A DATA-DRIVEN APPROACH TO IDENTIFYING ASYMPTOMATIC C. DIFF CASES

Hankyu Jang, Philip M. Polgreen, Alberto M. Segre, Daniel K. Sewell, Sriram V. Pemmaraju

* For the CDC MInD-Healthcare Group

epiDAMIK @KDD2020

The University of Iowa
comp|epi
computational epidemiology research

Clostridioides difficile (C. diff)

- C. diff is a bacterium that cause diarrhea and colitis
- C. diff infection (CDI) is a major health threat and known to be a common hospital acquired infection
  - Nearly 500 K Americans suffer from CDI each year [1]
- C. diff are passed in feces of anyone with C. diff and can spread if they fail to wash their hands thoroughly

[1] Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE. Burden of Clostridium difficile infection in the United States. New England Journal of Medicine. 2015 Feb 26;372(9):825-34.

Asymptomatic C. diff carriers

- Substantial fraction of hospitalized patients could be asymptomatic C. diff carriers
  - Up to 10% of patients admitted to a hospital were asymptomatic C. diff carriers [2]
- Asymptomatic C. diff carriers may play a significant role in the spread of C. diff
  - 45% of CDI cases originated from sources other than symptomatic cases [3]
  - Only 17% of CDI cases at a hospital ward had direct contact with other symptomatic patients [4]
- Understanding the role of asymptomatic C. diff carriers in the spread of C. diff is critical in designing effective interventions

[2] Leekha S, Aronhalt KC, Sloan LM, Patel R, Orenstein R. Asymptomatic Clostridium difficile colonization in a tertiary care hospital: admission prevalence and risk factors. American journal of infection control. 2013 May 1;41(5):390-3.
[3] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH. Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med. 2013 Sep 26;369:1195-205.
[4] García-Fernández S, Frentrup M, Steglich M, Gonzaga A, Cobo M, López-Fresneña N, Cobo J, Morosini MI, Cantón R, Del Campo R, Nübel U. Whole-genome sequencing reveals nosocomial Clostridioides difficile transmission and a previously unsuspected epidemic scenario. Scientific reports. 2019 May 6;9(1):1-9.

Challenges in designing a prediction model

- No "ground truth" data on asymptomatic C. diff carriage
  - Guidelines recommend only testing patients with new onset and unexplained diarrhea [5]
- We overcome the missing label problem in the following ways:
  - We consider two hypotheses on the relationship between CDI cases and asymptomatic C. diff carriers
  - We build a CDI prediction model (Stage 1 model) which is used to identify asymptomatic C. diff carriers
  - We design another CDI prediction model (Stage 2 model) to evaluate the carriers identified in the Stage 1 model by using measures of exposure to these carriers as an extra set of features

[5] L Clifford McDonald, et. al, Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA), Clinical Infectious Diseases, Volume 66, Issue 7, 1 April 2018, Pages e1–e48

Overview

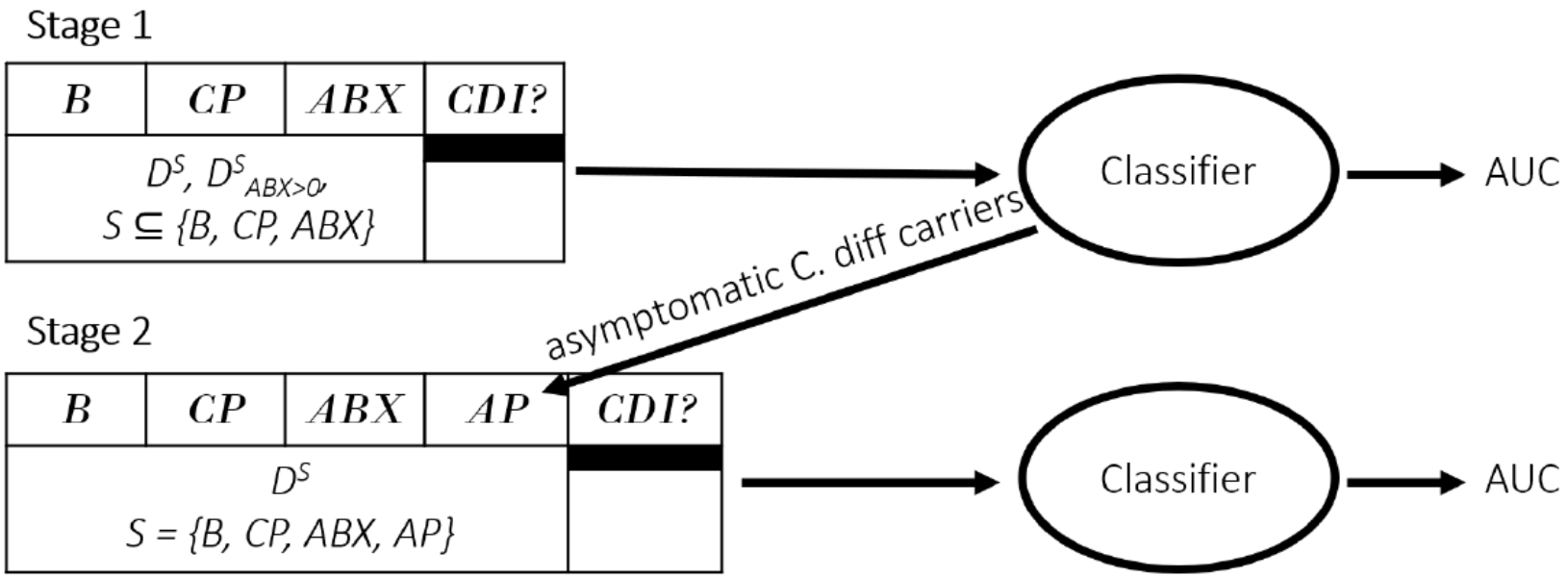
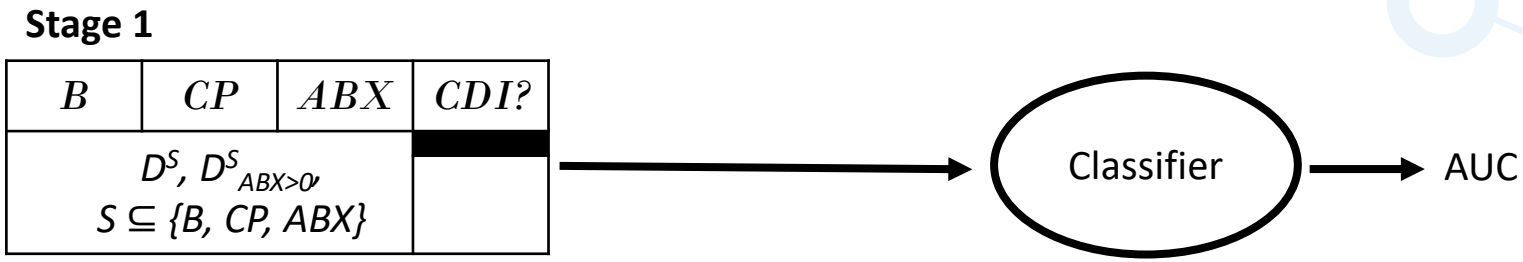


Figure 1: Diagram of the 2-stage model

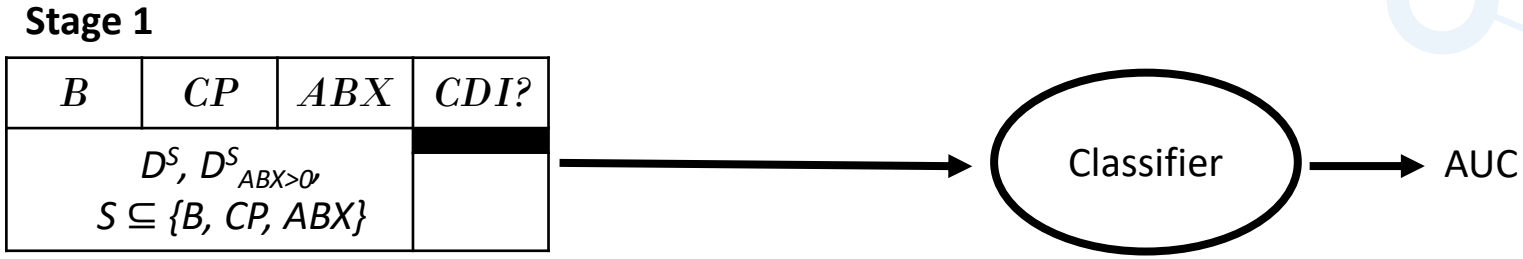
Stage 1 model: Infer asymptomatic C.diff carriers



- Patient data
  - Patient visits (154K) in 2007-2011 at the University of Iowa Hospitals and Clinics (UIHC)
    - $visit_{CDI}$ (0.8K): visits during which patients were tested positive for CDI
    - $visit_{CDIx}$ (115.3K): rest of the visits
    - We exclude short visits where patients were discharged within 48 hours to account for hospital acquired infection [6-7]
  - Generate daily instances (8.9K CDI instances, 988.8K CDIx instances)
    - We exclude CDI instances later than 72 hours before the CDI test positive date because there could be modifications to treatment during this period in response to potential CDI [8]

[6] Jeeheh Oh, Maggie Makar, Christopher Fusco, Robert McCaffrey, Krishna Rao, Erin E Ryan, LaraineWasher, Lauren RWest, Vincent B Young, John Guttag, et al. 2018. A generalizable, data-driven approach to predict daily risk of Clostridium difficile infection at two large academic health centers. infection control & hospital epidemiology 39, 4 (2018), 425–433
[7] Centers for Disease Control and Prevention. Jan, 2020 (accessed June 11, 2020). Identifying Healthcare-associated Infections (HAI) for NHSN Surveillance. https://www.cdc.gov/nhsn/pdfs/pscmanual/2psc_identifyinghais_nhsncurrent.pdf
[8] M.N. Monsalve, S.V. Pemmaraju, S. Johnson, and P.M. Polgreen. 2015. Improving Risk Prediction of Clostridium Difficile Infection Using Temporal Event-Pairs. In 2015 International Conference on Healthcare Informatics. 140–149.

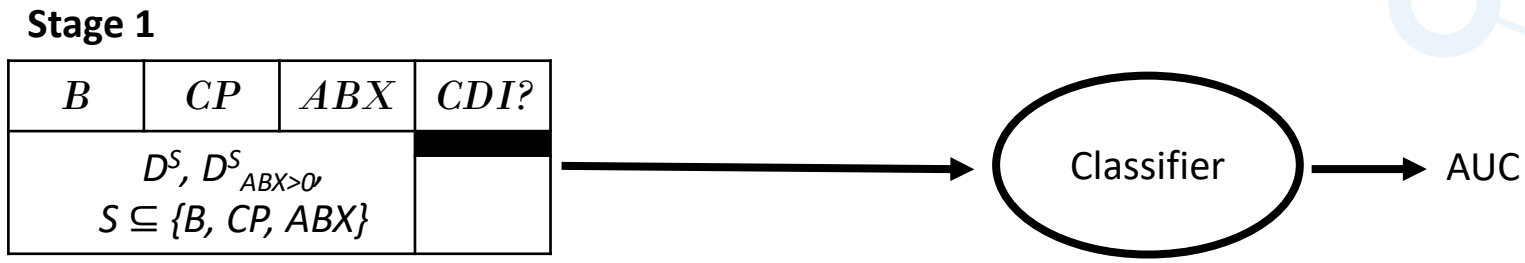
Stage 1 model: Infer asymptomatic C.diff carriers



- Individual risk factors for CDI [9-11]
  - Baseline (***B***): Length of stay of the visit, age, gender, previous visit, and gastric acid suppressors that are (i) H2-receptor antagonists, (ii) proton pump inhibitors
  - Antibiotics (***ABX***): high-risk antibiotics for CDI that are (i) Amoxillin or Ampicillin, (ii) Clindamycin (iii) Third generation Cephalosporin (iv) Fourth generation Cephalosporin (v) Fluoroquinolone
- Exposure risk factors for CDI [12]
  - CDI pressure (***CP***) : Exposure to CDI patients in room/unit level, daily
  - We assume that CDI patients are infectious 3 days before the positive result and up to 14 days after the test date

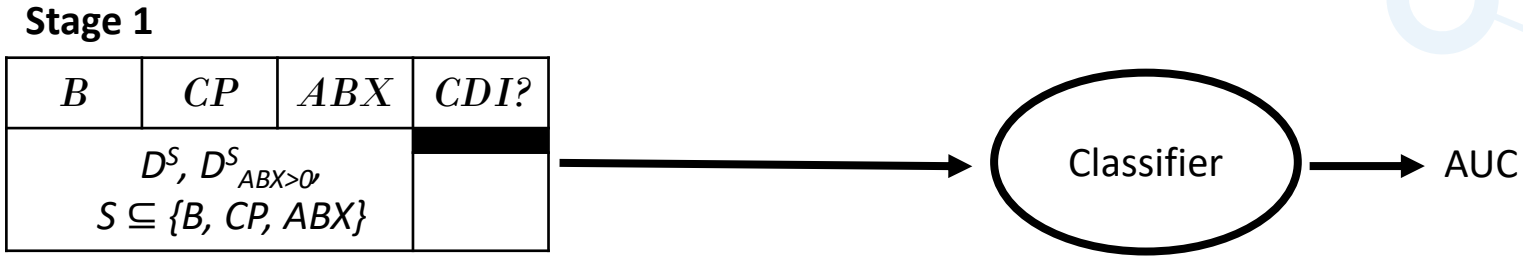
[9] Erik R Dubberke, Yan Yan, Kimberly A Reske, Anne M Butler, Joshua Doherty, Victor Pham, and Victoria J Fraser. 2011. Development and validation of a Clostridium difficile infection risk prediction model. Infection Control & Hospital Epidemiology 32, 4 (2011), 360–366
[10] Robert C Owens Jr, Curtis J Donskey, Robert P Gaynes, Vivian G Loo, and Carlene A Muto. 2008. Antimicrobial-associated risk factors for Clostridium difficile infection. Clinical Infectious Diseases 46, Supplement_1 (2008), S19–S31
[11] Sandra Dial, JAC Delaney, Alan N Barkun, and Samy Suissa. 2005. Use of gastric acid–suppressive agents and the risk of community-acquired Clostridium difficile–associated disease. Jama 294, 23 (2005), 2989–2995
[12] E. R. Dubberke, K. A. Reske, M. A. Olsen, K. M. McMullen, J. L. Mayfield, L. C. McDonald, and V. J. Fraser. 2007. Evaluation of Clostridium difficile-Associated Disease Pressure as a Risk Factor for C difficile-Associated Disease. Archives of Internal Medicine 167, 10 (05 2007), 1092–1097

Stage 1 model: Infer asymptomatic C.diff carriers



- Obstacle: No "ground truth" labels of asymptomatic C. diff carriers
- Overcome this by relying on two hypotheses which define a relationship between CDI and asymptomatic C. diff carriers
- Hypothesis 1: Asymptomatic C. diff carriers and CDI cases have similar risk profiles
  - D^B, $D^{B,CP}$, $D^{B,ABX}$, $D^{B,ABX,CP}$
- Hypothesis 2: The mechanism for acquiring CDI consists of the patient first being an asymptomatic C. diff carrier and then being prescribed high-risk ABX
  - $D_{ABX>0}{}^{B}$, $D_{ABX>0}{}^{B,CP}$, $D_{ABX>0}{}^{B,ABX}$, $D_{ABX>0}{}^{B,ABX,CP}$

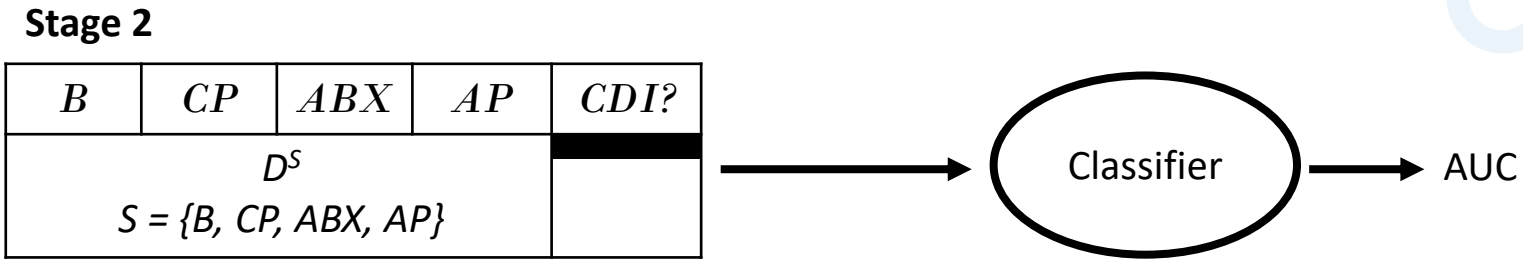
Stage 1 model: Infer asymptomatic C.diff carriers



- For each dataset, we build five prediction models, each model obtained by training on a 4-year subset, with one year excluded
  - Training: instances in 4 years (20% for validation)
  - Testing: instances in the remaining year
- For each instance (a day during a patient visit), we get a probability that we interpret to be the likelihood of that patient being an asymptomatic C. diff carrier on that day
- A visit in $visit_{CDIx}$ typically have multiple days: we select the maximum probability from daily instances and set it as the probability of that visit
- We mark top 10% [2], 5%, and 3% of these visits in $visit_{CDIx}$ as asymptomatic C.diff carrier visit: $visit_{ACDI10\%}$, $visit_{ACDI5\%}$, $visit_{ACDI3\%}$

[2] Leekha S, Aronhalt KC, Sloan LM, Patel R, Orenstein R. Asymptomatic Clostridium difficile colonization in a tertiary care hospital: admission prevalence and risk factors. American journal of infection control. 2013 May 1;41(5):390-3

Stage 2 model: evaluation



- Add exposure to asymptomatic C. diff carriers (***AP,*** asymptomatic C. diff pressure) that are identified in the Stage 1 model
- Investigate if this Stage 2 model improved the performance of the prediction model
- 24 sets of ***AP*** are computed (8 sets each for $visit_{ACDI10\%}$, $visit_{ACDI5\%}$, $visit_{ACDI3\%}$)

Stage 1 model results

Table 2: AUC on Stage 1 models

	D^{B}	$D^{B,ABX}$	$D^{B,CP}$	$D^{B,ABX,CP}$	$D^{B}_{ABX>0}$	$D^{B,ABX}_{ABX>0}$	$D^{B,CP}_{ABX>0}$	$D^{B,ABX,CP}_{ABX>0}$
AUC	0.676	0.635	0.704	*0.719	0.594	0.584	0.672	0.648

[a] AUC with asterisk denote best performer for $D^{S}, S \subseteq \{B, CP, ABX\}$

- AUCs reported here is averaged over the five years of test AUC values
- Best performing Stage 1 model uses all the standard risk factors for CDI

Stage 2 model results

Table 2: AUC on Stage 1 models

	D^{B}	$D^{B,ABX}$	$D^{B,CP}$	$D^{B,ABX,CP}$	$D^{B}_{ABX>0}$	$D^{B,ABX}_{ABX>0}$	$D^{B,CP}_{ABX>0}$	$D^{B,ABX,CP}_{ABX>0}$
AUC	0.676	0.635	0.704	*0.719	0.594	0.584	0.672	0.648

[a] AUC with asterisk denote best performer for $D^{S}, S \subseteq \{B, CP, ABX\}$

Table 3: AUC on Stage 2 models

	$D^{B,ABX,CP,AP}$							
AP	D^{B}	$D^{B,ABX}$	$D^{B,CP}$	$D^{B,ABX,CP}$	$D^{B}_{ABX>0}$	$D^{B,ABX}_{ABX>0}$	$D^{B,CP}_{ABX>0}$	$D^{B,ABX,CP}_{ABX>0}$
10%	0.712	0.687	*0.733	0.710	0.700	0.724	0.697	0.703
5%	0.701	0.690	*0.727	0.685	0.693	0.714	0.689	0.702
3%	0.689	0.698	*0.729	0.690	0.710	0.704	0.686	0.711

[a] AUC with asterisk denote best performer

- The Stage 2 model that uses ***B*** and ***CP*** but **not *ABX*** to identify asymptomatic C. diff carriers seems to accurately identify these carriers
  - Antibiotics use is not listed as a risk factor for asymptomatic C. diff carriage [13]
- This result implies that asymptomatic C. diff carriers contribute to CDI spread, confirming a conjecture from the CDI literature [3-4]

[3] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH. Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med. 2013 Sep 26;369:1195-205.
[4] García-Fernández S, Frentrup M, Steglich M, Gonzaga A, Cobo M, López-Fresneña N, Cobo J, Morosini MI, Cantón R, Del Campo R, Nübel U. Whole-genome sequencing reveals nosocomial Clostridioides difficile transmission and a previously unsuspected epidemic scenario. Scientific reports. 2019 May 6;9(1):1-9.
[13] Kong LY, Dendukuri N, Schiller I, Bourgault AM, Brassard P, Poirier L, Lamothe F, Béliveau C, Michaud S, Turgeon N, Toye B. Predictors of asymptomatic Clostridium difficile colonization on hospital admission. American journal of infection control. 2015 Mar 1;43(3):248-53.

Spatio-temporal clustering of cases

- CDI case proximity graph G_{CDI} [13]
  - CDI cases as nodes, and edges if two cases occurred within 14 days apart and within 30 m apart from each other (defined as spatio-temporal proximity)
- Revealed CDI case proximity graph $G_{RCDIx\%}$ for *x* in 3, 5, 10
  - CDI cases and the set of asymptomatic C. diff cases in $visit_{ACDIx\%}$ as nodes, and edges if two cases are in spatio-temporal proximity

Table 6: Network statistics

	G_{CDI}	$G_{RCDI3\%}$	$G_{RCDI5\%}$	$G_{RCDI10\%}$	
$\|V\|$	783	4241	6546	12310	$\|V\|$: number of nodes
$\|E\|$	120	4150	10630	37842	$\|E\|$: number of edges
$\langle k \rangle$	0.307	1.957	3.248	6.148	$\langle k \rangle$: average degree
k_{max}	4	18	31	47	k_{max}: maximum degree
std	0.581	2.095	3.145	5.195	std: standard deviation of degrees
cc	0.013	0.306	0.443	0.561	cc: clustering coefficient
$avg(\|E_{cpnt}\|)$	0.179	2.262	5.502	21.141	$avg(\|E_{cpnt}\|)$: average number of edges in connected components
$\|V_{giant}\|$	8	118	245	1239	$\|V_{giant}\|$: number of nodes in the giant component
$\|E_{giant}\|$	10	232	738	6393	$\|E_{giant}\|$: number of edges in the giant component

[13] Shreyas Pai, Philip M. Polgreen, Alberto Maria Segre, Daniel K. Sewell, and Sriram V. Pemmaraju. 2020. Spatiotemporal clustering of in-hospital Clostridioides difficile infection. Infection Control & Hospital Epidemiology 41, 4 (2020), 418–424.

Spatio-temporal clustering of cases

- To determine if asymptomatic C. diff cases lead to more clustering, we compared G_{CDI} and $G_{RCDIx\%}$ on the four different measures of density
- p-value of the Knox test [14] was 0, indicating spatio-temporal clustering of cases

Table 7: Network density

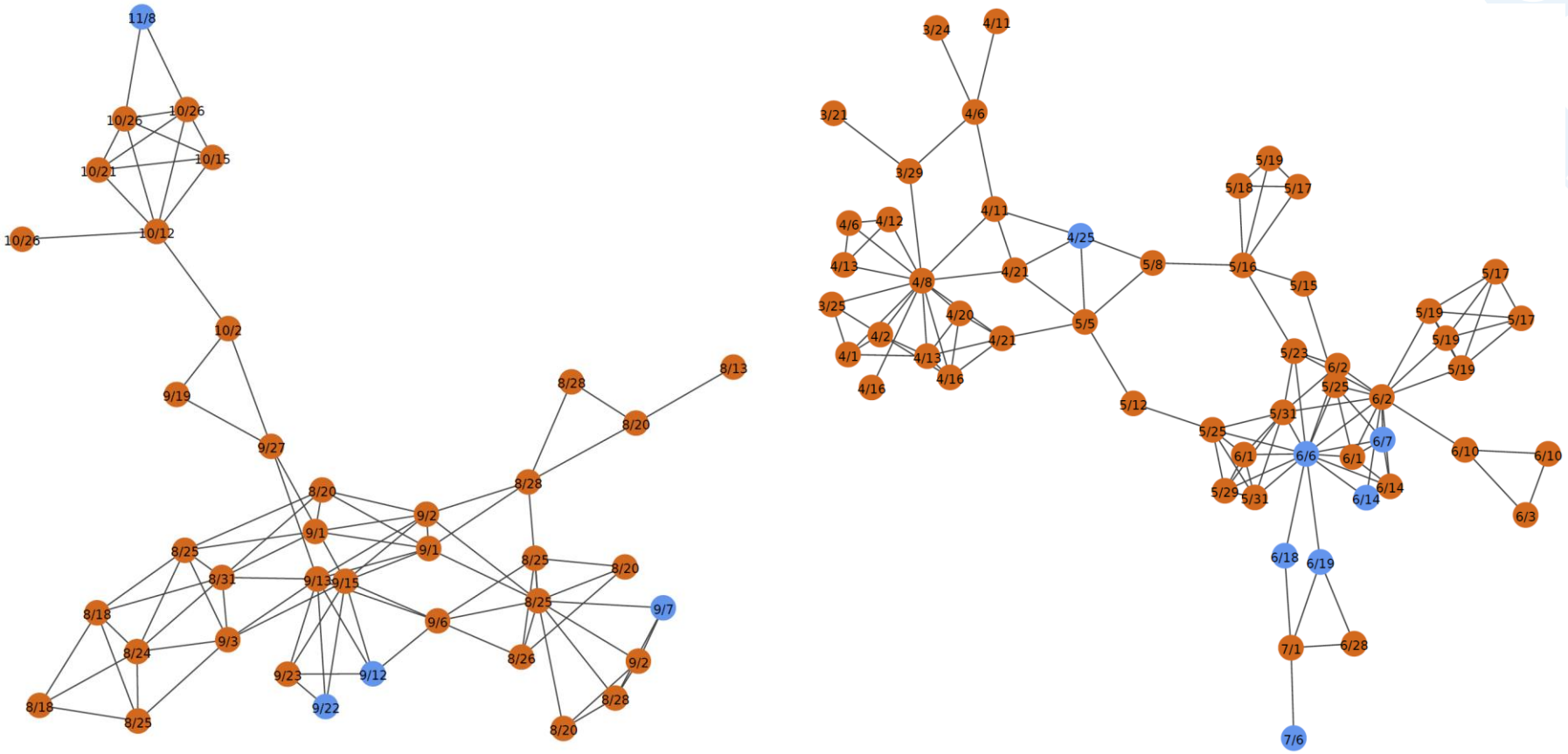
	G_{CDI}	$G_{RCDI3\%}$	$G_{RCDI5\%}$	$G_{RCDI10\%}$
$\|E\|/\|E*\|$	3.92e-04	4.62e-04	4.96e-04	4.99e-04
$\|E\|/\|V\|$	1.53e-01	9.79e-01	1.62e+00	3.07e+00
$\|E_{giant}\|/\|V\|$	1.28e-02	5.47e-02	1.13e-01	5.19e-01
$avg(\|E_{cpnt}\|)/\|V\|$	2.29e-04	5.33e-04	8.41e-04	1.72e-03

$|V|$: number of nodes, $|E|$: number of edges,
$|E*|$: number of possible edges,
$|E_{giant}|$: number of edges in the giant component,
$avg(|E_{cpnt}|)$: average number of edges in connected components

[14] Martin Kulldorff and Ulf Hjalmars. 1999. The Knox Method and Other Tests for Space-Time Interaction. Biometrics 55, 2 (1999), 544–552.

Spatio-temporal clustering of cases

- Illustration of two connected components in $G_{RCDI10\%}$



Discussion and future work

- Our finding suggest that risk factors for asymptomatic C. diff carriage include most of the risk factors for CDI except for high-risk antibiotics
  - This finding needs to be made more precise and tested by gathering prospective clinical data
- We did not consider a chain of infections that involves sequences of asymptomatic C. diff carriers
  - Combining more complicated exposure chains is another avenue of this work
- Yet another direction is to use deep embedding approaches such as graph convolutional networks
- Our asymptomatic C. diff carrier model can be applied to detect asymptomatic carriers in other infectious diseases where exposure plays a role in disease transmission