relationship exists for observed CDI cases.

Yet no such relationship exists for AP cases.

The latter is as expected, since AP is not an infectious disease.

The results strongly suggest direct or environmental transmissiom of CDI.

The results argue against endogenous (asymptomatic, self-colonized) transmission of CDI, contradicting Walker (2010), although it does not entirely rule out this pathway.

It does confirm hand hygiene and deep cleaning of patient rooms are critical defenses against nosocomial CDI.