ProtEGOnist: Visual Analysis of Interactions in Small World Networks Using Ego-graphs

N. Brich[1,*] ID, T.A. Harbig[1,*] ID, M. Witte Paz[1] ID, K. Nieselt[1] ID, and M. Krone[1] ID

[1] Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany
* These authors contributed equally and should be regarded as joint first authors.
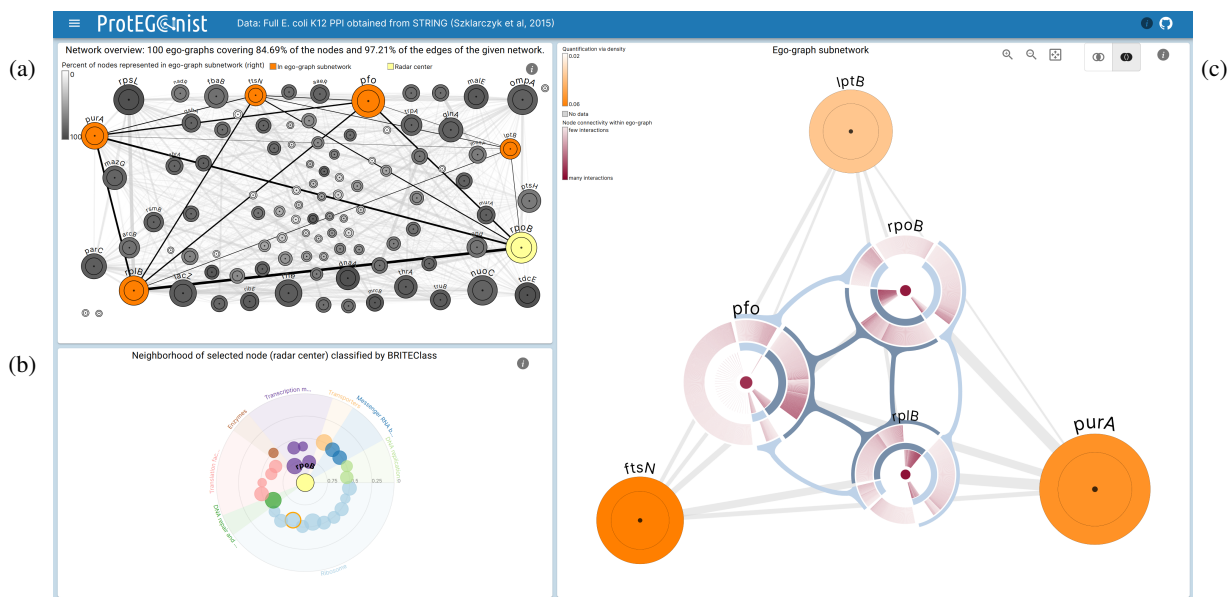
Figure 1: ProtEGOnist *uses ego-graphs for the visualization of small-world networks (in this case, the Escherichia coli protein-protein interaction network). (a) The network of ego-graphs visualized as aggregated glyphs gives an overview of the network with edges visualizing similarity. (b) A radar chart shows the similarity between a selected ego-graph (center, yellow node in (a)) and its neighbor ego-graphs grouped by metadata. Circles representing ego-graphs that are included in the subnetwork are highlighted by an orange outline. (c) An ego-graph subnetwork selected from the overview (orange nodes in (a)) provides details. Three ego-graph nodes are deaggregated forming an ego-graph group for a detailed comparison. The ego-graphs contain all nodes with a maximum distance of two.*

Abstract

Visualizing small-world networks such as protein-protein interaction networks or social networks often leads to visual clutter and limited interpretability. To overcome these problems, we present ProtEGOnist*, a visualization approach designed to explore small-world networks.* ProtEGOnist *visualizes networks using ego-graphs that represent local neighborhoods. Ego-graphs are visualized in an aggregated state as a glyph where the size encodes the size of the neighborhood and in a detailed version where the original network nodes can be explored. The ego-graphs are arranged in an ego-graph network, where edges encode similarity using the Jaccard index. Our design aims to reduce visual complexity and clutter while enabling detailed exploration and facilitating the discovery of meaningful patterns. To achieve this, our approach offers a network overview using ego-graphs, a radar chart for a one-to-many ego-graph comparison and meta-data integration, and detailed ego-graph subnetworks for interactive exploration. We demonstrate the applicability of our approach on a co-author network and two different protein-protein interaction networks. A web-based prototype of* ProtEGOnist *can be accessed online at* https://protegonist-tuevis.cs.uni-tuebingen.de/.

CCS Concepts
• *Applied computing* → *Biological networks;* • *Human-centered computing* → *Graph drawings; Information visualization;*

1. Introduction

Networks are used to model a wide array of systems. Depending on the underlying data, networks can differ in their parameters, size, density, and connectivity. Many networks such as social networks, biological networks, transportation networks, or citation networks exhibit the *small-world property*, which means that most nodes can be reached from any other node in a small number of steps [Mil67; WS98]. A famous example of this property is the idea of the 6-handshakes rule also known as *six degrees of separation*, stating that every person in the world only has a distance of a maximum of six handshakes from any other person [Kar29; NBW06].

The small world property is also prevalent in many biological networks [BO04] like *protein-protein interaction* (PPI) networks, which play a crucial role in modeling and understanding the intricate mechanisms governing cellular processes. In PPIs, proteins are seen as nodes, while interactions are represented by edges. Traditionally, two proteins are considered interacting if they bind physically. However, the concept is often extended to other indirect connections, such as the spatial proximity of the corresponding genes in the genome or co-occurrence validated in the literature.

Typical visualizations for small-world networks include node-link diagrams or matrix representations [FAM23]. Links in our case represent any type of interaction, e.g., interacting proteins or "interacting" researchers co-authoring a paper. Visualizing complete networks with many thousands of nodes as node-link diagrams typically results in a cluttered, hairball-like structure, especially when using standard force layouts [NOB14]. Moreover, the sheer number of nodes and interactions makes it challenging to find nodes of interest and do a targeted comparison of their neighborhoods.

Oftentimes, single nodes serve as starting points when analyzing small-world networks, such as oneself or a famous person in a social network. Usually, it is meaningful to inspect not only immediate contacts but also indirect ones. Social science studies have shown that in social networks such indirect contacts can affect, e.g., a person's happiness [FC08] and their ability to find a job [Pel10]. In PPI networks, a node of interest could be a protein that is the research focus of a biologist. Indirect connections are studied, e.g., when analyzing metabolic pathways. This is, e.g., important in PPIs showing physical interactions. It has been shown that proteins with the same interaction partners rarely interact directly [KLS*19]. Common path lengths for PPI networks are between four and five [XBBY11], thus, contacts with a distance higher than two tend to cover very large portions of the network [ARR*14].

To study such local subnetworks around nodes of interest, ego-graphs can be used [Spr99]. Originally developed for the study of social networks, this approach focuses on the local neighborhood of an individual node, instead of showing all nodes and interactions. An ego-graph consists of a central node of interest—the *ego*—and its local neighborhood in the network—the *alters*. Degree-1 alters are alters with a direct connection to the ego. Degree-2 alters have direct connections to degree-1 alters, but not to the ego. That is, 1-level ego-graphs only consider degree-1 alters, and 2-level ego-graphs consider degree-2 alters as well. Typically, 2-level ego-graphs are used [ZGC*16], i.e., "friends-of-friends" networks.

We developed *ProtEGOnist*, a novel visualization approach for the exploration of small-world networks that uses 2-level ego-graphs to aggregate local neighborhoods represented by glyphs. (Figure 1). Initially, *ProtEGOnist* was submitted as a contribution to the Bio+MedVis Challenge 2023 [23a] for redesigning the visualization of a specific PPI network by Gonçalves *et al.* [GPC*22]. The challenge dataset included a PPI network together with protein-drug associations predicted using a deep learning approach developed by the authors called *DeeProM*. The original visualization was a static figure (Figure 7F in original publication) showing the PPI network as a node-link diagram with 8,395 nodes and 66,721 edges. The main point of the challenge was to provide a less cluttered view of the network that includes overviews as well as detailed visualizations that are enriched with metadata. Moreover, the challenge specified the need to focus on specific proteins and explore their relationship. Based on the visualization of this dataset, *ProtEGOnist* was awarded as the best contribution to the Bio+MedVis Challenge 2023.

Our contributions can be summarized as follows: we present a universal approach for the exploration of small-world networks with many thousands of nodes. Our approach focuses on ego-graphs, i.e., placing the individual in the center as the ego and thus making it the "protagonist" of the graph (Section 3.1). To achieve this, *ProtEGOnist* creates a network of ego-graphs from an input interaction network and uses specifically designed glyphs to visualize these ego-graphs (Section 3.2). The edges in the network of ego-graphs are weighted by the similarity of the ego-graphs, which is calculated as the Jaccard index of the node sets for each pair of nodes. This concept allows for an exploration from an overview level down to analyzing subsets of ego-graphs, comparing up to three ego-graphs in detail, and inspecting single ego-graphs (Section 3.3). Using the taxonomy proposed by Filipov *et al.* [FAM23], we would position *ProtEGOnist* as a *group network visualization*, since the ego-graphs represent groups of nodes based on neighborhood. Using the *Vertex Group Structure Taxonomy* by Vehlow *et al.* [VBW17], we would describe *ProtEGOnist* as an *overlapping hierarchical* structure, which they found only in a single approach.

We demonstrate the effectiveness of our approach using three exemplary use cases. The first one is a co-author network built from IEEE VIS authors [IHK*17], i.e., a social network (Section 4.1). The other two show that our approach can be used for domain-specific datasets, such as visualizing PPI networks: we applied *ProtEGOnist* to a PPI network of *Escherichia coli* (Section 4.2) and a human PPI with metadata on drug-protein associations provided for the Bio+MedVis Challenge 2023 (Section 4.3).

2. Related Work

Simple node-link diagrams are the most commonly used visualization techniques for networks [NMSL19]. They are, e.g., popular for visualizing PPIs and are used by *STRING* [SFW*15] and the well-known network visualization tool *Cytoscape* [SMO*03]. Although node-link diagrams are conceptually intuitive and powerful for visual analysis, they quickly suffer from overdraw, layout problems, and clutter for larger networks [NOB14].

Therefore, approaches beyond force-directed node-link diagrams, different layouts [SKL*14] including hierarchical layouts [AMA07], on-node encodings [VV14], and hybrid network

visualizations [HFM07; ADM*19] have been developed. As an in-depth review of the vast literature on network visualization is beyond the scope of our paper, we refer to the recent state-of-the-art reports by Filipov *et al.* [FAM23] and Nobre *et al.* [NMSL19].

To reduce the amount of clutter created by the edges, various approaches are used. Edges can be partially indicated [BVKW12] or even omitted completely. An omission is only viable if edges are implicit or not of interest in the applied layout, for example, when applying containment to cliques. Both approaches can be enriched by drawing the edges in full on-demand [SA06]. Moreover, edge bundling can be used to use topological or semantic information to merge edges into bundles [Hol06; ZPYQ13]

A general approach to reduce the clutter in a node-link diagram is to reduce the number of elements via grouping, clustering, or aggregation. Grouping can be utilized using underlying semantic information to generate containment for groups sharing specific attributes [SA06; BLGS06]. These approaches depend on semantic metadata suitable to the applied grouping method and the underlying goal of the analysis. Alternatively, networks can be grouped purely by topological measures, for example, grouping into subnetworks of densely connected nodes or by creating ego-graphs for a set of manually or computationally determined nodes of interest.

Aggregation is often used by merging nodes into distinct glyphs, increasing readability [DS13]. In one such example, Vehlow *et al.* [VRW13] visualize multiple overlapping hierarchical networks using node-link diagrams. In their *fuzzy-communities* approach, they display an overview using multiple levels of abstraction. Depending on the chosen level, some or all nodes are collapsed into meta-nodes, which encode network membership-heterogenicity using the fuzziness of the shape.

Alternative techniques go further by substituting glyphs for other standalone visualization types, like adjacency-matrices [HFM07], chord diagrams [ADM*19] or customizable plots such as line- and bar-charts [VV14]. While these approaches aim at visualizing groups in general, specialized visualization types have been developed for ego-graphs, aiming at displaying their inherent hierarchical structure with an ego and alters. The *EgoComp* approach uses a hybrid network visualization for comparing ego-graphs in social networks [LGD*17]. It applies both an implicit hierarchical layout for the visualization of ego-graphs and a conventional node-link layout for linking identical nodes between the compared graphs. For the visualization of ego-graphs, nodes are placed around a center in partial circles according to their distance from the ego. The half-circles of the two compared ego-graphs are facing each other. Since two ego-graphs can contain the same nodes, edges connect the respective nodes to express identity.

Ego-graph visualizations are extensively used in the domain of dynamic graphs, as shown in a recent review by Kale *et al.* [KSP23]. While dynamic graphs represent a data structure with specific tasks and use cases, some of the visualization concepts apply to static graphs. Visualizations of ego-graphs in dynamic networks include node-link diagrams using a stress-majorization layout [QHS*15] circular glyphs [WPZ*16], radial layouts [PHW17; Rei10], linear layouts [ZGC*16; FMW*21] as well as combinations of the aforementioned approaches [SWW*15]. EgoSlider aggregates an ego-graph into a pie- or bar-chart glyph- They encode

changes in the properties of the graph at different time points. The authors use the circular variant as default, arguing that the design serves as a "metaphor of an ego surrounded by alters" [WPZ*16]. This metaphor corresponds to a radial approach to visualizing target-based graphs, where the center node of the node-link diagram effectively considers the graph as the root of a rooted tree. It is an approach commonly found in the literature [Wil99; DLR09; XWH*23], especially in literature focused on ego-graph visualization [BP10; FHQ11; LGD*17]. Differently, EgoLines is an example of a technique applying non-circular layouts when comparing an ego-graph at different states of a dynamic network [ZGC*16]. It visualizes ego-graphs as adjacency matrices, but as they grow quadratically in size with an increasing number of nodes, they are hard to interpret for large ego-graphs. The ego is placed in a central position with alters placed outwards, inducing a hierarchy within the alters even if not desired.

In addition to the general approaches, we also consider ego-graph approaches for PPI networks relevant to our work as a contribution to the Bio+MedVis Challenge [23a]. The *STRING* database [vMHJ*03] is a popular resource for PPI networks. It uses different interaction types to calculate a confidence score for each protein-protein interaction. Ego-graphs are shown when searching for a protein of interest [SFW*15]. The alters, are the proteins that are directly connected with a query protein – the ego. This 1-level ego-graph is shown as a node-link diagram with a force-directed layout. By default, it displays only the 10 highest-scoring interactions to reduce the network size. Optionally, a second shell of interactions can be displayed, showing the highest-scoring direct neighbors of the interactors of the target query (2-level ego-graph). In contrast to *STRING*, *BioLinker* [DMF17] visualizes the entire network in an overview, where nodes of interest can be selected and visualized as a subnetwork consisting of ego-networks in a separate view. Moreover, *BioLinker* highlights the egos such that they are visually distinct from the alters.

Another application using ego-graphs in biological networks is the *EgoNet* algorithm [YBQY14], which identifies disease subnetworks. *EgoNet* can be applied to PPIs where each protein is associated with protein abundances (also called *protein expression*) at different clinical outcomes, e.g., when comparing protein abundances in healthy and cancerous cells. Starting with an ego, the tool iteratively adds alters and calculates if the contained proteins suffice to predict the clinical outcome. This approach shows how ego-graphs are used as a data structure to computationally reduce the network size by focusing only on the most relevant nodes.

3. Approach

Based on the Bio+MedVis Challenge 2023 [23a] and aided by the task taxonomy for graph visualization by Lee *et al.* [LPP*06], we identified the following tasks for the development of *ProtEGOnist*:

Overviews can provide a starting point for the analysis [Shn96], especially in previously unexplored datasets and when there is no clear hypothesis about the data. For this, we want to simplify the network and declutter it by aggregating groups of nodes into meta-nodes. Then we can exploit the small-world property to facilitate an overview showing the important meta-nodes, like those covering a

large part of the interaction network. This can be described as an *Overview Task*.

Often the global context of entities of interest identified before exploration, such as specific known individuals within a social network or proteins within a PPI is of special interest. Van Ham and Perer [vP09] present an approach applying the *"Search, Show Context, Expand on Demand"* principle, which focuses on nodes of interest that can be interactively added to the visualization and shown in the context of the graph. A central aspect of *ProtEGOnist* should be the selection of nodes-of-interest, which is best described as an *Attribute-Based Task - On the Nodes*. For this, we want to create the meta-nodes based on the neighborhoods of nodes of interest and need to find the nodes accessible from these nodes (*Topology-Based Task - Accessibility*).

Moreover, we want to empower users to find similarities between nodes, for example, by estimating the overlap between meta-nodes. We also want to allow for a more meaningful and in-depth comparison of meta-nodes. A user might want to find the nodes shared between the corresponding neighborhoods. Both actions correspond to a *Topology-Based Task - Common Connection*. Finally, we also want to allow the users to utilize the metadata layer to find nodes fulfilling domain-specific criteria. For example, metadata should be used for filtering nodes of interest and mapped to visual channels. This can again be described as an *Attribute Task - On the Nodes*.

Based on these described tasks, we identified the following requirements for *ProtEGOnist*:

R1 **Overview:** Apply filtering and aggregation techniques to provide a comprehensive overview showing the most relevant meta-nodes(e.g., representing numerous interactions).

R2 **Subnetwork context:** Viewing meta-nodes in the local and global network context.

R3 **Detail**: Allow a detailed analysis of meta-nodes, such as finding shared nodes in subnetworks.

R4 **Metadata:** Provide the integration of further metadata on the network, such as categories or measurements for the instances represented by nodes.

We want to develop a layout that satisfies these requirements and enables the defined tasks and thus results in a less cluttered visualization in comparison to force-directed node-link diagrams like the one of the Bio+MedVis Challenge.

3.1. Ego-graph Concept & Visualization Design

We address the requirements defined above with *ProtEGOnist* using ego-graphs. Interaction networks consist of nodes representing entities, such as proteins in PPI networks or people in social networks, and edges representing interactions between them. Instead of visualizing every node and interaction individually, *ProtEGOnist* groups nodes and interactions into ego-graphs and represents them as circular glyphs. Similarity values are calculated for every pair of ego-graphs using the Jaccard index of the sets of contained nodes, i.e., the intersection size divided by the size of the union. Using the similarity values, an *ego-graph network* is created, where the nodes are visualized using the ego-graph glyphs. The small world property can be exploited to create an overview since a comparatively small set of ego-graphs is sufficient to cover a relatively large
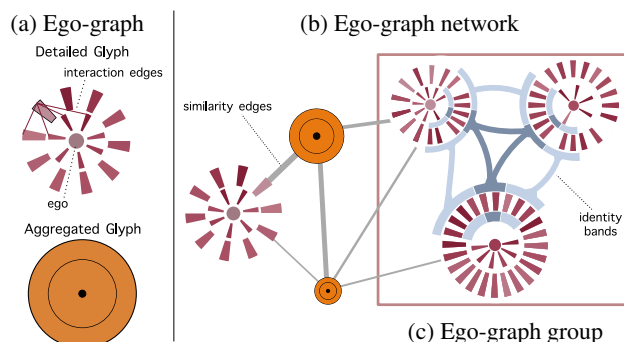

Figure 2: *Ego-graphs and ego-graph networks, the concept of* ProtEGOnist. *(a) A single ego-graph can be visualized in detail or aggregated. The detailed view shows the ego in the middle of a circular graph layout. Degree-1 alters are placed on the inner circle, and degree-2 alters on the outer circle. Nodes are colored corresponding to their number of interaction edges (few interactions ▬ many interactions). The interaction edges are only displayed when hovering over a node. The aggregated view of an ego-graph encodes the number of alters via the area of the glyph. (b) An ego-graph subnetwork consists of single ego-graphs and ego-graph groups. In this view, similarity edges connect single ego-graphs. Their width and opacity encode the Jaccard index between the respective ego-graphs. (c) Visualization of an ego-graph group. Ego-graph groups are arrangements within the ego-graph network of up to three detailed ego-graphs with identity bands connecting identical nodes. A darker blue indicates nodes occurring in all three ego-graphs, while the lighter blue indicates only pairwise intersections.*

set of nodes of the original network (**R1**). The glyphs provide additional details about the ego-graphs contained in the network and their relation to each other (**R2**,**R3**).

We use 2-level ego-graphs, i.e., all alters have at most a distance of two to the ego, to achieve a reasonable reduction of the original network as well as to offer visually feasible comparisons between any two ego-graphs. Each node can be chosen as the ego of an ego-graph. To represent a single ego-graph, we have designed two types of radial glyphs: a detailed one and an aggregated one (Figure 2a). The detailed glyph (Figure 2a, top) visualizes the alters as ring segments in two circular levels around the ego (**R3**). The circular layout highlights the central role of the ego and represents a space-efficient layout for its alters. The first, inner level contains all degree-1 alters, while the second, outer level contains all degree-2 alters. The ego is represented by a filled circle in the center. To visualize the connectedness of the alters, the ring segments representing the alters as well as the ego circle are colored to represent their node-degree in the network (*few interactions* ▬ *many interactions*). To avoid clutter, the *interaction edges* of the alters are only shown on hover. The aggregated glyph (Figure 2a, bottom) is a simplified, abstract version of the detailed glyph. It consists of two concentric circles to symbolize the two levels and a black dot in the center to represent the ego. The background of the glyph can be colored to represent a certain property of the underlying ego-graph. The size of both the detailed and the aggregated glyph can

be scaled to illustrate the number of elements in the ego-graph, that is, the cardinality of the ego-graph itself. For the detailed glyph, this is a deliberate double encoding that makes it easier to compare the size of two ego-graphs instead of counting the circle segments representing the nodes. Optionally, circular text labels on top of the glyph can show the name of the ego (Figure 1). The font size is scaled with the size of the glyph, and the text labels are automatically discarded if the glyph is too small.

The glyphs can also be used as the nodes of an ego-graph network (**R2**), where the size of each node encodes the cardinality of the ego-graph and the edge widths encode the similarity using the Jaccard index (*similarity edges*, Figure 2b). By default, ego-graphs are represented as aggregated glyphs that can be expanded on demand to show the detailed glyph for an in-depth analysis.

Connected ego-graphs can be selected to form an *ego-graph group* to show in detail which alters are shared between the ego-graphs or unique to an ego-graph (**R3**). The groups show the ego-graphs as detailed glyphs and the numbers of shared alters as *identity bands* (Figure 2c). We restrict groups to three ego-graphs to eliminate crossing bands. In the case of an ego-graph group with three ego-graphs, we divide each ego-graph circle into four sections: one for alters unique to the respective ego, two for alters shared between any two ego-graphs, and one for alters shared between all three ego-graphs. The three detailed ego-graph glyphs are placed to form an equilateral triangle. The sections shared between all three graphs are arranged to face towards the center of the imaginary triangle (dark blue), while the pairwise sections face towards each other (light blue), and the section containing the unique nodes faces away from the triangle center. The shared sections are illustrated by contour arcs covering the corresponding nodes of the detail glyphs. The arcs on the glyph surfaces are connected via *identity bands* to visualize that the corresponding sections in the ego-graphs contain the same nodes. These curved bands are optimized to avoid sharp angles or crossings by positioning them off-center to the corresponding arc. The colors of the bands match the arc color and facilitate distinguishing the portions of nodes shared between two and three ego-graphs. If the group only consists of two ego-graphs, only a single section is generated for the shared nodes, and the two ego-graphs are placed on a horizontal line.

The alters in the detailed ego-graphs are sorted separately. Within the sections, they are sorted by three criteria: (i) Their distance to the ego, (ii) their average distance to the other egos in the ego-graph group, and (iii) their node degree (**R2**). Thus, distinct subsections within the sections emerge, facilitating the location of shared nodes with a specific distance to the different egos.

3.2. Glyph and Ego-Graph Group Redesign

As an initial idea for the submission to the Bio+MedVis Challenge, we followed the approach implemented in *egoComp* [LGD*17], in which alters shared between two ego-graphs are connected using edges (Figure 3a). While this is feasible for comparing two ego-graphs in detail, we encountered several issues when using this in the ego-graph network and for ego-graph groups (**R3**).

To avoid edge crossings, the sort order of shared alters in ego-graph groups had to be identical in each graph. This caused some
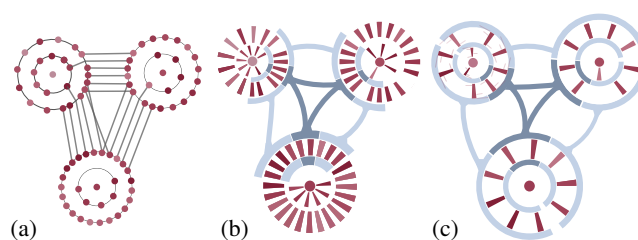

Figure 3: *Different concepts for visualizing ego-graph groups: (a) First concept as submitted to the Bio+MedVis Challenge 2023. Nodes are encoded as circles and connected to the instances of the same node with identity edges. (b) Redesigned concept, where nodes are encoded as ring segments and identical instances are connected with identity bands. (c) Shared-only mode, where alters not shared with other ego-graphs in the group are filtered out.*

proportions of the ego-graphs to remain in a non-logical order concerning the node degree, and the general distribution of node degrees could not be deduced visually. Moreover, we could not use the entire circle for displaying alters shared between ego-graphs but only a portion to avoid edges crossing the nodes. In the case of an ego-graph group of size three, it was hard to visually distinguish alters that are shared between all ego-graphs and alters shared between only two ego-graphs. In addition, identity edges could not easily be distinguished from similarity edges. Furthermore, alters were encoded by circles in the previous version. For large ego-graphs, the available space to arrange alters around the ego is limited, leading to tiny radii when displaying circles. This in turn caused a very poor "ink-to-space" ratio, which then made it very hard to properly distinguish single nodes.

With the introduction of colored curved identity bands (Figure 3b), we addressed all of these issues. The usage of identity bands leads to the circle being split into sections, effectively creating a donut chart-like visualization of the grouping of nodes. Identity bands can be distinguished from the similarity edges through the colors and the organic shape. By drawing bands instead of individual edges, we can now use the entire circle to display shared nodes, allowing the creation of a second view mode that shows only nodes shared by any of the detailed ego-graph instead of all nodes for all ego-graphs (Figure 3c, *shared-only mode*). Lastly, the problem of a low "ink-to-space" ratio was tackled using ring segments instead of circles to visualize the alters, as explained in Section 3.1. Furthermore, as individual identity edges are no longer drawn, more advanced sorting criteria could be introduced for the segments leading to a natural partitioning into subsections.

3.3. Visual Interface & Application Design

ProtEGOnist uses three main visualization components (Figure 1): a simplified *overview* of the original network showing a static ego-graph network (Figure 1a), a *radar chart* showing information about ego-graphs similar to one specific ego-graph (Figure 1b), and an *ego-graph subnetwork* (Figure 1c), which applies the concept of dynamically de-aggregating ego-graphs to a user-defined subset of the ego-graph network for a detailed analysis and comparison.
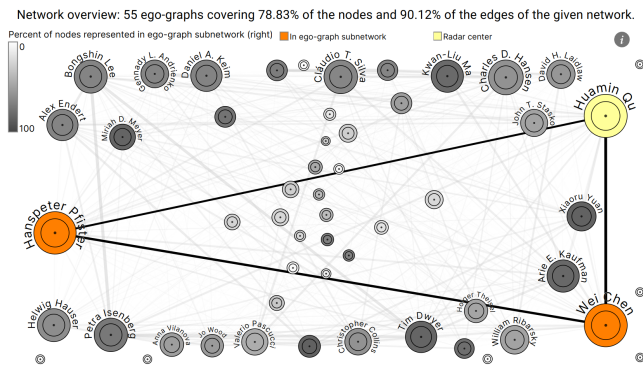
Figure 4: *Network overview using a set of 100 ego-graphs representing authors in a co-authorship network, covering 83% of the nodes and 95% of the edges of the original network. The color scale from white to gray maps to the percentage of nodes in the ego-graph currently visualized in the ego-graph subnetwork (0% ▭ 100%). A node is colored orange ● when it has been selected for visualization in the subnetwork view. A yellow node ● represents the current ego selected for the radar chart visualization.*

3.3.1. Overview

The *overview* shows a network of the most relevant ego-graphs (Figure 1a, **R1**). Depending on the dataset, the set of most relevant ego-graphs is already known (Section 4.3). For a general solution, we propose the following algorithm to extract an informative subnetwork of relevant nodes: provided that the input network has the small-world property, it is possible to cover a large portion of nodes and edges with a comparatively small subset of ego-graphs. That is, the problem can be translated into the Set Cover problem. Since this is an NP-hard problem [KV12], we use a heuristic approach. We calculate the ego-graphs for every node in the network and sort them by their cardinality. Then, we take the largest ego-graph and remove the covered edges from the remaining ego-graphs. We repeat this step until either a specified threshold of interaction coverage (default: 90 %) or a predefined maximum number of ego-graphs (default: 100) is reached.

The resulting overview network of relevant ego-graphs is visualized using the aggregated ego-graph glyphs (Figure 4). The percentage of nodes and edges in the original network covered by the resulting overview ego-graph network is displayed as a text label at the top. Following the "show context" and "details on demand" principles, each node in the overview can be selected for further inspection in the other views (**R2**). Moreover, the coloring of the aggregated glyphs in the overview provides context for the current selection for the visualizations. Glyphs are colored orange if the corresponding ego is visualized in the ego-graph subnetwork, and yellow if it is visualized in the radar chart. Ego-graph glyphs in the overview that are not selected are colored using a white-to-gray gradient, illustrating the percentage of nodes in the ego-graph that are contained in the ego-graph subnetwork (0% ▭ 100%). This allows users to either focus on ego-graphs that have a high overlap with the current selection (dark gray) or highly dissimilar ones (white or light gray), depending on their current analysis task.
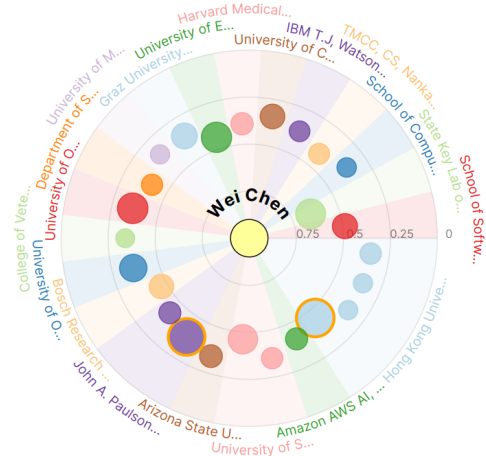

Figure 5: *Similar ego-graphs to a reference ego-graph (radar center) in a co-authorship network classified by affiliation. The radar chart shows the 25 ego-graphs most similar to the ego-graph in the center. The distance to the center corresponds to the Jaccard index. In addition to the similarity, categorical metadata is visualized. In this example, each circle represents the ego-graph of an author, while the colors represent their affiliation. Circles with an orange outline correspond to ego-graphs selected in the ego-graph subnetwork.*

3.3.2. Radar Chart

The *radar chart* provides information about a metadata attribute of egos whose ego-networks are similar to the one of the selected ego. (Figure 1b, **R4**). Similar to the aggregated glyphs, each circle represents an ego-graph, with the area corresponding to its cardinality. The radial distance to the center encodes the Jaccard index between the ego-graphs, i.e., the closer a node is to the center, the more alters it shares with the selected ego. This places the radar chart in close relation to the concept of *monadic exploration* [DCD14]. The core *monadic exploration* is to take the viewpoint of a subnetwork and display other subnetworks with overlapping relevance radially around it. Topics of higher relevance are placed closer to the center than topics of lower relevance. To avoid clutter, we only show the *n* ego-graphs with the highest Jaccard index (default $n = 25$). The colors of the nodes represent the metadata associated with the egos, such as author affiliation in a co-author network or the BRITE functional hierarchy in the case of proteins [KAG*07]. Ego-graphs that belong to the same category are put next to each other, and the corresponding circular segment of the radar chart is colored semitransparently with the same color. Additionally, text labels naming the categories corresponding to the circular segments are put around the radar chart. Users can select ego-graphs in the radar chart to add them to the ego-graph subnetwork. Ego-graphs in the radar chart that are also shown in the ego-graph subnetwork view have an orange outline, as shown in Figure 5.

3.3.3. Ego-graph Subnetwork

Ego-graphs selected in the overview or the radar chart are visualized in the *ego-graph subnetwork* (Figure 1c), showing different
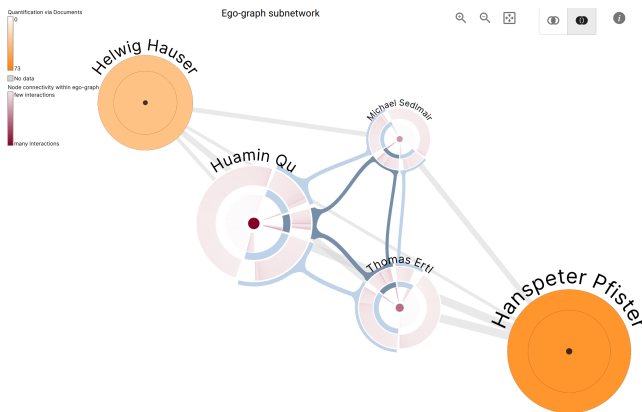
Figure 6: *Ego-graph subnetwork of a co-author network with five ego-graphs. A group of three ego-graphs is shown in its detailed view. The width of the gray similarity edges encodes the similarity between ego-graphs outside ego-graph groups, while the blue identity bands link identical nodes within an ego-graph group.*

levels of details of the respective ego-graphs. As mentioned in Section 3.1, the ego-graphs are initially visualized using the aggregated glyphs, but can be de-aggregated to the detailed glyphs on demand (Figure 6, **R3**). The color of each aggregated ego-graph glyph in this view encodes a quantitative metadata value associated with the ego (*min value* ▭ *max value*). Up to three connected ego-graphs can form an ego-graph group, as explained in Section 3.1.

3.3.4. Selection Table

Groups of nodes in ego-graphs or intersections can be selected for investigation in the selection table (Figure 7), shown on demand using a menu button. The table contains additional attributes for each node, such as metadata (**R4**) and information on the nodes, e.g., whether they are present in the overview and the ego-graph subnetwork. The rows can be sorted by any of the columns containing the attributes. The user can select any node for visualization in the ego-graph network and the radar chart. For a detailed analysis of intersections between ego-graphs, the user can select the corresponding intersection band in the ego-graph subnetwork, which allows to filter and sort the table by this subset.

	Radar	nodeID	institution	Documents ↓	Citations	Found in Ov...
☐	◎	Kwan-Liu Ma	University of Califo...	73	1931	Yes
☐	◎	M. Eduard Gröller	TU Wien, Austria	69	1680	No
☑	◎	Huamin Qu	Hong Kong Univer...	68	2615	Yes
☐	◎	Arie E. Kaufman	Department of Co...	61	788	Yes
☑	◎	Hanspeter Pfister	John A. Paulson S...	60	3568	Yes
☐	◎	Daniel A. Keim	University of Konst...	58	2304	Yes

Figure 7: *Excerpt of the selection table showing the top 5 entries of the co-author dataset (sorted by* Documents, *i.e., number of published papers). Only a subset of the columns is shown. The checkbox to the left adds the entry to the ego-graph subnetwork view.*

3.4. Implementation

We implemented *ProtEGOnist* as a web-based application with a client-server architecture. The server backend was written in Python using Flask [10]. The user interface and the visualizations in the frontend were mainly implemented in TypeScript using *React* [Fac13], *Jotai* [Kat23], *Material-UI* [23b], and *D3* [BOH11]. In the backend, we use the Python library *networkX* [HSS08] for the extraction of relevant features from the graph structure. *ProtEGOnist* is available at `https://protegonist-tuevis.cs.uni-tuebingen.de/`.

4. Use Cases

In this section, we demonstrate the applicability of our approach using three use cases. The first one shows the utility of *ProtEGOnist* and the interaction of its components for exploring a co-author network. The other two use cases show how it can be applied to PPI networks. The PPI network of *E. coli* serves as a well-known example dataset for domain experts from biology and highlights the advantages of the glyph design. The second PPI network stems from the Bio+MedVis Challenge 2023 and illustrates the application of *ProtEGOnist* to metadata-enriched datasets.

4.1. Co-author network

To showcase the usefulness of our *ProtEGOnist* approach for exploring social networks, we applied it to the Visualization Publications Dataset [IHK*17]. This dataset contains all publications of the IEEE VIS conference (SciVis, InfoVis, VAST) and its predecessor symposia and conferences. The metadata for each entry includes, e.g., the authors and the number of publications. The resulting co-author network has 6,610 nodes and 22,220 edges. The network and the metadata were extracted directly from the data, and the citation count provided by CrossRef [HTLF20] was used.

A typical starting question when exploring a co-author network could be to find out who the most well-connected researchers are, and whether they are also the most prolific ones in terms of publications. Investigating the ego-graph network using the *Network Overview*, the user can determine that the nodes for Huamin Qu, Hanspeter Pfister, and Wei Chen are the largest, indicating that they have the largest number of 1st and 2nd-degree coauthors (Figure 4, **R1**). We selected these three nodes for the *ego-graph subnetwork* view, which helps to visually compare node sizes (**R2**). The color mapping (*0* ▭ *max*) in the *ego-graph subnetwork* reveals that they all have a high number of publications (**R4**). Sorting the *Selection Table* by the number of documents (i.e., co-authored publications) allows for a quantitative assessment of the number of publications: all three are high-ranking, with Qu and Pfister being #3 and #5, respectively. Interestingly, Chen is only #13 (Figure 7, Supplementary Figure S2), despite having a high number of co-authors. Adding the two top-ranking researchers concerning their number of publications—Kwan-Liu Ma and M. Eduard Gröller—for an in-depth comparison reveals that Chen has a larger network than Gröller but also a higher percentage of unique co-authors that are not shared by the two (Figure 8, **R3**). One reason for the comparably large co-author ego-graph of Chen might be his joint publications with Qu and Pfister, thus benefitting from their large networks.
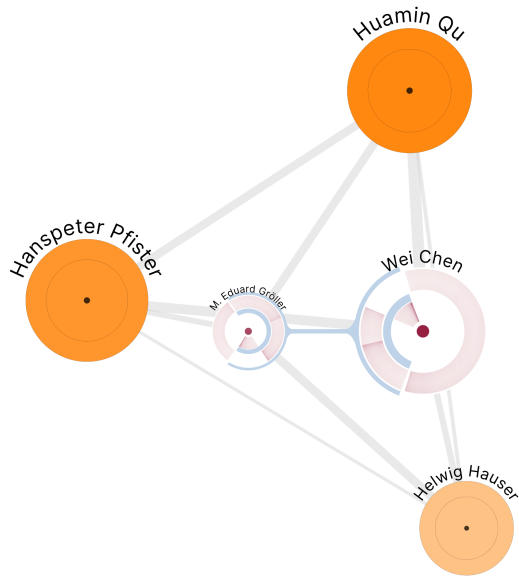
Figure 8: *Ego-graph subnetwork visualization with a detailed ego-graph group of co-author networks of Gröller and Chen. The detailed glyphs reveal that Chen has not only a larger network in total but also a higher percentage of unique second-level co-authors.*

Exploring the radar chart reveals that Chen has also published with other well-connected researchers like David Ebert, Benjamin Bach, or Yingcai Wu (identified by hovering the largest nodes in the radar chart shown in Figure 5). It also shows that the ego-graph of Chen contains researchers from institutions from all over the world.

4.2. *lac* operon in *E. coli* Protein-Protein Interaction Network

For protein-protein interactions, specific proteins and their context are often of interest. In the bacterium *E. coli*, the lactose operon (short *lac* operon) is a well-studied set of proteins that is required for the metabolism of lactose. It is active if glucose, the preferred energy source, is not available but only lactose.

Here, we analyze the PPI network of the K12 strain of *E. coli* as found in the *STRING* database [SFW*15] and demonstrate the effect of the ego-graph layout for analyzing three proteins in detail (**R3**). As a baseline, we loaded the PPI network into *Cytoscape* [SMO*03]. Figure 9a shows the proteins lacZ, lacY, and lacA of the *lac* operon and their degree-1 and degree-2 alters in a simple node-link diagram created using the *Cytoscape StringApp* [DMGJ19]. The node-link diagram forms a hairball-like structure due to the high number of nodes and edges. We can see that there is only a comparatively low number of degree-1 alters to the three proteins of interest (black nodes in Figure 9). Moreover, an edge connecting lacA and lacY—indicating a direct interaction between the two proteins—is visible. Any other conclusions about the connectivity between the lac proteins or about the sizes of the individual neighborhoods cannot be made due to occluding edges.

In comparison, with *ProtEGOnist* the neighborhood of the lac operon proteins can be grouped into three ego-graphs (Figure 9b). Strikingly, we can see that lacZ has by far the largest ego-graph of

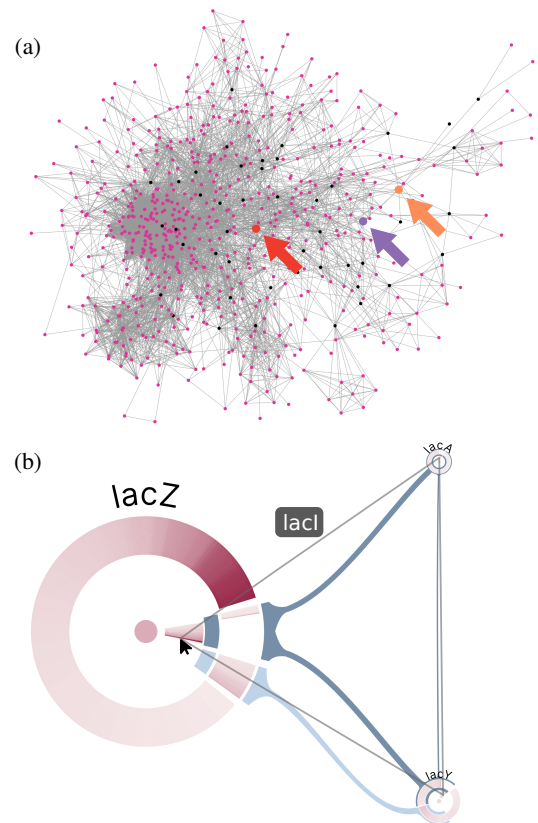**Figure 9:** *Visualizing the lac operon of E. coli using a node-link diagram created with* Cytoscape *(a). lacZ, lacA, and lacY are colored in* ●, ● *and* ●, *respectively. Nodes of distance one are colored* ● *while the ones of distance two are colored* ●. *The network consists of 653 nodes and 8,435 edges. Only interactions with a confidence score higher than 0.75 were considered. Visualizing the same proteins in* ProtEGOnist *(b). The node corresponding to lacI is hovered and shown in all ego-graphs.*

which most degree-2 alters are unique. This shows that lacZ also interacts with proteins not directly involved in the *lac* operon, indicating that it has a more central role in the PPI network compared to the other two proteins lacA and lacY. In contrast to lacZ, lacA has no unique alters, indicating a role more restricted to the operon.

From the band coloring, we can conclude that a large proportion of proteins is shared between all three ego-graphs. Notably, most of them have a distance of one to lacZ, while they have a distance of two to the other proteins. In fact, by hovering over the proteins, we find that the only degree-1 alter shared by all three proteins is lacI, which serves as the repressor for the operon.

Using the degree-2 alters, more distant associations can be investigated, for example, the relationship of the lac operon and the citrate cycle. The citrate cycle is one of the central metabolic pathways providing energy to the cell. When comparing the ego graphs of *lacZ* and *aceF*, a pivotal enzyme in the citrate cycle, by investigating the respective degree-1 alters we can see that they only have multi-degree associations (Supplementary Figure S4).

4.3. Human *DeeProM* Protein-Protein Interaction Network

We used the current version of *ProtEGOnist* to analyze the dataset by Gonçalves *et al.* [GPC*22] originally provided for the Bio+MedVis Challenge 2023. Proteins are common drug targets, i.e., drugs modify proteins to cause changes in the cell. In the case of cancer, drugs aim at disturbing the molecular pathways in cancer cells while leaving non-cancerous cells widely unharmed. Gonçalves *et al.* used a deep-learning approach to identify associations between drugs and proteins.

4.3.1. Analysis Using *ProtEGOnist*

For the overview, the ego-graphs of 91 proteins identified in the original publication to have relevant drug-protein associations were chosen (Supplementary Figure S1). This is an alternative approach to the other use cases, where the ego-graphs were selected via our set cover heuristic. Our analysis revealed that the union of proteins contained in these ego-graphs covers 57.3% of the proteins (nodes) and 91.6% of the interactions (edges) in the original PPI network. That is, the ego-graphs based on the proteins identified by DeeProM reflect most of the interactions in the original PPI network (**R1**). The metadata loaded into *ProtEGOnist* containted the drug-protein associations and the BRITE classification of the proteins.

Using *ProtEGOnist*, the results of DeeProM can be explored, opening up the black-box deep-learning model. Users can explore the proteins in the overview network in more detail, e.g., by selecting those associated with one drug of interest and viewing their BRITE functional classification. For the drug *Ara-G*, which prevents the elongation of DNA of cancer cells, four associated proteins are found in the overview network. Further inspection of these proteins in the subnetwork reveals three highly connected ego-graphs and a more distant one. The three most highly connected ego-graphs were selected as an ego-graph group (Figure 10). The lesser connected protein *SMARCC1* has been identified as a suppressor in some types of cancers [XLY*21], while the others act as possible drug targets or cancer biomarkers [YPY*23; WQH*20; XNK22]. All three proteins are associated with the BRITE class *Spliceosome*, i.e., these proteins are involved in the maturation of mRNA before translation [WWL09]. This association is even more prominent when inspecting the proteins *PPIH* and *RBM39* using the radar chart (Supplementary Figure S3, **R4**). In the ego-graph group, similarities of the highly connected ego-graphs can be explored in more detail by viewing shared proteins when selecting the intersection of all three ego-graphs (**R3**). From this point, users could, e.g., continue analyzing the found proteins using the KEGG pathway annotations to see in detail how they relate functionally.

4.3.2. Expert Feedback

Due to the positive outcome of the Bio+MedVis Challenge 2023 [23a], we contacted the authors of the DeeProM dataset to get their expert feedback on *ProtEGOnist*. We demonstrated our application to three of them and got a very positive response. One of the authors volunteered to test our application and to provide feedback. It consisted first of the free exploration and visualization of the dataset using *ProtEGOnist*. Subsequently, we provided a structured questionnaire of ten questions based on the System Usability Scale (SUS) framework [Bro96] and further open-ended questions.
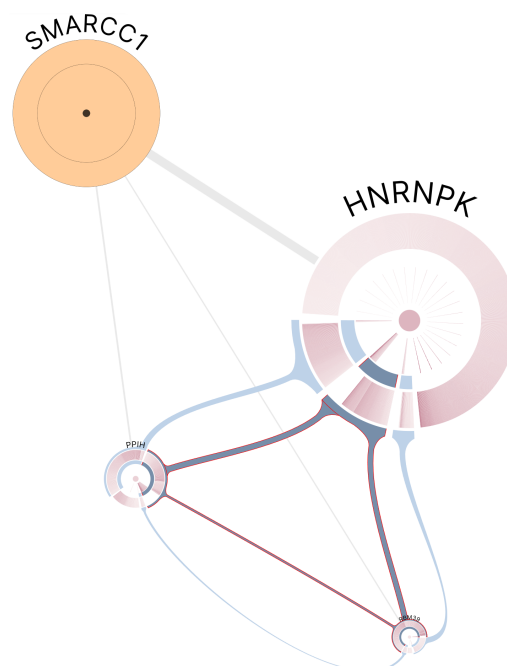

Figure 10: *Ego-graph subnetwork for all proteins of the overview network associated with the drug ARA-G. The three most highly connected ego-graphs (PPIH, HNRNPK, and RBM39) are shown as an ego-graph group. The intersection between all three ego-graphs has been selected (red outline).*

The expert that evaluated our tool had explored similar datasets using visualizations provided by *STRING* [SFW*15] and *Reactome* [FJM*18]. *ProtEGOnist* was assessed as slightly cumbersome, but they considered the complexity of our approach necessary, and its learning curve not steep. Overall, the expert enjoyed the exploration using ego-graphs. Nevertheless, they missed the integration of further information, e.g., which pathways a protein is involved in. As *ProtEGOnist* is easily extendable with additional arbitrary metadata, the pathway annotations were added as a new column to the selection table by adapting the input data.

5. Discussion

We presented *ProtEGOnist*, an interactive approach that applies ego-graphs to small-world interaction networks. Ego-graphs are a concept often encountered in real life, for example, when thinking about own friends or friends of friends in social networks. Therefore, the application of ego-graph in other domains, such as biological networks, employs a well-known mental model. In our case studies, we showed that this concept can be applied to datasets for a broad audience, like social networks, as well as to more domain-specific problems, such as PPI networks.

As shown in the co-author use case, the approach can be used to explore the network from an overview level down to detailed groups of ego-graphs. Furthermore, for the overview, we exploit the small-world property, which states that the maximal distance between two nodes is small compared to the network size. There-

fore, it is possible to show a relatively small set of ego-graphs as an overview while covering a large portion of the original interactions. Although the co-author network consists of 6,610 nodes and 22,220 edges, 100 ego-graphs suffice to cover more than 90 % of interactions and almost 80 % of nodes. Conceptually, even larger networks could be displayed using *ProtEGOnist*, as long as it fulfills the small world property. However, for very large networks, the overview might not be able to cover a large proportion of the network as the average minimum distance between the nodes could be too large. While 2-level ego-graphs are commonly used in practice, in the case of larger networks, ego-graphs with further levels of alters might be more appropriate.

In contrast to the *overview-first* approach of the first use case, in the second use case, we started the analysis with previous knowledge and analyzed the *lac* operon of *E. coli* in the context of the network. Here, we demonstrated the scalability of the glyph layout. In contrast to the conventional node-link diagram, *ProtEGOnist* allows us to immediately assess the size of the ego-graphs, and thus the centrality of the protein in the network. Even though the *lacZ* ego-graph contains 620 nodes, the layout effectively groups the nodes into sections containing unique or shared nodes, and alter levels. The sorting within the sections subsets the data even further and visualizes the distribution of the node degrees. We also illustrated the usefulness of 2-level ego-graphs for inspecting distant associations.

As exemplified in the third use case, our approach can easily be used with a user-defined set of nodes in the ego-graph overview network. Moreover, this example shows how network exploration in *ProtEGOnist* can be enhanced with different kinds of metadata. For PPI networks, even more data can be included in the analysis for visualization in the radar char or the aggregated ego-graphs of the subnetwork. For example, further omics data, like gene expression data or genomic data on mutations, could be used to analyze the proteins in more detail. This flexibility concerning the input data shows that *ProtEGOnist* can be generalized to a wide variety of application areas in which the small world property is fulfilled, e.g., linguistics [CS01], computer networking [XG03], or transportation networks [LM01].

Apart from *ProtEGOnist*, only few approaches have been proposed for comparing ego-graphs. However, their underlying goals are only remotely related to our approach. Out of those approaches, many are tailored to the visualization of dynamic networks. *EgoLines* [ZGC*16], among others [SWW*15; FMW*21], utilizes a linear layout of subsequent stages of the same ego-graph for a direct comparison of a stage with its predecessor or successor. *ProtEGOnist* aims at comparing different ego-graphs in a non-dynamic context. While the comparison in dynamic networks is often focused on the gradual changes of a single ego-graph, the differences when comparing multiple ego-graphs can be substantial. *EgoComp* [LGD*17] tackles this task for two ego-graphs. Our tool extends this approach to compare up to three ego-graphs simultaneously and puts them in context with other ego-graphs. We deliberately chose the comparison of three ego-graphs, as this allows us to use a layout of the setwise intersections without edge crossings. Increasing the amount of ego-graphs would be possible but incur edge crossings. Moreover, due to the usage of bands instead of a

conventional node-link diagram, *ProtEGOnist* has a much clearer layout for large ego-graphs.

6. Outlook & Conclusion

In the future, we plan to extend *ProtEGOnist* with more ways to incorporate metadata into the analysis process. One direction would be to associate edges and bands with metadata and to include separate visualizations for metadata, as well as network filtering options based on the metadata.

Moreover, we plan to allow the upload of user-generated data. By providing a network structure, metadata, and, if available, a set of nodes of interest, our approach can then be used for many other small-world network cases, such as transportation networks. We also plan to generalize our approach to accept different distance metrics to substitute the Jaccard index as the distance value. One of the improvements we identified from the expert feedback was the lack of connection between the visualization components and the tabular view. To enhance this connection, we plan to include a *pop-up* interaction for the selection table, where it is shown automatically when selecting a specific ego-graph or intersections between ego-graphs in the ego-graph subnetwork.

With *ProtEGOnist*, we provide a layout focused on ego-graphs omitting edges and aggregating subnetworks into single glyphs. We believe that the novel layout is one of the main reasons that our approach was rated as being a bit cumbersome to use. Adding a conventional node-link diagram to visualize one or more ego-graphs (similar to STRING [vMHJ*03]) could support the exploration process by providing a less abstract visualization as a detail view. These node-link diagrams could be shown either for the currently selected subnetwork or for a single ego-graphs, similar to BioLinker [DMF17]. Without specialized layout techniques, however, even a small number of ego-graphs can lead to unreadable visualizations due to a high number of nodes and edges (Figure 9 a).

To conclude, *ProtEGOnist* fills a gap in the network research space by combining established concepts for the analysis of small-world networks to a novel visualization approach. While it was initially intended for a specific domain with well-defined tasks based on the Bio+MedVis Challenge 2023, we show that this approach is applicable to different small-world networks from various domains.

References

[10] *Flask — A Micro Web Framework Written in Python*. 2010 7.

[23a] *Bio+MedVis Challenge @ IEEE VIS*. http://biovis.net/2023/biovisChallenges_vis/. 2023 2, 3, 9.

[23b] *MUI: The React Component Library You Always Wanted*. 2023 7.

[ADM*19] ANGORI, L., DIDIMO, W., MONTECCHIANI, F., et al. "ChordLink: A New Hybrid Visualization Model". *Graph Drawing and Network Visualization*. Ed. by ARCHAMBAULT, D. and TÓTH, C. D. Vol. 11904. Cham: Springer International Publishing, 2019, 276–290. ISBN: 978-3-030-35801-3 978-3-030-35802-0. DOI: 10.1007/978-3-030-35802-0_22 3.

[AMA07] ARCHAMBAULT, D., MUNZNER, T., and AUBER, D. "Topo-Layout: Multilevel Graph Layout by Topological Features". *IEEE Transactions on Visualization and Computer Graphics* 13.2 (Mar. 2007), 305–317. ISSN: 1077-2626. DOI: 10.1109/TVCG.2007.46 2.

[ARR*14] ALI, W., RITO, T., REINERT, G., et al. "Alignment-Free Protein Interaction Network Comparison". *Bioinformatics* 30.17 (Sept. 2014), i430–i437. ISSN: 1367-4811, 1367-4803. DOI: 10.1093/bioinformatics/btu447 2.

[BLGS06] BILGIC, M., LICAMELE, L., GETOOR, L., and SHNEIDERMAN, B. "D-Dupe: An Interactive Tool for Entity Resolution in Social Networks". *2006 IEEE Symposium On Visual Analytics Science And Technology*. Baltimore, MD, USA: IEEE, Oct. 2006, 43–50. ISBN: 978-1-4244-0591-6. DOI: 10.1109/VAST.2006.261429 3.

[BO04] BARABÁSI, A.-L. and OLTVAI, Z. N. "Network Biology: Understanding the Cell's Functional Organization". *Nature Reviews Genetics* 5.2 (Feb. 2004), 101–113. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg1272 2.

[BOH11] BOSTOCK, M., OGIEVETSKY, V., and HEER, J. "D³ Data-Driven Documents". *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), 2301–2309. ISSN: 1941-0506. DOI: 10.1109/TVCG.2011.185 7.

[BP10] BRANDES, U. and PICH, C. "More Flexible Radial Layout". *Graph Drawing*. Ed. by HUTCHISON, D., KANADE, T., KITTLER, J., et al. Vol. 5849. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, 107–118. ISBN: 978-3-642-11804-3 978-3-642-11805-0. DOI: 10.1007/978-3-642-11805-0_12 3.

[Bro96] BROOKE, J. *SUS -A Quick and Dirty Usability Scale*. 1996 9.

[BVKW12] BURCH, M., VEHLOW, C., KONEVTSOVA, N., and WEISKOPF, D. "Evaluating Partially Drawn Links for Directed Graph Edges". *Graph Drawing*. Ed. by VAN KREVELD, M. and SPECKMANN, B. Vol. 7034. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, 226–237. ISBN: 978-3-642-25877-0 978-3-642-25878-7. DOI: 10.1007/978-3-642-25878-7_22 3.

[CS01] CANCHO, R. F. I. and SOLÉ, R. V. "The Small World of Human Language". *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268.1482 (Nov. 2001), 2261–2265. ISSN: 0962-8452, 1471-2954. DOI: 10.1098/rspb.2001.1800 10.

[DCD14] DÖRK, M., COMBER, R., and DADE-ROBERTSON, M. "Monadic Exploration: Seeing the Whole through Its Parts". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto Ontario Canada: ACM, Apr. 2014, 1535–1544. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557083 6.

[DLR09] DRAPER, G. M., LIVNAT, Y., and RIESENFELD, R. F. "A Survey of Radial Methods for Information Visualization". *IEEE Transactions on Visualization and Computer Graphics* 15.5 (Sept. 2009), 759–776. ISSN: 1941-0506. DOI: 10.1109/TVCG.2009.23 3.

[DMF17] DANG, T., MURRAY, P., and FORBES, A. "BioLinker: Bottom-up Exploration of Protein Interaction Networks". *2017 IEEE Pacific Visualization Symposium (PacificVis)*. Apr. 2017, 265–269. DOI: 10/gm5wzh 3, 10.

[DMGJ19] DONCHEVA, N. T., MORRIS, J. H., GORODKIN, J., and JENSEN, L. J. "Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data". *Journal of Proteome Research* 18.2 (Feb. 2019), 623–632. ISSN: 1535-3893, 1535-3907. DOI: 10.1021/acs.jproteome.8b00702 8.

[DS13] DUNNE, C. and SHNEIDERMAN, B. "Motif Simplification: Improving Network Visualization Readability with Fan, Connector, and Clique Glyphs". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Paris France: ACM, Apr. 2013, 3247–3256. ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2466444 3.

[Fac13] FACEBOOK. *React - A JavaScript Library for Building User Interfaces*. 2013 7.

[FAM23] FILIPOV, V., ARLEO, A., and MIKSCH, S. "Are We There Yet? A Roadmap of Network Visualization from Surveys to Task Taxonomies". *Computer Graphics Forum* 42.6 (Sept. 2023), e14794. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.14794 2, 3.

[FC08] FOWLER, J. H. and CHRISTAKIS, N. A. "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis over 20 Years in the Framingham Heart Study". *BMJ* 337.dec04 2 (Dec. 2008), a2338–a2338. ISSN: 0959-8138, 1468-5833. DOI: 10.1136/bmj.a2338 2.

[FHQ11] FARRUGIA, M., HURLEY, N., and QUIGLEY, A. "Exploring Temporal Ego Networks Using Small Multiples and Tree-ring Layouts". (2011) 3.

[FJM*18] FABREGAT, A., JUPE, S., MATTHEWS, L., et al. "The Reactome Pathway Knowledgebase". *Nucleic Acids Research* 46.D1 (Jan. 2018), D649–D655. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkx1132 9.

[FMW*21] FU, K., MAO, T., WANG, Y., et al. "DyEgoVis: Visual Exploration of Dynamic Ego-Network Evolution". *Applied Sciences* 11.5 (Mar. 2021), 2399. ISSN: 2076-3417. DOI: 10.3390/app11052399 3, 10.

[GPC*22] GONÇALVES, E., POULOS, R. C., CAI, Z., et al. "Pan-Cancer Proteomic Map of 949 Human Cell Lines". *Cancer Cell* 40.8 (Aug. 2022), 835–849.e8. ISSN: 1535-6108, 1878-3686. DOI: 10.1016/j.ccell.2022.06.010 2, 9.

[HFM07] HENRY, N., FEKETE, J.-D., and MCGUFFIN, M. J. "NodeTrix: A Hybrid Visualization of Social Networks". *IEEE Transactions on Visualization and Computer Graphics* 13.6 (Nov. 2007), 1302–1309. ISSN: 1077-2626. DOI: 10.1109/TVCG.2007.70582 3.

[Hol06] HOLTEN, D. "Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data". *IEEE Transactions on Visualization and Computer Graphics* 12.5 (Sept. 2006), 741–748. ISSN: 1077-2626. DOI: 10.1109/TVCG.2006.147 3.

[HSS08] HAGBERG, A. A., SCHULT, D. A., and SWART, P. J. "Exploring Network Structure, Dynamics, and Function Using NetworkX". *Proceedings of the 7th Python in Science Conference*. Ed. by VAROQUAUX, G., VAUGHT, T., and MILLMAN, J. Pasadena, CA USA, 2008, 11–15 7.

[HTLF20] HENDRICKS, G., TKACZYK, D., LIN, J., and FEENEY, P. "Crossref: The Sustainable Source of Community-Owned Scholarly Metadata". *Quantitative Science Studies* 1.1 (Feb. 2020), 414–427. ISSN: 2641-3337. DOI: 10.1162/qss_a_00022 7.

[IHK*17] ISENBERG, P., HEIMERL, F., KOCH, S., et al. "Vispubdata.Org: A Metadata Collection About IEEE Visualization (VIS) Publications". *IEEE Transactions on Visualization and Computer Graphics* 23.9 (Sept. 2017), 2199–2206. ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2615308 2, 7.

[KAG*07] KANEHISA, M., ARAKI, M., GOTO, S., et al. "KEGG for Linking Genomes to Life and the Environment". *Nucleic Acids Research* 36.Database (Dec. 2007), D480–D484. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkm882 6.

[Kar29] KARINTHY, F. "Láncszemek". *Minden másképp van* (1929) 2.

[Kat23] KATO, D. *Jotai — Primitive and Flexible State Management for React*. 2023 7.

[KLS*19] KOVÁCS, I. A., LUCK, K., SPIROHN, K., et al. "Network-Based Prediction of Protein Interactions". *Nature Communications* 10.1 (Mar. 2019), 1240. ISSN: 2041-1723. DOI: 10.1038/s41467-019-09177-y 2.

[KSP23] KALE, B., SUN, M., and PAPKA, M. E. "The State of the Art in Visualizing Dynamic Multivariate Networks". *Computer Graphics Forum* 42.3 (June 2023), 471–490. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.14856 3.

[KV12] KORTE, B. and VYGEN, J. *Combinatorial Optimization: Theory and Algorithms*. Vol. 21. Algorithms and Combinatorics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-24487-2 978-3-642-24488-9. DOI: 10.1007/978-3-642-24488-9 6.

[LGD*17] LIU, D., GUO, F., DENG, B., et al. "egoComp: A Node-Link-Based Technique for Visual Comparison of Ego-Networks". *Information Visualization* 16.3 (July 2017), 179–189. ISSN: 1473-8716. DOI: 10.1177/1473871616667632 3, 5, 10.

[LM01] LATORA, V. and MARCHIORI, M. "Efficient Behavior of Small-World Networks". *Physical Review Letters* 87.19 (Oct. 2001), 198701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.87.198701 10.

[LPP*06] LEE, B., PLAISANT, C., PARR, C. S., et al. "Task Taxonomy for Graph Visualization". *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. BELIV '06. Venice Italy: ACM, May 2006, 1–5. ISBN: 978-1-59593-562-5. DOI: 10.1145/1168149.1168168 3.

[Mil67] MILGRAM, S. "The Small World Problem". *Psychology today* 2.1 (1967), 60–67 2.

[NBW06] NEWMAN, M., BARABÁSI, A.-L., and WATTS, D. J. *The Structure and Dynamics of Networks*. Vol. 12. Princeton Studies in Complexity. Princeton University Press, 2006. ISBN: 978-1-4008-4135-6. DOI: 10.1515/9781400841356 2.

[NMSL19] NOBRE, C., MEYER, M., STREIT, M., and LEX, A. "The State of the Art in Visualizing Multivariate Networks". *Computer Graphics Forum* 38.3 (June 2019), 807–832. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.13728 2, 3.

[NOB14] NOCAJ, A., ORTMANN, M., and BRANDES, U. "Untangling Hairballs: From 3 to 14 Degrees of Separation". *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Ed. by BAYRO-CORROCHANO, E. and HANCOCK, E. Vol. 8827. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, 101–112. ISBN: 978-3-319-12567-1 978-3-319-12568-8. DOI: 10.1007/978-3-662-45803-7_9 2.

[Pel10] PELLIZZARI, M. "Do Friends and Relatives Really Help in Getting a Good Job?". *ILR Review* 63.3 (Apr. 2010), 494–510. ISSN: 0019-7939, 2162-271X. DOI: 10.1177/001979391006300307 2.

[PHW17] PORTENOY, J., HULLMAN, J., and WEST, J. D. "Leveraging Citation Networks to Visualize Scholarly Influence Over Time". *Frontiers in Research Metrics and Analytics* 2 (Nov. 2017), 8. ISSN: 2504-0537. DOI: 10.3389/frma.2017.00008 3.

[QHS*15] QINGSONG LIU, HU, Y., SHI, L., et al. "EgoNetCloud: Event-based Egocentric Dynamic Network Visualization". *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Chicago, IL, USA: IEEE, Oct. 2015, 65–72. ISBN: 978-1-4673-9783-4. DOI: 10.1109/VAST.2015.7347632 3.

[Rei10] REITZ, F. *A Framework for an Ego-centered and Time-aware Visualization of Relations in Arbitrary Data Repositories*. Sept. 2010. arXiv: 1009.5183 [cs] 3.

[SA06] SHNEIDERMAN, B. and ARIS, A. "Network Visualization by Semantic Substrates". *IEEE Transactions on Visualization and Computer Graphics* 12.5 (Sept. 2006), 733–740. ISSN: 1077-2626. DOI: 10.1109/TVCG.2006.166 3.

[SFW*15] SZKLARCZYK, D., FRANCESCHINI, A., WYDER, S., et al. "STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life". *Nucleic Acids Research* 43.Database issue (Jan. 2015), D447–452. ISSN: 1362-4962. DOI: 10.1093/nar/gku1003 2, 3, 8, 9.

[Shn96] SHNEIDERMAN, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". *Proceedings 1996 IEEE Symposium on Visual Languages*. Boulder, CO, USA: IEEE Comput. Soc. Press, 1996, 336–343. ISBN: 978-0-8186-7508-9. DOI: 10.1109/VL.1996.545307 3.

[SKL*14] STOLPER, C. D., KAHNG, M., LIN, Z., et al. "GLO-STIX: Graph-Level Operations for Specifying Techniques and Interactive eXploration". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), 2320–2328. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2014.2346444 2.

[SMO*03] SHANNON, P., MARKIEL, A., OZIER, O., et al. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks". *Genome Research* 13.11 (Nov. 2003), 2498–2504. ISSN: 1088-9051. DOI: 10/b7kgpg 2, 8.

[Spr99] SPREEN, M. "Sampling Personal Network Structures: Statistical Inference in Ego-Graphs". PhD thesis. s.n. / University of Groningen, 1999 2.

[SWW*15] SHI, L., WANG, C., WEN, Z., et al. "1.5D Egocentric Dynamic Network Visualization". *IEEE Transactions on Visualization and Computer Graphics* 21.5 (May 2015), 624–637. ISSN: 1077-2626. DOI: 10.1109/TVCG.2014.2383380 3, 10.

[VBW17] VEHLOW, C., BECK, F., and WEISKOPF, D. "Visualizing Group Structures in Graphs: A Survey". *Computer Graphics Forum* 36.6 (Sept. 2017), 201–225. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.12872 2.

[vMHJ*03] Von MERING, C., HUYNEN, M., JAEGGI, D., et al. "STRING: A Database of Predicted Functional Associations between Proteins". *Nucleic Acids Research* 31.1 (2003), 258–261. DOI: 10.1093/nar/gkg034 3, 10.

[vP09] VAN HAM, F. and PERER, A. ""Search, Show Context, Expand on Demand": Supporting Large Graph Exploration with Degree-of-Interest". *IEEE Transactions on Visualization and Computer Graphics* 15.6 (Nov. 2009), 953–960. ISSN: 1941-0506. DOI: 10.1109/TVCG.2009.108 4.

[VRW13] VEHLOW, C., REINHARDT, T., and WEISKOPF, D. "Visualizing Fuzzy Overlapping Communities in Networks". *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), 2486–2495. ISSN: 1077-2626. DOI: 10.1109/TVCG.2013.232 3.

[VV14] VAN DEN ELZEN, S. and VAN WIJK, J. J. "Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 2014), 2310–2319. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2014.2346441 2, 3.

[Wil99] WILLS, G. J. "NicheWorks—Interactive Visualization of Very Large Graphs". *Journal of Computational and Graphical Statistics* 8.2 (June 1999), 190–212. ISSN: 1061-8600, 1537-2715. DOI: 10.1080/10618600.1999.10474810 3.

[WPZ*16] WU, Y., PITIPORNVIVAT, N., ZHAO, J., et al. "egoSlider: Visual Analysis of Egocentric Network Evolution". *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Jan. 2016), 260–269. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2468151 3.

[WQH*20] WANG, Z., QIU, H., HE, J., et al. "The Emerging Roles of hn-RNPK". *Journal of Cellular Physiology* 235.3 (Mar. 2020), 1995–2008. ISSN: 0021-9541, 1097-4652. DOI: 10.1002/jcp.29186 9.

[WS98] WATTS, D. J. and STROGATZ, S. H. "Collective Dynamics of 'Small-World' Networks". *Nature* 393.6684 (June 1998), 440–442. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/30918 2.

[WWL09] WAHL, M. C., WILL, C. L., and LÜHRMANN, R. "The Spliceosome: Design Principles of a Dynamic RNP Machine". *Cell* 136.4 (Feb. 2009), 701–718. ISSN: 00928674. DOI: 10.1016/j.cell.2009.02.009 9.

[XBBY11] XU, K., BEZAKOVA, I., BUNIMOVICH, L., and YI, S. V. "Path Lengths in Protein–Protein Interaction Networks and Biological Complexity". *PROTEOMICS* 11.10 (May 2011), 1857–1867. ISSN: 1615-9853, 1615-9861. DOI: 10.1002/pmic.201000684 2.

[XG03] XIAO FAN WANG and GUANRONG CHEN. "Complex Networks: Small-world, Scale-Free and Beyond". *IEEE Circuits and Systems Magazine* 3.1 (2003), 6–20. ISSN: 1531-636X. DOI: 10.1109/MCAS.2003.1228503 10.

[XLY*21] XIAO, Z.-M., LV, D.-J., YU, Y.-z., et al. "SMARCC1 Suppresses Tumor Progression by Inhibiting the PI3K/AKT Signaling Pathway in Prostate Cancer". *Frontiers in Cell and Developmental Biology* 9 (June 2021), 678967. ISSN: 2296-634X. DOI: 10.3389/fcell.2021.678967 9.

[XNK22] XU, Y., NIJHUIS, A., and KEUN, H. C. "RNA-binding Motif Protein 39 (RBM39): An Emerging Cancer Target". *British Journal of Pharmacology* 179.12 (June 2022), 2795–2812. ISSN: 0007-1188, 1476-5381. DOI: 10.1111/bph.15331 9.

[XWH*23] XUE, M., WANG, Y., HAN, C., et al. "Target Netgrams: An Annulus-Constrained Stress Model for Radial Graph Visualization". *IEEE Transactions on Visualization and Computer Graphics* 29.10 (Oct. 2023), 4256–4268. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2022.3187425 3.

[YBQY14] YANG, R., BAI, Y., QIN, Z., and YU, T. "EgoNet: Identification of Human Disease Ego-Network Modules". *BMC Genomics* 15.1 (Dec. 2014), 314. ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-314 3.

[YPY*23] YE, J., PANG, Y., YANG, X., et al. "PPIH Gene Regulation System and Its Prognostic Significance in Hepatocellular Carcinoma: A Comprehensive Analysis". *Aging* (Oct. 2023). ISSN: 1945-4589. DOI: 10.18632/aging.205134 9.

[ZGC*16] ZHAO, J., GLUECK, M., CHEVALIER, F., et al. "Egocentric Analysis of Dynamic Networks with EgoLines". *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, 5003–5014. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858488 2, 3, 10.

[ZPYQ13] ZHOU, H., PANPAN XU, YUAN, X., and QU, H. "Edge Bundling in Information Visualization". *Tsinghua Science and Technology* 18.2 (Apr. 2013), 145–156. ISSN: 1007-0214. DOI: 10.1109/TST.2013.6509098 3.