Sains Malaysiana 55(1)(2026): 103-116
http://doi.org/10.17576/jsm-2026-5501-08

Topological Analysis of Age-Related Proteins in Protein-Protein Interaction Networks *via* Local Persistent Homology

(Analisis Topologi Protein Berkaitan Umur dalam Rangkaian Interaksi Protein-Protein melalui Homologi Berterusan Tempatan)

ABDUL SYUKOR HAZRAM, SAKHINAH ABU BAKAR* & FATIMAH ABDUL RAZAK

Department of Mathematical Science, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Received: 8 November 2024/Accepted: 25 December 2025

ABSTRACT

Ageing is a complex biological process that gradually alters cellular function and patterns of protein interaction. Standard network-based measures such as degree, betweenness and clustering coefficient are widely used in protein–protein interaction networks (PPINs), but these metrics may overlook subtle changes within local neighbourhoods. This study applies Local Persistent Homology (LPH) to characterise age-related differences in the local topology of PPINs, providing structural information that is not captured through global or node-level analyses. For each protein, a level 2 ego network is constructed and its H_0 and H_1 features are summarised using persistence diagrams (PDs). The Wasserstein distance between PDs from adult and elderly networks is then computed to quantify topological variation across age groups. The Wasserstein distance for each protein was compared with its degree, betweenness, and local clustering coefficient to examine how local topological structure relates to standard centrality measures. Proteins with many topological components tend to exhibit higher degree and betweenness but lower clustering, while proteins in simpler neighbourhoods show longer average persistence and more stable structural patterns. By integrating LPH results with gene-disease association data, 25 proteins with notable age-related topological differences are identified, including several associated with neurodegenerative diseases. Overall, LPH deepens the analysis of PPIN architecture by exposing subtle, age-linked structural patterns that remain undetected using network centralities.

Keywords: Ageing; local persistent homology; network centrality

ABSTRAK

Penuaan merupakan suatu proses biologi kompleks yang mengubah fungsi sel dan corak interaksi protein secara beransur-ansur. Pengukuran rangkaian sedia ada seperti pemusatan darjah, pengantaraan dan pekali gugusan tempatan digunakan dalam rangkaian interaksi protein–protein (RIPP), namun metrik ini mungkin tidak mampu menangkap perubahan halus yang berlaku dalam kejiranan tempatan. Kajian ini menggunakan Homologi Gigih Tempatan (HGT) untuk mencirikan perbezaan berkaitan usia dalam topologi tempatan RIPP, sekali gus menyediakan maklumat struktur yang tidak dapat ditangkap melalui analisis peringkat global atau nod. Bagi setiap protein, rangkaian ego aras 2 dibina dan ciri H_0 serta H_1 diringkaskan melalui rajah gigih (PD). Jarak Wasserstein antara PD bagi rangkaian dewasa dan warga emas kemudiannya dikira untuk mengukur variasi topologi merentas kumpulan umur. Nilai jarak Wasserstein bagi setiap protein dibandingkan dengan pemusatan darjah, pengantaraan dan pekali gugusan tempatan untuk menilai hubungan antara struktur topologi tempatan dan pengukuran rangkaian tempatan. Protein dengan komponen topologi yang tinggi cenderung mempunyai nilai pemusatan darjah dan pengantaraan yang lebih tinggi tetapi pekali gugusan yang lebih rendah, manakala protein dalam kejiranan yang lebih ringkas menunjukkan purata jangka hayat yang lebih panjang dan struktur yang lebih stabil. Dengan menggabungkan hasil HGT bersama data hubungan gen-penyakit, sebanyak 25 protein dikenal pasti menunjukkan perbezaan topologi berkaitan usia yang ketara, termasuk beberapa yang berkaitan dengan penyakit neurodegeneratif. Secara keseluruhannya, HGT memperkukuh analisis struktur RIPP dengan mendedahkan pola halus yang berkait dengan usia, yang tidak dapat dikesan menggunakan pemusatan rangkaian.

Kata kunci: Homologi gigih tempatan; pemusatan rangkaian; penuaan

INTRODUCTION

Ageing is recognised as a complex biological process that involves a gradual decline in physiological function and an increased susceptibility to age-related diseases. These changes arise through a range of molecular, cellular and systemic alterations that occur throughout the body (Liu et al. 2019; Ni et al. 2022). At the cellular level, ageing influences protein behaviour and disrupts normal biological activity. The process is also shaped by genetic, environmental and lifestyle factors, which reflects its multifactorial nature (Partridge, Deelen & Slagboom 2018). Understanding the molecular mechanisms involved in ageing is therefore important for supporting healthy ageing and reducing disease risk.

Protein–protein interaction networks (PPINs) provide a useful representation of how proteins interact within the cell. By describing proteins as nodes and interactions as edges, PPINs make it possible to examine signalling pathways, regulatory processes and functional modules (Mooney, Morgan & McAuley 2016). When PPINs are studied across different age groups, they may show key proteins and pathways involved in age-associated functional decline or disease development (Calabrese, Molzahn & Mayor 2022).

Several commonly used analytical strategies have been applied to identify age-related proteins in PPINs. These include network-centric strategies such as functional enrichment and standard centrality measures, and standard network centrality measures such as degree, betweenness and closeness (Ashtiani et al. 2018; Faisal, Zhao & Milenkovic 2015). Such measures quantify different aspects of protein importance, including connectivity, communication flow and proximity to other nodes. They have been shown to be helpful for detecting influential proteins associated with biological ageing (Faisal & Milenković 2014; Syukor & Sakhinah 2019).

Changes in network connectivity and structure also contribute to the identification of age-related proteins. As protein interactions can shift over time, these changes may be reflected through differences in network topology, which includes the arrangement of nodes, edges, local clusters and small-scale structural features (Faisal & Milenković 2014; Teulière et al. 2023). Although node-level centrality analysis has been widely used to rank proteins according to their connectivity or clustering properties (Kosch & Schreiber 2004), these metrics do not fully capture neighbourhood-level structural variation. Some centrality measures may also be sensitive to small perturbations in the network and may implicitly assume uniform influence across nodes (Aktas, Akbas & El Fatmaoui 2019). These limitations highlight the need to incorporate additional structural information when analysing PPINs.

Topological data analysis (TDA) has emerged as a useful framework for describing structural patterns in complex data. One of its main tools, Persistent Homology (PH), captures topological features such as connected components and loop-like structures across multiple scales (Aktas, Akbas & El Fatmaoui 2019). PH has been successfully applied to biological networks, including brain networks, PPINs and genetic interaction studies (Hazram, Bakar & Razak 2024; Ignacio & Darcy 2019; Li et al. 2021; Song 2023). Its ability to identify subtle structural patterns makes it well-suited for the analysis of biological networks.

In this study, PH is applied in a local context through Local Persistent Homology (LPH). Instead of examining the global PPIN, LPH focuses on neighbourhood subnetworks to characterise the local topology surrounding each protein. Both H_0 (connected components) and H_1 (loop-like structures) are extracted from these neighbourhoods, and persistence diagrams (PDs) are used to represent their topological characteristics. Age-related differences are then quantified using the Wasserstein distance to measure the dissimilarity between PDs obtained from adult and elderly PPINs.

Overall, this study aims to characterise age-dependent changes in local PPIN topology and identify proteins exhibiting pronounced structural variation across age groups. By integrating LPH-derived information with standard centrality measures and gene–disease associations, this approach offers a complementary perspective for understanding ageing-related molecular changes.

PERSISTENT HOMOLOGY OF BIOLOGICAL NETWORKS

Persistent Homology (PH) is used to capture intrinsic geometric and structural properties of data by quantifying the persistence of topological features across multiple spatial scales. In biological network analysis, PH offers several advantages over standard network analysis methods. It enables complex network structures to be examined in a consistent manner and allows topological features to be detected even when they are not easily observed through raw network representations (Masoomy et al. 2021). Specifically, PH facilitates the identification of robust and biologically meaningful patterns, such as strongly connected clusters and stable structural motifs within biological networks (Islambekov & Gel 2019).

An additional advantage of PH is that it is largely unaffected by specific choices of network representation or parameter settings. This allows biological networks to be analysed in a more objective manner, with results derived from topological properties rather than assumptions about node influence or network density.

The PH workflow generally consists of three main components: data preparation, construction of a filtration, and extraction of topological features. The complete process used in this study is described in detail in the Materials and Methods section. Understanding how ageing affects PPIN topology is important, as changes in protein interactions have been linked to altered physiological functions and increased risk of age-related diseases (Faisal & Milenković

2014; Faisal et al. 2015; Liu et al. 2019). Investigating these topological changes across different age groups may therefore provide useful insight for disease detection and drug targeting (Aktas, Akbas & El Fatmaoui 2019).

In this study, a computational framework based on PH is developed to analyse PPINs and identify proteins that exhibit notable age-related topological variation. These proteins are subsequently characterised based on their potential biological relevance and disease associations.

MATERIALS AND METHODS

The whole process of this study, from data collection to analysis, is depicted in Figure 1.

DATA PREPARATION AND PPIN FORMATION

The protein list used in this study was obtained from Berchtold et al. (2008), consisting of wet-lab samples from 55 individuals. A total of 172 probe arrays were available for protein expression, and these were categorised according to gender and age group. The age groups were defined as follows: a) The adult age group consists of individuals between the ages of 20 to 69 and b) Individuals between the ages of 70 and 99 are considered elderly.

The list of interacting proteins, which serves as the set of network edges, was retrieved from the IntAct database (del Toro et al. 2022). IntAct contains detailed information on protein–protein interactions, including interaction type, supporting references and a Mutual Information (MI) score. The MI score ranges from 0 to 1 and reflects interaction confidence by incorporating the frequency of observed interactions, level of supporting evidence and experimental methodology. Higher values indicate greater experimental confidence and reproducibility. A total of 81,847 human protein–protein interactions with MI scores of 0.49 or higher were included in this study, as higher MI values indicate greater reliability and biological relevance (Sugis & Hermajakob 2019).

Four PPINs were constructed by merging the expressed protein lists with the interaction data, resulting in undirected and weighted networks, with MI scores assigned as edge weights. Only the largest connected component of each network was considered for further analysis, as proteins within this component are more likely to be functionally relevant and integrated within essential biological modules (Bhowmick & Seah 2016). These components accounted for approximately 86 percent of the full network. Small disconnected nodes, most of which appeared as isolated singletons, were excluded because they lack neighbourhood structure, which prevents meaningful local topological analysis.

The networks were constructed based on gender and age group, producing four categories: adult male, adult female, elderly male and elderly female. This stratification

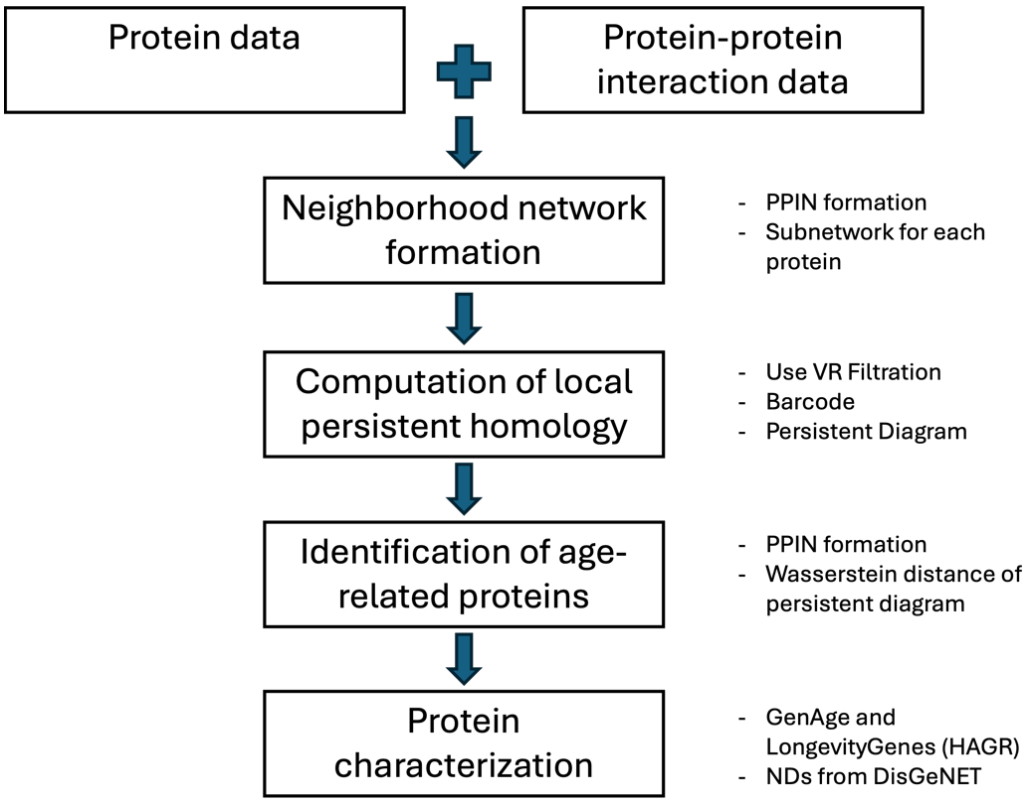

FIGURE 1. Workflow of the study

allows network characteristics to be examined within specific demographic subsets. Before analysing local neighbourhood structure, global network properties were evaluated to establish the overall organisation of each PPIN.

Several network-level metrics were computed, including the number of nodes (proteins), number of edges (interactions), network diameter, density, number of triangles, clustering coefficient and average edge weight. The diameter represents the longest shortest path between any two nodes, while density measures the proportion of possible edges that are present in the network.

Local structural features were assessed using metrics such as the number of triangles and the clustering coefficient, which describe the extent to which neighbouring nodes form interconnected clusters. The global clustering coefficient (GCC) was computed using the formula (Aguilar-Alarcón, Hernández-Gómez & Romero-Valencia 2023).

$$GCC = \frac{Number\ of\ triangles * 3}{Number\ of\ connected\ triples\ of\ vertices}$$

Average edge weight was also examined to indicate the overall strength of protein–protein interactions. Together, these metrics provide a comprehensive description of network behaviour across the four demographic groups.

NEIGHBOURHOOD NETWORK FORMATION

Local Persistent Homology (LPH) is defined as the computation of PH within a localised region or neighbourhood (Fasy & Wang 2016). The original computation is based on the point cloud data, and hence, H_0 and H_1 were extracted around the fixed proximity of every single point in the point cloud data. However, in a network, the LPH can still be computed, but in this context, the neighbourhood of the network can be defined around a fixed number of step(s) from every single node.

In this study, for every node in every network G, we constructed a level 2 ego network. The level 2 ego network is a subnetwork of G such that the node $v \in V$ is the centre of the subnetwork along with its immediate neighbours and the second neighbours from v. From now on, the neighbourhood in this study is referred to as a level 2 ego network of node v. The example of the neighbourhood of a node is depicted in Figure 2.

COMPUTATION OF LOCAL PERSISTENT HOMOLOGY AND FEATURE EXTRACTION

Homology refers to the topological characteristics of a given space. In the context of a topological space X, the homology groups H_0, H_1, and H_2 represent the components, holes, and voids of X, respectively. The construction of homology groups starts by considering a chain complex $C(X)$ that represents information about X. This chain complex consists of a sequence of Abelian groups $C_0(X)$, $C_1(X)$, $C_2(X)$,... connected by homomorphisms called boundary operators $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$. The k-th homology group $H_k(X)$ is defined as the kernel of the boundary operator ∂_k quotient by the image of the boundary operator ∂_{k-1}. Our main focus is on relative homology groups $H_k(X, A)$ (where $A \subset X$. These groups are defined using the same formula but with boundary maps on the quotient spaces $C_k(X)/C_k(A) \rightarrow C_{k-1}(X)/C_{k-1}(A)$.

The shape of X can be more accurately described using PH, which is a concept that incorporates multiple scales of homology. PH is computed through a sequence of nested spaces connected by inclusions, known as a filtration. That is, we can regard the finite sequence $\emptyset = X_0 \subseteq X_1 \subseteq \cdots \subseteq X_n = X$. On the other hand, LPH focuses on the local structure of the data. The k-th local homology group of X at a point $x_0 \in X$ is defined as the relative homology group $H_k(X, X - x_0)$. Alternatively, it can be defined as the limit of the homology of X with respect to all elements except a gradually decreasing neighbourhood around x, represented by the expression $\lim_{r \to 0} H_k(X, X \backslash U_r)$, where Ur is a neighbourhood of x_0 with a radius of r (Fasy & Wang 2016).

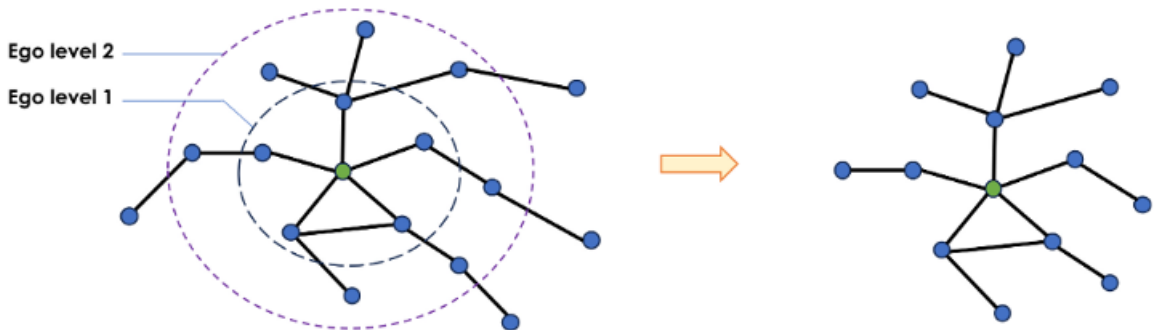

FIGURE 2. Example of level 2 ego network of a protein. The ego protein acted as the centre of the neighbourhood, marked by the green node

From a network perspective, these definitions have been revised using network terminologies. In this study, the variable X has been replaced by a graph object, specifically a weighted network denoted as $G = (V, E, W)$ such that V is a set of nodes, E is a set of edges with W is set of non-negative real values assigned on every single edge. The neighbourhood graph is defined as the induced subgraph consisting of all nodes located within a fixed number of graph-theoretic steps from the chosen node. The detailed process of computing and extracting topological features of the neighbourhood network is described in Figure 3.

For every single neighbourhood network, we used the Laplacian approach to convert the adjacency matrix into a distance matrix before embedding it into metric space. This is because the weights in the adjacency matrix are based on the MI score. The MI score does not satisfy metric properties such as the triangle inequality, so the adjacency matrix must be converted into a metric distance matrix before constructing the filtration. Prior to computing the LPH, we used Commute-Time Distance (CTD) to represent the neighbourhood adjacency matrix of the node. The formula for CTD for any pair of nodes x and y is given by:

$$CTD(x,y) = \left[\sum_{i=1}^{|V|-1} \frac{1}{\lambda_i}\left(\phi_i(x) - \phi_i(y)\right)^2\right]^{1/2}$$

such that $\{\lambda_i\}_{i=0}^{|V|-1}$ is the generalised eigenvalues and $\{\phi_i\}_{i=0}^{|V|-1}$ is the generalised eigenvector obtained from the Laplacian matrix of the PPIN, and V is the number of proteins in the PPIN (Hajij et al. 2018). Every pair of CTD distances of two proteins is recorded and represented in a distance matrix used to compute PH.

As mentioned, filtration is a process of extracting topological components from network data. Vietoris-Rips (VR) filtration is suitable for undirected and weighted networks and is defined as follows (Aktas, Akbas & El Fatmaoui 2019):

Definition 1. Consider an undirected and unweighted graph $G = (V, E)$ with the weight function $W: V \times V \to \mathbb{R}$ defined on E, and $G_\delta = (V_\delta, E_\delta) \subset G$ is a subgraph of G with $V_\delta = V$ and $E_\delta \in E$ containing edges with weight less than or equal to δ for any $\delta \in \mathbb{R}$. For any $\delta \in \mathbb{R}$, the VR complex is the simplicial complex of G_δ, Cl_δ, and the filtration is defined as $\{Cl(G_\delta) \to Cl(G_{\delta'})\}_{0 \le \delta \le \delta}$.

In other words, this filtration begins with the vertex set. Next, we rank edge weights from w_{min} to w_{max}, and we let the parameter δ increase from w_{min} to w_{max}. Each step involves adding edges and reforming the simplicial complex of the thresholded subgraph G_δ. This construction produces VR filtration in the network.

This study focuses on using LPH to identify up to 1-D topological components in a given dataset, which corresponds to the computation of H_0 and H_1 only. The 0-dimensional topological components are commonly referred to as connected components. On the other hand, 1-D topological components refer to any loop-like or hole-like structures that exist in the network, and 2-D topological components are represented by the void structure of the network. Usually, 2-D topological components are omitted as it is difficult to obtain this structure unless the network is high in density (Aktas, Akbas & El Fatmaoui 2019). Figure 4 displays several examples of topological components.

As the filtration process progresses, each H_0 and H_1 topological component that is captured will be returned as a barcode. This barcode represents the lines, and the length of each line corresponds to the persistency of the topological component. The lifetime of topological components in a network can be calculated by subtracting the birth time from the death time of each component. Besides, a Persistent Diagram (PD) is a graphical representation that illustrates the birth and death of topological components. Although both barcodes and persistence diagrams contain

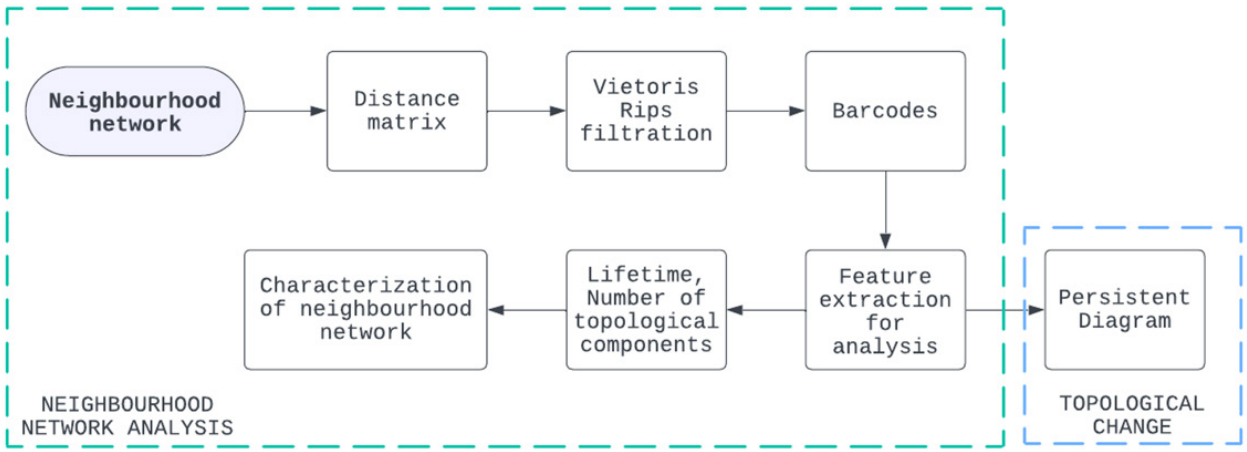

FIGURE 3. Flow of computing the LPH and extracting topological features

the same information, this study uses persistence diagrams exclusively for all subsequent analyses. Figure 5 illustrates the example of a simple network along with the VR process with barcodes and PD representation.

In this study, we will initially analyse the neighbourhood network to gain a deeper understanding of its properties. The lifetime of topological components is one of the first characteristics of LPH that are needed for the analysis. Subsequently, we use a PD to identify the proteins associated with ageing.

LOCAL TOPOLOGICAL CHANGE

The identification of age-related proteins will be conducted separately based on gender. We generated the neighbourhood networks for every protein in adult and elderly PPINs. For every protein, we obtained the PD resulting from both adult and elderly PPINs for computation. Given two persistence diagrams X and Y obtained from the neighbourhood network of the protein in both adult and elderly PPIN, respectively, we can calculate the similarity (or dissimilarity) between the topological components using the Wasserstein distance. The formula for Wasserstein distance, W_q is as follows:

$$W_q(X,Y) = \left[\inf_{\eta: X \to Y} \sum_{x \in X} ||x - \eta(x)||_\infty^q \right]^{1/q}$$

such that η is the bijection between points in the diagrams. In this study, the value of q is 2 (Hajij et al. 2018). In other words, the Wasserstein distance quantifies persistence diagram similarity by considering the total distance between the matched pair of points.

Subsequently, age-related proteins were identified by selecting proteins whose Wasserstein distances fall within the 90th percentile of the distribution. Proteins in this range are considered to exhibit the largest topological differences between adult and elderly PPINs. Since there is no established threshold for determining when a topological change becomes biologically significant, the 90th percentile was used as a practical and systematic cutoff to capture proteins with notably high variation.

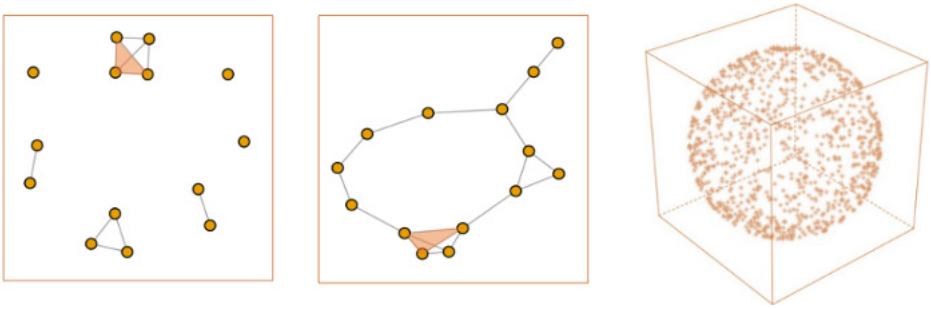

FIGURE 4. Example of topological components. From left: 0-D (connected components), 1-D (loop-like structure), and 2-D (void) topological components

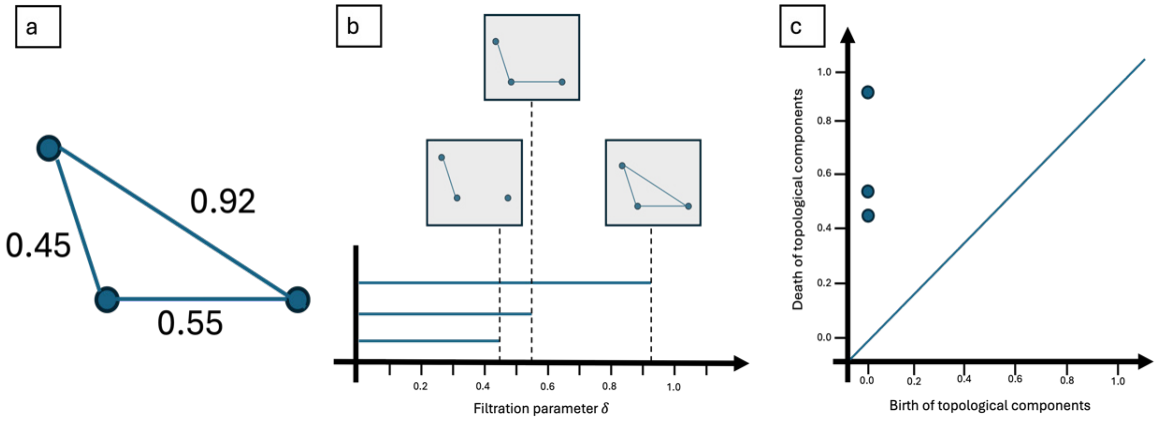

FIGURE 5. Example of (a) network with weight, (b) VR filtration process on the network, and (c) PD, the graphical representation of barcodes, with birth filtration value as the x-axis and death filtration value as the y-axis

For example, if a network contains 1,200 proteins, the Wasserstein distances are first arranged in ascending order. The top 90th percentile of these values, corresponding to the highest 120 proteins, are then selected. Proteins within this group are regarded as having substantial differences in their neighbourhood topology across age groups and are prioritised for further biological interpretation.

AGE-RELATED PROTEIN CHARACTERIZATION AND VALIDATIONS

After potential age-related proteins had been identified using LPH, the next step involved verifying their essentiality in relation to ageing. The list of proteins associated with ageing was retrieved from the Human Ageing Genomics Resource (HAGR) (De Magalhães & Toussaint 2004). More than 800 age-related proteins were identified, including proteins linked to longevity, cellular senescence and ageing-related genetic pathways (Budovsky et al. 2013; de Magalhães et al. 2009; Tacutu et al. 2018, 2013).

In addition, information on neurodegenerative diseases was obtained from the Disease Ontology database. The analysis focused on the relationship between proteins with high local topological differences between age groups, as determined through LPH, and several major neurodegenerative conditions. These include Huntington's disease, Parkinson's disease, Alzheimer's disease and dementia (Piñero et al. 2020).

RESULTS

This section will elaborate on the properties of all four PPINs to get the basic ideas of the overall network structure before further discussion on the local network topology. The properties of the network are included in Table 1.

In comparison, the number of nodes and edges decreased between age classes for male PPIN while growing for female PPIN. This led to a decrease in network density for male PPINs. Even when the number of nodes and edges for female PPIN grows, the density of PPIN across age groups decreases. This suggests that PPIN organisation undergoes measurable structural shifts with age (Corriveau-Lecavalier et al. 2023).

The clustering coefficient is another observable property. Across age groups, the Global Clustering Coefficient (GCC) values for both male and female PPINs are decreasing. There are several causes contributing to the drop in clustering coefficient values, as stated herewith: (a) Decrease local connectivity. The clustering coefficient is calculated primarily using the number of triangles. The decrease in local connectivity in the network can be attributed to the nodes' connections, which create fewer triangles or closed loops, and (b) Modularity changes. The clustering coefficient is closely related to network modularity, meaning the network's degree can be separated into distinct modules or communities. As a result, the elderly PPIN's modular partitioning or network structure is less modular than adults. This explanation is also consistent with the proportion of triangles in the network between adult and elderly PPINs (Erciyes 2023). a) Network evolution and reconfiguration. Networks, especially biological networks, are dynamic entities. Thus, the PPIN's transformation has been demonstrated to change over time, as evidenced by the number of nodes and edges in both male and female PPINs. The number of proteins expressed is not the only element that influences evolution. The number and weight of edges are also important considerations, which may result in a drop in clustering coefficient value due to an alteration in protein interaction (Teulière et al. 2023).

While global network analysis provides an overview of the overall PPIN, node-level analysis is required to understand the precise roles and functions of individual proteins within the network. Both approaches are complementary and contribute to a better understanding cellular processes and disease molecular mechanisms. Based on global network change across age groups, we intend to study the local H_0 and H_1 topological features of each protein expressed in the network.

Identifying the proteins that contribute to changes in the network topology may provide additional characterisation of the protein. By introducing LPH into the study, the local ranking value is based on the subnetwork's topological qualities, which include information about direct connections, the number of triangles and closed loops, as well as the edge attributes.

TABLE 1. Characteristics of all the PPIN according to gender and age class

Properties	Male adult	Male elderly	Female adult	Female elderly
Nodes	2920	2914	2869	2292
Edges	6790	6686	6622	6714
Number of triangles	3324	3165	3144	3180
Diameter	11.11000	11.11000	11.11000	11.11000
Density	0.001593	0.001575	0.001610	0.001573
Global Clustering Coefficient	0.026472	0.025862	0.026104	0.026035
Average edge weight	3.754932	3.713802	3.737985	3.714114

LOCAL NETWORK CHARACTERIZATION

To characterise the structural dynamics surrounding each protein, it is important to learn more about the LPH to illuminate the subtle yet profound topological changes occurring within localised regions. Therefore, this section will elaborate on the persistence of the topological components for every protein. Since our approach is novel, evaluating its relevance in relation to centrality methods is crucial. Through a comparative analysis with existing centrality measures, our objective is to show how LPH analysis may be related to existing network centralities. Several network centralities involved are:

- Degree centralities are measured by the number of edges a protein has with its immediate neighbours.
- Betweenness centralities, measures of the frequency of a protein being a mediator in pathways of other pairs of proteins, using the following formula (Brandes 2001):

$$BC(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

such that $\sigma_{st}(v)/ \sigma st$ is the proportion of the shortest path between points s and t through v and the overall shortest path from s to t.

- The local clustering coefficient, used to determine the tendency of a protein to form a clique and calculated by dividing the number of edges between a protein's neighbour by the number of edges that could possibly exist, in which calculated using the following formula:

$$LCC(v) = \frac{2E_v}{k_v(k_v - 1)}$$

such that $LCC(v)$ is the local clustering coefficient of node v, E_v is the number of edges in between the immediate neighbours of v, and k_v is the degree of node v (Aguilar-Alarcón, Hernández-Gómez & Romero-Valencia 2023).

The features extracted from LPH consist of the number of topological components and the persistence of these components, which is represented by their lifetime. The term lifetime refers to the difference between the birth and death values of each topological feature. In this study, the topological components considered are H_0 (connected components) and H_1 (loop-like structures). The barcode serves as a visual representation of these lifetimes, while the numerical features used in the analysis are obtained from the lifetimes themselves. These numerical features include the average lifetime, maximum lifetime and the total number of topological components. The average lifetime represents the mean duration of all H_0 and H_1 features, whereas the maximum lifetime corresponds to the longest persisting feature in the neighbourhood. The number of topological components is determined by counting all bars in the barcode, which correspond to the total number of H_0 and H_1 features extracted from the filtration. Figure 6 displays the correlation coefficients among the variables, illustrating the relationships among different local characteristics.

The correlation analysis for every subnetwork in all PPINs shows a weak negative association between maximum and average persistence. This indicates that subnetworks with a longer maximum lifetime tend to have a shorter average lifetime. While a subnetwork may contain a highly persistent topological feature, it may also include many features with shorter persistence values. A moderately negative correlation was also observed between the average lifetime and the number of topological components, suggesting that subnetworks with fewer components tend to have longer average lifetimes. In contrast, the correlation between the maximum lifetime and the number of components was moderately positive, indicating that subnetworks with more components are more likely to contain at least one long-lasting feature.

Overall, the local topology of each protein indicated that proteins with high local connectivity tended to possess highly persistent topological features and greater variability in their lifetimes. From the PPIN perspective, these patterns are influenced by both node connectivity and the edge weights derived from the MI score. A persistent feature generally reflects strong or well-supported interactions, suggesting that proteins with greater local structural richness are more biologically relevant.

PH also offers a useful node-level characterisation that complements standard centrality measures such as degree, betweenness and the local clustering coefficient. The persistence of topological components captures structural prominence within the network. Figures 7 and 8 show the relationship between the number of topological components and the degree or betweenness of each protein. Each point in the plot represents the level 2 ego network of a single protein and the horizontal axis reflects the centrality value of the ego protein, while the vertical axis captures the corresponding topological feature count.

Moreover, the local clustering coefficient, which measures the level of clustering or connectivity between a protein's neighbouring nodes, can also be deduced from the persistence of H_0 and H_1 topological components. Proteins with a high average persistent of these components may exhibit higher local clustering coefficients, indicating their participation in closely connected subnetworks or functional modules within the PPIN. Figure 9 indicates that proteins with a high average persistency tend to have a low number of topological components. It can be demonstrated that proteins with a low number of topological components generally exhibit a higher LCC in comparison to proteins with a high number of topological components, as depicted in Figure 9.

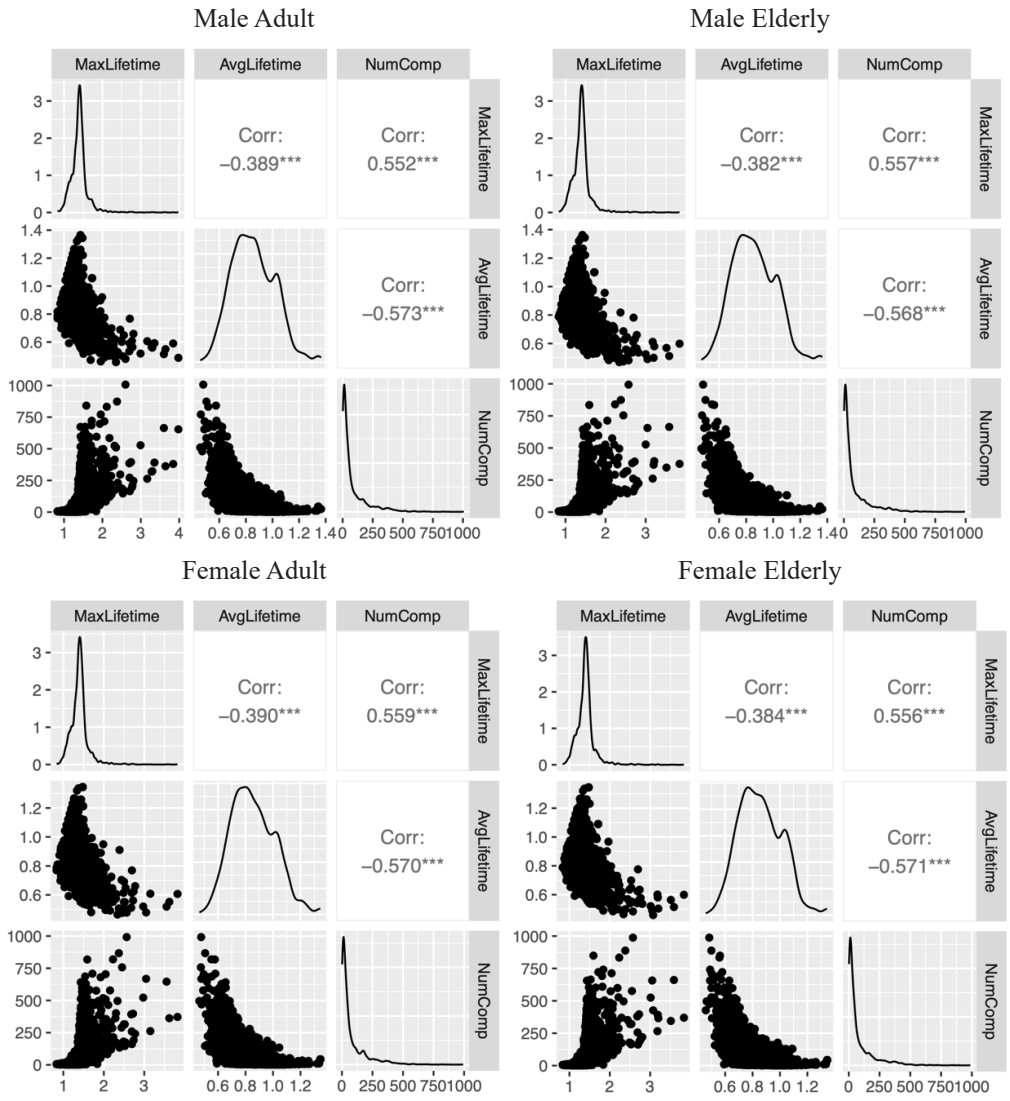

FIGURE 6. Correlogram of the topological properties for every subnetwork in all four PPINs

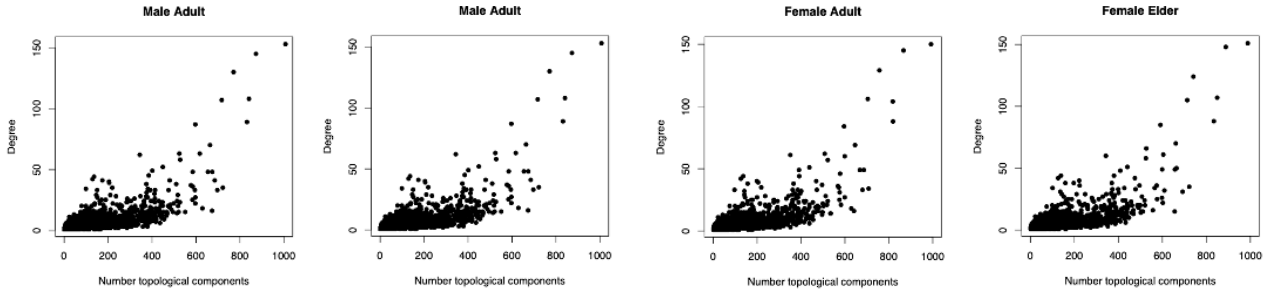

FIGURE 7. Scatter plot representing the number of topological components in the ego network against the degree of the ego protein

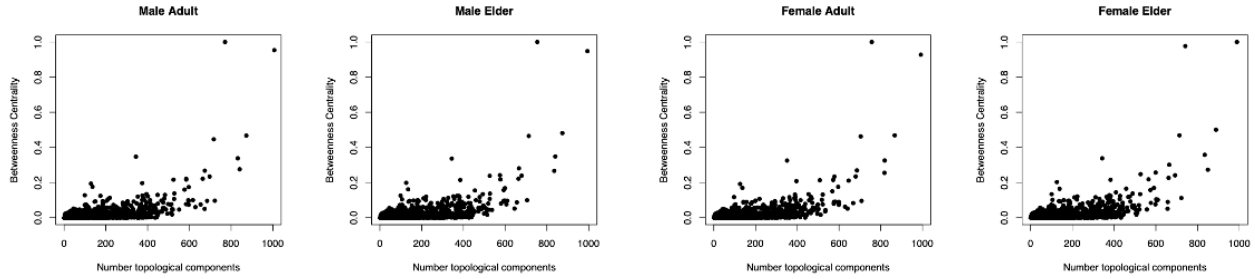

FIGURE 8. Scatter plot representing the number of topological components in the ego network against the betweenness of the ego protein

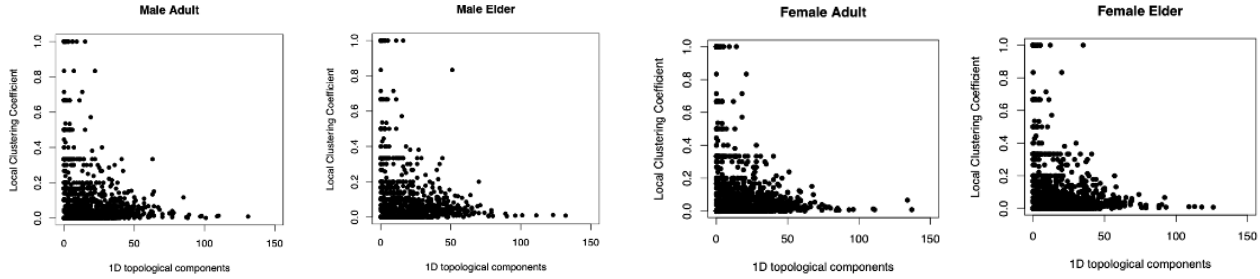

FIGURE 9. Scatter plot representing the number value of the local clustering coefficient of the proteins against the number of 1-D topological components in the neighbourhood

Overall, every subnetwork in the PPINs lies in the spectrum based on the number of topological components, where the high number of topological components may potentially have a high degree of centrality, high betweenness as well as low clustering coefficient. A subnetwork with a low number of topological components, on the other hand, may have a low degree, low betweenness centrality, and high topological persistency.

These results highlight the usefulness of PH in capturing important information about local network topology. This leads to a thorough understanding of PPINs and their biological importance in terms of identifying the dynamic change of local topology. The study's goal is to identify any topological changes via Wasserstein distance that may be associated with age-related properties.

LOCAL TOPOLOGICAL CHANGE OF PROTEINS PPINS AND DISEASE ASSOCIATION

For each protein expressed in the male and female PPINs, topological information was extracted from its persistence diagram (PD). The PD obtained from the adult PPIN was compared with the PD obtained from the elderly PPIN for the same protein. The Wasserstein distance was used to measure the dissimilarity between the two PDs. In cases where a protein was not present in either the adult or the elderly PPIN, one of the PDs was empty. When this occurred, the empty PD was compared to the diagonal by matching each point in the existing PD to its projection on the diagonal. For example, a feature with birth 0.3 and death 0.8 would be matched to the diagonal point (0.3, 0.3) (Cohen-Steiner, Edelsbrunner & Harer 2007; Dey, Shi & Wang 2015).

Once the Wasserstein distance for each protein had been obtained, the protein names were matched to entries in the DisGeNET disease ontology database (de Magalhães et al. 2009). Each protein was then assigned the corresponding Gene Disease Association (GDA) values provided by DisGeNET. These values represent the level of evidence supporting the association between a gene and a disease.

The analysis focused on four neurodegenerative diseases, namely Alzheimer's disease, Parkinson's disease, Huntington's disease and dementia. The GDA score for each protein was retrieved for all four diseases. Figure 10 illustrates the relationship between the Wasserstein distance and the GDA score for these neurodegenerative diseases.

We expect that proteins exhibiting substantial variations in their structure will be strongly correlated to Neurological Diseases (NDs) in both males and females. Nevertheless, proteins with high GDA scores do not always demonstrate correspondingly high Wasserstein distances. Thus, there is a possibility that local connectivity may vary or remain constant with age. This also implies that several proteins associated with ND have a topological structure in the subnetwork that remains relatively stable across different age groups.

However, it is crucial to consider external variables when examining this finding. For example, while the general topological structure may remain the same, there may still be changes in other factors, such as protein expression levels or post-translational modifications, that contribute to the progression of disease (Dong et al. 2023; Hwang, Lee & Kho 2022; Kurtishi et al. 2019). In addition, a thorough study is needed to determine why there are no big changes in the structural differences between protein networks in young and older people linked to age-related diseases.

By focusing on the 90th percentile of proteins with the most pronounced topological changes, we can identify potential candidates that may play important roles in age-related biological processes. To support this approach, information from the GenAge and LongevityMap databases in the HAGR resource was used as external evidence (de Magalhães et al. 2009). The analysis initially aimed to detect proteins exhibiting substantial topological shifts between age groups, with the Wasserstein distance used to quantify the degree of dissimilarity in local network structure.

For each gender, the Wasserstein distances between the adult and elderly PPINs were computed for all proteins expressed in either age group. The 90th percentile of these distances was then determined separately for males and females. Proteins with Wasserstein distance values at or above their respective percentile thresholds were selected as candidates with the largest age-related topological differences. After combining the male and female candidate sets and removing duplicates, a total of 25 unique proteins were obtained. These proteins, summarised in Table 2, represent those with the most pronounced changes in local PPIN topology across age groups.

Some proteins remain highly stable in their topological features across a wide range of age groups, which contrasts with the general expectation that protein–protein interactions change with age. This behaviour, in which certain proteins maintain consistent H_0 and H_1 features derived from persistent homology, suggests a degree of functional resilience. For example, proteins involved in essential cellular processes such as DNA replication, transcription and translation often preserve their structural roles, leading to persistent H_0 connectivity patterns or stable H_1 loop-like features that reflect long-lasting functional organisation (Ogrodnik, Salmonowicz & Gladyshev 2019).

In addition, several ageing-related proteins may possess structural or functional properties that reduce their sensitivity to age-associated changes in the cellular environment. Heat shock proteins and other molecular chaperones, for instance, stabilise protein folding and maintain proteostasis. These proteins may show stable H_0 and H_1 topological patterns because their interactions help prevent misfolding or aggregation commonly linked to ageing (Hipp, Park & Hartl 2014). Other proteins may form durable protein complexes or functional modules that preserve their interaction profiles over time. Proteins embedded in stable complexes tend to retain similar neighbourhood connectivity across age groups, supporting the maintenance of core cellular processes as the organism ages (Alberts et al. 2007).

Finally, our integrative approach, which relies on network centrality measures and topological analysis, sheds light on the molecular underpinnings of ageing. We identify proteins with high topological variability across age groups, showing potential targets for future research and therapeutic intervention in age-related diseases. Our findings help advance our understanding of the complex

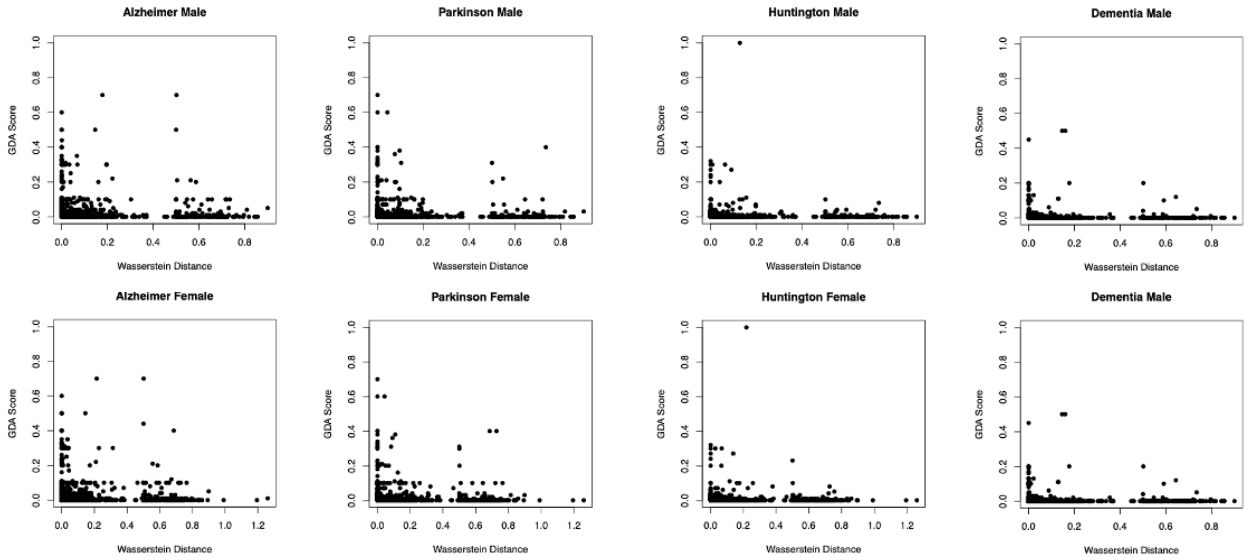

FIGURE 10. Scatter plot of Wasserstein distance against GDA Score for Alzheimer's, Parkinson's, Huntington's, and Dementia for every protein according to gender

TABLE 2. List of age-related proteins that exhibit high topological change from adult to elderly PPINs

UniprotID	Gender	Sources
O00206	Both	Longevity Map
O14920	Male	Gene Age
O43464	Female	Gene Age
O75771	Female	Longevity Map
O95229	Both	Longevity Map
P00533	Male	Gene Age
P01137	Female	Both
P02649	Both	Both
P02654	Male	Longevity Map
P09601	Female	Longevity Map
P10599	Male	Gene Age
P23025	Male	Gene Age
P25445	Both	Both
P27695	Female	Gene Age
P42229	Both	Gene Age
P42345	Female	Both
P45984	Male	Gene Age
P50402	Male	Gene Age
P61586	Female	Longevity Map
P63165	Male	Gene Age
Q08050	Both	Gene Age
Q09472	Female	Gene Age
Q8N122	Female	Longevity Map
Q92889	Male	Gene Age

interplay between protein networks and ageing processes, paving the way for developing precision medicine strategies to combat age-related disorders.

CONCLUSIONS

This study applied Local Persistent Homology to protein–protein interaction networks to characterise how local network structure changes with age. By focusing on level 2 ego networks and extracting H_0 and H_1 features from persistence diagrams, we quantified age related differences in local topology using the Wasserstein distance. These topological features were then related to standard network centrality measures, providing a complementary node level characterisation of proteins within PPINs.

The results showed that proteins with many topological components in their neighbourhoods tend to have higher degree and betweenness centrality but lower local clustering coefficients. In contrast, proteins embedded in simpler neighbourhoods with fewer components often display higher average persistence and higher clustering, suggesting the presence of more stable and tightly organised local structures. These findings indicate that LPH can show aspects of local organisation that are not fully captured by degree, betweenness or clustering coefficient alone.

By combining LPH based measurements with gene and disease association databases, we identified 25 unique proteins that exhibit pronounced age-related topological changes, including several associated with neurodegenerative diseases. At the same time, some proteins showed stable H_0 and H_1 patterns across age groups, which may reflect functional resilience in core cellular processes. Together, these observations suggest that both highly variable and structurally stable proteins can be important for understanding ageing mechanisms.

Future work could extend this framework by incorporating additional biological information, such as expression levels or post translational modifications, and by applying LPH to other age related or disease specific PPINs. Overall, this study demonstrates that local persistent homology provides a useful and interpretable tool for studying age related changes in protein interaction

networks and for generating hypotheses about proteins that may play key roles in ageing and age associated diseases. These results demonstrate that LPH provides a powerful complementary framework for uncovering subtle molecular patterns associated with ageing and disease.

ACKNOWLEDGMENTS

This work is part of a research project under the Ministry of Higher Education, Malaysia grant scheme FRGS/1/2020/STG06/UKM/02/6.

REFERENCES

Aguilar-Alarcón, J.J., Hernández-Gómez, J.C. & Romero-Valencia, J. 2023. The clustering coefficient for graph products. *Axioms* 12(10): 968.

Aktas, M.E., Akbas, E. & El Fatmaoui, A. 2019. Persistence homology of networks: Methods and applications. *Applied Network Science* 4: 1-28.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. & Walter, P. 2007. *Molecular Biology of the Cell*. 5th ed. New York: W.W. Norton & Company.

Ashtiani, M., Salehzadeh-Yazdi, A., Razaghi-Moghadam, Z., Hennig, H., Wolkenhauer, O., Mirzaie, M. & Jafari, M. 2018. A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology* 12(1): 80.

Berchtold, N.C., Cribbs, D.H., Coleman, P.D., Rogers, J., Head, E., Kim, R., Beach, T., Miller, C., Troncoso, J., Trojanowski, J.Q., Zielke, H.R. & Cotman, C.W. 2008. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences of the United States of America* 105: 15605-15610.

Bhowmick, S.S. & Seah, B.S. 2016. Clustering and summarizing protein-protein interaction networks: A survey. *IEEE Transactions on Knowledge and Data Engineering* 28(3): 638-658.

Brandes, U. 2001. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* 25(2): 163-177.

Budovsky, A., Craig, T., Wang, J., Tacutu, R., Csordas, A., Lourenço, J., Fraifeld, V.E. & De Magalhães, J.P. 2013. LongevityMap: A database of human genetic variants associated with longevity. *Trends in Genetics* 29(10): 559-560.

Calabrese, G., Molzahn, C. & Mayor, T. 2022. Protein interaction networks in neurodegenerative diseases: From physiological function to aggregation. *Journal of Biological Chemistry* 298(7): 102062.

Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. 2007. Stability of persistence diagrams. *Discrete and Computational Geometry* 37(1): 263-271.

Corriveau-Lecavalier, N., Gunter, J.L., Kamykowski, M., Dicks, E., Botha, H., Kremers, W.K., Graff-Radford, J., Wiepert, D.A., Schwarz, C.G., Yacoub, E., Knopman, D.S., Boeve, B.F., Ugurbil, K., Petersen, R.C., Jack, C.R., Terpstra, M.J. & Jones, D.T. 2023. Default mode network failure and neurodegeneration across aging and amnestic and dysexecutive Alzheimer's disease. *Brain Communications* 5(2): fcad058.

de Magalhães, J.P., Budovsky, A., Lehmann, G., Costa, J., Li, Y., Fraifeld, V. & Church, G.M. 2009. The human ageing genomic resources: Online databases and tools for biogerontologists. *Aging Cell* 8(2): 65-72.

de Magalhaes, J.P., & Toussaint, O. 2004. GenAge: a genomic and proteomic network map of human ageing. *FEBS letters* 571(1-3): 243-247.

del Toro, N., Shrivastava, A., Ragueneau, E., Meldal, B., Combe, C., Barrera, E., Perfetto, L., How, K., Ratan, P., Shirodkar, G., Lu, O., Mészáros, B., Watkins, X., Pundir, S., Licata, L., Iannuccelli, M., Pellegrini, M., Martin, M.J., Panni, S., Duesbury, M., Vallet, S.D., Rappsilber, J., Ricard-Blum, S., Cesareni, G., Salwinski, L., Orchard, S., Porras, P., Panneerselvam, K. & Hermjakob, H. 2022. The IntAct database: Efficient access to fine-grained molecular interaction data. *Nucleic Acids Research* 50(D1): D648-D653.

Dey, T.K., Shi, D. & Wang, Y. 2015. Comparing graphs via persistence distortion. *Leibniz International Proceedings in Informatics (LIPIcs)* 34: 491-506.

Dong, X., Li, Y., Li, Q., Li, W. & Wu, G. 2023. Identification of immune signatures in Parkinson's disease based on co-expression networks. *Frontiers in Genetics* 14: 1090382.

Erciyes, K. 2023. Graph-theoretical analysis of biological networks: A survey. *Computation* 11(10): 188.

Faisal, F.E. & Milenković, T. 2014. Dynamic networks reveal key players in aging. *Bioinformatics* 30(12): 1721-1729.

Faisal, F.E., Zhao, H. & Milenkovic, T. 2015. Global network alignment in the context of aging. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 12(1): 22-32.

Faisal, F.E., Meng, L., Crawford, J. & Milenković, T. 2015. The post-genomic era of biological network alignment. *Eurasip Journal on Bioinformatics and Systems Biology* https://doi.org/10.1186/s13637-015-0022-9

Fasy, B.T. & Wang, B. 2016. Exploring persistent local homology in topological data analysis. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings.* pp. 6430-6434.

Hajij, M., Wang, B., Scheidegger, C. & Rosen, P. 2018. Visual detection of structural changes in time-varying graphs using persistent homology. *IEEE Pacific Visualization Symposium* pp. 125-134.

Hazram, A.S., Bakar, S.A. & Razak, F.A. 2024. Discovering the important proteins through persistent homology in aging protein-protein interaction networks. *Commun. Math. Biol. Neurosci.* 2024: 115.

Hipp, M.S., Park, S.H. & Hartl, U.U. 2014. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology* 24(9): 506-514.

Hwang, J.T., Lee, A. & Kho, C. 2022. Ubiquitin and ubiquitin-like proteins in cancer, neurodegenerative disorders, and heart diseases. *International Journal of Molecular Sciences* 23(9): 5053.

Ignacio, P.S.P. & Darcy, I.K. 2019. Tracing patterns and shapes in remittance and migration networks via persistent homology. *EPJ Data Science* 8: 1.

Islambekov, U. & Gel, Y.R. 2019. Unsupervised space–time clustering using persistent homology. *Environmetrics* 30: e2539.

Kosch, D. & Schreiber, F. 2004. Comparison of centralities for biological networks. *Proc. German Conf. Bioinformatics* 2004: 199-206.

Kurtishi, A., Rosen, B., Patil, K.S., Alves, G.W. & Møller, S.G. 2019. Cellular proteostasis in neurodegeneration. *Molecular Neurobiology* 56(5): 3676-3689.

Li, J., Bian, C., Luo, H., Chen, D., Cao, L. & Liang, H. 2021. Multi-dimensional persistent feature analysis identifies connectivity patterns of resting-state brain networks in Alzheimer's disease. *Journal of Neural Engineering* 18(1): 016012.

Liu, Y., Weng, W., Gao, R., Liu, Y. & Monacelli, F. 2019. New insights for cellular and molecular mechanisms of aging and aging-related diseases: Herbal medicine as potential therapeutic approach. *Oxidative Medicine and Cellular Longevity* 2019: 4598167.

Masoomy, H., Askari, B., Tajik, S., Rizi, A.K. & Jafari, G.R. 2021. Topological analysis of interaction patterns in cancer-specific gene regulatory network: Persistent homology approach. *Scientific Reports* 11(1): 16414.

Mooney, K.M., Morgan, A.E. & McAuley, M.T. 2016. Aging and computational systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 8(2): 123-139.

Ni, Y.Q., Xu, H. & Liu, Y.S. 2022. Roles of long non-coding RNAs in the development of aging-related neurodegenerative diseases. *Frontiers in Molecular Neuroscience* 15: 844193.

Ogrodnik, M., Salmonowicz, H. & Gladyshev, V.N. 2019. Integrating cellular senescence with the concept of damage accumulation in aging: Relevance for clearance of senescent cells. *Aging Cell* 18(1): e12841.

Partridge, L., Deelen, J. & Slagboom, P.E. 2018. Facing up to the global challenges of ageing. *Nature* 561: 45-56.

Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F. & Furlong, L.I. 2020. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 48(D1): D845-D855.

Song, E. 2023. Persistent homology analysis of type 2 diabetes genome-wide association studies in protein–protein interaction networks. *Frontiers in Genetics* 14: 1270185.

Sugis, E. & Hermajakob, H. 2019. Human PPI from IntAct database (IAH). *Figshare. Dataset.* 10.6084/m9.figshare.10200136

Syukor, H. & Sakhinah, A.B. 2019. Characterization of essential protein in aging based on network analysis. *Journal of Physics: Conference Series* 1366: 012026.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., DIana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V.E. & De Magalhães, J.P. 2018. Human ageing genomic resources: New and updated databases. *Nucleic Acids Research* 46(D1): D1083-D1090.

Tacutu, R., Craig, T., Budovsky, A., Wuttke, D., Lehmann, G., Taranukha, D., Costa, J., Fraifeld, V.E. & De Magalhães, J.P. 2013. Human ageing genomic resources: Integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Research* 41(D1): D1027-D1033.

Teulière, J., Bernard, C., Corel, E., Lapointe, F.J., Martens, J., Lopez, P. & Bapteste, E. 2023. Network analyses unveil ageing-associated pathways evolutionarily conserved from fungi to animals. *GeroScience* 45(2): 1059-1080.

*Corresponding author; email: p94850@siswa.ukm.edu.my