2021 IEEE 9th International Conference on Healthcare Informatics (ICHI) | 978-1-6654-0132-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICHI52183.2021.00036

Identification of Disease-Disease Network Communities in Subpopulations of Patients with Prostate Cancer

Ali Jazayeri
College of Computing and Informatics
Drexel University
Philadelphia, USA
ali.jazayeri@drexel.edu

Niusha Jafari
College of Computing and Informatics
Drexel University
Philadelphia, USA
nj396@drexel.edu

Nikita Nikita
Department of Medical Oncology
Thomas Jefferson University
Philadelphia, USA
fnu.nikita@jefferson.edu

Christopher C. Yang
College of Computing and Informatics
Drexel University
Philadelphia, USA
chris.yang@drexel.edu

Grace Lu-Yao
Department of Medical Oncology
Thomas Jefferson University
Philadelphia, USA
grace.luyao@jefferson.edu

***Abstract*—Prostate cancer (PCa) is ranked as one of the most common cancer diagnoses among men worldwide. Different research studies show that the current diagnosis and treatment strategies have improved the health condition of patients with PCa. Nonetheless, the number of new cases and the morbidity associated with PCa remain high. In this study, using the Medicare claims of patients identified from the SEER-Medicare database, we analyzed the disease-disease interactions in patients at different stages of PCa. Also, we implemented community detection to identify common co-occurring diseases in different subpopulations, identified by PCa stages. The similarity analysis of complication subgroups was performed to identify the most distinct complication subgroups among subpopulations. The results of the study show that the level of experiencing co-occurring diseases is different among subpopulations. Identifying distinct disease-disease interactions can inform the prediction of hospitalization rate and frequency and mortality rate among different subpopulations and enhance our understanding of the pathological correlations among diseases.**

***Index Terms*—Disease-Disease Networks, Community Detection, Prostate Cancer**

I. Introduction

Prostate cancer is ranked as the second most common cancer diagnosis among men worldwide [1]. It is shown that by using survivorship plans [2], and as a result of the progress made in the diagnosis and treatment strategies, the health conditions of PCa patients and their stratification have improved over the past years [3]. Nonetheless, the data provided as part of the GLOBOCAN project in collaboration with the World Health Organization shows that PCa has been the second most common cancer based on the number of new cases and the fifth cancer type with the highest mortality rate worldwide among men in 2020 [4].

On the other hand, it is shown that comorbidities play an important prognostic role and can improve decision making in cancer patients' care [5]–[7]. The analysis of comorbidities and disease-disease interactions has rich literature. For example, in [8], using protein interaction networks and pathway databases, a disease-disease similarity network is constructed. In [9], CytoCom is proposed for analysis, visualization, and exploration of human disease comorbidity network. CytoCom is created based on the patient medical records and International Classification of Diseases (ICD) codes collected from the US Medicare claims database. In [10], a network-based approach is adopted to human diseases modeling. Disease modules are defined as disease sets with localized regions of disease-related protein connections. It is shown that perturbation in one disease's pathways will disrupt another disease's pathways when both diseases are in the same modules.

In most of these studies, general disease-disease interaction networks are developed for all the patients without considering the potentially inherent differences among various subpopulations. In this study, we focus on patients with PCa. The association of comorbidities as pre-existing health conditions relative to PCa diagnosis or other complications as the cause or effect of prostate cancer diagnosis and treatment has been studied in numerous studies [11]–[15]. However, these studies generally focus on specific comorbidities to verify their associations with prostate cancer. The comorbidities in PCa patients at different levels of risks of mortality and prognosis can be investigated by adopting a data-driven approach. In this approach, the data collected from patients are used to identify the most common or significant comorbidities among patients' subpopulations. One analytical approach toward this problem is network modeling of comorbidities.

We also consider two control subpopulations with no history of prostate cancer and no history of cancer. In other words, in our study, we adopt a more in-depth analysis approach

978-1-6654-0132-6/21/$31.00 ©2021 IEEE
DOI 10.1109/ICHI52183.2021.00036

and focus on the differences of disease-disease interactions among different subpopulations. Using a network-based modeling approach, we identify disease classes co-occurring more frequently in different subpopulations of PCa patients. This study aims to assess whether co-occurring disease classes are different among subpopulations of prostate cancer, and also among PCa subpopulations with control cohorts. Furthermore, we identify the most dissimilar co-occurring disease classes among different subpopulations of PCa patients. Identification of such disease classes could enhance our understanding of the pathological correlations among diseases and their association with different levels of PCa severity.

II. Materials and Methods

A. Study population

We used data from the Surveillance, Epidemiology, end End Results (SEER)-Medicare linked data files to identify patients diagnosed with PCa between January 1, 2010, and December 31, 2015. The PCa patients were classified into two groups: patients diagnosed at localized or regional (L\R) stages, and patients diagnosed at the distant (D) stages of PCa. These two classes were further stratified as survivors and non-survivors. We also considered two control subpopulations, patients from a 5% population sample with no history of PCa, and patients with no history of cancer. In this study, we referred to the metastasis codes (CSMET) to stage PCa patients. If there was no distant metastasis in five years after diagnosis dates, we assumed the patient was a stage L\R; otherwise, the patient was categorized into stage D subpopulations. For patients with less than five years of data after the diagnosis date, we determined the PCa stages using the available data. The mortality data, if available, extracted from Medicare data, otherwise, it was obtained from SEER date of death related variables. We used the date of death to identify mortality status. A patient was flagged as survivor if he/she survived five years after the diagnosis date. The number of patients and the number of insurance claims of the subpopulations are provided in Table I.

TABLE I
NUMBER OF PATIENT AND INSURANCE CLAIMS FOR SUBPOPULATIONS.

Subpopulation	Number of patients	Number of insurance claims
Patient with no history of cancer	23,904	1,676,156
Patient with no history of PCa	22,223	2,636,910
PCa stage L\R survivors	19,528	1,744,946
PCa stage L\R non-survivors	3,318	386,340
PCa stage D survivors	1,843	182,237
PCa stage D non-survivors	2,477	249,759

B. Methods

Network development: We develop network models to represent the co-occurring complications in different subpopulations. The ICD-9 codes recorded at insurance claims are used as network nodes. The associations between complication pairs are represented as edges of the networks. The edges are weighted, and their weights represent the probability of complication pairs occurring concurrently. Two nodes (n_i and n_j) in the networks are connected if their associated ICD codes are observed in the same hospitalization record (or in insurance claims recorded almost at the same time). The weights of edges are computed at the subpopulation level based on the following formulation:

$$wt_{ij} = \sqrt{P(n_i|n_j) \times P(n_j|n_i)} = \sqrt{\frac{{f_{ij}^{11}}^2}{(f_{ij}^{11} + f_{ij}^{10}) \times (f_{ij}^{11} + f_{ij}^{01})}} \quad (1)$$

where:

- wt_{ij} is the weight of the edge connecting nodes n_i and n_j
- $P(n_i|n_j)$ is the probability of observing n_i complication given the n_j complication
- f_{ij}^{kl} ($k, l \in \{0, 1\}$, 0: no failure 1: failure) is the frequency of observing co-occurrence of complications represented by n_i and n_j (if $k = l = 1$) or failure of one complication without the other complication failing ($k \neq l$).

We create six networks representing the co-occurring disease complications in the six subpopulations. The edges of the networks are weighted using Equation 1. Then, these networks are used for community detection.

Community detection: Communities in a network are defined as densely interconnected nodes with sparse connections between communities [16]. In our study, we are interested in finding subgroups of complications co-occurring more frequently in different subpopulations of PCa. Modularity, Q, is represented by [17]:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2)$$

Here, A_{ij} denotes the edge's weight connecting nodes n_i and n_j, k_i and k_j represents the sum of the edges' weights connected to nodes n_i and n_j, respectively, and c_i and c_j are communities of nodes n_i and n_j. Also, δ is a function that returns 1 if both nodes are from the same communities, and 0 otherwise. Besides, m is defined as the total sum of weights of the network. The modularity is conceptually defined as a fraction of edges belonging to the same community minus the expected value of the same quantity if edges have been randomly distributed among communities, considering the degree of nodes in the same community [18]. For detection of complication subgroups, we use the networks created based on the approach discussed above and the popular algorithm proposed in [19]. This algorithm is based on the concept of modularity. In other words, this algorithm tries to optimize modularity gain initializing the communities with one node and merging them iteratively until no further improvement is possible. This study used a version of this algorithm implemented in the community API of Python NetworkX package

[20]. Also, we use the terms communities and complication subgroups interchangeably in this study.

Hierarchical clustering: The previous steps' output is a set of complication subgroups M^i for subpopulation i. Based on the definition adopted for modularity, each subgroup's complications co-occur more frequently with each other than other complications. Besides, the number of modules detected for each subpopulation might be different. Because each of the networks' nodes can be uniquely identified with the associated ICD code, we can represent the subgroups detected as edge lists (with probably different lengths) without losing the modules' structure. Using the value of 0 for edges not included in a module, we would be able to create equal-length edge lists:

$$\begin{aligned} M_m^{sp_k} &= [wt_{e_1}, wt_{e_2}, \ldots, 0, wt_{e_i}, \ldots, wt_{e_n}] \\ M_n^{sp_l} &= [wt_{e_1}, 0, \ldots, wt_{e_{i-1}}, wt_{e_i}, \ldots, wt_{e_n}] \end{aligned} \tag{3}$$

where $M_m^{sp_k}$ and $M_n^{sp_l}$ represents the subgroup m and n of complication subgroups in subpopulation sp_k and sp_l, respectively, and wt_{e_i} denotes the weight of edge e_i for both subpopulations. These edge lists can be used to measure the edge-lists similarities between different subpopulations. Here, we use cosine similarity to quantify the similarities between each pair of subgroups. These quantities can be used to create similarity matrices $SimMat^{sp_k sp_l}$ for each pair of subpopulations sp_k and sp_l. Finally, we implement hierarchical clustering over similarity matrices to cluster more similar complication subgroups between subpopulations. Different algorithms can be used for hierarchical clustering [21]. Also, different metrics can be used for measuring the distance or similarity between similarity vectors, such as Minkowski distance, Hamming distances, and cosine similarity. In this paper, we used Minkowski distance with a parameter of 2, which translates to Euclidean distance. We used algorithms discussed in [22], [23] and implemented in Python seaborn package [24].

Most dissimilar subgroups: The similarity matrices developed in the previous step can be used to identify the most dissimilar complication subgroups among subpopulations. Each row of the similarity matrix $SimMat^{sp_k sp_l}$ represents the similarities between one complication subgroup of sp_k and all the complication subgroups of sp_l. Denoting the set of six subpopulations by S, and $\bar{r}_m^{kl}$ the row-wise mean of similarities between subgroup m of subpopulation k and all subgroups of subpopulation l, we can find the most dissimilar subgroup Dis^{sp_k} of a subpopulation sp_k from all other subpopulations' subgroups using the following equation:

$$Dis^{sp_k} = \underset{m \in M}{\arg\min}(\sqrt{\sum(\bar{r}_m^{kl})^2}) \;\; k \in S, \forall l \in S \setminus k \tag{4}$$

The results of implementing the above methods on the dataset used in this study are provided in the next section.

III. Results

Based on the network development approach adopted, a network was created for each subpopulation of the study. The number of nodes, $|N|$, number of edges, $|E|$, and the density of these networks are shown in Table II. Figure 1 shows a boxplot related to the edges' weights for different subpopulations of the study.

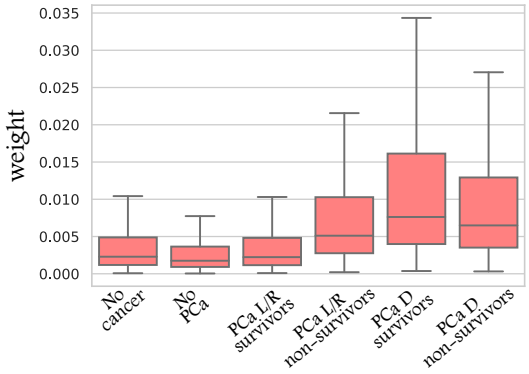

Fig. 1. The boxplot created from weights of edges in subpopulations' networks.

TABLE II
THE CHARACTERISTICS OF SUBPOPULATIONS' NETWORKS.

Subpopulation	$\|N\|$	$\|E\|$	Density
Patient with no history of cancer	10,044	1,451,797	0.029
Patient with no history of PCa	10,669	2,126,401	0.037
PCa stage L\R survivors	9,105	1,220,925	0.029
PCa stage L\R non-survivors	7,124	707462	0.028
PCa stage D survivors	5,886	347,086	0.020
PCa stage D non-survivors	6,439	530,138	0.026

The results of the implementation of the community detection algorithm on these six networks are shown in Table III. Because many of the subgroups (communities) identified were composed of isolated nodes, we decided to keep complication subgroups with more than 20 nodes. The last column of Table III shows the number of edges removed due to this decision, which in comparison with the total number of edges in all the subgroups, column $|E|$, is negligible.

In the next step, we used the complication subgroups edge-lists to compute the similarity between communities detected for different subpopulations. This step's results were 30 similarity matrices (each of six subpopulations can be compared with five other subpopulations). These similarity matrices were used for hierarchical clustering. Figure 2 shows the results of implementing the hierarchical clustering over the similarity matrices. Note the similarity heatmap provided for (sp_k, sp_l) is transposed for the heatmap provided for the (sp_l, sp_k). Considering that the number of subgroups might differ between each subpopulation pair, the matrices will not necessarily be square. Consequently, we are not able to find perfect diagonal matrices. However, in this figure, the more similar subpopulations have darker approximate diagonal entries.

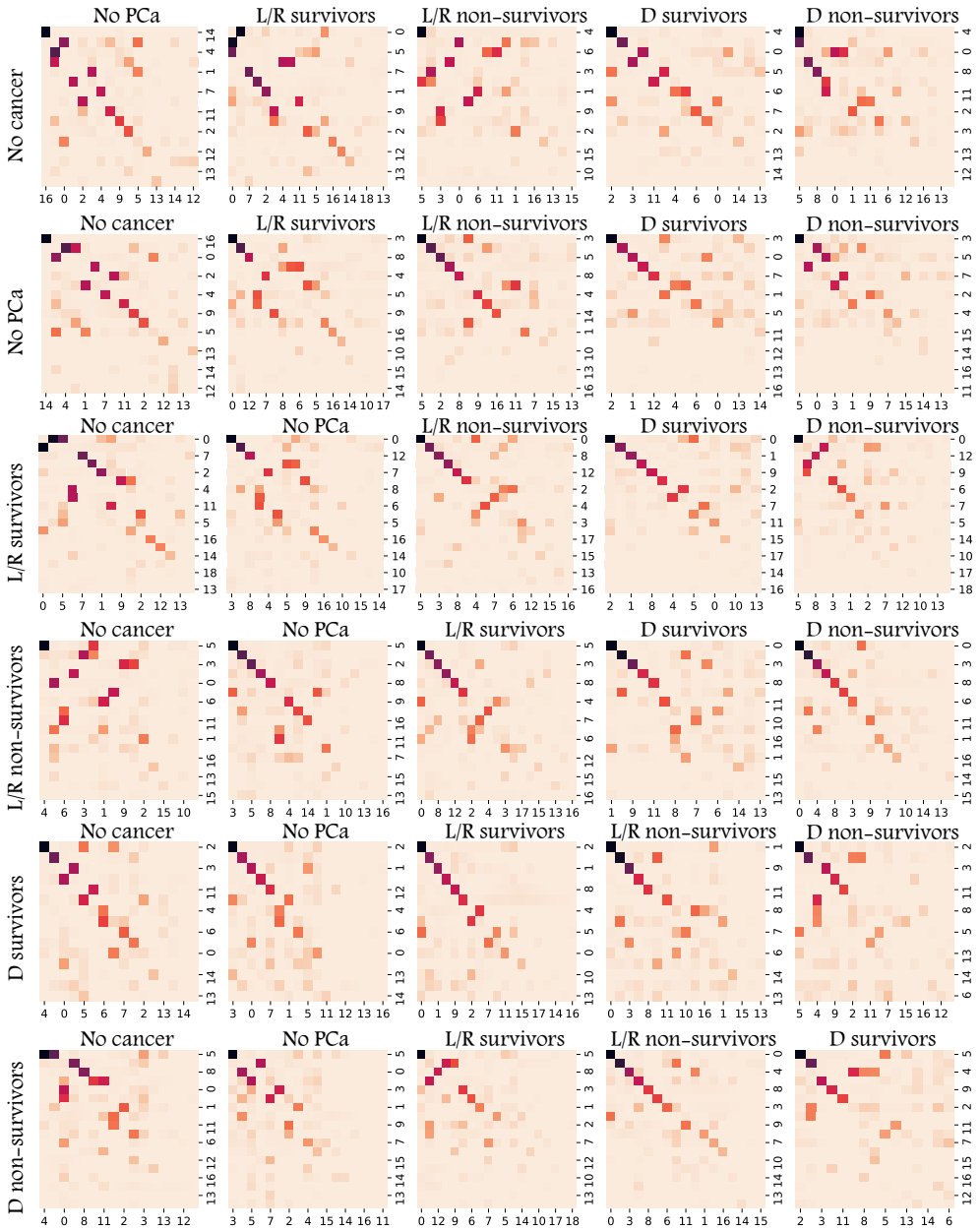

Fig. 2. The hierarchical clustering is implemented over similarity matrices for the six subpopulations. Each row includes the results for each subpopulation in comparison with all the other five subpopulations. The closer a heatmap to a diagonal matrix, the more similar are their corresponding subpopulations. For each pair of subpopulations, the darker the cells, the more similar are the related subgroups.

Finally, we used the similarity matrices to identify the most dissimilar subgroup in each subpopulation from all the other subgroups in other subpopulations. Figure 3 shows the most dissimilar subgroups identified for the subpopulations of patients with stage D PCa. In these networks, the nodes belong to the same chapter of ICD9 are shown with the same color. Also, the size of nodes is adjusted based on the nodes' degrees. The diseases that co-occur more frequently with other diseases would have larger nodes. In the next section, the findings of this study are discussed.

IV. Discussion

Table II shows that the densities of networks related to control subpopulations are higher than the networks of patients with more severe conditions of PCa. It implies that the absolute number of co-occurrences of diseases is more common in the control subpopulations. Although we identified fewer co-occurring complications for the patients with severe PCa conditions (Table II), the probability of experiencing co-occurring diseases is higher among these patients compared to the control patients (Figure 1). This observation also can

TABLE III

NUMBER OF COMPLICATION SUBGROUPS, TOTAL, AND WITH MORE THAN 20 NODES, DETECTED FOR SUBPOPULATIONS. THE LAST TWO COLUMNS SHOW THE TOTAL NUMBER OF EDGES IN THESE SUBGROUPS AND NUMBER OF EDGES REMOVED DUE TO REMOVAL OF SUBGROUPS WITH LESS THAN 20 NODES.

Subpopulation	$\|M\|$	$\|M^*\|$	$\|E\|$	$\|E^-\|$
Patient with no history of cancer	74	18	460,853	192
Patient with no history of PCa	76	17	529,055	239
PCa stage L\R survivors	53	16	396,082	199
PCa stage L\R non-survivors	34	17	178,085	45
PCa stage D survivors	30	15	85,991	97
PCa stage D non-survivors	40	18	146,860	111

$|M|$: number of subgroups detected in subpopulations' networks.
$|M^*|$: number of subgroups with more than 20 nodes.
$|E|$: total number of (inter-subgroup) edges kept from the original networks.
$|E^-|$: number of edges in subgroups with less than 20 nodes.

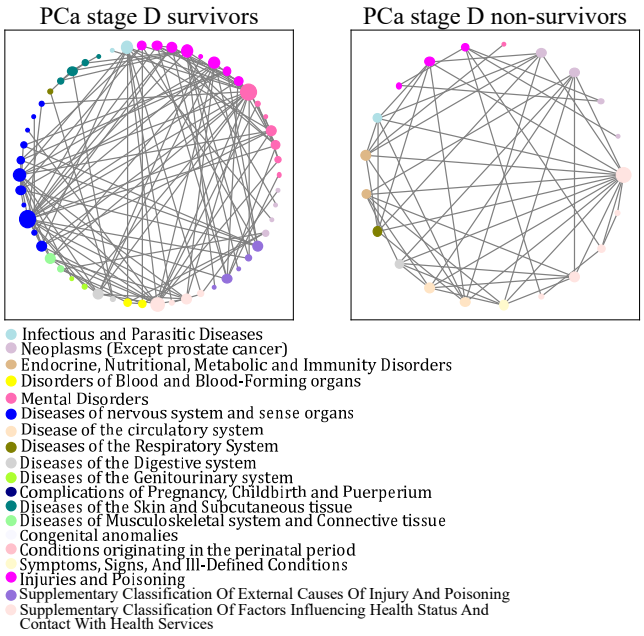

Fig. 3. The networks of most dissimilar subgroups in subpopulations of PCa stage D survivors and non-survivors.

be attributed to the unbalanced dataset of this study. Both control subpopulations and PCa stage L\R survivors are composed of many patients compared to the number of patients with stage D of PCa. Therefore, it is likely that some of the disease complications co-occur more frequently in the earlier subpopulations. However, Figure 1 shows that patients experiencing more severe conditions of PCa have experienced disease complications more concurrently.

Figure 2 provides the subpopulations similarities. The first row of this figure is related to the patients with no history of cancer. The heatmaps visualized in this row show that these patients' complication subgroups are not similar to any of the other subpopulations' heatmaps. These dissimilarities are more observable for non-survivor subpopulations versus patients with no history of cancer.

The complication subgroups identified for patients with no history of PCa show some similarities with the stage L\R non-survivors and stage D survivors. The patient with no history of PCa are patients with other types of cancer. These heatmaps show that these patients have experienced some of the co-occurring complications common in the stage L\R non-survivors and stage D survivors. However, stage D non-survivors have experienced co-occurring diseases relatively differently from patients with no history of PCa.

The heatmaps related to stage L\R survivors are more similar to the heatmaps of stage L\R non-survivors and stage D survivors. In other words, stage L\R survivors experienced some of the co-occurring complications common in stage L\R non-survivors. It is in line with our expectations since both subpopulations suffered the same stage of PCa. The stage L\R survivors also have some similarities with stage D survivors. The stage D-survivors can be considered the closest subpopulation to stage L\R survivors after stage L\R non-survivors. One reason could be the longer survival time, compared to non-survivor groups.

The heatmaps related to stage L\R non-survivors and stage D survivors show similarities with all other subpopulations except for the patients with no history of cancer. The second and third rows of Figure 2 show that the similarities between stage L\R survivors and patients with no history of PCa are relatively limited. These two observations suggest that stage L\R non-survivors and stage D survivors have experienced some of the co-occurring complications that both stage L\R non-survivors and patients with no history of PCa experienced. However, these co-occurring complications are not shared between stage L\R survivors and patients with no history of PCa.

The row related to the stage D non-survivors shows that these patients have experienced some of the co-occurring complications of stage L\R non-survivors and stage D survivors. Both subpopulations share a characteristic with stage D non-survivors.

The more in-depth examination of complication subgroups can inform the identification of distinct co-occurring patterns in different subpopulations. Figure 3 shows the most dissimilar subgroups identified for Stage D survivors and non-survivors. As seen, the density of the stage D survivors' network is higher than the network of the stage D non-survivors (with a larger number of complications). It can be attributed to the fact that the former has lived longer to experience more co-occurring complications. Also, this figure shows that different categories of co-occurring diseases are common in these two subpopulations. For example, multiple nodes represent disorders of blood and blood-forming organs (colored in yellow) and diseases of genitourinary systems (colored in light green) in stage D survivors. However, there are no representations of these disorders in the most dissimilar complication subgroups identified in stage D non-survivors. Similarly, we observed diseases related to endocrine, nutritional, metabolic, and immunity in stage D non-survivors, while these types of disorders are not identified in the most dissimilar complication subgroups of stage D

survivors. Taking the edge weights into account, dementia with Lewy bodies (LBD) and epistaxis had the highest weighted degrees in the most dissimilar complication subgroups detected for stage D survivors and non-survivors, respectively. The association between some of the treatment strategies with increased risk of dementia and recurrent epistaxis as a sign of metastatic prostate cancer have been reported earlier [25], [26]. This study's data-driven approach sheds light on some of the unknown or not very well-known associations among diseases in patients with PCa. A systematic approach to confirm these associations would be the next step.

V. Conclusion

In this study, we implemented network analysis for modeling disease-disease co-occurrences in patients with PCa. For PCa and control subpopulations, disease-disease networks were created from patients' Medicare insurance claims. Then, the results of community detection on these networks were used to measure the similarity among different subpopulations of PCa and control subpopulations. The patients with no cancer had the least similar complication subgroups with other subpopulations. The results showed that subgroups detected from survivor patients' networks at different stages of PCa and subgroups detected for non-survivor patients' networks at various stages of PCa are to some extent similar. Also, subgroups detected from the same stage (L\R and D) survivor and non-survivor patients' networks show similar patterns.

The distinct co-occurring complication patterns among different subpopulations can be used for prediction purposes or early detection of the initiation of different stages of PCa. The visualization and the network properties in this study can be useful in the explainability of the predictive models. It is the next step in our research direction. Also, these networks are limited to disease complications. In our future studies, we will focus on drug-drug and drug-disease interaction networks in patients with PCa.

Acknowledgment

This work was supported in part by the National Science Foundation under the Grant NSF-1741306, IIS-1650531, and Thomas Jefferson University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] P. Rawla, "Epidemiology of prostate cancer," *World Journal of Oncology*, vol. 10, no. 2, 2019. [Online]. Available: https://www.wjon.org/index.php/wjon/article/view/1191

[2] T. A. Skolarus, A. M. Wolf, N. L. Erb, D. D. Brooks, B. M. Rivers, W. Underwood III, A. L. Salner, M. J. Zelefsky, J. B. Aragon-Ching, S. F. Slovin, D. A. Wittmann, M. A. Hoyt, V. J. Sinibaldi, G. Chodak, M. L. Pratt-Chapman, and R. L. Cowens-Alvarado, "American cancer society prostate cancer survivorship care guidelines," *CA: A Cancer Journal for Clinicians*, vol. 64, no. 4, pp. 225–249, 2014.

[3] M. S. Litwin and H.-J. Tan, "The Diagnosis and Treatment of Prostate Cancer: A Review," *JAMA*, vol. 317, no. 24, pp. 2532–2542, 06 2017.

[4] "Ferlay j, ervik m, lam f, colombet m, mery l, piñeros m, znaor a, soerjomataram i, bray f (2020). global cancer observatory: Cancer today. lyon, france: International agency for research on cancer," https://gco.iarc.fr/today, accessed: 2020-03-03.

[5] J. F. Piccirillo, R. M. Tierney, I. Costas, L. Grove, and E. L. Spitznagel, Jr, "Prognostic Importance of Comorbidity in a Hospital-Based Cancer Registry," *JAMA*, vol. 291, no. 20, pp. 2441–2447, 05 2004.

[6] M. Extermann, "Measurement and impact of comorbidity in older cancer patients," *Critical reviews in oncology/hematology*, vol. 35, no. 3, pp. 181–200, 2000.

[7] W. A. Satariano, "Comorbidities and cancer," in *Cancer in the Elderly*. CRC Press, 2000, pp. 486–508.

[8] L. Yu and L. Gao, "Human pathway-based disease network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 4, pp. 1240–1249, 2019.

[9] M. A. Moni, H. Xu, and P. Liò, "CytoCom: a Cytoscape app to visualize, query and analyse disease comorbidity networks," *Bioinformatics*, vol. 31, no. 6, pp. 969–971, 11 2014.

[10] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabási, "Uncovering disease-disease relationships through the incomplete interactome," *Science*, vol. 347, no. 6224, 2015.

[11] G. Lu-Yao, N. Nikita, S. W. Keith, G. Nightingale, K. Gandhi, S. E. Hegarty, T. R. Rebbeck, A. Chapman, P. W. Kantoff, J. Cullen, L. Gomella, and W. K. Kelly, "Mortality and hospitalization risk following oral androgen signaling inhibitors among men with advanced prostate cancer by pre-existing cardiovascular comorbidities," *European Urology*, vol. 77, no. 2, pp. 158–166, 2020.

[12] J. W. Park, D. H. Koh, W. S. Jang, J. Y. Lee, K. S. Cho, W. S. Ham, K. H. Rha, W. H. Jung, S. J. Hong, and Y. D. Choi, "Age-adjusted charlson comorbidity index as a prognostic factor for radical prostatectomy outcomes of very high-risk prostate cancer patients," *PLOS ONE*, vol. 13, no. 6, pp. 1–11, 06 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0199365

[13] H. Kodama, S. Hatakeyama, M. Momota, K. Togashi, T. Hamaya, I. Hamano, N. Fujita, Y. Kojima, T. Okamoto, T. Yoneyama, H. Yamamoto, K. Yoshikawa, T. Yoneyama, Y. Hashimoto, and C. Ohyama, "Effect of frailty and comorbidity on surgical contraindication in patients with localized prostate cancer (frart-pc study)," *Urologic Oncology: Seminars and Original Investigations*, vol. 39, no. 3, pp. 191.e1–191.e8, 2021.

[14] S. M. Rice, J. L. Oliffe, M. T. Kelly, P. Cormie, S. Chambers, J. S. Ogrodniczuk, and D. Kealy, "Depression and prostate cancer: Examining comorbidity and male-specific symptoms," *American Journal of Men's Health*, vol. 12, no. 6, pp. 1864–1872, 2018, pMID: 29957106.

[15] K. L. Matthes, M. Limam, G. Pestoni, L. Held, D. Korol, and S. Rohrmann, "Impact of comorbidities at diagnosis on prostate cancer treatment and survival," *Journal of cancer research and clinical oncology*, vol. 144, no. 4, pp. 707–715, 2018.

[16] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[17] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, Feb 2004. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.69.026113

[18] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, p. 056131, Nov 2004.

[19] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, oct 2008.

[20] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx." [Online]. Available: https://www.osti.gov/biblio/960616

[21] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: an overview," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[22] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," 2011.

[23] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola, "Fast optimal leaf ordering for hierarchical clustering ," *Bioinformatics*, vol. 17, no. suppl_1, pp. S22–S29, 06 2001.

[24] M. Waskom and the seaborn development team, "mwaskom/seaborn," Sep. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.592845

[25] H. Lim, A. Agarwal, N. Agarwal, and J. Ward, "Recurrent epistaxis as a presenting sign of androgen-sensitive metastatic prostate cancer," *Singapore Medical Journal*, vol. 50, no. 5, p. e178, 2009.
[26] K. Nead, S. Sinha, and P. Nguyen, "Androgen deprivation therapy for prostate cancer and dementia risk: a systematic review and meta-analysis," *Prostate cancer and prostatic diseases*, vol. 20, no. 3, pp. 259–264, 2017.