

A bi-objective network design approach for discovering functional modules linking Golgi apparatus fragmentation and neuronal death

Eduardo Álvarez-Miranda ^{*1}, Hesso Farhan^{†2,3}, Martin Luipersbeck ^{‡4}, and Markus Sinnl^{§4}

¹*Department of Industrial Engineering, Universidad de Talca, Curicó, Chile*

²*Biotechnology Institute Thurgau, Kreuzlingen, Switzerland*

³*Department of Biology, University of Konstanz, Konstanz, Germany*

⁴*Department of Statistics and Operations Research, Faculty of Business, Economics and Statistics, University of Vienna, Vienna, Austria*

Abstract

Experimental records show the existence of a biological linkage between neuronal death and Golgi apparatus fragmentation. The comprehension of such linkage should help to understand the dynamics undergoing neurological damage caused by diseases such as Alzheimer's disease or amyotrophic lateral sclerosis.

In this paper, the bi-objective minimum cardinality bottleneck Steiner tree problem along with an ad-hoc exact algorithm are proposed to study such phenomena. The proposed algorithm is based on integer programming and the so-called ϵ -constraint method. A key feature of the devised approach is that it allows an efficient integer programming formulation of the problem.

The obtained results show that it is possible to obtain additional evidence supporting the hypothesis that alterations of the Golgi apparatus structure and neuronal death interact through the biological mechanisms underlying the outbreak and progression of neurodegenerative diseases. Moreover, the function of cellular response to stress as a biological linkage between these phenomena is also investigated further. Complementary, computational results on a synthetic dataset are also provided with the aim of reporting the performance of the proposed algorithm.

1. Introduction and motivation

The rapid increase in the prevalence of neurodegenerative diseases, such as Alzheimer's diseases, urges for having a better understanding of their causes and the dynamics underlaying the observed symptomatology. According to data surveyed by [44], the prevalence of Alzheimer's disease in the 65+ years old population in Europe is up to 4.4%. In the United States, this number goes up to 9.7%. The situation in other areas of the world is not different. A recent study of the World Health Organization [see 14], reports that more than 30 million people in the world are currently affected by Alzheimer's disease, with around 5 million new cases occurring every year [5, 38, 46].

It is widely accepted that complex molecular networks and cellular pathways are involved in disease susceptibility and disease progression. Nowadays, one of the biggest challenges is to understand how the genes comprising these networks function in a coordinated manner in the presence of a disease [see, .e.g., 54, 55].

^{*}ealvarez@utalca.cl

[†]hesso.farhan@uni-konstanz.de

[‡]martin.luipersbeck@univie.ac.at

[§]markus.sinnl@univie.ac.at

In the last decade, starting with the seminal work of [26], *network optimization-based approaches* have been proposed to analyze the complex relationships between proteins, genes and/or metabolic reactions [see, .e.g., 3, 12, 24, 39, 57, and the references therein]. Roughly speaking, biologists have realized that well-known network optimization problems such as theSteiner Tree Problem (StT), the Prize-Collecting Steiner Tree Problem (PCStT), or the Maximum Weight Connected Subgraph Problem (MWCS) [1, 18], enable reasonable representations of the dynamics among proteins when analyzing a particular biological process. More precisely, the solutions obtained from those problems can be interpreted as *active subnetworks* or *functional modules*, i.e., coordinated protein clusters that govern a specific biological function within a much larger set of interactions. This a prominent example of Operations Research (OR) tools applied to Bioinformatics.

Considering the experience regarding the application of OR techniques in the analysis of biological processes, the focus of this paper is on finding more evidence respect to how, and to what extent, structural rearrangements of the Golgi apparatus and neuronal death, are related through signaling pathways relevant for neurodegeneration. Although is known that the Golgi apparatus fragments in the pathogenesis of neurodegeneration, the mechanism itself is still unknown exactly. The hypothesis is that the functional relationship among alterations in Golgi apparatus morphology and neuronal death is likely to be triggered within a particular set of proteins interactions associated with neurodegenerative diseases. Finding such set of interactions and analyzing its dynamics, shall help to have a better understanding of the process that neurodegenerative diseases activate, which ultimately lead towards neuronal death.

Contribution The main contribution of this work is to provide further insights in the identification and characterization of functional connections between regulators of the fragmentation of Golgi apparatus, and proteins associated with neuronal death. This is achieved by defining, modeling and solving a novel bi-objective network design problem coined as bi-objective minimum cardinality bottleneck Steiner tree problem (2MCBStT). The computational complexity of the proposed problem is investigated, and a specially tailored exact algorithm is designed to tackle the problem. The proposed algorithm is based on integer programming and the so-called ϵ -constraint method. A key feature of the devised approach is that it allows an efficient integer programming formulation of the problem. Note that this general recipe could be used to design more efficient approaches to tackle other bi-objective problems. Afterwards, the obtained results are analyzed in-silico by searching for relevant *enriched* pathways (connected proteins that present correlated behaviors in terms of expression levels, signal intensities, allele occurrences, etc.) associated with neurodegenerative diseases and other related biological processes. The performance of the proposed algorithm is also assesed using synthetic instances from the well-known SteinLib[32] of Steiner tree instances.

Outline The paper is organized as follows. In Section 2, the bi-objective minimum cardinality bottleneck Steiner tree problem is presented, along with the corresponding algorithmic approach based on integer linear programming and some details about bi-objective optimization. Basic biological background of the involved topics are presented in Section 3. The inference and analysis of a network linking neuronal death pathways with regulators of Golgi morphology is presented in Section 4, as well as the computational results for the synthetic benchmark instances. Finally, concluding remarks and paths for future research are presented in Section 5.

2. A bi-objective network design approach

The detailed description of the human protein-protein *interactome*, i.e., the whole set of protein-protein interactions in the human cell, is an ongoing project. Several databases provide accurate information about interactions among proteins at a genomic scale [see, e.g. 6, 45, 51]; this information is typically obtained by searching over tens of thousands of research articles and applying discrimination criteria for mining the appropriate data. [see, e.g. 52]. Intuitively, such interactome can be abstracted into a network $N = N(V, E)$, where V is the set of nodes (proteins) and E is the set of edges (interactions). Additionally, these databases typically provide a *confidence* of each of the reported interactions, i.e., how likely these interactions actually occur within the interactome. Consequently, consider that $c : E \rightarrow \mathbb{N}$ is the confidence function, such that

$c_{ij} > 0$, $\{i, j\} \in E$, is the confidence that there exists functional interaction between protein $i \in V$ and protein $j \in V$. More details about the interactome used in this paper can be found in Section 4.

At the protein level, any given biological process, say \mathcal{P} , may be characterized by a set of functional protein interactions. From a modeling perspective, this requires a comprehensive knowledge of the set of proteins, say $P \subset V$, that build such protein-protein interaction network. This has been proven to be a difficult task for biologists in the past. However, the emergence of systems biology approaches in the past decade largely facilitated the identification of so called *hits*; i.e., specialized proteins that are likely to regulate a biological process of interest (e.g., a type of cancer or an autoimmune response).

Nowadays, RNAi screening is a wide-spread method for finding such hits [see, e.g. 49, and the references therein]. Once that a list of hits is obtained via high-throughput experiments, one is typically interested in knowing if such proteins form a stand-alone functional module or if they are part of a more complex one. Likewise, it is typically important to know if P is part of signaling pathways that are associated to other processes, in order to have insights of how the studied process \mathcal{P} interacts with other biological functions.

Because the structure of the Golgi is altered in neurodegenerative diseases, a key question is whether it is possible to construct a network from regulators (hits) of Golgi structures together with regulators of neuronal cell death. If the answer is yes, then one would like to know whether it is possible to learn from this network new insights about neurodegeneration. Hence, suppose that, for a given representation N of the human protein interactome, a set $G \subset V$ is the set of *hits* associated with the first process, and a set $D \subset V$ is the set of *hits* associated with latter process. Given these two sets, an intuitive question would be: is it possible to reconstruct a functional protein network relating G and D on N ? In network optimization language, the question is whether exists a connected component, say $X = (V(X), E(X)) \subseteq N$, containing G and D ($G \subset V(X)$ and $D \subset V(X)$).

Generally speaking, X should be comprised by highly probable pathways, which depends on the confidence of the interactions $E(X)$. Although this approach maximizes the likelihood of the connecting pathways, it typically leads to a inferred subnetwork containing many *surplus* proteins that might not be known to be related with the process of interest. Hence, it is desirable that, at the same time, the number of such surplus proteins should be as small as possible. Aiming at achieving these two goals simultaneously, is crucial for the inference of protein networks that enable a more accurate biological analysis [see, e.g. 56, 57, for a detailed argumentation on these ideas].

In order to achieve this goal, the resulting problem as bi-objective network design problem is reformulated using the concept of Pareto optimality (see Definition 1). Hence, the goal is not to find a single solution, but a set of Pareto optimal solutions, which are then analyzed in detail with respect to their biological implications. Note that usually, two conflicting objectives are combined in a single objective using a weighted-sum approach, and the resulting single-objective problem is then solved. For instance, the PCStT combines cost and revenue into a single objective [see 39, for another recent network design problem, where two (conflicting) objectives are combined using a weighted sum].

Definition 1 [see, e.g., 11, 15] Let $z_1(X)$ and $z_2(X)$ be the two objective functions of a bi-objective optimization problem and assume both objective are in minimization form. A feasible solution X^* is Pareto optimal, iff there exists no other feasible solution X' such that $z_1(X') \leq z_1(X^*)$ and $z_2(X') \leq z_2(X^*)$ with at least one strict inequality. A feasible solution X^* is weakly Pareto optimal, iff there is no solution X' , where both inequalities are strict.

Then the task of finding a functional network X that exhibit the aforementioned properties can be accomplished by solving the following bi-objective minimum cardinality bottleneck Steiner tree problem (2MCBStT),

$$z_1(X) = \max \min_{e \in E(X)} c_e, \quad z_2(X) = \min |V(X)| \quad (1)$$

$$\text{s.t.} \quad G \subset V(X), D \subset V(X), \text{ and } X \subset N \text{ is connected.} \quad (2)$$

The first objective function $z_1(X)$ forces maximizing the *minimum* confidence of the interactions comprising the sought solution. Complementary, the second part objective $z_2(X)$ aims at minimizing the cardinality of

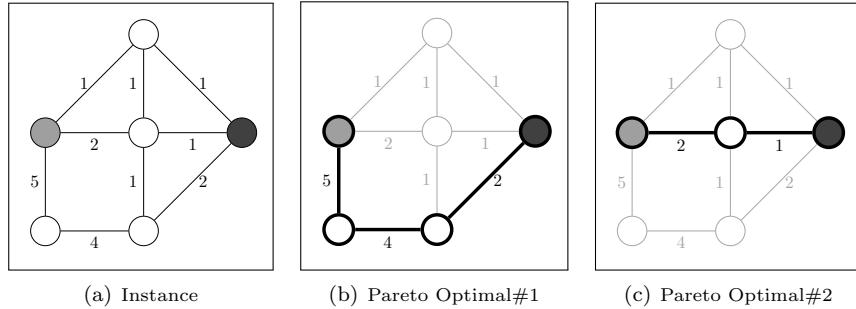


Figure 1: Example of an instance and two Pareto optimal solutions

the inferred network. The topological conditions required for X are represented in (2); any feasible network should contain the hits in sets G and D , and it must be a connected component of N .

In Figure 1(a) an example of an instance of the problem is displayed. Sets G and D are comprised by one node only; gray and dark gray, respectively. The solution X_1 shown in Figure 1(b) is Pareto optimal; there does not exist an alternative solution X'_1 such that $z_1(X'_1) > z_1(X_1) = 2$ and $z_2(X'_1) \leq z_2(X_1) = 4$. Likewise, the solution X_2 displayed in Figure 1(c) is also Pareto optimal; it is not possible to find another solution X'_2 such that $z_1(X'_2) \geq z_1(X_2) = 1$ and $z_2(X'_2) < z_2(X_2) = 3$.

Considering only objective function $z_1(X)$, one would end up with a bottleneck-type Steiner tree problem. Usually, bottleneck Steiner tree problems consider a $\min \max_{e \in E(X)} c_e$ objective function, and they can be solved in polynomial time [see, e.g., 13]. Likewise, the $z_1(X) = \max \min_{e \in E(X)} c_e$ objective function appearing in the 2MCBStT also induces a problem which can be solved by a polynomial time algorithm. Let $\epsilon = \max_{e \in E} c_e$. Consider the graph $N(\epsilon)$ obtained by deleting all edges with $c < \epsilon$. Check with a breadth-first search if there exists a connected subgraph $N' \subseteq N(\epsilon)$ which contains G and D . If yes, the process is over and the optimal solution value is ϵ ; if not, decrease ϵ to the next highest c_e and repeat the procedure, until a connected graph is obtained.

On the other hand, the problem induced by $z_2(X)$ corresponds to a special case of the node-weighted Steiner tree problem, which is NP-hard [see, e.g., 42]. Therefore, the 2MCBStT is NP-hard.

A point $(z_1(X^*), z_2(X^*)) \in \mathbb{R}^2$ corresponding to a Pareto optimal solution is called non-dominated. Observe that two Pareto optimal solutions X_1^* , X_2^* may give the same non-dominated points, i.e., $(z_1(X_1^*), z_2(X_1^*)) = (z_1(X_2^*), z_2(X_2^*))$. The set of all non-dominated points is called *Pareto front* (or also *Pareto frontier*). It is customary in multi-objective optimization to search for the Pareto front, i.e., to discover one Pareto optimal solution for every non-dominated point [see, e.g., 7, 35]. In the context of a biological problem as considered in this paper, it could be useful to discover more than one Pareto optimal solution for every point on the Pareto front to give a decision maker even more useful solutions to analyze. The presented approach is capable of doing so.

Finally, consider two non-negative values λ_1, λ_2 , which enable to combine the two objectives (1) into a single weighted-sum objective, i.e., $z(X) = \max (\lambda_1 \min_{e \in E(X)} c_e - \lambda_2 |V(X)|)$. The optimal solution of such model will also be in the Pareto front, since it can be determined by simply evaluating all Pareto optimal solutions with the given λ_1, λ_2 . This follows from a well-known result for multi-objective optimization [see, e.g., 11, 15].

In the remainder of this section, the algorithmic scheme to obtain the Pareto front of the 2MCBStT is presented.

2.1 Solution algorithm and ILP formulation

In the first part of this subsection a detailed description of the algorithm for solving the 2MCBStT is given. In the second part, an ILP formulation of the auxiliary problem that must be solved within the algorithm is presented along with details of the method to solve it.

Algorithm 1 ϵ -Constraint Method

```

1:  $Sol \leftarrow \emptyset$ 
2:  $\epsilon \leftarrow \min_{e \in E} c_e$ 
3:  $infeasible \leftarrow false$ 
4: while  $\epsilon \leq \max_{e \in E} c_e$  and  $infeasible = false$  do
5:   solve MinCardSTtT( $\epsilon$ )
6:   if MinCardSTtT( $\epsilon$ ) is feasible then
7:      $X^* \leftarrow$  optimal solution of MinCardSTtT( $\epsilon$ )
8:      $Sol \leftarrow Sol \cup X^*$ 
9:      $\epsilon \leftarrow z_1(X^*) + \Delta$ 
10:  else
11:     $infeasible \leftarrow true$ 

```

Finding the Pareto front: the ϵ -constraint method To discover the Pareto front, the ϵ -constraint method is used [see, e.g., 22]. The ϵ -constraint method is an iterative exact solution method for multi-objective optimization problems. Iterative means, that a single-objective problem is repeatedly solved. In the bi-objective case, one objective is put as constraint, while the other is kept. For the problem under study, by putting z_1 as constraint and keeping z_2 in the objective, one obtains the following minimum-cardinality Steiner tree problem (MinCardStT) with ϵ -constraint (MinCardSTtT(ϵ)),

$$\min |V(X)| \tag{3}$$

$$\text{s.t. } G \subset V(X), D \subset V(X), \text{ and } X \subset N \text{ is connected} \tag{4}$$

$$\min_{e \in E(X)} c_e \geq \epsilon. \tag{5}$$

The idea of the method is to start at one end of the Pareto front, and by systematically modifying (i.e., increasing or decreasing) the value of ϵ and resolving the problem MinCardSTtT(ϵ), the whole Pareto front is determined. One can easily see that in this problem the values that $z_1(X)$ can take are contained between $\min_{e \in E} c_e$ and $\max_{e \in E} c_e$. Let Δ be the greatest-common-divisor of all confidence values c , and let Sol be the set containing the solutions on the Pareto front. Additionally, let X^* be an optimal solution obtained by solving MinCardSTtT(ϵ). The problem MinCardSTtT(ϵ) is solved by formulating it as an ILP problem, and using a branch-and-cut approach to tackle it (further details are given in the second part of this subsection).

Algorithm 1 gives an overview of the ϵ -constraint method applied to the 2MCBStT. In line 5 and instance of the MinCardSTtT(ϵ) is solved for a value of ϵ (which is initialized to $\min_{e \in E} c_e$ in line 2). If such instance can be solved, then a Pareto optimal solution X^* is found, and the value of ϵ is updated in line 9. The process is repeated as long as $\epsilon \leq \max_{e \in E} c_e$ and the MinCardSTtT remains feasible.

The Pareto front resulting from the application of Algorithm 1 to one of the instances considered in this paper is shown in Figure 2. The instance contains 19,245 nodes and 8,545,750 edges, and the confidence ranges from 1 to 1000; besides $|G \cup D| = 541$. Despite the enormous size of this network, the proposed method is able to compute in no more than 20 seconds each of the solutions of the front. From the plot in the figure one can see more than 20 points; each of them corresponds to a non-dominated pair $(z_1(X^*), z_2(X^*))$.

Despite the correctness of Algorithm 1, one issue must be considered; it allows to discover only one Pareto optimal solution for every non-dominated point. In order to find all Pareto optimal solutions, one can modify Algorithm 1 by incorporating an additional loop in which so-called *no-good* constraints are iteratively added to the MinCardSTtT(ϵ) ILP model. This enables to find all the minimum cardinality solutions associated with a given ϵ . The revised algorithm corresponds to Algorithm 2. As can be seen from this pseudocode, the main difference between the two algorithms corresponds to the addition of the loop in lines 12-19. Essentially, it is within this loop that the algorithm finds all Pareto optimal solutions associated with a given ϵ , i.e., all the minimum cardinality solutions X^* with the same objective value $z_2(X^*)$ that verify $\min_{e \in E(X^*)} c_e = \epsilon$. The details of how this is achieved are provided below.

Observe that there are many other exact methods for bi-objective optimization [see, e.g., 7, 35, and the references therein for recent computational methods]. Nonetheless, in this case the ϵ -constraint method

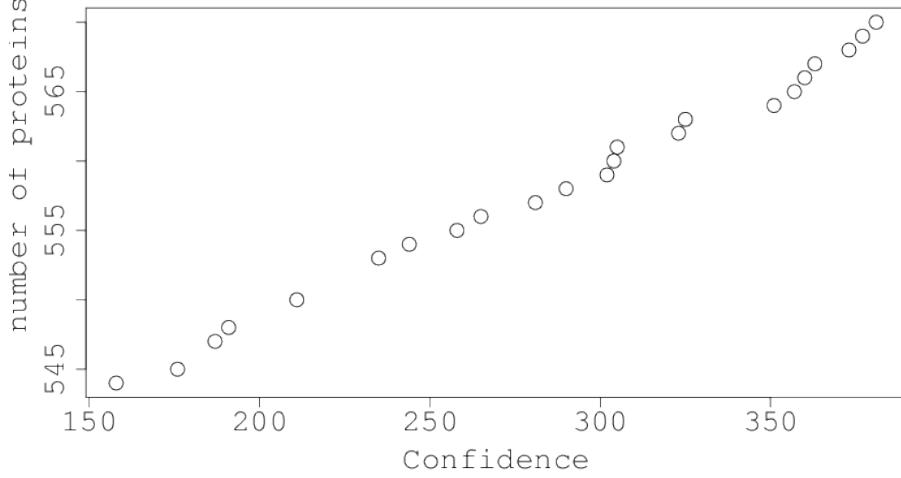


Figure 2: Pareto front of one of the studied biological networks

Algorithm 2 Modified ϵ -Constraint Method

```

1:  $Sol \leftarrow \emptyset$ 
2:  $\epsilon \leftarrow \min_{e \in E} c_e$ 
3:  $infeasible \leftarrow false$ 
4: while  $\epsilon \leq \max_{e \in E} c_e$  and  $infeasible = false$  do
5:    $Sol(\epsilon) \leftarrow \emptyset$ 
6:   solve MinCardSTtT( $\epsilon$ )
7:   if MinCardSTtT( $\epsilon$ ) is feasible then
8:      $X^* \leftarrow$  optimal solution of MinCardSTtT( $\epsilon$ )
9:      $Sol(\epsilon) \leftarrow X^*$ 
10:     $z^* = z_2(X^*)$ 
11:     $stop = false$ 
12:    repeat
13:      solve MinCardSTtT( $\epsilon, Sol(\epsilon)$ )
14:      if MinCardSTtT( $\epsilon, Sol(\epsilon)$ ) is feasible and  $z_2(X^*) = z^*$  then
15:         $X^* \leftarrow$  optimal solution of MinCardSTtT( $\epsilon, Sol(\epsilon)$ )
16:         $Sol(\epsilon) \leftarrow Sol(\epsilon) \cup X^*$ 
17:      else
18:         $stop = true$ 
19:      until  $stop = true$ 
20:    else
21:       $infeasible \leftarrow true$ 
22:     $Sol \leftarrow Sol \cup Sol(\epsilon)$ 
23:     $\epsilon \leftarrow z_1(X^*) + \Delta$ 

```

allows a convenient reformulation of the problem when modeling it as an ILP. Although the 2MCBStT is such that one objective is related to the edges in the solution, and the other objective is related to the nodes, the problem can be formulated using only node-variables. This is crucial when dealing with large-scale networks such as those encountered in biology (e.g., interactomes consisting of over a million of edges), since one can get rid of these millions (binary) edge-variables. More details about this formulation are presented below.

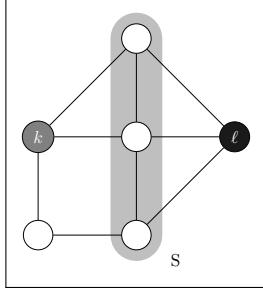


Figure 3: Example of an (k, ℓ) -node-separator

Solving the MinCardSTtT As already described, in each iteration of Algorithm 2, instances of the MinCardSTtT(ϵ) and MinCardSTtT($\epsilon, Sol(\epsilon)$) are solved. For the ILP formulation of these problems, the concept of *node-separators* is used in this paper to ensure connectivity [see, e.g., 2, 18, for more details on models based on node-separators]. Consider the following definition.

Definition 2 For two different nodes k and ℓ from V , a subset of nodes $S \subseteq V \setminus \{k, \ell\}$ is called a (k, ℓ) -node-separator if and only if after removing S from V there is no (k, ℓ) -path in N .

In Figure 3 an example of a node separator is shown; as can be seen, if the three nodes comprising set S (within the gray cover) are removed, then no path between k and ℓ can be established. Let $\mathcal{S}(k, \ell)$ denote the family of all (k, ℓ) -separators. Let $\mathcal{S}_\ell = \cup_{k \neq \ell} \mathcal{S}(k, \ell)$ be the family of all node separators with respect to $\ell \in V$ that it will be referred to as ℓ -separators.

Let $r \in T = G \cup D$ be an arbitrary node from T that will be regarded as *root* node. Let $N_\epsilon = (V, E_\epsilon)$ be the graph induced by a given ϵ , i.e., all edges $e \in E$ with $c_e < \epsilon$ are removed. For any $i \in V$, let $\mathcal{K}(\epsilon)r, i$ be the set of all (r, i) -node-separators in N_ϵ between r and i .

Finally, let $\mathbf{x} \in \{0, 1\}^{|V|}$ be vector of binary variables such that $x_i = 1$ if node $i \in V$ is taken in X , and $x_i = 0$ otherwise. Therefore, for a given vector \mathbf{x} and a given ϵ , the corresponding subnetwork $X = (V(X), E_\epsilon(X))$ is given by $V(X) = \{i \in V \mid x_i = 1\}$ and $E_\epsilon(X) = \{\{i, j\} \in E_\epsilon \mid i, j \in V(X), i \neq j\}$.

Using this notation, the ϵ -constraint MinCardSTtT, MinCardSTtT(ϵ), can be formulated as follows using only node-variables

$$z_2^*(\epsilon) = \min \sum_{i \in V} x_i \quad (\epsilon\text{-CARD.1})$$

$$\sum_{v \in K} x_v \geq 1, \quad \forall K \in \mathcal{K}(\epsilon)_{r,v}, \forall v \in T \setminus \{r\} \quad (\epsilon\text{-CARD.2})$$

$$x_v = 1, \quad \forall v \in T \quad (\epsilon\text{-CARD.3})$$

$$\mathbf{x} \in \{0, 1\}^{|V|}. \quad (\epsilon\text{-CARD.4})$$

Constraints (ϵ -CARD.2) ensure that there is a path between r and all other nodes in T , while (ϵ -CARD.3) impose that all nodes in T must be taken in the network.

Note that when using the above model within the algorithmic framework, one has to account with the fact that constraints (ϵ -CARD.2) are of an exponential number; therefore, they shall be separated and added in a cutting-plane fashion. Such separation is usually performed using a max-flow algorithm [see, e.g., 2, 18]. However, for large-scale networks such max-flow separation can very time-consuming [see the computational results section in 18]. Thus, the separation of (ϵ -CARD.2) is only performed when the branch-and-cut solver gives back an integer solution \mathbf{x}^* ; in this case the separation reduces to a breadth-first-search on the graph induced by \mathbf{x}^* .

Additionally, one can improve formulation (ϵ -CARD.1)-(ϵ -CARD.4), by adding the following so-called

flow-balance constraints;

$$\sum_{j:\{i,j\} \in E} x_j \geq 2x_i, \quad \forall i \in V \setminus T.$$

These constraints are not necessary for correctness of the model, but have been shown to be computationally helpful [18]. They are based on the observation that in an optimal solution, a non-terminal node will never be a leaf-node.

The above described model is used to solve the MinCardSTtT(ϵ) as appears in line 6 of Algorithm 2. Once this problem is solved and an optimal solution, say X^* , is found, one can attempt to find an alternative Pareto optimal solution (lines 12-19) by adding a so-called no-good cut. Such cut is given by

$$\sum_{i \in V(X \setminus X^*)} x_i \geq \kappa. \quad (\text{NG})$$

This constraint imposes that the next solution must differ by at least $\kappa \in \mathbb{Z}_{\geq 1}$ nodes with respect to X^* . If such solution exists and it induces the same objective function value than X^* (see line 14), then an additional no-good cut (NG) is generated and added (see lines 15 and 16), by appending the current solution X^* to the set $Sol(\epsilon)$. The process is repeated until no alternative Pareto optimal solution can be found (i.e., the problem with the added no-good cuts is either infeasible or gives a higher objective value). Note that for $\kappa = 1$, one would obtain *all* the solutions comprising the Pareto front.

3. Golgi fragmentation and neuronal death: Recent discoveries

The Golgi apparatus plays a crucial role in the cellular activity; it is a major sorting and dispatch station for protein trafficking. It receives proteins, processes them (e.g, by *attaching* sugar monomers to them), packages them into membrane-bound vesicles, and sends these vesicles either for extra-cellular secretion or for their use inside the cell [41]. Moreover, recent evidence has shown that the Golgi apparatus acts as a signaling *hub*; it responds to environmental factors, and it modulates the outcome of signaling cascades [see 8, 17].

Current biological evidence shows that the major part of the pool of neurons that comprise the human central nervous system is generated before birth. Later, neurons are post-mitotic, i.e. they do not divide anymore [see, e.g. 25]. Therefore, the brain has a limited capacity for regeneration [33]. Inappropriate neuronal cell death (i.e., inappropriate activation of the *apoptosis* process), may result from pathological causes related to neurodegenerative diseases such as Alzheimer’s disease, amyotrophic lateral sclerosis or Parkinson’s disease [58].

In the following, biological background and up-to-date discoveries regarding the studied processes will be provided. In particular, the elements comprising sets of hits D and G will be presented.

Regulators of neuronal death The proteins involved in neuron death are rather well identified [33]. Nowadays, there are several massive databases with constantly updated curated information regarding different types of biological processes; among them, death of neurons. In this paper, a total of 239 genes associated with this processes was obtained from the Gene Ontology database [21]. The list is available in Table 2 in the Appendix; these genes comprise set D (see §2). In Figure 5 in the Appendix one can see the function of these hits in the complete functioning of the cell; this network was produced using the expression analysis tools from the REACTOME project [45]. By looking at the highlighted clusters of the network, one can see that besides neuronal death (programmed cell death), these genes are involved, mainly, in the mechanism associated with developmental biology, immune system, disease and cellular response to stress. While the association with first three of these processes does not offer any real useful insight, the interplay role of these proteins in the cellular response to stress shall attract more efforts due to the possible connections with exogenous stimuli which might finally end up with apoptosis. In fact, this later issue is investigated in §4.

The Golgi apparatus fragmentation in neurodegenerative diseases The fragmentation of the Golgi apparatus has been observed in several neurodegenerative diseases [see 16, 43, and the references therein].

Functional linkages between amyotrophic lateral sclerosis (ALS) and Golgi fragmentation has been reported using ALS mice models [see, e.g., 4]. As a matter fact, it is well known that a mutation of the Superoxide Dismutase 1, Soluble (SOD1) gene is present in patients with ALS [see, e.g. 19]. Recent discoveries have shown that mutations of the OPTN gene are also associated with Golgi fragmentation in ALS patients [see, e.g. 29].

Likewise, efforts have been devoted for establishing how Alzheimer’s disease is associated with Golgi fragmentation. The characterization of a clear signaling pathway affecting the Golgi matrix proteins (mainly via phosphorylation) and its later destruction have been reported in [9, 27, 28]. The principal phenomenon studied in those paper is how a particular gene, the Cyclin-Dependent Kinase 5 (CDK5) intervenes in the structure of the Golgi apparatus by means of different mechanisms. Nonetheless, while the case of the CDK5 is reasonably well understood, there is lack of a broader and systematic understanding of more signaling pathways that regulate Golgi morphology under conditions induced by neurodegenerative diseases.

Discovering regulators of the Golgi morphology Over the last years, several studies have been carried out for understanding the genetic interactions underlying the process of rearrangements of the Golgi apparatus morphology. Recently, in [10, 40, 50] authors have identified genes that, when depleted, result in a structural alteration of the Golgi apparatus.

The first full-genome screen in human cells for regulators of secretion was presented in [50]. Although over 600 hits were identified, only 78 of them were tested for their impact on Golgi morphology (see Table 3 in the Appendix). In [10], the authors screened the human kinome and phosphatome in the seek for regulators of the Golgi. Their results show 178 genes involved with the regulation of the Golgi structure (see Table 4 in the Appendix). Besides this finding, the authors discovered enriched subnetworks associated with the stress activated mitogen protein kinase (MAPK). As said above, a more recent list of Golgi fragmentation hits has been provided in [40]. Although the main goal of that work was the identification of new regulators of cell migration, one can establish that 63 of them can be included in the list for proteins that regulate the Golgi morphology (see Table 5 in the Appendix).

These three lists comprise set G (see §2). In Figure 6 in the Appendix, similar to Figure 5, the function of these hits in the complete functioning of the cell is shown. As expected, these proteins are mainly involved in the signal transduction processes, and, with the exception of the immune system regulation, they seem to play a marginal role in the other processes considered in the REACTOME expression analysis.

4. Results and analysis

In this section, the results obtained when applying the method outlined in §2 to the data provided in §3 are presented. Complementary, the proposed bi-objective problem is also solved, on a dataset comprised of instances adapted from the SteinLib data collection [see 32], which are originally created for the Steiner tree problem. The computational experiments have been made using a single-core of an Intel E5-2670v2 with 2.5 GHz and 64 GB RAM using CPLEX 12.6 as branch-and-cut framework.

4.1 Results on biological data

As said before, the goal is to computationally infer a network linking the hits related to neuronal death (set D comprised by the elements in Table 2 in the Appendix), with the hits related to fragmentation of the Golgi apparatus (set G comprised by the elements in Tables 3, 4 and 5 in the Appendix). This linkage network is sought within an interactome N ; in this paper the underlying interactome was created by combining the information available in [6] and [51]. In its raw form, the retrieved human interactome was comprised by 19,245 proteins and 8,545,750 interactions. These interactions correspond to all those reported in these databases, i.e., either observed or predicted interactions by the methods of Neighborhood, Gene Fusion, Co-occurrence, Co-expression, Experiments, Databases, and Textmining [see 51, 52, for further details].

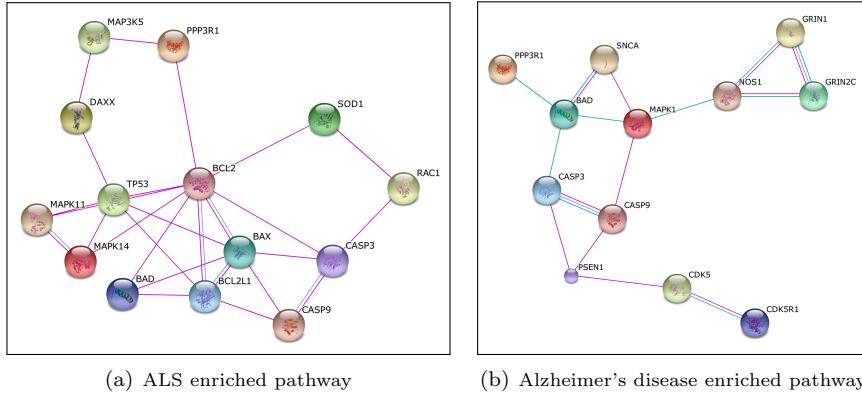


Figure 4: Enriched pathways associated with neurodegenerative diseases

Although in §2 some results using this large interactome have been reported, they are likely to fail in having a real biological meaning. This is due to the presence of many connections that do not have experimental evidence. Nonetheless, those experiments have been performed exclusively as a proof of concept and verify how effective is the algorithmic scheme when tackling such a big instance.

For the purposes of this work, only those interactions that have been experimentally proved to yield binding relations with a confidence greater than 500 have been considered (confidence ranges from 1 to 1000). This allows to obtain networks whose interactions have sufficient evidence of actually be present in the studied process, and it is a typical value use for this type of biological studies. This yields a connected interactome N comprised by 18,378 proteins and 1,056,914 interactions; which is still quite huge considering the benchmark instances of the Steiner tree and other related network design problems [18].

Seeking for relevant enriched pathways As explained in §2, when having multiple objective functions it is possible to find not only one, but many Pareto-optimal solutions. In particular, for the above described interactome N and considering the sets D and G described before, the proposed algorithm was able to obtain a Pareto front comprised of two points (P_1 and P_2); P_1 is associated with $z_1(X^*) = 500$ (confidence) and $z_2(X^*) = 587$ (cardinality), and P_2 is associated with $z_1(X^*) = 505$ and $z_2(X^*) = 588$. For each of these points, Algorithm 2 was used to compute 5 different Pareto optimal solutions using $\kappa = 10$. Hence 5 solutions were calculated verifying that the minimum confidence is 500 (resp. 505) and the cardinality is 587 (resp. 588), and all of them differ among each other by at least 10 nodes. Finding each of these 10 solutions required, in average, no more than 20 seconds.

For these two sets solutions, the enrichment pathway tool available in [51] was used to find relevant processes in each of the corresponding networks. In both cases, the analysis revealed the following outcomes: (i) a connected pathway associated with the ALS is enriched within the network with a p -value in the order of 10^{-15} ; (ii) a connected pathway associated with the Alzheimer's disease is enriched within the network with a p -value in the order of 10^{-15} ; (iii) the obtained networks presented exactly the same enriched pathways for both processes. Such small p -values allow to conclude that there is enough empirical evidence to support the hypothesis that both ALS and Alzheimer's disease are part of the intracellular dynamics binding Golgi apparatus fragmentation and neuronal death.

The pathway associated with ALS is shown in Figure 4(a). This subnetwork is comprised by proteins from sets G (MAPK11, MAPK14, PPP3R1, RAC1) and D (CASP9, MAP3K5, BCL2L1, CASP3, SOD1, BAX, BCL2, BAD, DAXX, TP53). This verifies that the ALS is indeed a functional linkage between Golgi apparatus fragmentation and neuronal death. Considering the recent discoveries [see, e.g. 47, and the references therein], these 14 proteins are all involved in the pathological development of ALS. Moreover, by analyzing the KEGG network diagram of the pathway associated to ALS [30]¹, which is shown in Figure 7 in

¹KEGG network diagrams are manually drawn detailed molecular interaction/reaction schemes corresponding to different

the Appendix, one can see that most of the reactive/interactive dynamics related to ALS are comprised by the 14 proteins found out within the solutions associated with P_1 and P_2 . This supports the hypothesis that some proteins associated with neurodegenerative diseases, such as ALS, comprise the functional relationship between alterations in Golgi apparatus morphology and neuronal death.

The connected component associated with enriched pathway corresponding to Alzheimer's disease found within the networks associated with P_1 and P_2 is shown in Figure 4(b). This subnetwork is comprised by proteins from sets G (GRIN2C, MAPK1, PPP3R1) and D (BAD, CASP3, CASP9, CDK5, CDK5R1, GRIN1, NOS1, SNCA, PSEN1). The existence of this pathway linking sets G and D is coherent with the evidence that Golgi apparatus fragmentation is part of processes triggered by Alzheimer's disease [9, 27, 28]. In Figure 8 in the Appendix, the KEGG diagram of the Alzheimer's diseases pathway is shown. One can see that the proteins found to be functionally linking G and D via Alzheimer's disease play all a key role in the development of this disease which ultimately leads to neuronal death.

Cellular response to stress: A biological linkage The previous discussion on the ALS and Alzheimer's disease pathways found within P_1 and P_2 , support the evidence that neuronal death and Golgi fragmentation shall be systematically related through, or because of, neurodegenerative diseases. The question that arises is whether there is any biological process that embeds along with the Golgi fragmentation and neurodegeneration into a tripartite function.

For answering this question a so-called BiNGO analysis was performed [36], in order to determine which Gene Ontology terms [21] are significantly overrepresented in the proteins comprising the solution associated with P_1 and P_2 . Obviously, the most overrepresented terms were all related with regulation of cell death. Nonetheless, one of the most statistically overrepresented processes corresponded to the regulation of cellular response to stress (p -value, in average, in the order of 10^{-40}); this was the case for P_1 and P_2 . This means that for each of the 10 solutions under consideration, an enriched submodule (i.e., a subnetwork) associated with the regulation of cellular response to stress can be retrieved after performing a BiNGO analysis. These subnetworks were actually different when considering different solutions. Out of the 10 retrieved modules, 2 were comprised by 105 nodes, 2 by 106 nodes, 3 by 107 nodes, and 3 by 108 nodes. While the union of these 10 sets of nodes was comprised by 111 different nodes, the intersection of them is actually comprised by 104. This shows that the 10 enriched modules are basically constituted by a common body of proteins; this is consistent with the observation made in §3 in relation to the presence of stress response regulators among apoptosis regulators. Moreover, this supports the evidence provided by [10] regarding the presence of enriched subnetworks associated with the stress activated MAPK pathways within the hits associated with Golgi apparatus fragmentation.

The detailed list of the 111 proteins encompassing these 10 subnetworks is reported in Table 6 in the Appendix. From this table one can see that 39 proteins belong exclusively to the set of hits associated with Golgi fragmentation, 59 belong exclusively to the set of hits associated with neuronal death, 4 belong to both sets, and 9 proteins are so-called Steiner nodes, i.e., surplus proteins required for connectivity of the sought networks.

The key of the result represented by the network in Figure 9 in the Appendix that it emphasizes the conjecture that (i) neurodegenerative diseases set up conditions that cells perceive as stressors, (ii) responsive mechanisms of the Golgi to these stimuli induce its structural rearrangement, and (iii) this finally triggers the death of the neuron. In other words, this unbiased way of analysis, allows to infer that cellular stress is a potential player in the crosswalk between Golgi structural alteration and neuronal cell death.

As said before, additional proteins were required to ensure the connection between the elements of set D and those of set G in each of the 10 networks under study. The list of these additional proteins and the number of networks in which they appeared is the following; GRB2 (appeared in 10 out of 10 networks), STAT5B (10 out of 10), HERPUD1 (8 out of 10), PPAR γ (6 out of 10), TOP2A (2 out of 10), ITPR1 (1 out of 10), NUP153 (1 out of 10), PPP1CA (1 out of 10), TRIM28 (1 out of 10). Clearly, the first 4 of these proteins are worth to be analyze since they systematically appear in the linkage among cellular response to stress, Golgi apparatus structure rearrangement and neuronal apoptosis. Understanding the nature and role of these proteins is decisive for learning from this linkage.

processes of interest; they are made available by the Kyoto Encyclopedia of Genes and Genome [KEGG 30]

HERPUD1 (also known as Herp) has been shown to enhance the so called Endoplasmatic Reticulum-associated degradation of proteins [31]. This is a process to degrade misfolded proteins that would otherwise stress the cell and result in its death. As a matter of fact, the involvement of HERPUD1 in neurodegeneration was shown previously by demonstrating that it enhances the formation of beta-amyloid [48]. However, the mechanism behind this effects remains unclear and requires further investigations. Strikingly, HERPUD1 was shown to localize prominently to the Golgi [53]. Whether alteration of the Golgi in neurodegeneration also affects the function of HERPUD1 is an exciting thing to test. Based on this, one might hypothesize that the alteration of the Golgi is not only a bystander phenomenon in neurodegenerative diseases, but is an active part of the pathologic process.

Complementary, the GRB2 is an important hub for cellular signaling in many processes [20]. Nonetheless, it is difficult at the current stage to dissect a definitive role for it in the interaction among response to stress, Golgi fragmentation, and neuronal death.

With respect to STAT5B, no role for this protein in neurodegeneration has been reported so far. Interestingly, genetic depletion of STAT5B was shown to result in strong alteration of Golgi structure [34]. It will require more work to determine whether STAT5B is altered in neurodegenerative disorders and whether the Golgi alteration also occurs as a consequence of this, or if other stress-related signaling pathways (like CDK5) are the dominating factors.

Finally, the presence of PPAR γ in the studied interaction is quite interesting. PPAR γ is part of the peroxisome proliferator-activated receptors (PPARs); these receptors act as transcription factors regulating the expression of genes. In humans, PPARs play essential roles in the regulation of cellular differentiation, development, and metabolism. However, they also may be involved in the pathogenesis of several disorders of the central nervous system, such as ALS, Alzheimer's and Parkinson's disease [see 23, for further details]. In particular, empirical evidence has shown that the activation of these transcription factors has been associated with potent anti-inflammatory as well as anti- β -amyloidogenic effects in Alzheimer's disease animal models. Moreover, results on mice treated with PPAR γ antagonist suggest that PPAR γ -based therapies may have clinical utility in treating this disease [see 37]. The suggestion that PPAR γ might control the inflammation and β -amyloidogenesis, both associated to neurodegeneration, via cellular response to stress opens new possibilities for developing more effective therapies.

4.2 Results on synthetic instances from SteinLib [32]

To test the computational performance of the proposed approach in a more general setting, some experiments on synthetic data were also performed. The datasets P4E and P4Z from the well-known **SteinLib** dataset library [32] were used. These groups of instances correspond to a collection of Steiner tree instances in complete graphs. In P4E, the edge weights (which are used as confidence values) are euclidean (rounded to the next integer); while in P4Z, they are random integers. Note that some preliminary tests using other instances from **SteinLib** were also performed; however, most of instances often have very special characteristics. In particular, most of instances from the **SteinLib** library, have integral edge weights between one and ten; this makes them not too useful for the purposes of this study, since the range of the edge-weights (confidence values) is an upper bound on the size of the Pareto front, as discussed before.

In Table 1 results of the carried out experiments are reported. Complementary, some characteristics of the considered instances are also given. Note that in these experiments, only the Pareto front was calculated (i.e., one solution for every point on the front). Notwithstanding, preliminary experiments showed that there were sometimes hundreds of solutions associated with every point of the front. Interestingly, even though the confidence values have a large range in these instances, the sizes of the Pareto fronts are rather small. The largest is comprised by 16 points and it corresponds to instance p407. One can see that instances from group P4Z have larger Pareto fronts; moreover, three instances from the group P4E have a Pareto front consisting of only two points. The runtime to determine each of these Pareto fronts is quite short; the longest runtime is under 13 seconds. The number of terminals and also the size of the graph seem to influence the runtime.

<i>name</i>	$ V $	$ E $	$ T $	$\min c_e$	$\max c_e$	$ Front $	$t[s]$
p455	100	4950	5	4	1101	3	0.30
p456	100	4950	5	4	1101	4	0.25
p457	100	4950	10	4	1101	4	0.53
p458	100	4950	10	4	1101	5	0.54
p459	100	4950	20	4	1101	4	1.57
p460	100	4950	20	4	1101	2	1.32
p461	100	4950	50	4	1101	2	4.53
p463	200	19900	10	4	1213	3	0.89
p464	200	19900	20	4	1213	2	2.62
p465	200	19900	40	4	1213	3	7.67
p466	200	19900	100	4	1213	3	12.72
<hr/>							
p401	100	4950	5	1	1500	11	0.40
p402	100	4950	5	1	1500	13	0.77
p403	100	4950	5	1	1500	8	0.33
p404	100	4950	10	1	1500	13	1.05
p405	100	4950	10	1	1500	7	0.65
p406	100	4950	10	1	1500	12	1.37
p407	100	4950	20	1	1500	16	2.41
p408	100	4950	20	1	1500	6	2.07
p409	100	4950	50	1	1500	5	8.71
p410	100	4950	50	1	1500	8	8.79

Table 1: Results for instances P4E and P4Z obtained from [32]; $|V|$: number of nodes, $|E|$: number of edges, $|T|$: number of terminals, $\min c_e$: minimum edge confidence in the instance, $\max c_e$: maximum edge confidence in the instance, $|Front|$: number of solutions in the Pareto front, $t[s]$: timed required time (in seconds) to compute the front.

5. Concluding remarks and future work

In this paper, the bi-objective minimum cardinality bottleneck Steiner tree problem is introduced for the task of discovering functional links between Golgi apparatus fragmentation and neuronal death. A specially tailored exact algorithmic scheme has been devised; the scheme is based on integer programming and the so-called ϵ -constraint method. The approach allows an efficient integer programming formulation of the problem and it could be extended to similar problems.

The inferred networks bring information that complement prior biological evidence regarding the pathological development of neurodegenerative diseases and the consequent observation of Golgi fragmentation.

Moreover, when investigating whether this relation is somehow triggered by a natural biological function of the cell, an in-silico analysis shows that the cellular response to stress is a statistically predominant functional theme within the discovered network. Hence, this allows to conjecture that (i) neurodegenerative diseases set up conditions that cells perceive as stressors, (ii) the responsive mechanisms of the Golgi to these stimuli induce its fragmentation, and (iii) this finally results in the death of the neuron. Furthermore, the functional linkage of Golgi fragmentation and neuronal death via cellular response to stress seems to require the expression of four proteins (HERPUD1, GRB2, STAT5B and PPAR γ).

The proposed approach was also tested on a set of instances from the well-known **SteinLib** instance library. The obtained results show that the algorithm is able to solve all the considered instances in few seconds.

Finally, it is important to remark that, to the best of the knowledge of the authors, this is the first time that the search of alternative optimal solutions is performed in this type of Bioinformatics setting. The common practice is to simply compute an optimal solution and analyze what this single solution brings.

Nonetheless, even in single objective problems [as those studied in 3, 24, 39, 57], it is crucial to investigate other optimal solutions to verify the correctness of the drawn conclusions. This is addressed in [12], although in a single objective context; they look for sub-optimal solutions that are, at least, at a pre-defined Hamming distance, say k nodes. This is particularly important for future research, especially if one takes into account that biological networks are huge so the presence of many optimal solutions is plausible.

Acknowledgements The authors would like to acknowledge the anonymous reviewers for their insightful and constructive comments which helped to improve the quality of the paper. E. Álvarez-Miranda acknowledges the support of the Chilean Council of Scientific and Technological Research, CONICYT, through the grant FONDECYT N.11140060 and through the Complex Engineering Systems Institute (ICM:P-05-004-F, CONICYT:FBO16). H. Farhan is supported by the Swiss National Science Foundation, the German Science Foundation, the Young Scholar Fund of the University of Konstanz and by the Canton of Thurgau. M. Sinnl is supported by the Austrian Research Fund (FWF, Project P 26755-N19). M. Luipersbeck acknowledges the support of the University of Vienna through the uni:docs fellowship programme.

References

- [1] E. Álvarez-Miranda, I. Ljubić, and P. Mutzel. The maximum weight connected subgraph problem. In M. Jünger and G. Reinelt, editors, *Facets of Combinatorial Optimization*, pages 245–270. Springer, 2013.
- [2] E. Álvarez-Miranda, I. Ljubić, and P. Mutzel. The rooted maximum node-weight connected subgraph problem. In C. Gomes and M. Sellmann, editors, *Proceedings of CPAIOR 2013*, volume 7874 of *LNCS*, pages 300–315. Springer, 2013.
- [3] C. Backes, A. Rurainski, G. Klau, O. Müller, D. Stöckel, A. Gerasch, J. Küntzer, D. Maisel, N. Ludwig, M. Hein, A. Keller, H. Burtscher, M. Kaufmann, E. Meese, and H. Lenhof. An integer linear programming approach for finding deregulated subgraphs in regulatory networks. *Nucleic Acids Research*, 1: 1–13, 2011.
- [4] S. Bellouze, M. Schäfer, D. Buttigieg, G. Baillat, C. Rabouille, and G. Haase. Golgi fragmentation in pmn mice is due to a defective arf1/tbce cross-talk that coordinates copi vesicle formation and tubulin polymerization. *Human molecular genetics*, 23(22):5961–5975, 2014.
- [5] L. Bertram and R. Tanzi. The genetic epidemiology of neurodegenerative disease. 115(6):1449–1457, 2005.
- [6] BioGRID^{3.4}. Biological general repository for interaction datasets, 2015. URL <http://thebiogrid.org/>.
- [7] N. Boland, H. Charkhgard, and M. Savelsbergh. A criterion space search algorithm for biobjective mixed integer programming: The rectangle splitting method. *Optimization Online*, 2015.
- [8] J. Cancino and A. Luini. Signaling circuits on the golgi complex. *Traffic*, 14(2):121–134, 2013.
- [9] J. Castro-Alvarez, S. Uribe-Arias, D. Mejía-Raigosa, and G. Cardona-Gómez. Cyclin-dependent kinase 5, a node protein in diminished tauopathy: a systems biology approach. *Frontiers in aging neuroscience*, 6, 2014.
- [10] J. Chia, G. Goh, V. Racine, S. Ng, P. Kumar, and F. Bard. RNAi screening reveals a large signaling network controlling the Golgi apparatus in human cells. *Molecular systems biology*, 8(1):629, 2012.
- [11] A. Chinchuluun and P. Pardalos. A survey of recent developments in multiobjective optimization. *Annals of Operations Research*, 154(1):29–50, 2007.

- [12] M. Dittrich, G. Klau, A. Rosenwald, T. Dandekar, and T. Muller. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics*, 24(13):i223–i231, 2008.
- [13] Cees W Duin and Anton Volgenant. The partial sum criterion for steiner trees in graphs and shortest paths. *European journal of operational research*, 97(1):172–182, 1997.
- [14] B. Duthey. Alzheimer disease and other dementias. *World Health Organization: Priority Medicines for Europe and the World "A Public Health Approach to Innovation"*, 2013.
- [15] M. Ehrgott and M. Wiecek. Multiobjective programming. In *Multiple criteria decision analysis: State of the art surveys*, pages 667–708. Springer, 2005.
- [16] J. Fan, Z. Hu, L. Zeng, W. Lu, X. Tang, J. Zhang, and T. Li. Golgi apparatus and neurodegenerative diseases. *International Journal of Developmental Neuroscience*, 26(6):523–534, 2008.
- [17] H. Farhan and C. Rabouille. Signalling to and from the secretory pathway. *Journal of Cell Science*, 124(2):171–180, 2011.
- [18] M. Fischetti, M. Leitner, I. Ljubić, M. Luipersbeck, M. Monaci, M. Resch, D. Salvagnin, and M. Sinnl. Thinning out steiner trees a node-based model for uniform edge costs. *Workshop of the 11th DIMACS Implementation Challenge*, 2014.
- [19] Y. Fujita and K. Okamoto. Golgi apparatus of the motor neurons in patients with amyotrophic lateral sclerosis and in mice models of amyotrophic lateral sclerosis. *Neuropathology*, 25(4):388–394, 2005.
- [20] Gene Cards: Human Gene Database. GRB2 Genecard, 2015. URL <http://www.genecards.org/cgi-bin/carddisp.pl?gene=GRB2>.
- [21] Gene Ontology Consortium . The Gene Ontology Project, 2015. URL <http://geneontology.org/>.
- [22] Y. Haimes, L. Lasdon, and D. Wismer. On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems Man and Cybernetics*, (1):296–297, 1971.
- [23] M. Heneka, E. Reyes-Irisarri, M. Hüll, and M. Kummer. Impact and therapeutic potential of ppars in alzheimers disease. *Current Neuropharmacology*, 9(4).
- [24] S. Huang and E. Fraenkel. Integration of proteomic, transcriptional, and interactome data reveals hidden signaling components. *Science Signaling*, 2(81):ra40, 2009.
- [25] J. Hutchins and S. Barger. Why neurons die: Cell death in the nervous system. *The Anatomical Record*, 253:79–90, 1998.
- [26] T. Ideker, O. Ozier, B. Schwikowski, and A. Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18(Supplement 1):S233–S240, 2002.
- [27] Q. Jiang, L. Wang, Y. Guan, H. Xu, Y. Niu, L. Han, Y. Wei, L. Lin, J. Chu, Q. Wang, Y. Yang, L. Pei, J. Wang, and Q. Tian. Golgin-84-associated golgi fragmentation triggers tau hyperphosphorylation by activation of cyclin-dependent kinase-5 and extracellular signal-regulated kinase. *Neurobiology of aging*, 35(6):1352–1363, 2014.
- [28] G. Joshi, Y. Chi, Z. Huang, and Y. Wang. A β -induced golgi fragmentation in Alzheimer's disease enhances a β production. *Proceedings of the National Academy of Sciences*, 111(13):E1230–E1239, 2014.
- [29] M. Kamada, Y. Izumi, T. Ayaki, M. Nakamura, S. Kagawa, E. Kudo, W. Sako, H. Maruyama, Y. Nishida, H. Kawakami, H. Ito, and R. Kaji. Clinicopathologic features of autosomal recessive amyotrophic lateral sclerosis associated with optineurin mutation. *Neuropathology*, 34(1):64–70, 2014.

- [30] KEGG. Kyoto encyclopedia of genes and genomes, 2015. URL <http://www.genome.jp/kegg/>.
- [31] T. Kim, E. Kim, S. Yoon, and J. Yoon. Herp enhances er-associated protein degradation by recruiting ubiquilins. *Biochemical and Biophysical Research Communications*, 369(2):741 – 746, 2008.
- [32] T. Koch, A. Martin, and A. Voß. *SteinLib: An updated library on Steiner tree problems in graphs*. Springer, 2001.
- [33] M. Kristiansen and J. Ham. Mprogrammed cell death during neuronal development: the sympathetic neuron model. *Cell Death and Differentiation*, 21:1025–1035, 2014.
- [34] J. Lee, Y. Yang, F. Liang, D. Gough, D. Levy, and P. Sehgal. Nongenomic STAT5-dependent effects on Golgi apparatus and endoplasmic reticulum structure and function. *American Journal of Physiology - Cell Physiology*, 302(5):C804–C820, 2012.
- [35] M. Leitner, I. Ljubić, and M. Sinnl. A computational study of exact approaches for the bi-objective prize-collecting steiner tree problem. *INFORMS Journal on Computing*, 27(1):118–134, 2014.
- [36] S. Maere, K. Heymans, and M. Kuiper. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21:448–3449, 2005.
- [37] T. Malm, M. Mariani, L. Donovan, L. Neilson, and G. Landreth. Activation of the nuclear receptor ppar is neuroprotective in a transgenic mouse model of alzheimers disease through inhibition of inflammation. *Journal of Neuroinflammation*, 12(7):1–15, 2015.
- [38] R. Mayeux. Epidemiology of neurodegeneration. *Annual Review of Neuroscience*, 26(1):81–104, 2003.
- [39] A. Mazza, I. Gat-Viks, H. Farhan, and R. Sharan. A minimum-labeling approach for reconstructing protein networks across multiple conditions. *Algorithms for Molecular Biology*, 9(1):1–8, 2014.
- [40] V. Millarte, G. Boncompain, K. Tillmann, F. Perez, E. Sztul, and H. Farhan. Phospholipase c gamma 1 regulates early secretory trafficking and cell migration via interaction with p115. *Mol Biol Cell*, 26: 2263–2278, 2015.
- [41] A. Mironov and M. Pavelka, editors. *The Golgi Apparatus: State of the art 110 years after Camillo Golgi's discovery*. Springer, 1st edition, 2008.
- [42] A. Moss and Y. Rabani. Approximation algorithms for constrained node weighted steiner tree problems. *SIAM Journal on Computing*, 37(2):460–481, 2007.
- [43] S. Nakagomi, M. Barsoum, E. Bossy-Wetzel, C. Sütterlin, V. Malhotra, and S. Lipton. A golgi fragmentation pathway in neurodegeneration. *Neurobiology of Disease*, 29(2):221–231, 2008.
- [44] C. Qiu, M. Kivipelto, and E. von Strauss. Epidemiology of Alzheimers disease: occurrence, determinants, and strategies toward intervention. *Dialogues in Clinical Neuroscience*, 11(2).
- [45] REACTOME. A curated pathway database, 2015. URL <http://www.reactome.org/>.
- [46] C. Reitz, C. Brayne, and R. Mayeux. Epidemiology of alzheimer disease. 7(3):137–152, 2011.
- [47] A. Renton, A. Chiò and B. Traynor. State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, 17:17–23, 2015.
- [48] X. Sai, Y. Kawamura, K. Kokame, H. Yamaguchi, H. Shiraishi, R. Suzuki, T. Suzuki, M. Kawaichi, T. Miyata, T. Kitamura, B. De Strooper, K. Yanagisawa, and H. Komano. Endoplasmic reticulum stress-inducible protein, Herp, enhances presenilin-mediated generation of amyloid β -protein. *Journal of Biological Chemistry*, 277(15):12915–12920, 2002.

- [49] R. Sharma. RNAi screening: tips and techniques. *Nature immunology*, 10(8):799–804, 2009.
- [50] J. Simpson, B. Joggerst, V. Laketa, F. Verissimo, C. Cetin, H. Erfle, M. Bexiga, V. Singan, J. Hériché, B. Neumann, A. Mateos, J. Blake, S. Bechtel, V. Benes, S. Wiemann, J. Ellenberg, and R. Pepperkok. Genome-wide rnai screening identifies human proteins with a regulatory function in the early secretory pathway. *Nat Cell Biol*, 14:764–774, 2012.
- [51] STRING¹⁰. Known and predicted protein-protein interactions, 2015. URL <http://string-db.org>.
- [52] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. Jensen, and C. von Mering. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 39(suppl 1):D561–D568, 2011.
- [53] S. Tuvia, D. Taglicht, O. Erez, I. Alroy, I. Alchanati, V. Bicoviski, M. Dori-Bachash, D. Ben-Avraham, and Y. Reiss. The ubiquitin E3 ligase POSH regulates calcium homeostasis through spatial control of Herp. *The Journal of Cell Biology*, 177(1):51–61, 2007.
- [54] I. Ulitsky, R. Karp, and R. Shamir. Detecting disease-specific dysregulated pathways via analysis of clinical expression profiles. In M. Vingron and L. Wong, editors, *Research in Computational Molecular Biology*, volume 4955 of *LNCS*, pages 347–359. Springer, 2008.
- [55] I. Ulitsky, A. Krishnamurthy, R. Karp, and R. Shamir. DEGAS: de novo discovery of dysregulated pathways in human diseases. *PLoS One*, 5(10):e13367, 2010.
- [56] N. Yosef, L. Ungar, E. Zalckvar, A. Kimchi, M. Kupiec, E. Ruppin, and R. Sharan. Toward accurate reconstruction of functional protein networks. *Molecular Systems Biology*, 5(248), 2009.
- [57] N. Yosef, E. Zalckvar, A. Rubinstein, M. Homilius, N. Atias, L. Vardi, I. Berman, H. Zur, A. Kimchi, E. Ruppin, and R. Sharan. ANAT: A tool for constructing and analyzing functional protein networks. *Science Signaling*, 4(196):pl1–pl1, 2011.
- [58] J. Yuan, M. Lipinski, and A. Degterev. Diversity in the mechanisms of neuronal cell death. *Neuron*, 40:401–413, 2003.

Appendix

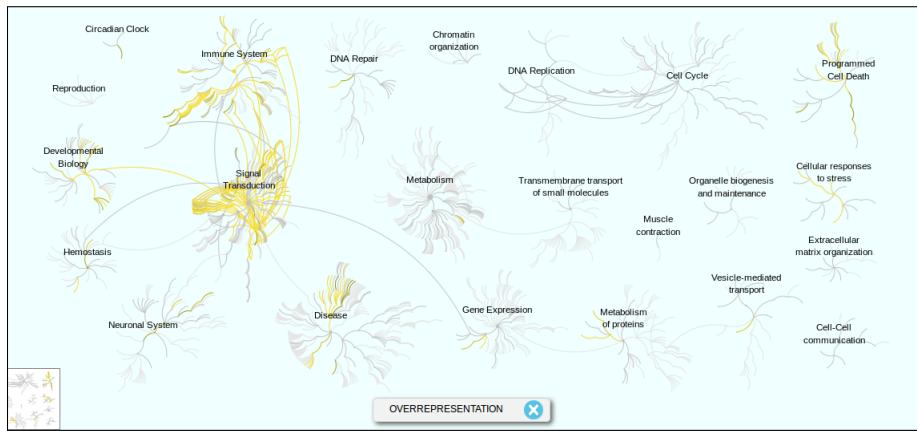


Figure 5: Pathways in which the neuronal death hits appear to be statically meaningful

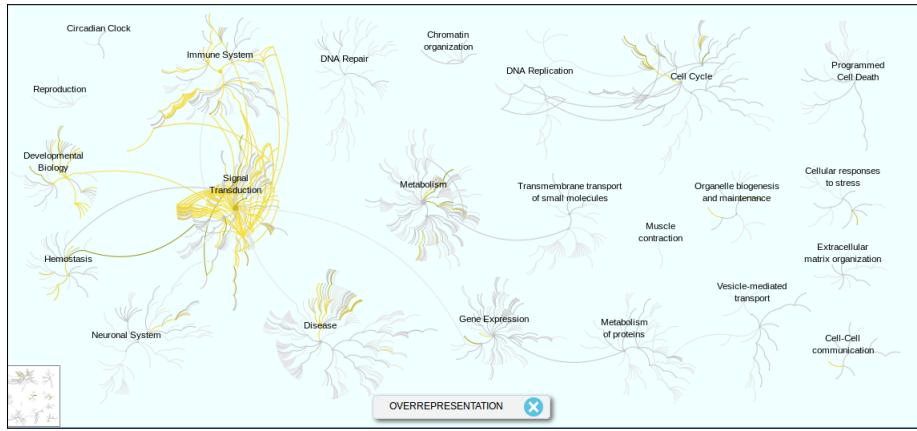


Figure 6: Pathways in which the Golgi fragmentation hits appear to be statically meaningful

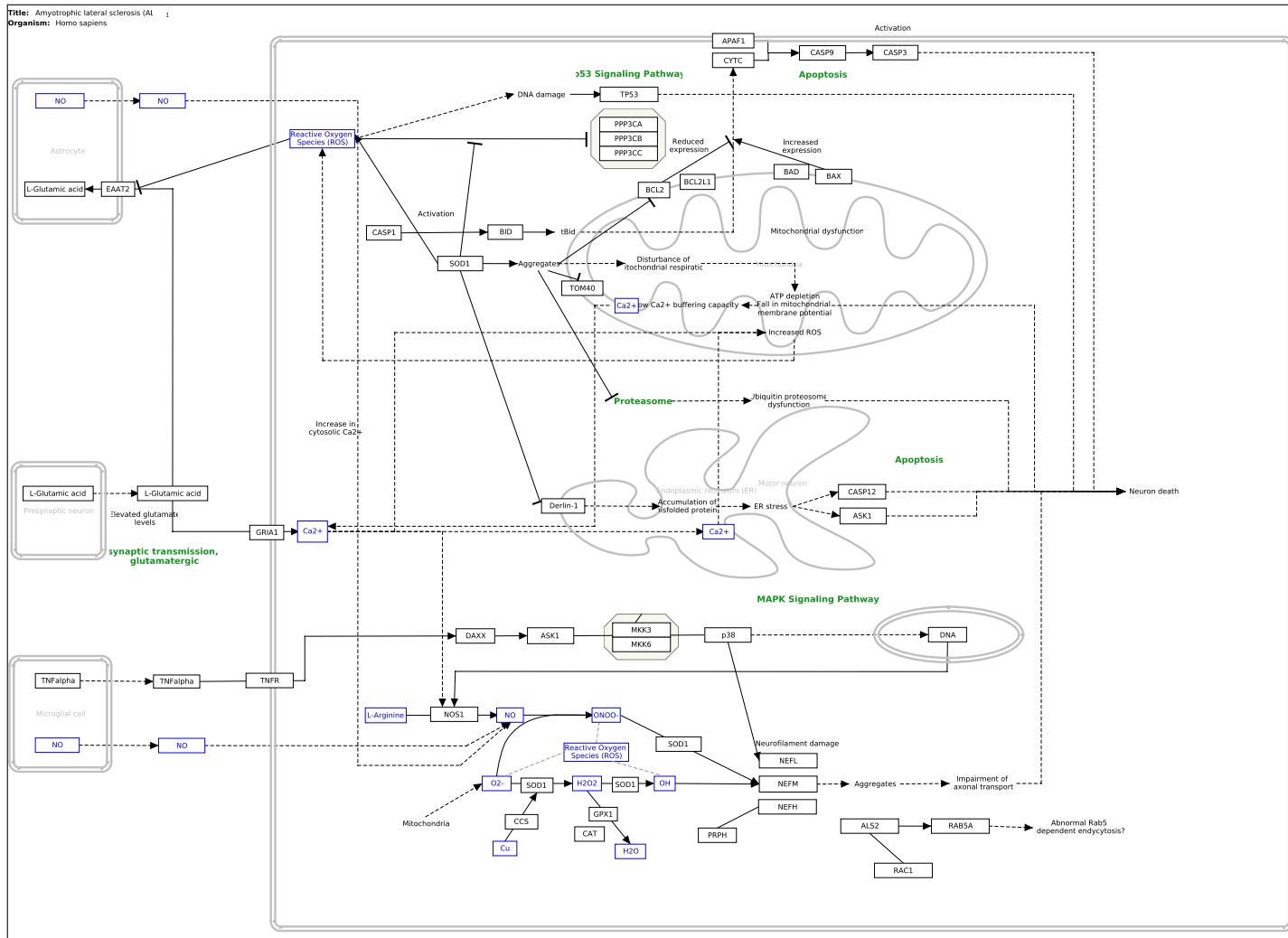


Figure 7: KEGG pathway associated with ALS [obtained from 30]

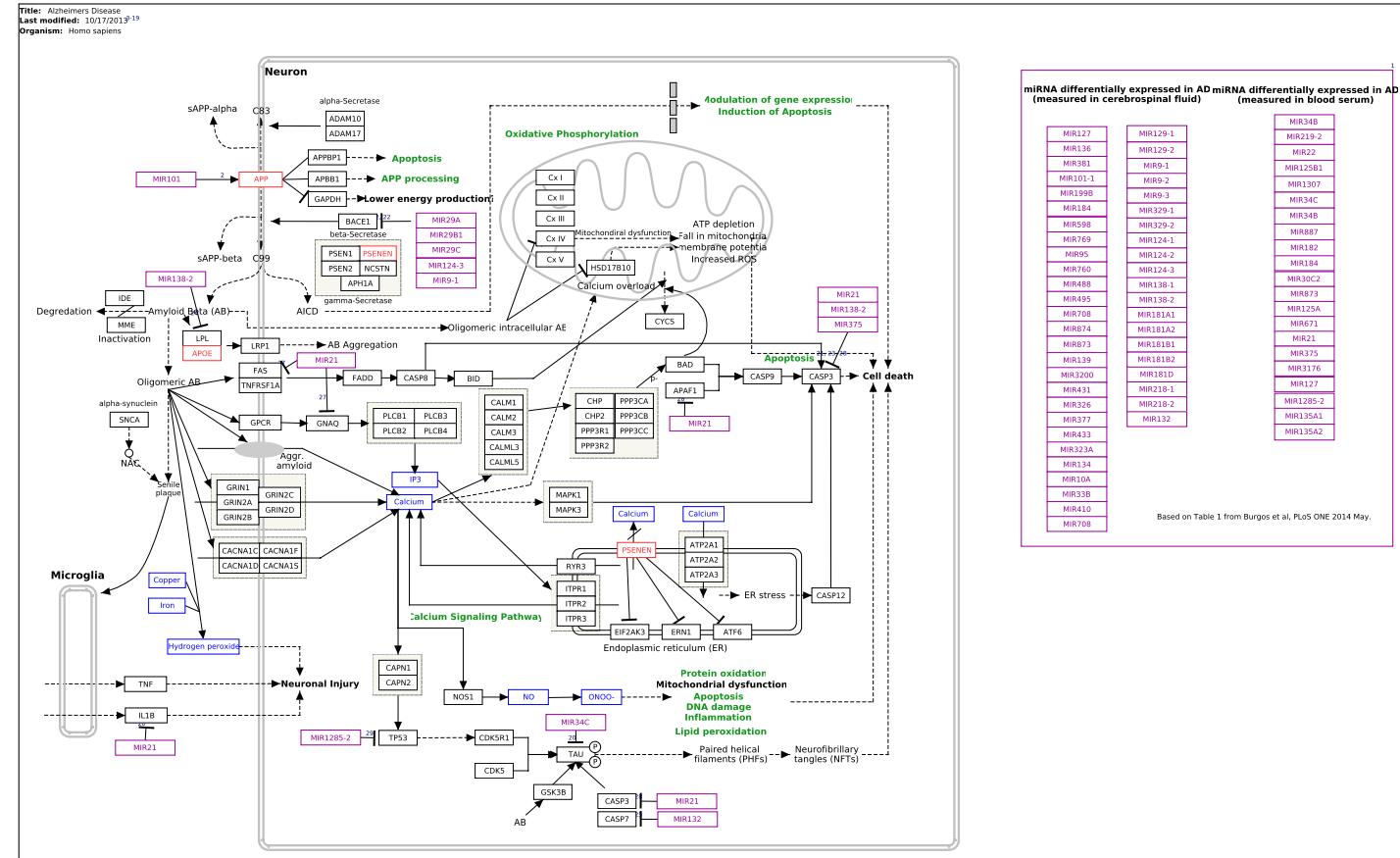


Figure 8: KEGG pathway associated with Alzheimer's disease [obtained from 30]

AARS	CCL5	FAIM2	IKBKG	NQO1	SET
ABL1	CD34	FAM134B	IL13	NR3C1	SIGMAR1
ADAM8	CDC34	FASLG	IL6	NR4A2	SIRT1
ADNP	CDC42	FBXW7	ILK	NR4A3	SIX1
AGAP2	CDK5	FGF8	ISL1	NRBP2	SIX4
AGER	CDK5R1	FGFR3	ITGA1	NRP1	SNCA
AGRN	CEPB	FOXB1	ITSN1	NTF3	SNCB
AGT	CHL1	FOXO3	JAK2	NTRK1	SOD1
AIMP2	CHMP4A	FYN	JUN	NTRK2	SOD2
AKT1	CHMP4B	G6PD	KCNB1	OPTN	SORL1
AKT1S1	CHP1	GABRA5	KCNIP3	OXR1	SRPK2
AMBRA1	CITED1	GABRB2	KDM2B	PAK3	STAMBP
ANGPT1	CLCF1	GABRB3	KIF14	PARK2	STAR
APOE	CLN3	GATA3	KRAS	PARK7	STAT3
ASCL1	CLU	GCLC	LGNN	PAWR	STXBP1
ATF4	CNTF	GCLM	LIG4	PCSK9	TBK1
ATM	CNTFR	GDF5	LRP1	PICALM	TFAP2A
ATP7A	CORO1A	GDNF	LRRK2	PIK3CA	TFAP2B
AXL	CPEB4	GFRAL	MAP3K11	PIN1	TFAP2D
BAD	CRHR1	GHR	MAP3K5	PINK1	TGFB2
BAG1	CRLF1	GNB2L1	MCL1	PM20D1	TGFB3
BARHL1	CSF3	GPI	MDK	PMAIP1	TOX3
BAX	CTNNB1	GPR75	MECP2	POU4F1	TP53
BBC3	DAXX	GPX1	MEF2C	PPARA	TP53BP2
BCL2	DCC	GRID2	MSH2	PPARGC1A	TP63
BCL2L1	DDIT3	GRIK2	MT3		TP73
BCL2L11	DDIT4	GRIK5	MTNR1B	PPT1	TRAF2
BDNF	DHCR24	GRIN1	MUSK	PRKCG	TRIM2
BHLHB9	DLX1	GRM4	NAE1	PRKCI	TYRO3
BIRC5	DNAJC5	HDAC4	NAIP	PSEN1	UBB
BRAF	DRAXIN	HIF1A	NCF2	PTK2B	UBE2M
BTBD10	EGLN2	HIPK2	NDNF	RAPSN	UBE2V2
BTG2	EGLN3	HMOX1	NELF	RASA1	UCN
C5AR1	EGR1	HRAS	NES	REL	UNC5B
CACNA1A	ELK1	HSP90AB1	NF1	RILPL1	VEGFB
CASP2	EPHA7	HSPD1	NGF	ROCK1	WFS1
CASP3	EPHB1	HTRA2	NGFR	RRAS2	XRCC2
CASP9	EPOR	HTT	NLRC4	SARM1	ZNF259
CCL2	ERBB3	HYOU1	NONO	SCT	ZNF746
CCL3	F2R	IKBKB	NOS1	SERPINF1	

Table 2: Proteins associated with Neuronal Death [21]

ABHD1	CKAP2	GRIN2C	MAPK15	RIOK3	TM4SF19
ABHD5	CLPB	GSG2	MAST3	RXRA	TMTM1
ACTR3	CLSTN3	IL18R1	MXD4	SIRT2	TRIM41
AKAP5	COPB2	INADL	NT5C	SOS2	TSPAN1
AQPEP	COPG1	INPP5J	NUDT4	SPRYD4	TTC37
AR	CTGF	KIF1C	PLEKHM2	SRSF1	TYRO3
ARHGAP12	DENND4C	KIF26A	PML	SST	UBE2E1
ARHGAP32	ADORA2A	KIFAP3	PPFIA1	STC1	WDR75
ARHGAP44	EPDR1	KIRREL	PROC	STPG1	YY2
B3GAT2	FAM177B	KRT6B	PRR4	SYT1	ZNF503
C1S	GLRB	LRP4	PTBP1	SYT7	ZNF512
CCDC124	GPBP1	LUC7L3	RADIL	TALDO1	ZNF830
CCR4	GPT	MAGIX	RHOU	TANC2	

Table 3: Proteins regulating Golgi apparatus fragmentation [50]

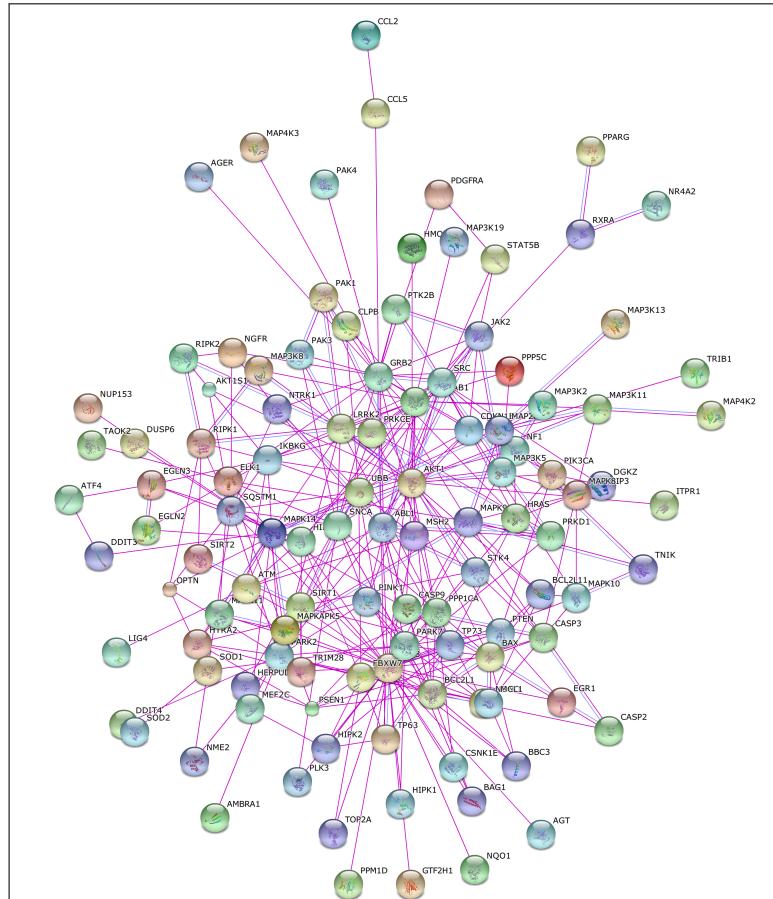


Figure 9: Network associated with the response to stress pathway (obtained via BiNGO analysis [36] and displayed using [51] web tool)

ACPT	CSNK1E	HCK	MKNK1	PPM1L	RPS6KB2
ACYP1	CSNK1G1	HIPK1	MKNK2	PPP1R11	SCYL3
ADCK1	CSNK1G2	HIPK3	MPP2	PPP2CA	SGK223
ADCK5	CSNK2B	HIPK4	MTMR1	PPP2R2B	SHPK
AK4	CXCL10	HK1	MYO3B	PPP2R5E	SPHK1
AK7	DCLK2	IGBP1	NEK11	PPP3R1	SQSTM1
ALPK2	DGKD	IGF1R	NEK2	PRKAG1	SRMS
ALPP	DGKQ	IKBKE	NLK	PRKAG3	SRPK2
ANGPT4	DGKZ	INPP1	NME2	PRKCE	STK32A
ANP32E	DLG3	IPMK	NPR2	PRKCSH	STK36
AURKB	DMPK	ITK	NRBP1	PRKX	STK4
AXL	DUSP2	ITPKA	NRG3	PRPS1L1	STK40
BCKDK	DUSP22	ITPKB	PAG1	PTEN	TAOK2
BMPR1B	DUSP6	KHK	PAK1	PTK2	TBCK
BMX	DUSP8	LAMTOR3	PAK3	PTK7	TESK1
BUB1	EIF2AK2	LCP2	PANK3	PTP4A1	TGFBR1
CAMK1	ENPP7	MALT1	PAPSS1	PTP4A3	TNIK
CDC25A	EPHA1	MAP2K7	PAPSS2	PTPN14	TNS3
CDC42BPA	EPHA8	MAP3K13	PDGFRA	PTPRA	TRIB1
CDC42BPG	EPHB1	MAP3K19	PDK4	PTPRD	TRPM7
CDK1	EPM2A	MAP3K2	PFKP	PTPRF	TSKS
CDK11A	ERBB3	MAP3K8	PHKG1	PTPRN2	TTK
CDK20	EXOSC10	MAP4K2	PHKG2	PTPRT	TWF2
CDKL2	FBP2	MAP4K3	PI4KA	PTPRU	TXK
CDKN1B	FGFR1	MAPK11	PI4KB	PXK	ULK4
CHKA	FGFR2	MAPK15	PINK1	RIPK2	VRK3
CKM	FLT3	MAPKAPK5	PIP5K1A	ROCK1	WNK3
CLK1	GALK1	MARK4	PKLR	ROS1	YWHAH
COL4A3BP	GAP43	MAST1	PPM1D	RPRD1A	
CSNK1A1L	GTF2H1	MECOM	PPM1F	RPS6KB1	

Table 4: Proteins regulating Golgi apparatus fragmentation [10]

ABL1	CLTC	MAPK14	PIK3C2B	PTPRN	TMED7
AKAP14	DCLK2	MAPK8IP3	PLCE1	RAC1	TPD52L3
BCR	DUSP23	MAPK9	PLCG1	RAF1	TRRAP
BLK	FAM48A	MAPRE1	PLCG2	RGS2	TSSK2
BMP2K	FYN	MBL2	PLK3	RIPK1	TSSK3
CASP10	GOLGA2	NCOA3	PRKACA	RYK	UNC119
CBL	HRAS	NEDD9	PRKCA	SH3RF1	ZAP70
CCM2	KIT	NUAK2	PRKCZ	SRC	
CD44	LYN	PAK4	PRKD1	SSH1	
CDC42	MAPK1	PARD6A	PSPH	STYX	
CDK4	MAPK10	PDCD10	PTPRC	TLR4	

Table 5: Proteins regulating Golgi apparatus fragmentation [40]

AGER (D)	FBXW7 (D)	PARK7 (D)	HIPK1 (G)	RIPK1 (G)
AGT (D)	HIF1A (D)	PIK3CA (D)	MAP2K7 (G)	RIPK2 (G)
AKT1 (D)	HIPK2 (D)	PPP5C (D)	MAP3K13 (G)	RXRA (G)
AKT1S1 (D)	HMOX1 (D)	PSEN1 (D)	MAP3K19 (G)	SIRT2 (G)
AMBRA1 (D)	HSP90AB1 (D)	PTK2B (D)	MAP3K2 (G)	SQSTM1 (G)
ATF4 (D)	HTRA2 (D)	SIRT1 (D)	MAP3K8 (G)	SRC (G)
ATM (D)	IKBKG (D)	SNCA (D)	MAP4K2 (G)	STK4 (G)
BAG1 (D)	JAK2 (D)	SOD1 (D)	MAP4K3 (G)	TAOK2 (G)
BAX (D)	LIG4 (D)	SOD2 (D)	MAPK10 (G)	TNIK (G)
BBC3 (D)	LRRK2 (D)	TP53 (D)	MAPK11 (G)	TRIB1 (G)
BCL2L1 (D)	MAP3K11 (D)	TP63 (D)	MAPK14 (G)	GRB2 (S,10)
BCL2L11 (D)	MAP3K5 (D)	TP73 (D)	MAPK8IP3 (G)	STAT5B (S,10)
CASP2 (D)	MCL1 (D)	UBB (D)	MAPK9 (G)	HERPUD1 (S,8)
CASP3 (D)	MEF2C (D)	ABL1 (D&G)	MAPKAPK5 (G)	PPAR γ (S,6)
CASP9 (D)	MSH2 (D)	HRAS (D&G)	NME2 (G)	TOP2A (S,2)
CCL2 (D)	NF1 (D)	PAK3 (D&G)	PAK1 (G)	ITPR1 (S,1)
CCL5 (D)	NGFR (D)	PINK1 (D&G)	PAK4 (G)	NUP153 (S,1)
DDIT3 (D)	NONO (D)	CDKN1B (G)	PDGFRA (G)	PPP1CA (S,1)
DDIT4 (D)	NQO1 (D)	CLPB (G)	PLK3 (G)	TRIM28 (S,1)
EGLN2 (D)	NR4A2 (D)	CSNK1E (G)	PPM1D (G)	
EGLN3 (D)	NTRK1 (D)	DGKZ (G)	PRKCE (G)	
EGR1 (D)	OPTN (D)	DUSP6 (G)	PRKD1 (G)	
ELK1 (D)	PARK2 (D)	GTF2H1 (G)	PTEN (G)	

Table 6: Proteins associated with the response to stress