On Atomic Cliques in Temporal Graphs[1]

Yajun Lu

Department of Management & Marketing, Jacksonville State University

Joint work with:

Zhuqi Miao[a], Parisa Sahraeian[b], and Baski Balasundaram[c]

[a]School of Business, State University of New York at New Paltz
[b]iHeartMedia, Inc.
[c]School of Industrial Engineering & Management, Oklahoma State University

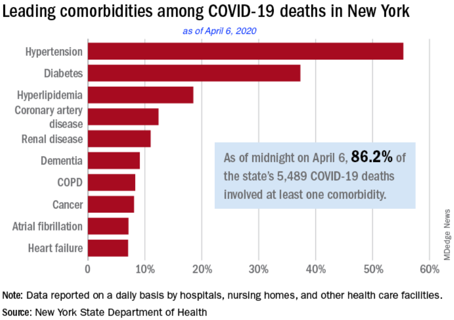
March 9, 2023

[1] Lu, Y., Miao, Z., Sahraeian, P., & Balasundaram, B. (2023). On atomic cliques in temporal graphs. *Optimization Letters*, 1-16.

Outline

Outline

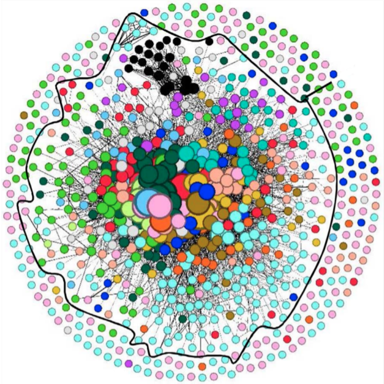
Comorbidity

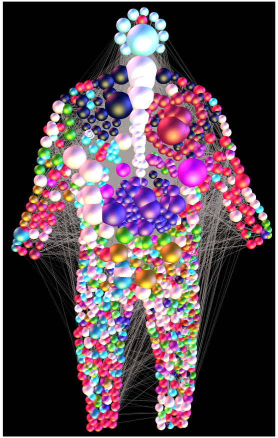


Comorbidity refers to two or more coexisting diseases or medical conditions in a patient (Feinstein, 1970; Gijsen et al., 2001) :

- worse medical outcomes
- more complex clinical treatments
- increased medical costs

Comorbidity Network

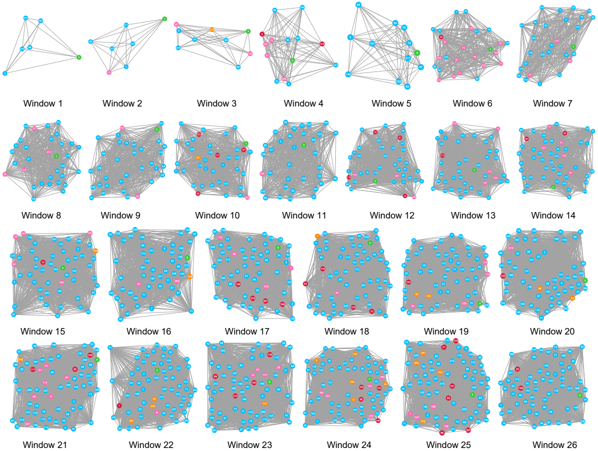




- Better presentation of disease associations (Divo et al., 2015; Warner et al., 2015)

Source: Kalgotra et al. (2017); Kalgotra and Sharda (2021)

Comorbidity Over Time



An example of temporal disease networks (TDNs) and each window spans 12 hours (Lu et al., 2021).

- Comorbidity networks constructed based on *Cerner Health Facts®* EHR data:
  - female patients aged 65 or older with the onset of C.Diff between November 1999 and August 2017
  - 2,229,051 inpatient hospital visits
- Comorbidity progression over 2 weeks

Atomic Clique

Notations:

- $\mathscr{G}$: a collection of (simple and undirected) graphs[2]
- G^0: the support graph of the collection $\mathscr{G}$; i.e., the minimal super-graph that contains every graph $G \in \mathscr{G}$
- $V(G^0)$: the vertex set of support graph G^0

[2]The vertex sets of the graphs in $\mathscr{G}$ are not assumed to be identical.

Atomic Clique

Notations:

- $\mathscr{G}$: a collection of (simple and undirected) graphs[2]
- G^0: the support graph of the collection $\mathscr{G}$; i.e., the minimal super-graph that contains every graph $G \in \mathscr{G}$
- $V(G^0)$: the vertex set of support graph G^0

Definition (Lu et al. (2021))

Given a collection of graphs $\mathscr{G}$ with support graph G^0, a subset of vertices $S \subseteq V(G^0)$ is called an *atomic clique* if one of the following conditions hold in every graph $G \in \mathscr{G}$:

[2]The vertex sets of the graphs in $\mathscr{G}$ are not assumed to be identical.

Atomic Clique

Notations:

- $\mathscr{G}$: a collection of (simple and undirected) graphs[2]
- G^0: the support graph of the collection $\mathscr{G}$; i.e., the minimal super-graph that contains every graph $G \in \mathscr{G}$
- $V(G^0)$: the vertex set of support graph G^0

Definition (Lu et al. (2021))

Given a collection of graphs $\mathscr{G}$ with support graph G^0, a subset of vertices $S \subseteq V(G^0)$ is called an *atomic clique* if one of the following conditions hold in every graph $G \in \mathscr{G}$:

1. $S \subseteq V(G)$ and forms a clique in G, or
2. $S \cap V(G) = \emptyset$.

[2]The vertex sets of the graphs in $\mathscr{G}$ are not assumed to be identical.

Atomic Clique Example

Definition (Lu et al. (2021))

Given a collection of graphs $\mathscr{G}$ with support graph G^0, a subset of vertices $S \subseteq V(G^0)$ is called an *atomic clique* if one of the following conditions hold in every graph $G \in \mathscr{G}$:

1. $S \subseteq V(G)$ and forms a clique in G, <u>or</u>
2. $S \cap V(G) = \emptyset$.

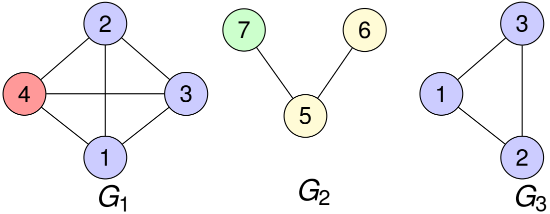


Figure 1: Three graphs $\{G_1, G_2, G_3\}$.

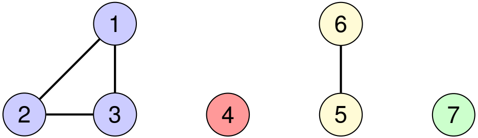
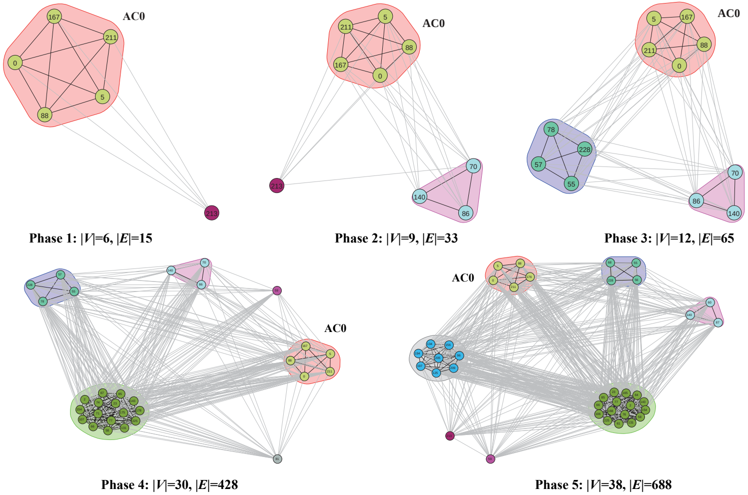


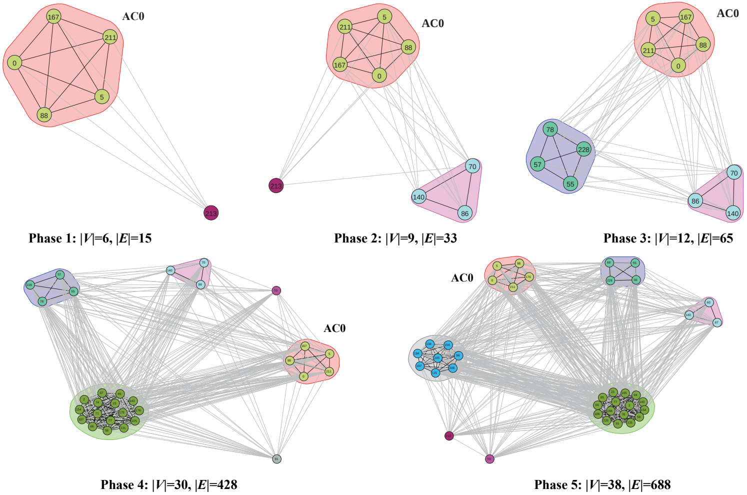
Figure 2: Four atomic cliques $\{1,2,3\}$, $\{4\}$, $\{5,6\}$, and $\{7\}$ across three graphs $\{G_1, G_2, G_3\}$

Streamlined Visualization by Atomic Clique Partition



- Some diseases (acute renal failure–node 5, fluid and electrolyte disorder–node 88, other gastrointestinal disorders–node 167, and septicemia–node 211) along with C. Diff (node 0) form an atomic clique (marked as AC0) that occurs persistently across all phases.
- Many clinical studies (Bauer et al., 2012; Doshi et al., 2018) have reported similar findings that these diseases are highly associated with C. Diff.

Streamlined Visualization by Atomic Clique Partition



- New diseases appeared at later phases tend to occur together.
- For example, urinary tract infection (UTI, node 228) appears in Phases 3–5 and forms an atomic clique along with cardiac dysrhythmias (node 55), chronic kidney disease (node 57), and disorders of lipid metabolism (node 78).
- This result echoes a previous study finding that UTI is associated with prolonged hospitalization of C. Diff patients (Warner et al., 2013).

Research Gaps

- Lu et al. (2021) presented an integer programming (IP) based heuristic to partition $V(G^0)$ into atomic cliques, but no exact algorithms were proposed.
- No IP formulations for atomic cliques exist.

Outline

Edge Peeling for Atomic Cliques

Algorithm 1: Edge Peeling: Generic Version for Atomic Cliques.

Input: a collection of graphs $\mathscr{G}$

while $\exists uv \in E(G) \setminus E(H)$ *for some* $G, H \in \mathscr{G}$ *and* $V(H) \cap \{u, v\} \neq \emptyset$ **do**

  delete edge *uv* from every graph that contains it

return $\mathscr{G}$

Consistent Connected Components

Lemma 1

Algorithm 1 produces a consistent set of connected components after edge peeling. That is, if $J \in \mathrm{cc}(G)$ and $K \in \mathrm{cc}(H)$ for graphs $G, H \in \mathscr{G}$, where $\mathscr{G}$ is the output of Algorithm 1, then one of the following conditions holds:

1. *Either $V(J) \cap V(K) = \emptyset$; or,*
2. *J and K are identical graphs, i.e., $V(J) = V(K)$ and $E(J) = E(K)$.*

Transformation from Atomic Clique to Clique

Theorem 1

Let $\mathscr{G}'$ and $\mathscr{G}$ be the input and output of Algorithm 1, respectively. Let $\widehat{G}$ be the (auxiliary) graph whose connected components are precisely the union of the consistent set of connected components of the graphs in the collection $\mathscr{G}$. In other words, we let $V(\widehat{G}) := \bigcup_{G \in \mathscr{G}} V(G)$ and $E(\widehat{G}) := \bigcup_{G \in \mathscr{G}} E(G)$. Then, S is an atomic clique of $\mathscr{G}'$ if and only S is a clique of $\widehat{G}$.

IP Formulation for the Maximum Atomic Clique Problem (MACP)

$$\max \sum_{u \in V(G^0)} x_u \tag{1a}$$

$$x_u + x_v \leq 1 \qquad \forall uv \in E(\overline{G}), G \in \mathscr{G} \tag{1b}$$

$$x_u + x_v \leq 1 \qquad \forall u \in V(G), v \notin V(G), G \in \mathscr{G} \tag{1c}$$

$$x_u \in \{0, 1\} \qquad \forall u \in V(G^0) \tag{1d}$$

Computational Experiments

- **Goal:** Gauge the effectiveness of edge peeling in conjunction with a maximum clique solver in solving the maximum atomic clique problem
- **Test-bed:** Real-life temporal graph collections from the Stanford Large Network Dataset Collection (SNAP) (Leskovec and Krevl, 2014) and graph collections generated from DIMACS Clique Challenge benchmarks (Johnson and Trick, 1996)
- **Software:** Gurobi™ Optimizer v9 and implemented in C++
- **Hardware:** 64-bit Linux® compute node with dual intel® Skylake 6130 processors and 96 GB RAM at the High Performance Computing Center at Oklahoma State University

Outline

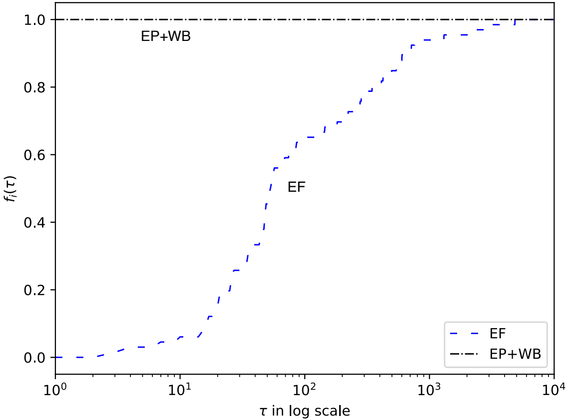
Results for Stanford Large Network Benchmarks

Comparing edge peeling followed by solving maximum clique against directly solving the maximum atomic clique problem IP formulation on SNAP temporal graph benchmarks.

						Wall-clock time (sec)	
Name	$\lvert V(G^0)\rvert$	$\lvert\mathscr{G}\rvert$	$\sum_{G\in\mathscr{G}}\lvert E(G)\rvert$	$\lvert E(\hat{G})\rvert$	Objective	EP+WB	EF
CollegeMsg_new	1,899	7	15,714	146	4	0.01	116.46
sx-mathoverflow_new	24,818	8	213,564	1,429	3	0.13	LPNS
sx-askubuntu_new	159,316	8	464,237	23,238	3	0.64	MEM
sx-superuser_new	194,085	9	734,144	19,227	3	0.87	MEM
wiki-talk-temporal_new	1,140,149	8	2,872,615	112,510	5	3.61	MEM
sx-stackoverflow_new	2,601,977	9	28,879,562	220,687	4	37.05	MEM

- EP: Edge Peeling
- WB: An effective max clique solver for sparse graphs by Walteros and Buchanan (2020)
- EF: Enhanced formulation
- The entry “LPNS” means that the root LP relaxation was not solved to optimality under the one–hour time limit. The entry “MEM” indicates that the solver did not terminate gracefully due to a memory–related crash.

Results for Instances Based on DIMACS Clique Challenge Benchmarks



Performance profile comparing the two approaches for solving the maximum atomic clique problem.

Outline

Concluding Remarks

- Atomic clique is a new network model used for analyzing disease progression.
- We presented a polynomial-time algorithm that transforms the maximum atomic clique problem to the maximum clique problem on an auxiliary graph.
- Computational results demonstrate the effectiveness of this transformation in solving the maximum atomic clique problem in comparison to direct integer programming based approaches.
- The proposed approach is also applicable when solving variants like the minimum atomic clique partitioning problem or the maximum weighted atomic clique problem.

Code Shared on Github

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

Reference I

M. P. Bauer, M. P. Hensgens, M. A. Miller, D. N. Gerding, M. H. Wilcox, A. P. Dale, W. N. Fawley, E. J. Kuijper, and S. L. Gorbach. Renal failure and leukocytosis are predictors of a complicated course of clostridium difficile infection if measured on day of diagnosis. *Clinical Infectious Diseases*, 55(suppl_2):S149–S153, 2012.

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.

R. Doshi, J. Desai, Y. Shah, D. Decter, and S. Doshi. Incidence, features, in-hospital outcomes and predictors of in-hospital mortality associated with toxic megacolon hospitalizations in the United States. *Internal and Emergency Medicine*, 13(6):881–887, 2018.

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.

D. Johnson and M. Trick, editors. *Cliques, Coloring, and Satisfiablility: Second* DIMACS *Implementation Challenge*, volume 26 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*. American Mathematical Society, Providence, RI, 1996.

Reference II

P. Kalgotra and R. Sharda. When will i get out of the hospital? modeling length of stay using comorbidity networks. *Journal of Management Information Systems*, 38(4):1150–1184, 2021.

P. Kalgotra, R. Sharda, and J. M. Croff. Examining health disparities by gender: a multimorbidity network analysis of electronic medical record. *International journal of medical informatics*, 108:22–28, 2017.

J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

Y. Lu, S. Chen, Z. Miao, D. Delen, and A. Gin. Clustering temporal disease networks to assist clinical decision support systems in visual analytics of comorbidity progression. *Decision Support Systems*, 148: 113583, 2021.

J. L. Walteros and A. Buchanan. Why is maximum clique often easy in practice? *Operations Research*, 68 (6):1866–1895, 2020. doi: 10.1287/opre.2019.1970. URL `https://doi.org/10.1287/opre.2019.1970`.

J. L. Warner, A. Zollanvari, Q. Ding, P. Zhang, G. M. Snyder, and G. Alterovitz. Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. *Journal of the American Medical Informatics Association*, 22:e281–e287, 2013.

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.

Backup slides

Algorithm 2: Edge Peeling for Atomic Cliques.

Input: $\mathscr{G}$

Construct support graph G^0

$\mathscr{I}(v) \leftarrow \{G \in \mathscr{G} : v \in V(G)\}$ $\quad \forall v \in V(G^0)$

$\mathscr{J}(uv) \leftarrow \{G \in \mathscr{G} : uv \in E(G)\}$ $\quad \forall uv \in E(G^0)$

contain$[v, G] \leftarrow \mathtt{false}$ $\quad \forall v \in V(G^0), G \in \mathscr{G}$

for $v \in V(G^0)$ **do**

  for $G \in \mathscr{I}(v)$ **do**

    contain$[v, G] \leftarrow \mathtt{true}$

$V(\widehat{G}) \leftarrow V(G^0), E(\widehat{G}) \leftarrow \emptyset$

for $uv \in E(G^0)$ **do**

  preserve $\leftarrow \mathtt{true}$

  contain-edge$[G] \leftarrow \mathtt{false}$ $\quad \forall G \in \mathscr{G}$

  contain-edge$[G] \leftarrow \mathtt{true}$ $\quad \forall G \in \mathscr{J}(uv)$

  for $G \in \mathscr{G}$ **do**

    if contain-edge$[G] = \mathtt{false}$ **then**

      if contain$[u, G] = \mathtt{true}$ *or* contain$[v, G] = \mathtt{true}$ **then**

        preserve $\leftarrow \mathtt{false}$

        break

  if *preserve* = $\mathtt{true}$ **then**

    $E(\widehat{G}) \leftarrow E(\widehat{G}) \cup \{uv\}$

return $\widehat{G}$

Formulation refinements

$$
\begin{aligned}
\sum_{J \in cc(G)} y_J^G &\leq 1 && \forall G \in \mathscr{G} && \text{(2a)}\\
x_u &\leq y_J^G && \forall u \in V(J), J \in cc(G), G \in \mathscr{G} && \text{(2b)}\\
x_u + x_v &\leq 1 && \forall uv \in E(\bar{J}), J \in cc(G), G \in \mathscr{G} && \text{(2c)}\\
y_J^G &\in \{0,1\} && \forall J \in cc(G), G \in \mathscr{G} && \text{(2d)}
\end{aligned}
$$

$$
\begin{aligned}
x_v &\leq 1 - z_G && \forall v \notin V(G), G \in \mathscr{G} && \text{(3a)}\\
x_u &\leq z_G && \forall u \in V(G), G \in \mathscr{G} && \text{(3b)}\\
z_G &\in \{0,1\} && \forall G \in \mathscr{G} && \text{(3c)}
\end{aligned}
$$