Clustering Temporal Disease Networks to Assist Clinical Decision Support Systems in Visual Analytics of Comorbidity Progression

Yajun Lu

Department of Management & Marketing, Jacksonville State University
ylu@jsu.edu

Joint work with:

Suhao Chen[a,b], Zhuqi Miao[a], Dursun Delen[a,c], and Andrew Gin[a]

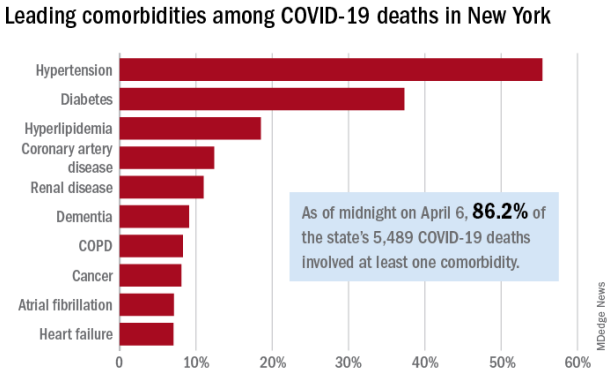
[a]Center for Health Systems Innovation, Oklahoma State University
[b]School of Industrial Engineering & Management, Oklahoma State University
[c]Department of Management Science & Information Systems, Oklahoma State University

Outline

Outline

Comorbidity

Comorbidity refers to one or more other health conditions coexisting among patients with a particular index disease (Feinstein, 1970; Gijsen et al., 2001).



Comorbidity Network

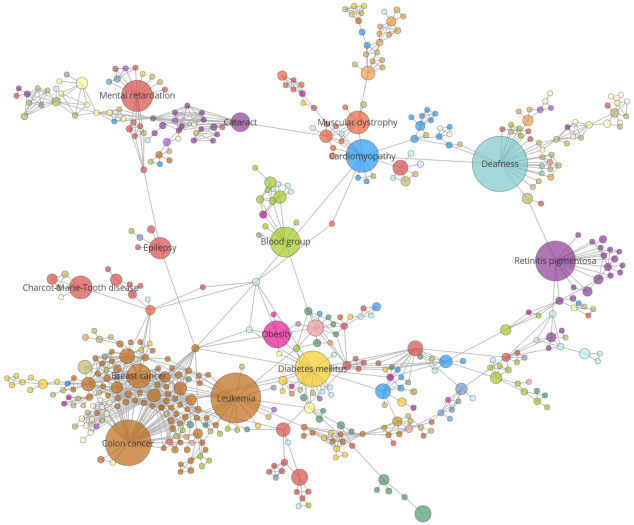
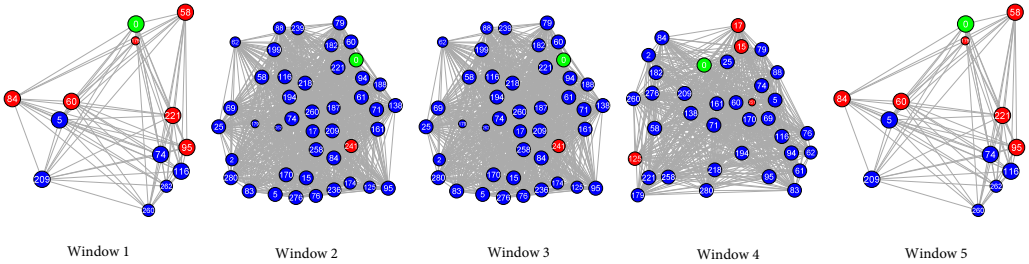


Image source: https://github.com/empet

- Better presentation of disease associations (Divo et al., 2015; Warner et al., 2015)
- Capability to incorporate additional biomarkers (Nam et al., 2019)
- Support for disease progression investigation (Chen et al., 2009; Chmiel et al., 2014)

Temporal Disease Network

By discretizing the entire time frame of the index disease into different windows, modeling comorbidity within each window as a disease network is referred to as temporal disease network (TDN).



A sequence of TDNs across five time windows

Research Gaps

In the area of TDN modeling and analysis, there has been limited capability to:

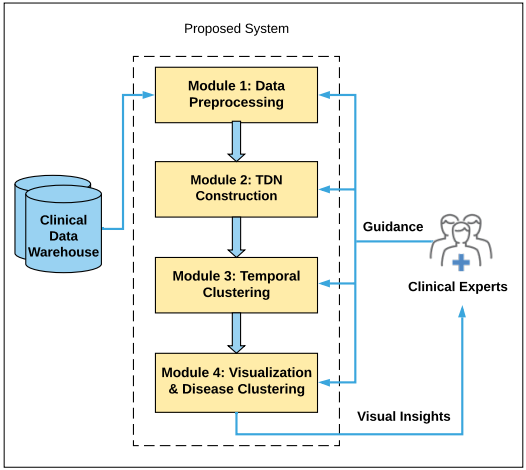
- Detect the progression timing.
  - Most TDN-related studies (Chen et al., 2009; Martel et al., 2016; McElroy et al., 2018) predefined a granularity parameter m
  - Then discretized the entire time frame of the study cohort into m windows of even length or even sample size, without providing algorithms that can detect at which window(s) notable changes of TDNs had occurred.

Research Gaps

In the area of TDN modeling and analysis, there has been limited capability to:

- Detect the progression timing.
  - Most TDN-related studies (Chen et al., 2009; Martel et al., 2016; McElroy et al., 2018) predefined a granularity parameter *m*
  - Then discretized the entire time frame of the study cohort into *m* windows of even length or even sample size, without providing algorithms that can detect at which window(s) notable changes of TDNs had occurred.
- Streamline the visualization.
  - TDNs are often large, dense networks that can result in disordered display in visualization.

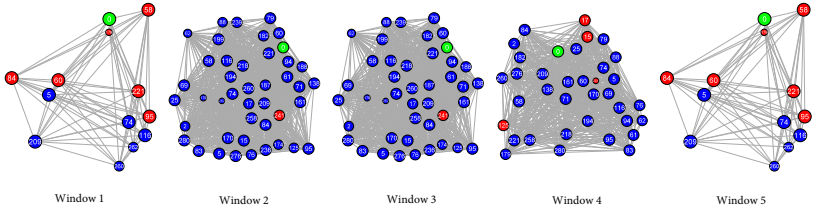
Research Goal



The proposed TDN-based clinical decision support system for pattern detection and visualization of comorbidity progression.

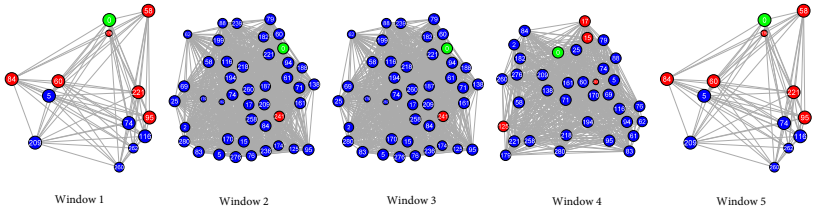
Outline

Consecutive *p*-Median Clustering (CPMP)



A sequence of TDNs across five time windows, of which the ones on Window 1 and Window 5 are identical and the ones through Window 2 to Window 4 are highly similar.

Consecutive p-Median Clustering (CPMP)



A sequence of TDNs across five time windows, of which the ones on Window 1 and Window 5 are identical and the ones through Window 2 to Window 4 are highly similar.

Problem: Consecutive p-median problem.
Input: A positive integer p, a collection of m objects $\mathscr{O} := \{O_1, O_2 \ldots, O_m\}$, and the distance between any two objects.
Output: From $\mathscr{O}$, find p objects with indices $\{j_1, j_2, \ldots, j_p\}$ as medians and assign the remaining $m-p$ objects to the medians such that

- The total summation of distances from each O_i to its assigned median is minimized, and
- When O_i is assigned to median O_{j_q}, if $i \geq j_q$ then O_k for all $j_q \leq k < i$ must be assigned to O_{j_q}, otherwise O_k for all $i < k \leq j_q$ must be assigned to O_{j_q}.

An Integer Programing (IP) Formulation for the CPMP

$$\min \sum_{i=1}^{m}\sum_{j=1}^{m} d(G_i, G_j)x_{ij} \tag{1}$$

$$\text{subject to: } \sum_{j=1}^{m} x_{jj} \leq p \tag{2}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \qquad \forall i \in \{1,2,\cdots,m\} \tag{3}$$

$$x_{ij} \leq x_{jj} \qquad \forall i,j \in \{1,2,\cdots,m\} \mid i \neq j \tag{4}$$

$$x_{ij} \leq x_{kj} \qquad \forall i \in \{1,2,\cdots,m-2\}, j \in \{i+2, i+3,\cdots,m\}, k \in \{i+1, i+2,\cdots,j-1\} \tag{5}$$

$$x_{ij} \leq x_{kj} \qquad \forall i \in \{3,4,\cdots,m\}, j \in \{1,2,\cdots,i-2\}, k \in \{j+1, j+2,\cdots,i-1\} \tag{6}$$

$$x_{ij} + x_{i+1,j} \leq 1 \qquad \forall j \in \{1,2,\cdots,m\}, i \in \{1,2,\cdots,m-1\} \mid d(G_i, G_{i+1}) \geq \tau \tag{7}$$

$$x_{ij} \in \{0,1\} \qquad \forall i,j \in \{1,2,\cdots,m\}. \tag{8}$$

- $x_{ij} = 1$ if and only if TDN G_i is assigned to median G_j, for any $i,j \in \{1,2,\ldots,m\}$ such that $i \neq j$, otherwise $x_{ij} = 0$.
- $x_{jj} = 1$ indicates that G_j is served as a median for any $j \in \{1,2,\cdots,m\}$.

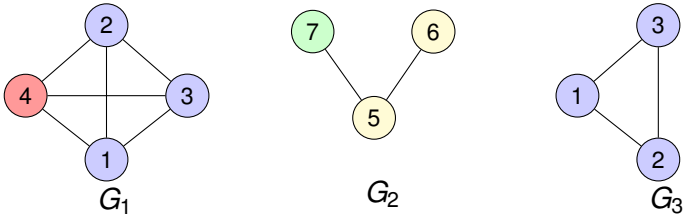
Outline

Atomic Clique

Given a network, a clique is a subset of nodes connecting with each other.

Definition 1 (Atomic Clique)

Given a collection of networks, $\mathscr{G} = \{G_1, G_2, \cdots, G_m\}$, a subset $S \subseteq \bigcup_{i=1}^{m} V(G_i)$ is called an atomic clique if S is a clique in $G_j, \forall j \in M$, but $S \cap V(G_k) = \emptyset, \forall k \notin M$, where $M = \{i \in \{1,2,\cdots,m\} \mid S \subseteq V(G_i)\}$.



Four atomic cliques $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, and $\{7\}$ across three networks $\{G_1, G_2, G_3\}$.

Algorithm 1: Atomic clique partition algorithm

Input: A collection of networks $\mathscr{G} = \{G_1, G_2, \cdots, G_m\}$.
Output: An atomic clique partition $\mathscr{K}$.

$\mathscr{K} \longleftarrow \emptyset$
while $\mathscr{G} \neq \emptyset$ **do**
  $M \longleftarrow \emptyset$
  $D \longleftarrow V(G_k)$, where $k = \min\{i \mid G_i \in \mathscr{G}\}$
  for $G_i \in \mathscr{G}$ **do**
    if $D \cap V(G_i) \neq \emptyset$ **then**
      $D \longleftarrow D \cap V(G_i)$
      $M \longleftarrow M \cup \{i\}$
  while $D \neq \emptyset$ **do**
    find a subset $K \subseteq D$ such that K is a clique in $G_i[D], \forall i \in M$ and $|K|$ is maximized
    $\mathscr{K} \longleftarrow \mathscr{K} \cup K$
    for $i \in M$ **do**
      $G_i \longleftarrow G_i[V(G_i) \setminus K]$
      if $V(G_i) = \emptyset$ **then**
        $\mathscr{G} \longleftarrow \mathscr{G} \setminus G_i$
    $D \longleftarrow D \setminus K$
return $\mathscr{K}$

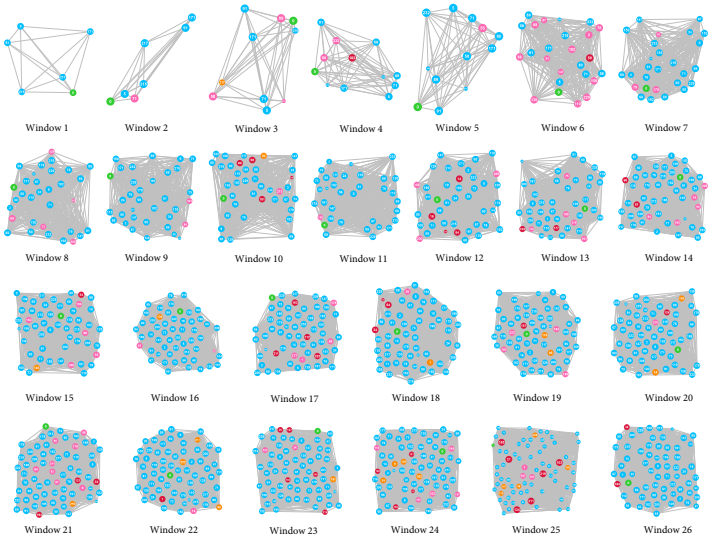
Outline

C. Diff & Stroke Background, Data Source, and Preparation

- C. Diff is a bacterial infection that can cause life-threatening diarrhea and colitis (Centers for Disease Control and Prevention, 2019).
- Stroke is one of the leading chronic conditions for death/disability in the U.S. (Members et al., 2016).
- Integrated Cerner Health Facts® EHR data warehouse as the data source into our system.
- Extracted all inpatient hospital encounters of female patients aged 65 or older with the onset of C. Diff/Stroke between November 1999 and August 2017.

C.Diff		Stroke	
Enct #	Diag #	Enct #	Diag #
2,229,051	20,677,407	117,262	571,186

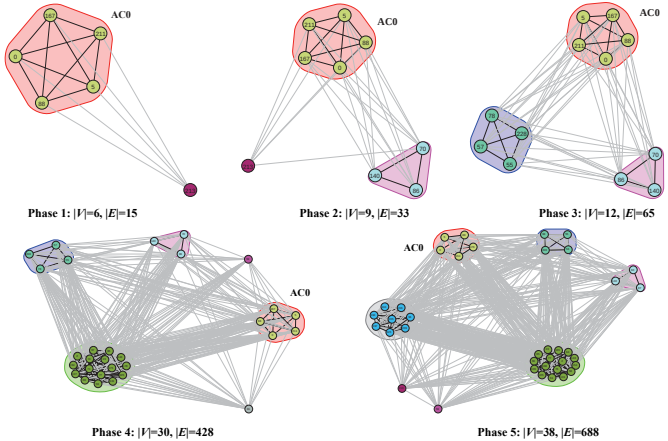
TDNs constructed for the senior female cohort (C.Diff)



Phases and corresponding windows and days

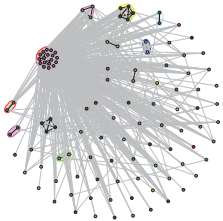
Cohorts	Time Unit	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
C. Diff	Window	1 – 3	4 – 5	6 – 11	12 – 20	21 – 26
	Day	2 – 3	3 – 4	4 – 7	7 – 11	12 – 14
Stroke	Window	1 – 8	9 – 15	16 – 26	–	–
	Day	2 – 5	6 – 9	9 – 14	–	–

Visualization of TDNs in Phases for C.Diff Cohort

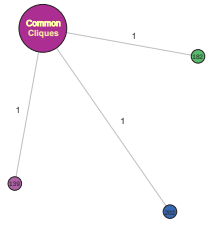


TDNs constructed on clustered phases of the senior female cohort.

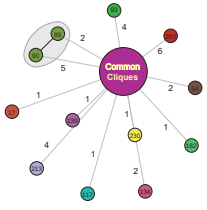
Visualization of TDNs in Phases for Stoke Cohort



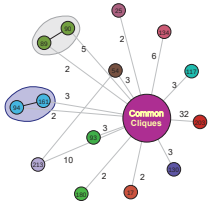
Common Cliques: $|V|$=109, $|E|$=696



Phase 1: $|V|$=112, $|E|$=699



Phase 2: $|V|$=121, $|E|$=882



Phase 3: $|V|$=123, $|E|$=1060

The common cliques are visualized in detail at the upper left part of the figure and simplified as a large node in the TDNs across the phases. The edge weight in the TDNs indicates how many nodes inside the set of common cliques are connected to a node outside the common cliques.

Outline

Concluding remarks

- Comorbidity is a prominent challenge in healthcare practice and research
- We modeled comorbidity progression as a sequence of TDNs, and designed a clinical decision support system
- The temporal and disease clustering technologies were developed to mine and visualize progression patterns from the TDN sequence
- Case studies of applying the system to C. Diff and Stroke demonstrates the effectiveness of the system

THANK YOU

Q & A

ylu@jsu.edu
http://yajunlu.com

Reference I

Centers for Disease Control and Prevention. Antibiotic resistance threats in the united states, 2019.

L. Chen, N. Blumm, N. Christakis, A.-L. BarabÃ¡si, and T. Deisboeck. Cancer metastasis networks and the prediction of progression patterns. *British Journal of Cancer*, 101(5):749–758, 2009.

A. Chmiel, P. Klimek, and S. Thurner. Spreading of diseases through comorbidity networks across life and gender. *New Journal of Physics*, 16(11):115013, 2014.

M. J. Divo, C. Casanova, J. M. Marin, V. M. Pinto-Plata, J. P. de Torres, J. J. Zulueta, C. Cabrera, J. Zagaceta, P. Sanchez-Salcedo, J. Berto, R. B. Davila, A. B. Alcaide, C. Cote, and B. R. Celli. Copd comorbidities network. *European Respiratory Journal*, 46:640–650, 2015.

A. R. Feinstein. The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, 23(7):455–468, 1970.

R. Gijsen, N. Hoeymans, F. G. Schellevis, D. Ruwaard, W. A. Satariano, and G. A. M. van den Bos. Causes and consequences of comorbidity: A review. *Journal of Clinical Epidemiology*, 54(7):661–674, 2001.

M. M. Martel, C. A. Levinson, J. K. Langer, and J. T. Nigg. A network analysis of developmental change in adhd symptom structure from preschool to adulthood. *Clinical Psychological Science*, 4(6):988–1001, 2016.

Reference II

E. McElroy, P. Fearon, J. Belsky, P. Fonagy, and P. Patalay. Networks of depression and anxiety symptoms across development. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(12): 964–973, 2018.

W. G. Members, D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, M. Cushman, S. R. Das, S. de Ferranti, J. Després, et al. Heart disease and stroke statistics-2016 update: a report from the american heart association. *Circulation*, 133(4):e38–e360, 2016.

Y. Nam, D.-g. Lee, S. Bang, J. H. Kim, J.-H. Kim, and H. Shin. The translational network for metabolic disease–from protein interaction to disease co-occurrence. *BMC Bioinformatics*, 20(576), 2019.

J. L. Warner, J. C. Denny, D. A. Kreda, and G. Alterovitz. Seeing the forest through the trees:uncovering phenomic complexity through interactive network visualization. *Journal of the American Medical Informatics Association*, 22:324–329, 2015.