

2023

Analysis of protein-protein interaction network comprising the mammalian target of rapamycin (mTOR) interactome

<https://hdl.handle.net/2144/48390>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY

ARAM V. CHOBANIAN & EDWARD AVEDISIAN SCHOOL OF MEDICINE

Thesis

**ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK
COMPRISING THE MAMMALIAN TARGET OF
RAPAMYCIN (MTOR) INTERACTOME**

by

MICHAEL PATRICK STIERER

B.A., Rhodes College, 2016

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science

2023

© 2023 by
MICHAEL PATRICK STIERER
All rights reserved

Approved by

First Reader

Beth Bragdon, Ph.D.
Assistant Professor of Orthopaedic Surgery

Second Reader

Kimberly Raab-Graham, Ph.D.
Professor of Physiology and Pharmacology
Wake Forest University, School of Medicine

Third Reader

Paul Laurienti, MD, Ph.D.
Professor of Radiology
Wake Forest University, School of Medicine

ACKNOWLEDGMENTS

I'd like to thank Dr. Kim Raab-Graham and Dr. Paul Laurienti for taking a chance on me as a student in their labs, and for their continuing mentorship.

I'd also like to thank LCBN programmer Rober Lyday for his time and patience in helping me to prepare my analysis scripts and in helping me to learn the Matlab language.

**ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK
COMRPISING THE MAMMALIAN TARGET OF RAPAMYCIN (MTOR)
INTERACTOME**

MICHAEL PATRICK STIERER

ABSTRACT

The mammalian target of rapamycin (mTOR) is a protein implicated in a variety of cellular processes involving growth and division. In the context of the brain, it regulates synaptic plasticity and axon elongation; its dysfunction is implicated in the pathogenesis of multiple complex, heterogeneous neurodegenerative diseases. These include, but are not limited to Alzheimer's Disease (AD), autism spectrum disorder (ASD), and epilepsy. mTOR boasts a deeply complex and far-reaching signalling cascade, and its activity affects the expression levels of a large number of proteins. As such, investigation of the proteins with whom mTOR interacts is a pertinent endeavor to the advancement of understanding the complex pathogenesis of neurodegenerative disease.

The complexity of this endeavor makes it a target well-poised for protein-protein interaction network (PPIN) analysis. Thus, using a previously recorded MS/MS dataset listing proteins whose expression levels change upon rapamycin administration, we set out to identify key proteins and characterize the properties of the mTOR interactome overall using a variety of topological measures and analytical techniques.

Using such techniques, we found that the in the PPIN created from our data, a certain subset of proteins subjected the network to particular fragility. Namely, the kinless hubs, which have high within-module degree as well as a large participation

coefficient, show vulnerability exceeding that of even conventionally defined hub. Some of these kinless hubs exhibit critical structural roles in the PPIN such that their removal damages the overall efficiency of communication within the network at an individually observable level. Work is ongoing to further investigate these proteins and the potential biological implications of their importance in the network described in the present study.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES	x
INTRODUCTION	1
Mammalian Target of Rapamycin (mTOR) and its Role in Disease.....	1
Protein-Protein Interaction Networks (PPIN).....	2
Terms Relevant to PPIN	2
Previous Investigation.....	5
Specific Aims of the Present Study	6
METHODS	7
Description of Measures	7
Network Construction.....	11
Network Thresholding	12
Network Statistics.....	14
Community Structure.....	14
Network Perturbation Analyses	15
Assortativity.....	16
RESULTS	17

Network Statistics and Hub Analysis	17
Modularity	19
Functional Cartography	21
Network Perturbation Analysis.....	24
Assortativity.....	30
DISCUSSION.....	31
Consistency with Previous Findings.....	31
Threshold Dependency of Results	31
Attack Susceptibility.....	33
FUTURE STUDIES.....	34
Rab28	34
Module Identification and Gene Ontology	34
CONCLUSION.....	36
APPENIDX A: Scaled Inclusivity Results at Sparsity $s=2.5$	37
APPENDIX B: Matlab Scripts Written for Analyses.....	43
BIBLIOGRAPHY.....	51
CURRICULUM VITAE.....	55

LIST OF TABLES

Table 1. Top 10% of Proteins by Network Statistics.....	17
Table 1 (continued) :Top 10% of Proteins by Network Statistics	18
Table 2. Assortativity Values of Selected Subnetworks.....	31

LIST OF FIGURES

Figure 1. Schematic representation of PPI network components	4
Figure 2. Network Topological Measures used to Determine Threshold.....	13
Figure 3. Modular Structure of PPIN at Threshold $s = 2.5$	20
Figure 4. Functional Cartography Classification of Hub Proteins.....	22
Figure 5. Topological Distribution of Functional Cartography Hubs.....	24
Figure 7. Dissolution of Network as Measured by Giant Component Ratio.....	25
Figure 8. Removal of Differentially Expressed Proteins Compared to Failure and Targeted Attack.....	27
Figure 9. Functional Cartography Hub-Based Network Attack.. ..	29

INTRODUCTION

Mammalian Target of Rapamycin (mTOR) and its Role in Disease

The mammalian target of rapamycin (mTOR), whose name derives from its direct inhibition by the immunosuppressive drug rapamycin, is a serine/threonine kinase that plays a crucial role in regulating various cellular processes, including cell growth, proliferation, and survival (Kim et al., 2002). mTOR combines with other proteins to form two complexes: mTORC1 and mTORC2, both of which mechanistically govern their own distinct realms of cellular function. Prior work has established mTOR as a key interlocutor in the balance of catabolism and anabolism, regulating many cell types' growth and division in response to varying environmental factors (Saxton & Sabatini, 2017).

In addition to its role in normal cellular processes, mTOR has been implicated in the pathogenesis of various diseases, including cancer, diabetes, and neurodegenerative diseases (Dormond, 2019). In particular, dysregulation of mTOR signaling has been associated with the development and progression of neurodegenerative diseases such as Alzheimer's disease (AD), autism spectrum disorder (ASD), and Epilepsy (Pei & Hugon, 2008; Swiech et al., 2008).

The heterogeneous nature of the diseases in which mTOR is implicated and the relative lack of clinical efficacy in direct pharmacological targeting of mTOR have made effective understanding of its complex interactome a challenging puzzle for the scientific community (Wood & Gutkind, 2022).

Protein-Protein Interaction Networks (PPIN)

In the study of diseases such as those in which mTOR is implicated, the construction and analysis of protein-protein interaction networks (PPIN) have become emerging tools to augment other *in vitro* and *in vivo* methods (Basu et al., 2021).

Interactions between proteins can either be detected manually through techniques such as proximity ligation or yeast 2-hybrid (Basu et al., 2021), or via software algorithms which, using large data inputs, compute and sort interactions between proteins (Kuzmanov & Emili, 2013). The latter, automated method has seen increasing use as a variety of user-friendly and readily accessible software packages are continually developed and updated (Basu et al., 2021).

Construction of PPIN grants investigators a means of modelling complex interactional schemata, and topological analysis of such networks has yielded findings that have helped to guide and focus research efforts: this includes identifying proteins that would not be detected through more conventional means, as well as isolating the key regulators at the core of a given set of protein interactions (Kuzmanov & Emili, 2013).

Terms Relevant to PPIN

Node

A *node* is an entity in a network, and in the context of PPIN, a single node constitutes a single protein.

Seed

Seeds refer to the proteins used in the initial construction of the network. A network can be comprised entirely of seeds, or of seeds and other interaction partners (often retrieved via interactional database). Constructing a PPIN is sometimes synonymous with *seeding* the network.

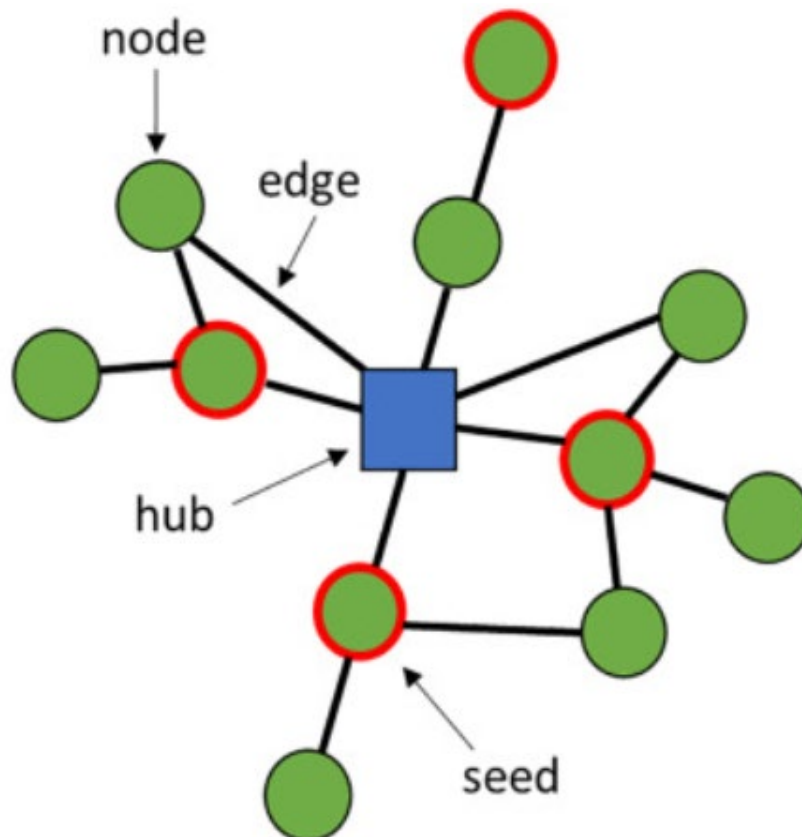
Edge

Edges connect nodes to other nodes: all PPIN exist fundamentally as a collection of nodes and edges. In PPIN, edges are typically defined as interactions between proteins. *Distance* in a network is referent to the number of edges separating two given nodes.

Hub

A *hub* is a node which is connected to more nodes than average. What constitutes a hub can be labile depending on context and/or stringency of the investigators' chosen definition.

Figure 1. Schematic representation of PPI network components. Seeds are the proteins used to start the analysis and around which the network is built (green nodes with red border), a protein interaction between 2 nodes is referred to as an edge (black connecting lines). Nodes can be classified based on topology, a hub for example is a node with a number of connections above average (represented in the diagram as a blue, square node). Note: Reprinted from “Advances in protein-protein interaction network analysis for Parkinson's disease” with Open Access via Creative Commons CC-BY license (Tomkins & Manzoni, 2021)



Previous Investigation

To better understand the changes in proteomic expression conferred by the inhibition and (by extension) activity of mTOR, Niere and colleagues (2016) administered rapamycin to a group of rats and measured rapid changes in cortical protein levels via tandem mass spectrometry. Proteins whose measured levels increased after rapamycin administration were termed “mTOR off” proteins, whereas proteins that decreased were deemed “mTOR on”.

In addition to this in vitro study, the researchers created a protein-protein interaction network (PPIN) using the BisoGenet (Martin et al., 2010) software package, which identified PARK7 as a gene of interest by virtue of its topological importance as a hub protein in their constructed network. Niere and team seeded their network with proteins found in their dataset of 744 proteins known to be associated with either ASD, AD, or epilepsy (2016), querying for interactions up to 2 proteins away from these seed proteins. This network highlighted PARK7 as a hub protein, possessing a large number of connections to other proteins in the network. This protein had yet to be implicated in any of three diseases being investigated, yet has since been shown to be involved in the pathogenesis of AD (Uneri et al., 2021). Niere and team isolated this protein by virtue of its large degree (see Description of Measures) in accordance with the “centrality-lethality rule” (Zotenko et al., 2008), which posits that centrally-distributed nodes in PPIN tend to be the critical regulatory proteins of their respective interactomes.

Specific Aims of the Present Study

We sought to build upon the work of Niere (2016) by using network analysis to drive the methods. Since publication of this prior work, new advances and tools have been made in the field of PPIN, and we intended to capitalize and make use of such advancements. Namely, using the same list of 744 proteins, we employed an entirely different network construction algorithm, in addition to introducing a number of analytical techniques not previously employed.

Rather than recapitulating the results of the previous study, we intended to construct a network under entirely different premises, with the lack of preference or bias for any protein or set of proteins being at the core of this intention. Instead of relying on the database which generates the PPIN to determine our network topology, we thresholded our network according to where it found its most stable topological measures. While we expect overlap as far as the contents of the network generated, we hope that our methods provide venues for further research and investigation in terms of mTOR's involvement in neurodegenerative pathogenesis.

METHODS

Description of Measures

Degree

Degree (k_i) refers to the number of connections to other nodes (also known as edges)

which exist for a given node. The total degree L of the network is given by:

$$L = \frac{1}{2} \sum_{i=1}^N k_i \quad (1)$$

where N =the total number of nodes in the network and i is a node in the network.

Betweenness Centrality

Betweenness centrality (C_b) is a measure of a node's ability to control information flow within a network (Freeman, 1977). Nodes which are placed optimally to link other nodes, especially to those placed on the periphery of the network, have higher values for this metric. It is given by:

$$C_b(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (2)$$

where i is a given node and $\sigma_{st}(i)$ is the number of shortest paths between node s and node t .

Eigenvector Centrality

Eigenvector centrality (C_e) measures a nodes the extent to which a given node exerts influence over the network (Newman, Mark E. J., 2018). It is calculated for an adjacency matrix A_{ij} :

$$C_e(i) = \frac{1}{\lambda} \sum_{j=1}^I A_{ij} x_j \quad (3)$$

where i is a given node, λ is a constant, j is a neighbor of node i , and x_j is a vector of node i 's centralities. Nodes can have high eigenvector centrality either by having many connections (high degree), or by being connected to other nodes of high eigenvector centrality (Newman, Mark E. J., 2018).

Global Efficiency

Global efficiency (E_{glob}) measures the average communication efficiency across a network, and is inversely proportional to average path length (Latora & Marchiori, 2001). The more nodes in distance that it takes to get from node to another (following the edges of the network), the lower the global efficiency. For a network V with N total nodes, global efficiency is calculated by:

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \frac{1}{d_{ij}} \quad (4)$$

where d_{ij} is the shortest path between nodes i and j .

Clustering Coefficient

Defined by Watts and Strogatz (1998), clustering coefficient (C_i) measures how connected the neighbor nodes of a given node are. For a node k :

$$C_i(k) = \frac{2m_i}{k_i(k_i-1)} \quad (5)$$

where m_i is the total degree among its neighbor nodes. If a node's neighbors are all

connected to each other, the resultant C_i value would be 1, and 0 if none its neighbors are connected to each other.

Participation Coefficient

Used primarily in the present study's functional cartography analysis, in addition to the subsequent network perturbation analysis, participation coefficient (P_i) defines to what extent a node i is connected to nodes within its own module versus outside the module (Guimerà & Nunes Amaral, 2005). Partition coefficient is calculated in terms of node i :

$$P_i = 1 - \sum_{s=1}^{N_M} \left(\frac{k_{is}}{k_i} \right)^2 \quad (6)$$

where k_{is} is the number of connections within module s and k_i is the total degree.

Assortativity

Assortative mixing, also known as assortativity (r_{jk}), is a measure of to what extent a network or subnetwork contains connections between nodes of the same degree (Newman, M. E. J., 2002). We calculated assortativity across the whole network, on the top 10% of nodes by degree and on all nodes classified as provincial, connector, and kinless hubs as found by functional cartography. Assortativity is the Pearson's correlation of nodes j_i and k connected by the i^{th} edge of the network:

$$r_{jk} = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (7)$$

with $i=1 \dots M$. If the nodes in the network are arranged such that connections occur between nodes of a similar degree, the correlation value r will be close to 1, and the network can be categorized as assortative (Newman, M. E. J., 2002). Disassortativity occurs when connections between nodes of unequal degree predominate, resulting in r values approaching -1.

Giant Component Ratio

A network whose majority of nodes are connected to at least two other nodes is said to possess a giant component (Newman, Mark E. J., 2018). In the context of analyses which perturb or disrupt the network, the giant component is observed until the critical point (F_c), at which point more the majority of nodes are within individual clusters and the giant component ceases to exist. In our perturbation analyses, we measured the fraction of nodes that remained in the giant component following node removal: we hence refer to this fraction as the giant component ratio (gcr).

Louvain Modularity

As one of the methods used to measure the community structure of a network, the Louvain Method for community detection is an unsupervised algorithm that groups nodes into assemblies (modules) (Blondel et al., 2008). Quantitatively, the algorithm assigns a modularity value (Q) on a per-node level for every node in a network. This value is derived for a given node pair ij by:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta_{c_i c_j} \quad (8)$$

where $A_{ij} = 1$ for connected nodes and 0 if no connection exists, δ is the Kronecker delta function (1 if true, 0 if false), m is the number of edges in the network, k is degree, and c is a node's community/module of membership.

Sparsity

Lastly, sparsity (s) is a unitless, inverse measure of the density of the network, which measures on a whole-network level the relationship between how many connections exist versus how many are possible (Telesford et al., 2011). We calculate it as a logarithmic relationship between the total nodes in the network and the average degree (k), and is given by:

$$s = \frac{\log(N)}{\log(k)} \quad (9)$$

where N is the total number of nodes and L is the total number of edges in the network. In the present study, we use sparsity predominantly to establish a threshold at which to perform subsequent analysis.

Network Construction

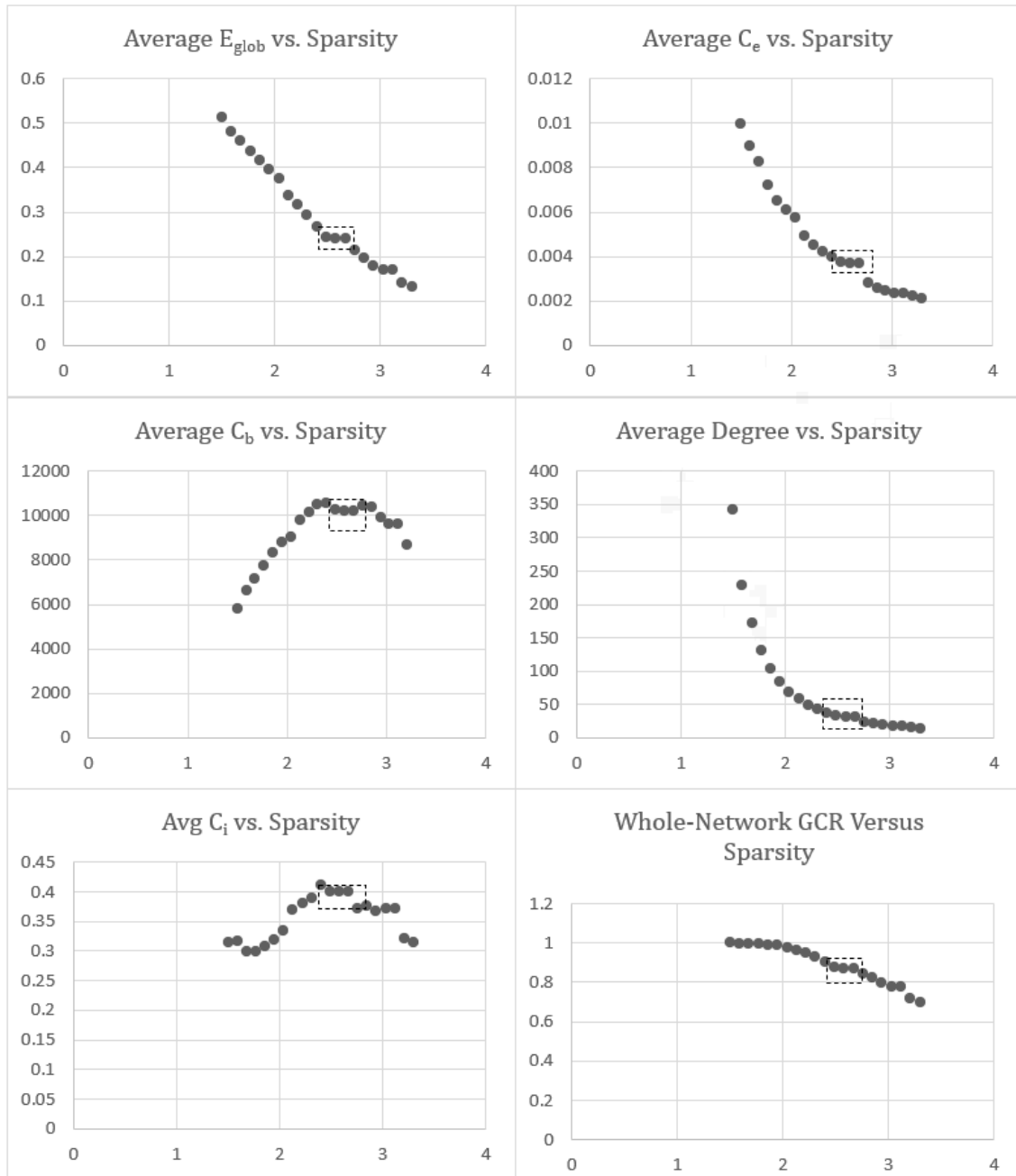
To construct a PPIN, we used all 744 proteins detected via tandem mass spectrometry upon administration of rapamycin (Niere et al., 2016) as seeds. The network was constructed using Cytoscape 3.9.1. and the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) add-in for Cytoscape (Szklarczyk et al., 2019). Interactions among these proteins were queried from the STRING database at the minimum confidence threshold of 01, resulting in a PPIN with 5601 proteins overall. This network

was then manually imported into Matlab as an edge list and subsequently converted into an unweighted binary adjacency matrix for further processing and statistical analysis.

Network Thresholding

As STRING employs a confidence-based scoring system to ascribe functional interactions between pairs of proteins, we set out to control for potentially spurious interactions while still maximizing the breadth of our analysis. Rather than thresholding based on STRING's built-in confidence value, we employed a fixed-density thresholding strategy (van Wijk et al., 2010). To do this, we removed nodes from the network systematically while recording its topological properties (Figure 2). The network was found to have optimal stability at a sparsity value of 2.5, which corresponds to 4,880 nodes and 88,406 edges. Subsequent topological, statistical, and perturbative analyses were performed on the network at this sparsity. Thresholding and all analyses were performed in Matlab Version R2020B unless otherwise noted. We used a combination of base Matlab functions, functions developed by the Laboratory of Complex Brain Networks (LCBN) at Wake Forest University School of Medicine (all prefixed "x_net"), and functions from the Brain Connectivity Toolbox (Rubinov & Sporns, 2010) to perform all necessary functions in Matlab. As the thresholded network comprises the putative interactome of mTOR-related proteins, it will henceforth be referred to interchangeably as both the "mTOR network" and the "thresholded network".

Figure 2. Network Topological Measures used to Determine Threshold. Sparsity, Global Efficiency, Eigenvector Centrality, Betweenness Centrality, Degree, Clustering Coefficient, and Giant Component Ratio were measured as random nodes were removed from the network. Maximum network stability was found to be at the sparsity value of 2.5.



Network Statistics

Across all measured thresholds, whole-network and nodal network statistics were calculated (see *Description of Measures* section for specifics). These include degree, eigenvector centrality, betweenness centrality, clustering coefficient, and giant/fractional component ratio. In addition to being used in the thresholding process, these calculations were also one of the methods used to identify proteins of interest in our network.

Community Structure

Once the threshold was set, the Louvain community detection algorithm (Blondel et al., 2008) was utilized to assign modular community membership to each node in the network. Scaled inclusivity (Steen et al., 2011) was also calculated to assess the threshold-dependent consistency of nodal community membership.

Lastly, functional cartography (Guimerà & Nunes Amaral, 2005) was computed with a hub cutoff value of $p < 0.05$. Functional cartography calculates the within-module degree of all nodes in a given network and for the nodes having within-module degree greater than the cutoff value (hubs), functionally classifies them based on their participation coefficient (P_i), which measures the proportion of connections to each hub that are within-module versus those outside of the node's assigned module. Based on this, the hubs are assigned designations of *kinless hubs* ($P_i > 0.75$), *connector hubs* ($0.3 < P_i \leq 0.75$), or *provincial hubs* ($0.3 \leq P_i$).

Assigning community structure to the network and the nodes within it can be helpful in ascribing a functional role to proteins in relation to the modular assembly of the network overall.

Network Perturbation Analyses

To measure whether specific subsets of proteins conferred particular structural importance within the thresholded network, we removed nodes from the network according to the below procedures whilst recording global efficiency. These procedures are similar to the process outlined by Crucitti (2003). For comparison purposes and to better characterize our network, we rewired it to resemble a random Erdős -Rényi (ER) lattice network (Erdős & Rényi, 1959).

Random Failure

Over 10 separate trials, all 4880 nodes were removed from the network in a randomized order.

Differentially Expressed Protein-Based Attack

Proteins that showed significant changes in proteomic expression levels upon mTORC1 inhibition (both up-regulated and down-regulated) (Niere et al., 2016) were removed in a random order over 100 trials.

Targeted Degree-Based Attack

At each step, degree was calculated for each node in the network and the node with the highest degree subsequently removed.

Functional Cartography Hub-Based Attack

All nodes determined to be kinless, connector, or provincial hubs were removed from the network both collectively and separately according to the above designations. In

each instance, nodes were removed in a descending degree-wise manner, similar to the above degree-based attack method.

Assortativity

We computed assortativity on the full mTOR network, as well as on the first-neighbor subnetworks of mTOR-on proteins, mTOR-off proteins, provincial hubs, connector hubs, and kinless hubs. This mirrors the selection criteria used in the network perturbation analysis. For comparison, we also computed assortativity on randomly selected nodes with identical degree distributions to these grouped proteins. Finally, we also repeated the assortativity measures above on the functional cartography hubs after removing the overlap between them and the hubs as defined by degree, with the intention of isolating these nodes' contribution to assortativity or the lack thereof.

RESULTS

Network Statistics and Hub Analysis

To identify hub proteins of potential biological significance, we measured the top 10% of proteins in the network at sparsity threshold $s = 2.5$ in terms of betweenness centrality, eigenvector centrality, and degree (Table 1). Accounting for overlap among the three network statistics, from the three individual lists of 51, this amounted to a final non-overlapped list of 113 unique proteins. Out of the 10 proteins determined by Niere et al. (2016) in their PPIN to be hubs, 5 (UBC, MAPK3, SNCA, DLG4, STX1A) of these hubs were found to also be hubs in our network.

Table 1. Top 10% of Proteins by Network Statistics. The top 51 proteins are shown as measured by betweenness centrality, eigenvector centrality, and degree.

Top 10% C_b Proteins	C_b Value	Top 10% C_e Proteins	C_e Value	Top 10% Degree Proteins	Degree Value
RPS27A	843482.3	RPS27A	0.120308	RPS27A	568
UBA52	744914.6	UBA52	0.111359	UBA52	543
RHOA	525523.1	RPS4X	0.102661	UBC	454
SRC	474711.6	RPS8	0.102648	UBB	449
UBC	471148.5	RPS2	0.102333	SRC	372
UBB	419311.4	RPS3	0.101522	MAPK1	317
RAB28	413362.4	RPS3A	0.101392	RAC1	305
PRKACA	404371.8	RPS20	0.100826	MAPK3	302
EGFR	403847.4	GNB2L1	0.100236	PIK3R1	290
HSPA8	360897.8	RPS16	0.099808	PRKACA	287
MAPK1	355260.5	RPS11	0.099405	RHOA	287
CTNNB1	349479.6	RPS18	0.099042	EGFR	284
PRKACB	348127.9	RPS5	0.098799	PRKACB	282
HSP90AA1	337409.5	RPS13	0.098439	PIK3CA	275
MAPK3	323766.4	RPSA	0.098353	PRKACG	274
PRKACG	319449.3	RPL9	0.098151	HRAS	266
RAC1	286890.6	RPS24	0.09789	CDC42	259
TP53	277848.8	RPL15	0.097466	GRB2	251

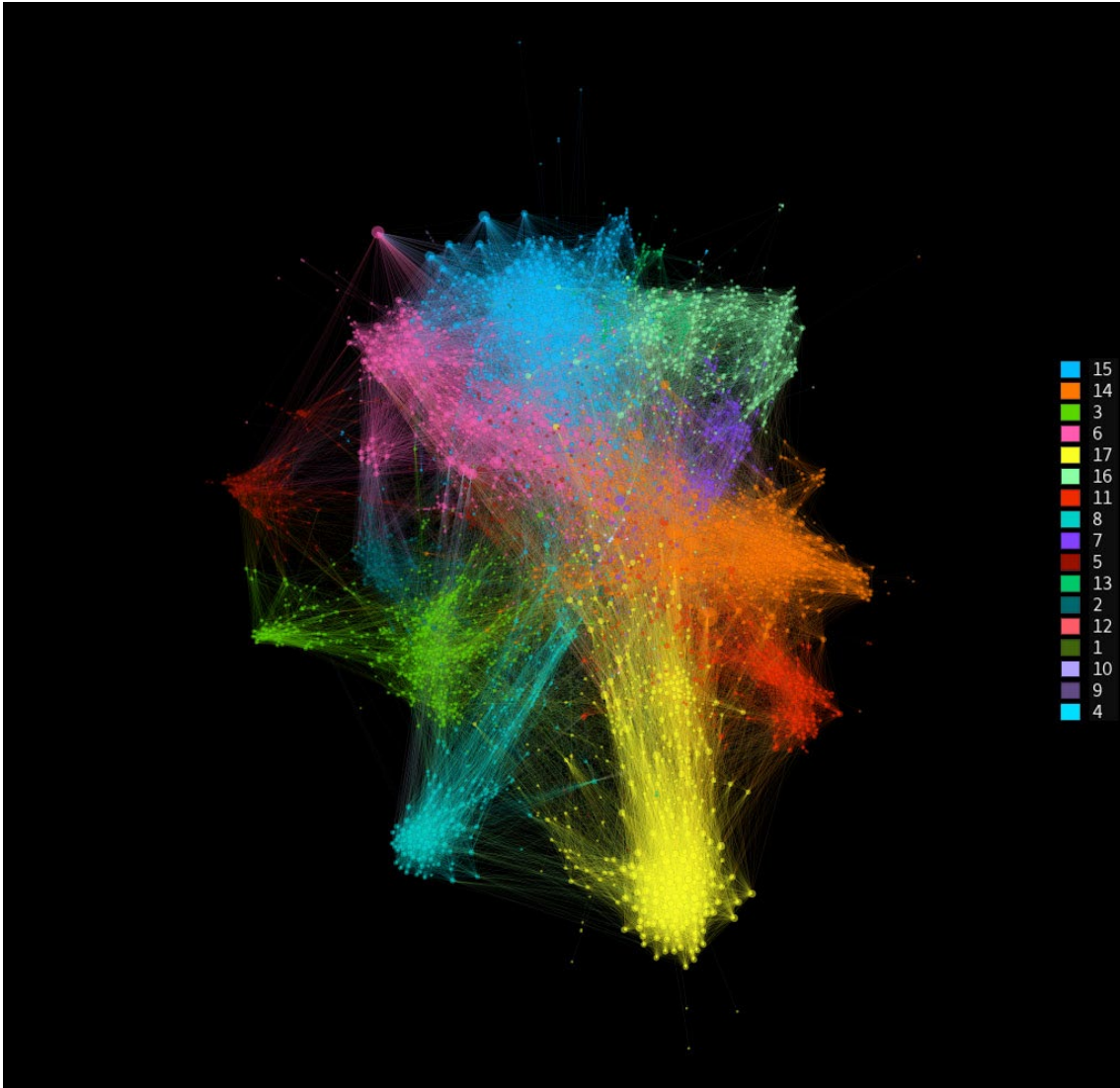
Top 10% C_b Proteins	C_b Value	Top C_e 10% Proteins	C_e Value	Top 10% Degree Proteins	Degree Value
ACTB	273823.3	RPS9	0.097403	EP300	242
CDC42	269652.8	RPS6	0.097058	AKT1	240
SOD1	256384.2	RPL5	0.096876	CTNNB1	238
AKT1	239238.5	RPS19	0.096673	PLCG1	233
EP300	229903.4	RPS7	0.096254	KRAS	229
GRB2	215741.3	RPS14	0.096198	CREBBP	227
PLCG1	212835.8	RPS26	0.09548	NRAS	223
VAMP2	201402.3	RPL11	0.095421	HSP90AA1	221
HRAS	198741.4	RPL18A	0.094763	FYN	221
CREBBP	195327.2	RPS25	0.094611	GNB1	218
PIK3R1	183463.6	RPL27	0.094184	TP53	217
APP	181243.1	RPS28	0.094176	ACTB	216
RXRA	170595.6	RPL27A	0.094105	PTK2	205
ESR1	169031.4	RPL35A	0.09409	GNB2L1	204
FAU	163541.9	RPS12	0.093872	CCT5	200
JUN	162287.7	RPL13	0.093733	CCT2	200
PRKCA	153862.6	RPS15A	0.093662	GNAI1	199
INS	152711.3	RPS23	0.093449	HSPA8	194
SNCA	150094	RPLP0	0.093272	CCT4	193
CDK1	148712	RPL8	0.093202	GNAI2	192
GNB2L1	143030.5	RPL6	0.093019	PRKCA	192
ITGB1	141542.4	RPL13A	0.0921	CCT7	190
RELA	141076.3	RPL23	0.092037	RELA	190
DLG4	136797.6	RPS15	0.092005	MAPK14	189
PIK3CA	135624.2	RPL3	0.091819	RPS3	184
KRAS	134712.3	RPL4	0.091631	RPS20	184
GNB1	130565.8	RPL31	0.091313	MAPK8	183
FYN	130229.5	RPL23A	0.091243	RPS2	182
SEC13	129072.7	RPS29	0.090791	GNAI3	180
PTGES3	127904	RPS27	0.090787	GNGT1	179
STX1A	123868.8	RPL30	0.090697	PPP2CA	179
ALDH3A2	123756.7	RPL28	0.090551	GNG2	175
HDAC1	123136.4	RPL14	0.089714	RPS18	175

Table 1 (continued) : Top 10% of Proteins by Network Statistics. The top 51 proteins are shown as measured by betweenness centrality, eigenvector centrality, and degree.

Modularity

Upon applying the Louvain Method for Community Detection (Blondel et al., 2008) at sparsity $s = 2.5$, the network was found to be divided into 17 distinct communities (Figure 3), with the largest in size being comprised of 1683 nodes (module 17) and the smallest containing only 17 nodes (module 4). The modules assigned in this analysis were used later in the calculation of functional cartography as well as that of scaled inclusivity (Appendix A).

Figure 3. Modular Structure of PPIN at Threshold $s = 2.5$. Nodes are colored according to the modules of which they are members. Colored lines indicate connections between nodes, and the size of each node corresponds to its degree (k_i). The legend on the right side of the figure lists the modules in descending order of nodes contained.



Functional Cartography

Functional cartography analysis (Guimerà & Nunes Amaral, 2005) resulted in the classification of 35 proteins as provincial hubs, 118 as connector hubs, and 69 as kinless hubs ($p < .05$, Figure 4, Figure 4). Of these, 5 provincial hubs, 23 connector hubs, and 33 kinless hubs overlapped with the conventionally (Sporns et al., 2007) defined hubs found in table 1.

Figure 4. Functional Cartography Classification of Hub Proteins. Hubs as determined by functional cartography are shown below the hub cutoff (p value) of .05. In all, 222 nodes out of the total 4480 had p values less than .05 and are shown here as colored points on the plot.

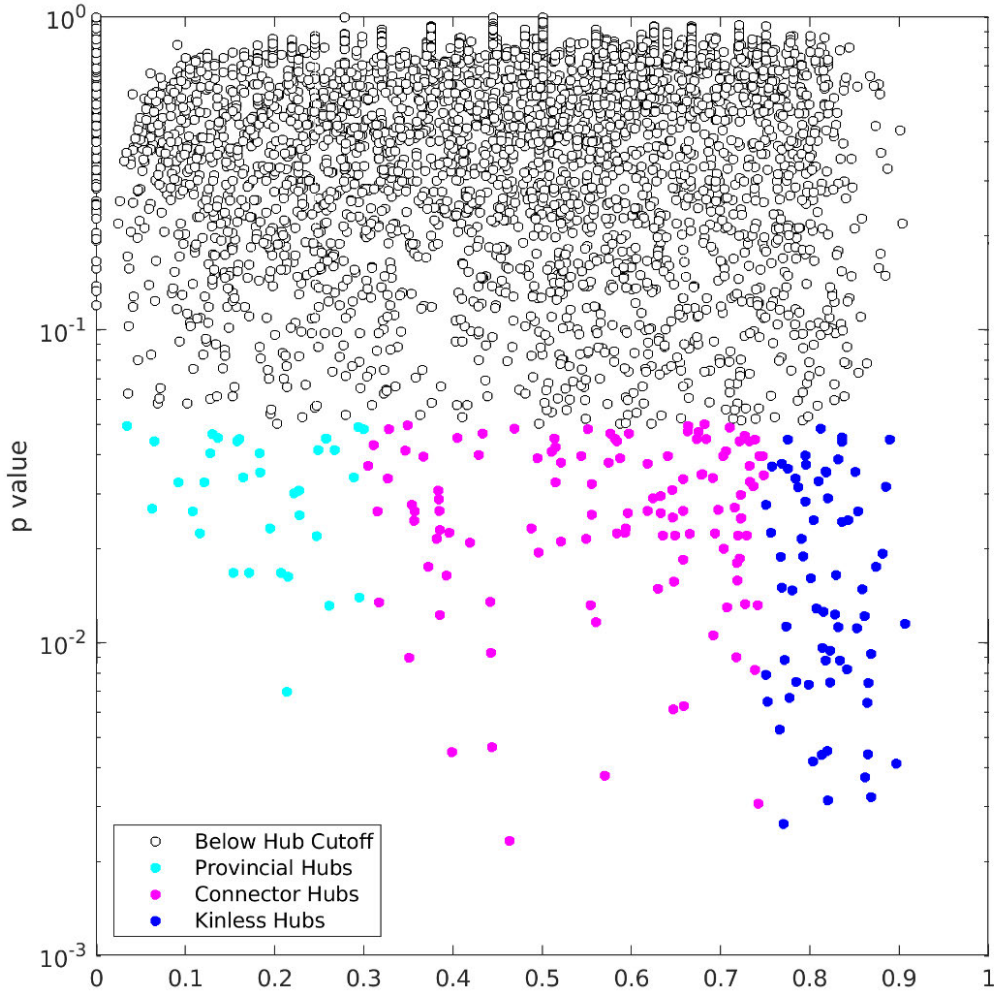
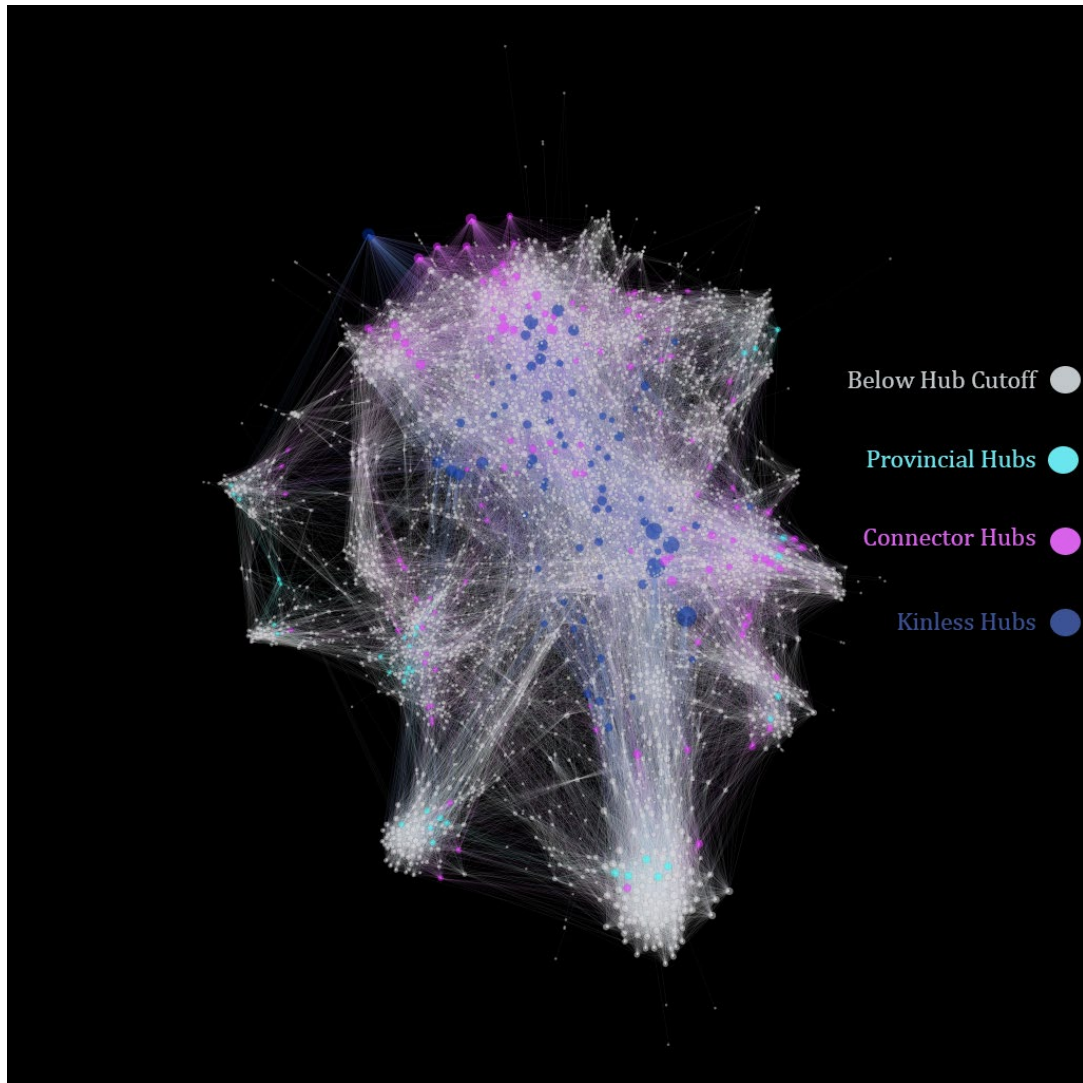


Figure 5. Topological Distribution of Functional Cartography Hubs Image depicts the thresholded network's arrangement of functional cartography hubs. Nodes are sized by degree (k_i).

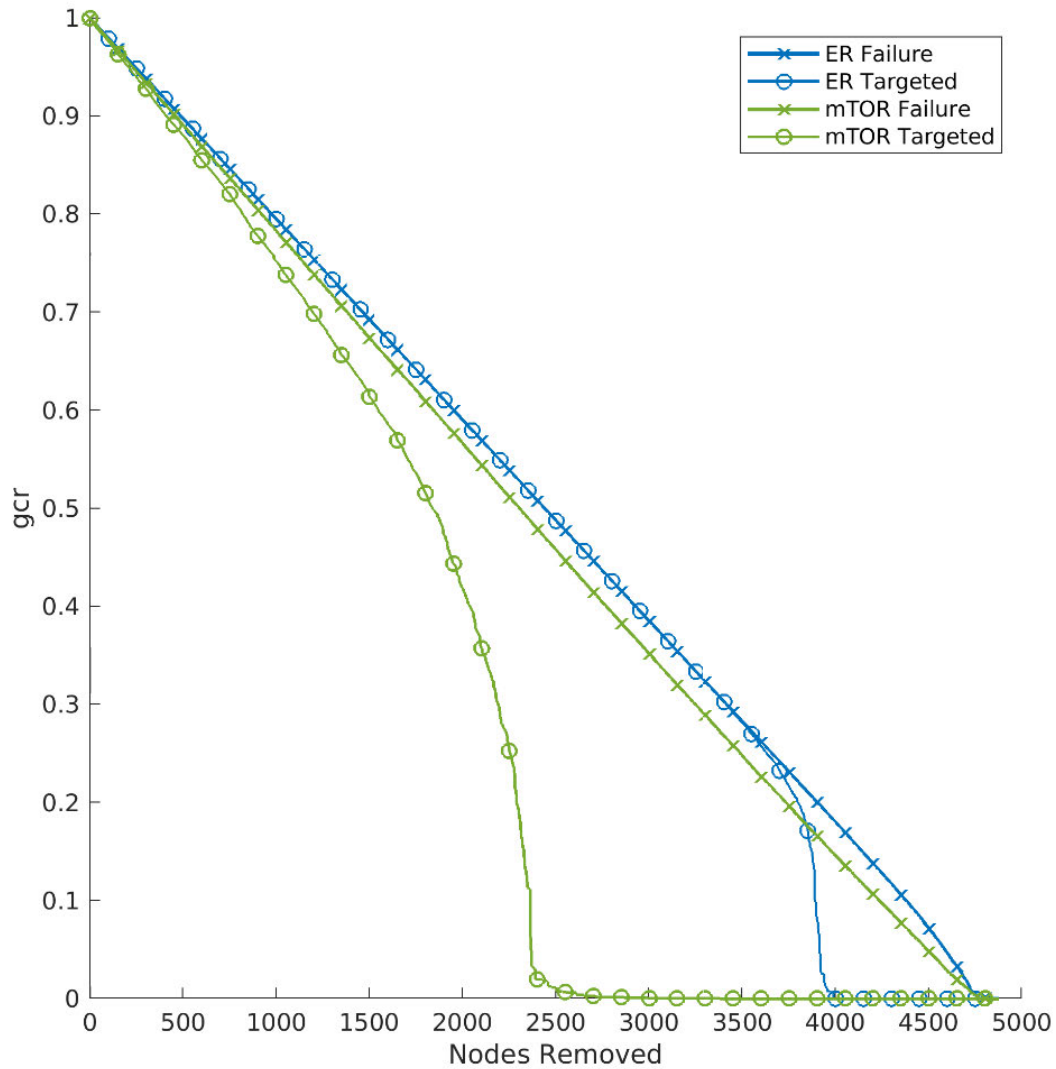


Network Perturbation Analysis

Dissolution of the Giant Component

Upon removing nodes in random and in order of descending k_i for both our thresholded network and its randomly rewired ER counterpart, we found that both the forms of attack resulted in earlier disappearance of the giant component for the thresholded network than for the ER network. This was especially true of the targeted attack, with the giant component disappearing once around the top 2500 nodes in terms of degree were removed, versus the same occurring around the 4000 node point for the ER network. This drastically earlier lethality for the mTOR network upon degree-based attack implies at some level of scale-free character exists within our network (Albert et al., 2000a), as is expected of a PPIN (Jeong et al., 2000).

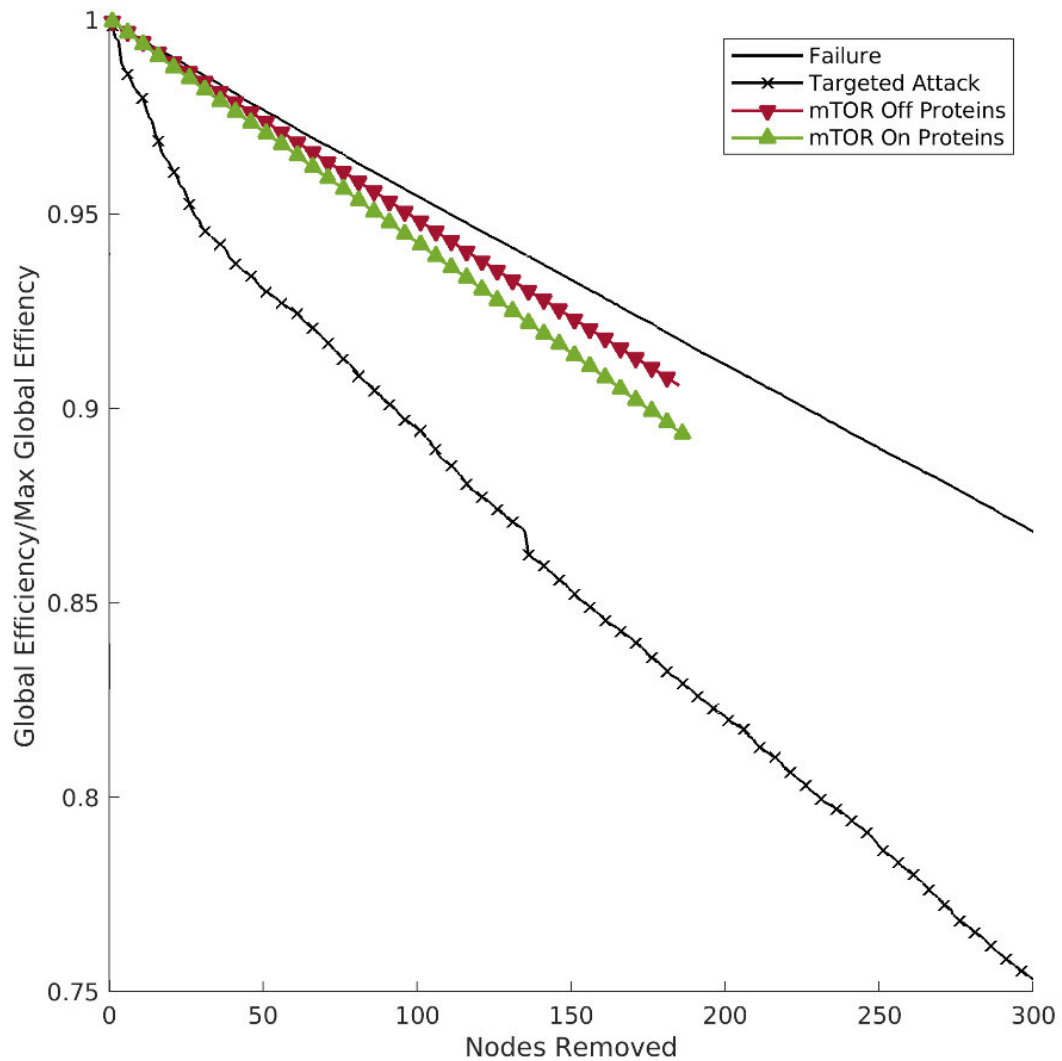
Figure 7. Dissolution of Network as Measured by Giant Component Ratio. Plots depict the fraction of nodes that remain within each network's giant component (gcr) as nodes are removed either randomly (failure) or by degree (targeted). This analysis, along with the degree distribution, established that the mTOR network exhibited scale-free character as far as the vulnerability of hub proteins to targeted attack.



Differentially Expressed Proteins -BasedAttack

The proteins whose log fold-changes were either significantly positive (“mTOR off”) or (“mTOR on”) upon administration of rapamycin (Niere et al., 2016) were more damaging to network global efficiency than random attack/failure. Despite having no distinguishing centrality values among the proteins on either list (such as high degree).

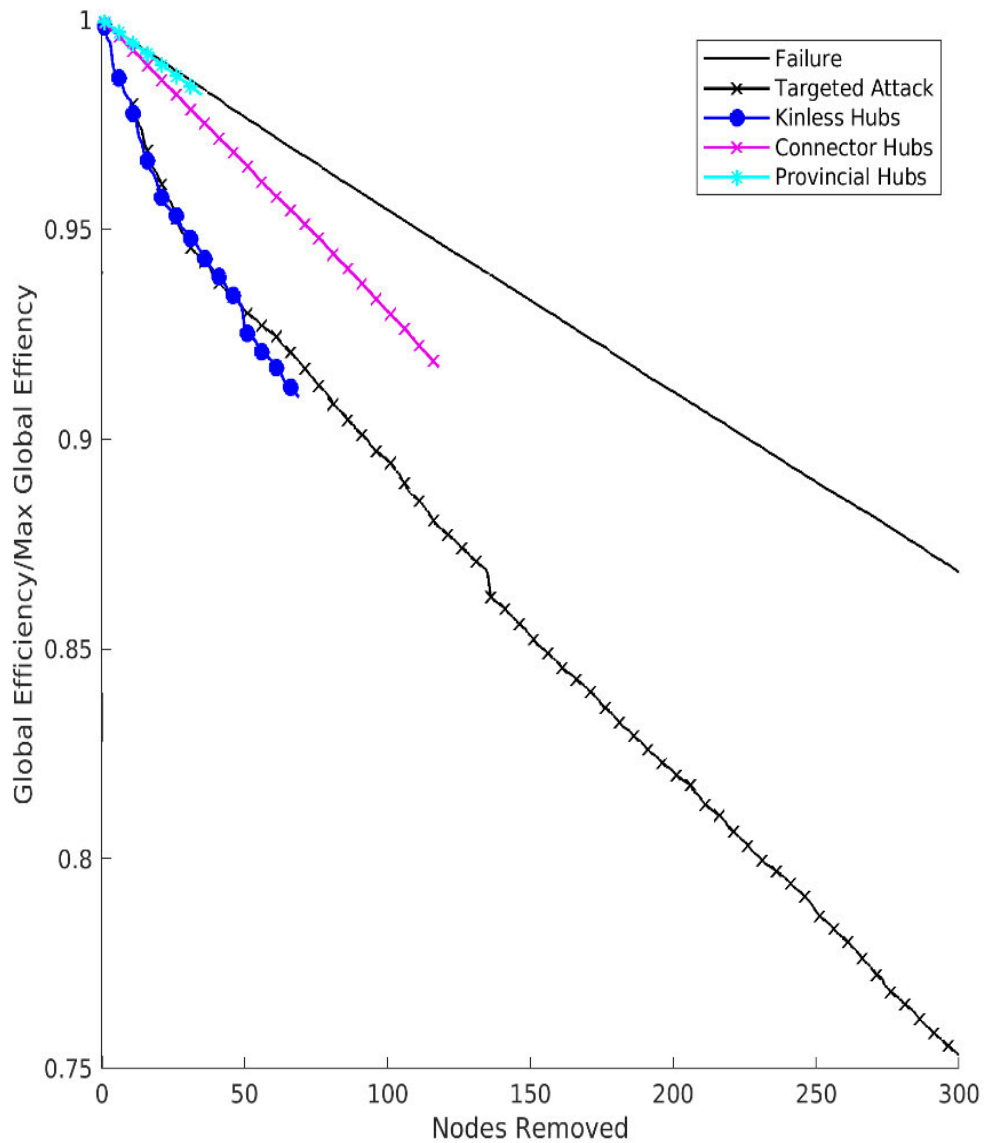
Figure 8. Removal of Differentially Expressed Proteins Compared to Failure and Targeted Attack. mTOR on and mTOR off proteins were removed in 10 randomized orders from the network and were more damaging to network signal propagation as measured by global efficiency than a random attack. Global efficiency was normalized by dividing each E_{glob} value by the maximum global efficiency of 0.3170.



Functional Cartography-Based Attack

Figure 9 shows the results of the functional cartography-based network perturbation analysis: in terms of detriment to communication efficacy, removal of the kinless hubs from the network proved to be the most damaging, surpassing the removal of sequentially highest-degree nodes in terms of global efficiency value after removal of the last kinless hub, with a final value of $E_{globkh}/E_{globmax} = 0.2885$ for the kinless hubs and vs. $E_{globta} / E_{globmax}=2915$ for degree-based/targeted attack E_{glob} value matched to the same number of nodes removed. There was no appreciable difference between provincial hub removal and random attack, and connector hubs lie somewhere between the kinless and provincial hubs in terms of their effect on network efficiency.

Figure 9. Functional Cartography Hub-Based Network Attack. Kinless, provincial, and connector hubs were removed from the network in order of descending k_i whilst recording global efficiency. Provincial hub-based attack showed similarity to random failure, whereas removal of connector hubs and kinless hubs were markedly damaging than network failure.



Assortativity

Assortativity analysis (Table 2) on the node neighborhoods of the same protein revealed that the neighborhood defined by kinless hubs and their first neighbors had the most disassortative value ($r_{jk} = -0.4145$), with all selected subnetworks showing a lower assortativity value than the whole network overall ($r_{jk} = 0.1094$). We found disassortativity among the top 69 nodes by degree ($r_{jk} = -0.2977$), with this number of nodes having been chosen for direct comparison with kinless hubs. However, with the overlap removed between the kinless hubs and these top degree hubs (33 proteins in common), the assortativity value of the top-degree proteins shifted to become positive ($r_{jk} = 0.0345$). Adjusting for the overlap between functional cartography hubs and hubs by virtue of degree, we found that kinless hubs were still disassortative even with the 33 common top degree hubs removed from the list of 69 kinless hubs, amounting to a value of $r_{jk} = -0.1367$.

Table 2. Assortativity Values of Selected Subnetworks. The full mTOR network, all functional cartography hubs, and mTOR on/off proteins were examined in terms of the subnetworks comprising these nodes and their first neighbors (all nodes one edge away in distance). The first 69 nodes of the degree-based attack were included as well to correspond with the total number of kinless hubs. For functional cartography hubs, we measured assortativity a second time after removing any nodes shared between each node list and nodes from the top 10% of degree.

Network / Subnetwork	Full Network	First 69 of Top Degree Nodes	Kinless Hub First Neighbor Network	Connector First Neighbor Network	Provincial Hub First Neighbor Network	Mtor-On Proteins	Mtor Off Proteins
r_{jk} Value	0.1094	-0.2977	-0.4145	-0.1087	-0.0708	0.0362	0.0505
r_{jk} w/ Overlap Removed	n/a	0.0345	-0.1367	0.1070	0.2131	n/a	n/a

DISCUSSION

Consistency with Previous Findings

As noted in the results, 5 of the 113 proteins in the top 10% of centrality were found to be among the 10 hubs by Niere and colleagues (2016). Though, it is impossible to reconstruct the network made in this previous study for direct comparison, as BisoGenet's servers are offline at the time of writing, we are satisfied that this constitutes satisfactory internal consistency given that the present and previous networks were constructed using fundamentally different methods and intentions.

Threshold Dependency of Results

It is worth noting that the vast majority of results were computed at the selected sparsity value of $s = 2.5$. Were we to run our analyses at a different threshold it is entirely

possible that we could have derived results and conclusions divergent from those offered in the present study. It has been noted that specifically, PPI network thresholds should be navigated cautiously and set in a context-dependent manner (Bozhilova et al., 2019). This is especially the case in measures like modularity, which can be used to assess groups of proteins from a functional standpoint: community structure is ultimately something that varies across thresholds and doesn't exist in any permanent state per se and so it is difficult to ascribe with great confidence the existence of any "true" module of proteins (Steen et al., 2011).

With that said, we thresholded the mTOR network using the combined approach of measuring network stats while increasing sparsity in addition to respecting the confidence/stringency values established by STRING's algorithm (Szklarczyk et al., 2019). Many studies using PPIN's to mine for data simply choose a stringency value set by the software of choice, and simply elaborate on the results provided by this somewhat arbitrary value (Bozhilova et al., 2019). By no means can we declare that our network provides a perfect, complete view of the mTOR interactome, but we can however state that our thresholding method was employed to minimize the variability of our community structure and was conducted with the knowledge of the aforementioned potential pitfalls. We used scaled inclusivity (Steen et al., 2011) to measure to what extent module membership varied at our given threshold, and to quantitatively establish values for this phenomenon (Appendix A).

Attack Susceptibility

The targeted and random attack analyses on our network served a twofold purpose. The first was to support or refute the hypothesis that our network contained hub proteins of structural significance. The second was to characterize the groupings of these proteins and to measure the relative importance of these groupings in terms of their lethality to network communication efficiency upon their removal. As global efficiency is the measure how easily any given node can “talk” to another node via the edges existent in the network, this translates in the biological context to the ease with which essential protein can interact with one another (Latora & Marchiori, 2001).

Of all the groups of proteins we removed to systematically attack our network, the targeted, degree-based attack expectedly provided the greatest detriment to the final global efficiency value (Figure 8), as the rapid dissolution of PPIN’s when removing hubs is well-documented and partially accounts for the establishment of the centrality-lethality rule (Albert et al., 2000b; Jeong et al., 2001). While this analysis reinforced the structural, and by extension, chemical importance of the hub proteins of high centrality, it also was valuable as a metaphorical measuring bar against which we could measure attack efficacy using protein/node groupings less well-established and obvious as degree, eigenvector centrality, and betweenness centrality.

As such, we found that, despite comprising only 69 proteins, removing the kinless hubs damaged network efficiency moreso than even the degree-based attack for the same amount of proteins. This finding was surprising, as stated previously, degree has a well-known role in terms of the structural integrity of PPIN’s (Jeong et al., 2001). This finding

led us to surmise that some topological feature outside the conventional measures of centrality were conferring such destructive power to these nodes.

FUTURE STUDIES

Rab28

One protein of particular interest to us in our follow-up experiments is the protein known as Rab28, which is a small Ras-related GTPase encoded by the gene of the same name (Brauers et al., 1996). This interest derives from the fact that, despite having a relatively low degree among top 10% of proteins by degree (117 vs. the highest value 568), removal of this protein constituted the largest drop in global efficiency (.0017) in both the kinless hub-based and degree-based attacks (Figure 8). Additionally, there are no currently known associations between this protein and any of the neurodegenerative diseases in which mTOR is implicated (RAB28 - ras-related protein rab-28 - homo sapiens (human) | UniProtKB | UniProt.), making it a good candidate for further research and study. To begin, we plan on using antibodies against Rab28 to measure expression levels in an animal model of Tuberous Sclerosis Complex (TSC), a disease that is well-characterized by its association with mTOR dysfunction (Inoki et al., 2005).

Module Identification and Gene Ontology

Though our modularity analysis produced results (Figure 3) that effectively categorized our proteins according to their communities of membership, we did not ascribe function to these modules in the present study. Ongoing work is currently being conducted to use gene ontology (Ashburner et al., 2000) to assign functional annotations to each protein and module. This information can supplement the findings

from the present study by allowing us to examine the role of specific proteins within their module, providing an additional context for data interpretation.

CONCLUSION

In accordance with our goals, we have used topological network features in order to identify proteins of potential regulatory and/or structural significance within the interaction network of proteins detected by Niere and team (2016). Conventional centrality metrics yielded a list of proteins that were themselves useful to the current study of mTOR's involvement in neurodegenerative diseases, and the combination of functionality cartography and assortativity measures helped us to refine and/or supplement these findings.

Network analysis remains a useful tool that complements and augments the increasingly complex field of molecular neuroscience. It is the author's hope that the contributions from the present study can help with the task of elucidating mechanisms for the complex, heterogeneous diseases in which mTOR is implicated through the use of the described techniques.

APPENIDX A: Scaled Inclusivity Results at Sparsity $s=2.5$

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
RPS27A	Top 10% Centrality	1.213519	TP53	Top 10% Centrality	2.16237	INS	Top 10% Centrality	0.381576
UBA52	Top 10% Centrality	1.390839	ACTB	Top 10% Centrality	1.63678	SNCA	Top 10% Centrality	0.914959
RHOA	Top 10% Centrality	1.882353	CDC42	Top 10% Centrality	1.882353	CDK1	Top 10% Centrality	1.793471
SRC	Top 10% Centrality	1.882353	SOD1	Top 10% Centrality	0.008235	GNB2L1	Top 10% Centrality	1.862622
UBC	Top 10% Centrality	1.390839	AKT1	Top 10% Centrality	0.577883	ITGB1	Top 10% Centrality	1.882353
UBB	Top 10% Centrality	1.390839	EP300	Top 10% Centrality	2.16237	RELA	Top 10% Centrality	1.390839
RAB28	Top 10% Centrality	0.746398	GRB2	Top 10% Centrality	1.882353	DLG4	Top 10% Centrality	0.420824
PRKACA	Top 10% Centrality	1.060065	PLCG1	Top 10% Centrality	1.882353	PIK3CA	Top 10% Centrality	1.882353
EGFR	Top 10% Centrality	1.882353	VAMP2	Top 10% Centrality	1.510187	KRAS	Top 10% Centrality	1.882353
HSPA8	Top 10% Centrality	1.512679	HRAS	Top 10% Centrality	1.882353	GNB1	Top 10% Centrality	1.060065
MAPK1	Top 10% Centrality	0.509371	CREBBP	Top 10% Centrality	2.16237	FYN	Top 10% Centrality	1.882353
CTNNB1	Top 10% Centrality	0.019999	PIK3R1	Top 10% Centrality	1.882353	SEC13	Top 10% Centrality	1.319965
PRKACB	Top 10% Centrality	1.060065	APP	Top 10% Centrality	2.208934	PTGES3	Top 10% Centrality	0.348247
HSP90AA1	Top 10% Centrality	0.346513	RXRA	Top 10% Centrality	2.16237	STX1A	Top 10% Centrality	0.12516
MAPK3	Top 10% Centrality	0.509371	ESR1	Top 10% Centrality	1.783163	ALDH3A2	Top 10% Centrality	2.755304

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
PRKACG	Top 10% Centrality	1.060065	FAU	Top 10% Centrality	1.862622	HDAC1	Top 10% Centrality	2.16237
RAC1	Top 10% Centrality	1.882353	JUN	Top 10% Centrality	2.16237	RPS4X	Top 10% Centrality	1.86262 2
INS	Top 10% Centrality	0.381576	PRKCA	Top 10% Centrality	0.509371	RPS8	Top 10% Centrality	1.86262 2
SNCA	Top 10% Centrality	0.914959	RPS3	Top 10% Centrality	1.862622	RPS26	Top 10% Centrality	1.86262 2
CDK1	Top 10% Centrality	1.793471	RPS3A	Top 10% Centrality	1.862622	RPL11	Top 10% Centrality	1.86262 2
GNB2L1	Top 10% Centrality	1.862622	RPS20	Top 10% Centrality	1.862622	RPL18A	Top 10% Centrality	1.86262 2
ITGB1	Top 10% Centrality	1.882353	RPS16	Top 10% Centrality	1.862622	RPS25	Top 10% Centrality	1.86262 2
RELA	Top 10% Centrality	1.390839	RPS11	Top 10% Centrality	1.862622	RPL27	Top 10% Centrality	1.86262 2
DLG4	Top 10% Centrality	0.420824	RPS18	Top 10% Centrality	1.862622	RPS28	Top 10% Centrality	1.86262 2
PIK3CA	Top 10% Centrality	1.882353	RPS5	Top 10% Centrality	1.862622	RPL27A	Top 10% Centrality	1.86262 2
KRAS	Top 10% Centrality	1.882353	RPS13	Top 10% Centrality	1.862622	RPL35A	Top 10% Centrality	1.86262 2
GNB1	Top 10% Centrality	1.060065	RPSA	Top 10% Centrality	1.862622	RPS12	Top 10% Centrality	1.86262 2
FYN	Top 10% Centrality	1.882353	RPL9	Top 10% Centrality	1.862622	RPL13	Top 10% Centrality	1.86262 2
SEC13	Top 10% Centrality	1.319965	RPS24	Top 10% Centrality	1.862622	RPS15A	Top 10% Centrality	1.86262 2
PTGES3	Top 10% Centrality	0.348247	RPL15	Top 10% Centrality	1.862622	RPS23	Top 10% Centrality	1.86262 2
STX1A	Top 10% Centrality	0.12516	RPS9	Top 10% Centrality	1.862622	RPLP0	Top 10% Centrality	1.86262 2
ALDH3A2	Top 10% Centrality	2.755304	RPS6	Top 10% Centrality	1.862622	RPL8	Top 10% Centrality	1.86262 2

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
HDAC1	Top 10% Centrality	2.16237	RPL5	Top 10% Centrality	1.862622	RPL6	Top 10% Centrality	1.862622
RPS4X	Top 10% Centrality	1.862622	RPS19	Top 10% Centrality	1.862622	RPL13A	Top 10% Centrality	1.862622
RPS8	Top 10% Centrality	1.862622	RPS7	Top 10% Centrality	1.862622	RPL23	Top 10% Centrality	1.862622
RPS2	Top 10% Centrality	1.862622	RPS14	Top 10% Centrality	1.862622	RPS15	Top 10% Centrality	1.862622
RPL3	Top 10% Centrality	1.862622	MAPK8	Top 10% Centrality	0.509371	CYP2E1	Provincial Hub	2.755304
RPL4	Top 10% Centrality	1.862622	GNAI3	Top 10% Centrality	1.060065	LPCAT4	Provincial Hub	2.688343
RPL31	Top 10% Centrality	1.862622	GNGT1	Top 10% Centrality	1.060065	PLA2G16	Provincial Hub	2.688343
RPL23A	Top 10% Centrality	1.862622	PPP2CA	Top 10% Centrality	0.577883	PLB1	Provincial Hub	2.688343
RPS29	Top 10% Centrality	1.862622	GNG2	Top 10% Centrality	1.060065	ATP5D	Provincial Hub	2.374063
RPS27	Top 10% Centrality	1.862622	AGXT2	Provincial Hub	2.755304	NDUFB7	Provincial Hub	2.374063
RPL30	Top 10% Centrality	1.862622	GOT2	Provincial Hub	2.755304	NDUFB10	Provincial Hub	2.374063
RPL28	Top 10% Centrality	1.862622	TKTL2	Provincial Hub	2.755304	NDUFB8	Provincial Hub	2.374063
RPL14	Top 10% Centrality	1.862622	AGXT	Provincial Hub	2.755304	CYC1	Provincial Hub	2.374063
NRAS	Top 10% Centrality	1.882353	GOT1L1	Provincial Hub	2.755304	UQCRQ	Provincial Hub	2.374063
PTK2	Top 10% Centrality	1.882353	TALDO1	Provincial Hub	2.755304	SNRPF	Provincial Hub	1.512679
CCT5	Top 10% Centrality	1.862622	CYP3A4	Provincial Hub	2.755304	SNRPE	Provincial Hub	1.512679
CCT2	Top 10% Centrality	1.862622	ACOX3	Provincial Hub	2.755304	HIST2H2AC	Provincial Hub	2.16237

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
GNAI1	Top 10% Centrality	1.060065	GSTO1	Provincial Hub	2.019778	HIST1H2AC	Provincial Hub	2.16237
CCT4	Top 10% Centrality	1.862622	GOT1	Provincial Hub	2.755304	YKT6	Provincial Hub	1.510187
GNAI2	Top 10% Centrality	1.060065	ALDH7A1	Provincial Hub	2.755304	RAB1B	Provincial Hub	1.507179
CCT7	Top 10% Centrality	1.862622	TKT	Provincial Hub	2.755304	RAB1A	Provincial Hub	1.507179
MAPK14	Top 10% Centrality	0.509371	GSTA2	Provincial Hub	2.021481	BSG	Connector Hub	0.047254
PKM	Connector Hub	0.244315	PLD1	Connector Hub	2.688343	SNRPD2	Connector Hub	1.512679
PKLR	Connector Hub	0.244315	PLA2G4A	Connector Hub	2.210305	HLA-G	Connector Hub	0.28709
ENPP1	Connector Hub	2.090174	APOA2	Connector Hub	2.22297	SNAP25	Connector Hub	0.514033
NME1-NME2	Connector Hub	0.893918	PPP2CB	Connector Hub	0.577883	STXBP1	Connector Hub	0.38579
ENPP3	Connector Hub	2.090174	PPP2R1B	Connector Hub	0.577883	DLG3	Connector Hub	0.420824
DLD	Connector Hub	2.062881	PPP2R1A	Connector Hub	0.577883	KCNAB2	Connector Hub	0.283801
ACO2	Connector Hub	1.327408	PPP1CA	Connector Hub	0.577883	CASK	Connector Hub	0.514033
EHHADH	Connector Hub	2.755304	PPP1CC	Connector Hub	0.577883	RBX1	Connector Hub	1.87572
ACOX1	Connector Hub	2.755304	PPP1CB	Connector Hub	0.577883	NEDD8	Connector Hub	0.955379
PDHB	Connector Hub	2.062881	TUBA1A	Connector Hub	1.720326	CCNA2	Connector Hub	1.793471
NOS2	Connector Hub	2.755304	DYNC1H1	Connector Hub	1.329085	HIST1H2BA	Connector Hub	2.16237
CS	Connector Hub	2.062881	DCTN1	Connector Hub	1.329085	PLK1	Connector Hub	1.793471

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
GAD1	Connector Hub	2.75530 4	PAFAH1B1	Connector Hub	1.10010 8	AURKB	Connector Hub	1.79347 1
GAD2	Connector Hub	1.59381 9	ATP5H	Connector Hub	2.37406 3	UBE2N	Connector Hub	1.39083 9
GSTP1	Connector Hub	2.75530 4	COX7C	Connector Hub	2.37406 3	RUVBL1	Connector Hub	0.95537 9
APOB	Connector Hub	2.22297	NDUFAB1	Connector Hub	2.37406 3	CUL1	Connector Hub	1.87572
APOA1	Connector Hub	2.22297	SNRPG	Connector Hub	1.51267 9	HIST1H2AJ	Connector Hub	2.16237
PLD2	Connector Hub	2.68834 3	NCBP2	Connector Hub	1.51267 9	HIST1H2AD	Connector Hub	2.16237
HIST2H2BE	Connector Hub	2.16237	PIK3CB	Connector Hub	1.88235 3	CLTC	Connector Hub	1.62640 6
UBE2D1	Connector Hub	1.39083 9	RAC3	Connector Hub	1.88235 3	PSMA3	Connector Hub	0.88294 1
PCNA	Connector Hub	2.16237	TLN1	Connector Hub	1.88235 3	EPRS	Connector Hub	1.86262 2
UBE2D2	Connector Hub	1.39083 9	LCK	Connector Hub	1.88235 3	IARS	Connector Hub	1.86262 2
TRAF6	Connector Hub	0.10008 5	SDC1	Connector Hub	1.88235 3	YWHAE	Kinless Hub	0.57788 3
H2AFX	Connector Hub	2.16237	BCAR1	Connector Hub	1.88235 3	YWHAB	Kinless Hub	0.57788 3
POLR2E	Connector Hub	1.63615 5	SHC1	Connector Hub	1.88235 3	CAMK2B	Kinless Hub	0.89929 2
POLR2A	Connector Hub	2.16237	NCK1	Connector Hub	1.88235 3	YWHAZ	Kinless Hub	0.57788 3
HIST1H2BB	Connector Hub	2.16237	ITGB3	Connector Hub	1.88235 3	PPP2R5D	Kinless Hub	0.57788 3
HIST1H2BN	Connector Hub	2.16237	ACTG1	Connector Hub	1.88235 3	HSP90AB1	Kinless Hub	0.34824 7
VCL	Connector Hub	1.88235 3	CLTA	Connector Hub	1.62640 6	MAPT	Kinless Hub	0.68437 6

Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value	Protein Name	Grouping	SI Value
PIK3R2	Connector Hub	1.88235 3	AP2M1	Connector Hub	1.62640 6	RAN	Kinless Hub	1.51267 9
RAC2	Connector Hub	1.88235 3	TMED10	Connector Hub	0.87987 1	HLA-A	Kinless Hub	0.28709
ITGAV	Connector Hub	1.88235 3	AP2A1	Connector Hub	1.62640 6	SKP1	Kinless Hub	1.87572
ITGA2B	Connector Hub	0.86903 2	SPTAN1	Connector Hub	0.88219 8	NR3C1	Kinless Hub	1.78237 1
PXN	Connector Hub	1.88235 3	GRIA1	Connector Hub	0.01524 2			
PAK1	Connector Hub	1.88235 3	ARF1	Connector Hub	1.62640 6			
ITGA1	Connector Hub	1.88235 3	SEC22B	Connector Hub	1.51018 7			
RAF1	Kinless Hub	1.41358 2	CCT6A	Kinless Hub	1.86262 2			
MAP2K1	Kinless Hub	1.88235 3	CCT3	Kinless Hub	1.86262 2			
SDC2	Kinless Hub	1.88235 3	TCP1	Kinless Hub	1.86262 2			
CASP3	Kinless Hub	1.88235 3	PFDN5	Kinless Hub	1.86262 2			
PAK2	Kinless Hub	1.40027 4	CCNB1	Kinless Hub	1.79347 1			
STAT1	Kinless Hub	1.88235 3	MDM2	Kinless Hub	1.22496 9			
MYO1C	Kinless Hub	0.50707	GSK3B	Kinless Hub	0.03082 4			
ARRB2	Kinless Hub	0.00621 9						
ARRB1	Kinless Hub	0.51034 5						
NFKB1	Kinless Hub	0.1343						

APPENDIX B: Matlab Scripts Written for Analyses

DATA IMPORT

```

%Load in edgelist
data=load('STRING network.txt');

%Isolate unique nodes in a descending vector
uniquenodes = unique(data(:,1));
aij = zeros(length(uniquenodes));

%Loop over nodes and convert vector into adjacency matrix
for x = 1:size(data,1)
    aij(uniquenodes==data(x,1),uniquenodes==data(x,2))=1;
end
%Sum adjacency matrix and its transposition to create final adjacency
matrix
aij=aij+aij';

%Save Outputs
save('matrixoutput.mat','aij','uniquenodes')

```

NETWORK CONSTRUCTION

```

clear all

%Load in edgelist and binarize such that the connection to each node is
bidirectional
Edgelist=readtable('Edgelist.csv');
EdgesFull=table2cell(Edgelist(:,2));
sourceedge=cell(size(EdgesFull));
targetedge=cell(size(EdgesFull));
for x=1:length(EdgesFull)
    temp=strsplit(EdgesFull{x},' (pp) ');
    sourceedge{x}=temp{1};
    targetedge{x}=temp{2};
end

%Find unique nodes and label their row/column locations from 1 to 5601
- do the same to their edges
uniquenode=unique([sourceedge,targetedge]);
sourceedgenum=zeros(size(EdgesFull));
targetedgenum=zeros(size(EdgesFull));

for x=1:length(uniquenode)
    x
    sourceedgenum(strcmp(sourceedge,uniquenode(x)))=x;
    targetedgenum(strcmp(targetedge,uniquenode(x)))=x;
end

%Save outputs

```

```

save('stringadj.mat','sourceedgenum','targetedgenum','uniquenode')

%Load in STRING edge weights (confidence values) and apply them to
existing adjacency matrix
edgeweights=table2array(Edgelist(:,[6 8 9]));
edgeweights(isnan(edgeweights))=0;
edgeweights=max(edgeweights,[],2);
aij=zeros(5601);
waij=zeros(5601);

for x=1:length(edgeweights)
    x
    waij(sourceedgenum(x),targetedgenum(x))=edgeweights(x);
    aij(sourceedgenum(x),targetedgenum(x))=1;
end

%Reconstruct adjacency matrix with imported edge weights and save out
new matrix
aij=aij+aij';
waij=waij+waij';
save('stringaij.mat','aij','waij')

THRESHOLDING

clear all
load('/isilon/datalake/lcbn_research/final/beach/Mike/stringaij.mat');

%Sparsity minimum, maximum, and interval
sparsities=[1.5:.1:15];

%Sort weighted adjacency matrix into descending vector
sortedaij=sort(waij(:),'descend');

%thresholding loop: remove edges to reach sparsities in above interval
in order of descending STRING confidence - measure gcr, fcr, and global
efficiency
    for x=1:length(sparsities)

%calculate average degree for each sparsity specified in above range
        averagedegree=10^(3.74826557267/sparsities(x));

%calculate number of edges needed to reach a given sparsity
        edgesneeded=ceil(averagedegree*5601);
        threshold=sortedaij(edgesneeded);

%apply threshold to weighted adjacency matrix
        aij=waij>=threshold;

%assign component membership to nodes as matrix is thresholded
        [comps,comp_sizes] = get_components(aij);
        [maxval,maxloc]=max(comp_sizes);
        gcr(x)=maxval/5601;
    end

```

```

    comp_sizes(maxloc)=[];
    fcr(x)=mean(comp_sizes)/(5601-maxval);

%global efficiency
    E_glob(x)=efficiency_bin(aij);

%local efficiency
    E_loc_temp=efficiency_bin(aij,1);
    E_loc(x)=mean(E_loc_temp);
end

%Save loop outputs as XLS andmat
output=array2table(sparsities');
output=addvars(output,E_glob,E_loc);
writetable(output,'output.xls')
save('thresholdinitial.mat','gcr','fcr');

ACROSS-THRESHOLD NETWORK STATS

clear all
load('/isilon/datalake/lcbn_research/final/beach/Mike/stringaij.mat');
%sparsity minimum, interval, and maximum
sparsities=[1.5:.09:3.3];

%sort highest confidence values among edges into descending vector
sortedaij=sort(waij(:),'descend');

%thresholding loop - same as previous script
for x=1:length(sparsities)
    sparsities(x)
    averagedegree=10^(3.74826557267/sparsities(x));
    edgesneeded=ceil(averagedegree*5601);
    threshold=sortedaij(edgesneeded);
    aij=waij>=threshold;
    fprintf('working on Global Efficiency... ')
    E_glob=efficiency_bin_Mike(aij);
    fprintf('done\n')

%Betweenness Centrality
    fprintf('working on Betweenness Centrality... ')
    BC=betweenness_bin(aij);
    fprintf('done\n')

%Clustering Coefficient
    fprintf('working on Clustering Coefficient... ')
    C=clustering_coef_wu(aij);
    fprintf('done\n')

%Degree
    fprintf('working on Degree... ')
    deg=degrees_und(aij);
    fprintf('done\n')

```

```

%Eigenvector Centrality
    fprintf('working on Eigenvector Centrality... ')
    v=eigenvector_centrality_und(aij);
    fprintf('done\n')

%Participation Coefficient
    fprintf('working on Participation Coefficient... ')
    p=participation_coef(aij,0);
    fprintf('done\n')

%Write output of network stats into tabular format
    output=array2table(E_glob');

%Add new variables from any added network stats below
    output=addvars(output,BC',C,deg',v,p);

%Save output as xls
writetable(output,['networkstats' num2str(sparsities(x)) '.xls'])

end

LOUVAIN MODULARITY

clear all
load('/isilon/datalake/lcbn_research/final/beach/Mike/stringaij.mat');
sparsities=[2.0:.25:3.0];
sortedaij=sort(waij(:),'descend');
for x=1:length(sparsities)
    sparsities(x)
    averagedegree=10^(3.74826557267/sparsities(x));
    edgesneeded=ceil(averagedegree*5601);
    threshold=sortedaij(edgesneeded);
    aij=waij>=threshold;
%Use Qcut function to define Louvain module membership
    fprintf('Working on Qcut... ')
    [cluster,q]=Qcut(aij,2:4,0,.25)
    clusterout=zeros(5601,1);

    for y=1: length(cluster)
        clusterout(cluster{y})=y;
    end

    fprintf('Done!\n')

%save out module membership and variables
qcutmodules=array2table(clusterout)
writetable(qcutmodules,['qcutmodules' num2str(sparsities(x))
'.xls'])
save(['qcutvariables' num2str(sparsities(x))
'.mat'],'clusterout','q');
end

FUNCTIONAL CARTOGRAPHY

```

```

clear all
load('/isilon/datalake/lcbn_research/final/beach/Mike/stringaij.mat');
sparsities=[2.0:.25:3.0];
sortedaij=sort(waij(:),'descend');
for x=1:length(sparsities)
    FAij{1}=['/isilon/datalake/lcbn_research/final/beach/Mike/thresh'
            num2str(sparsities(x)) '/traij' num2str(sparsities(x)) '.mat'];
    FMod{1}=['/isilon/datalake/lcbn_research/final/beach/Mike/thresh'
            num2str(sparsities(x)) '/modularity/clus.mat'];
    load (FMod{1});
    load (FAij{1});
    for y=1:length(clus)
        temp=clus{y};
        for z=1:length(temp)
            temp(z)=find(pointer==temp(z));
        end
        clus{y}=temp;
    end
    FMod{1}=['/isilon/datalake/lcbn_research/final/beach/Mike/thresh'
            num2str(sparsities(x)) '/modularity/clus_fc.mat'];
    save(FMod{1},'clus');
    saveflag=1;
    zflag=0;
    hubcutoff=.05;
%Functional Cartography Function
    hubClas = xnet_functional_cartography_mike(FAij,FMod,0,1,0);
end

NETWORK PERTURBATION ANALYSIS

%Frontend function
maxNumCompThreads(5)
load
/isilon/datalake/lcbn_research/final/beach/Mike/thresh2.5/traij2.5.mat;
parfor x=1:10
    randattackfull_backend(aij,x,'randomvars')
end

%Backend function
function randattackfull_backend(aij,x,basename)
datetime()
rdata=(randsample(4880,4880));
shaij=aij;

for y=1:4880
    fprintf('x=%d\ty=%d\n',x,y);

    shaij(rdata(y),:)=0;
    shaij(:,rdata(y))=0;

%giant component and fractional component size
    [comps,comp_sizes] = get_components(shaij);

```

```

numcomps(y)=length(comp_sizes);

[maxval,maxloc]=max(comp_sizes);
gcr(y)=maxval/4880;
comp_sizes(maxloc)=[];

fcr(y)=mean(comp_sizes);

fprintf('gcr=%d \n',gcr(y));

tic;
%global efficiency
fprintf('working on Global Efficiency... ')

[Eglob ENi] = xnet_calc_Eglob_all_Mike(shaij);
E_globout(y)=(Eglob);
fprintf('E_globout=%d \n',E_globout(y));

fprintf('done\n')
toc;
end
save([basename '_' num2str(x)
'.mat'],'rdata','gcr','fcr','numcomps','E_globout');

```

ASSORTATIVITY

```

clear all
load
/isilon/datalake/lcbn_research/final/beach/Mike/thresh2.5/traij2.5.mat;
%Kinless Overlapped
data=readtable('KHOverlapped.xlsx');
data=table2array(data);
for x=1:length(data)
    nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids,:)=1;
aijmask(:,nodeids)=1;
aij2=aij.*aijmask;
[rows,cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end
rjkkHO=corr(kirow,kicols)

%Top Degree Overlapped
data=readtable('TopDegreeOverlapped.xlsx');
data=table2array(data);
for x=1:length(data)

```

```

        nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids, :)=1;
aijmask(:, nodeids)=1;
aij2=aij.*aijmask;
[rows, cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end
rjkTDO=corr(kirow, kicols)

%Kinless Non-Overlapped
data=readtable('KHUnique.xlsx');
data=table2array(data);
for x=1:length(data)
    nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids, :)=1;
aijmask(:, nodeids)=1;
aij2=aij.*aijmask;
[rows, cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end
rjkKHU=corr(kirow, kicols)

%Connector Non-Overlapped
data=readtable('CHUnique.xlsx');
data=table2array(data);
for x=1:length(data)
    nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids, :)=1;
aijmask(:, nodeids)=1;
aij2=aij.*aijmask;
[rows, cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end

```

```

end
rjkCHU=corr(kirow,kicols)

%Top Degree Non-Overlapped
data=readtable('TopdegUnique.xlsx');
data=table2array(data);
for x=1:length(data)
    nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids,:)=1;
aijmask(:,nodeids)=1;
aij2=aij.*aijmask;
[rows,cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end
rjkTDU=corr(kirow,kicols)

%Within-Degree of Unique KH Non-Overlapped
data=readtable('DegreeCompareNodes.xlsx');
data=table2array(data);
for x=1:length(data)
    nodeids(x)=find(pointer==data(x));
end
aijmask=zeros(4880);
aijmask(nodeids,:)=1;
aijmask(:,nodeids)=1;
aij2=aij.*aijmask;
[rows,cols]=find(aij2);
ki=sum(aij);
kirow=zeros(size(rows));
kicols=zeros(size(cols));
for x=1:length(rows)
    kirow(x)=ki(rows(x));
    kicols(x)=ki(cols(x));
end
rjkCompare=corr(kirow,kicols)

```

BIBLIOGRAPHY

- Albert, R., Jeong, H., & Barabási, A. (2000b). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378-382. doi:10.1038/35019019
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25-29. doi:10.1038/75556
- Badkas, A., De Landtsheer, S., & Sauter, T. (2022). Construction and contextualization approaches for protein-protein interaction networks. *Computational and Structural Biotechnology Journal*, *20*, 3280-3290. doi:10.1016/j.csbj.2022.06.040
- Basu, A., Ash, P. E., Wolozin, B., & Emili, A. (2021). Protein Interaction Network Biology in Neuroscience. *Proteomics*, *21*(3-4), 1900311. doi:10.1002/pmic.201900311
- Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. doi:10.1088/1742-5468/2008/10/P10008
- Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., & Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinformatics*, *20*, 446. doi:10.1186/s12859-019-3036-6
- Brauers, A., Schürmann, A., Massmann, S., Mühl-Zürbes, P., Becker, W., Kainulainen, H., Joost, H. G. (1996). Alternative mRNA splicing of the novel GTPase Rab28 generates isoforms with different C-termini. *European Journal of Biochemistry*, *237*(3), 833-840. doi:10.1111/j.1432-1033.1996.0833p.x
- Crucitti, P., Latora, V., Marchiori, M., & Rapisarda, A. (2003). Efficiency of scale-free networks: error and attack tolerance - ScienceDirect. *Physica*, *320*, 622-642. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0378437102015455>
- Dormond, O. (2019). mTOR in Human Diseases. *International Journal of Molecular Sciences*, *20*(9), 2351. doi:10.3390/ijms20092351
- Erdős, P., & Rényi, A. (1959). On random graphs. I. *Publicationes Mathematicae (Debrecen)*, *6*(3-4), 290-297. doi:10.5486/PMD.1959.6.3-4.12
- Freeman, L. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry*, *40*, 35-41. doi:10.2307/3033543
- Guimerà, R., & Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature*, *433*(7028), 895-900. doi:10.1038/nature03288

- Inoki, K., Corradetti, M. N., & Guan, K. (2005). Dysregulation of the TSC-mTOR pathway in human disease. *Nature Genetics*, *37*(1), 19-24. doi:10.1038/ng1494
- Jeong, H., Mason, S. P., Barabási, A. -, & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41-42. doi:10.1038/35075138
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. -. (2000). The large-scale organization of metabolic networks. *Nature*, *407*(6804), 651-654. doi:10.1038/35036627
- Kim, D., Sarbassov, D. D., Ali, S. M., King, J. E., Latek, R. R., Erdjument-Bromage, H., Sabatini, D. M. (2002). mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell*, *110*(2), 163-175. doi:10.1016/s0092-8674(02)00808-5
- Kuzmanov, U., & Emili, A. (2013). Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Medicine*, *5*(4), 37. doi:10.1186/gm441
- Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701. doi:10.1103/PhysRevLett.87.198701
- Martin, A., Ochagavia, M. E., Rabasa, L. C., Miranda, J., Fernandez-de-Cossio, J., & Bringas, R. (2010). Bisogenet: a new tool for gene network building, visualization and analysis. *BMC Bioinformatics*, *11*, 91. doi:10.1186/1471-2105-11-91
- Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, *296*(5569), 910-913. doi:10.1126/science.1065103
- Newman, M. E. J. (2002). Assortative Mixing in Networks. *Physical Review Letters*, *89*(20), 208701. doi:10.1103/PhysRevLett.89.208701
- Newman, M. E. J. (2018). The Mathematics of Networks. *Networks* (Second edition ed.,). Oxford: Oxford University Press. Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=030377658&sequence=000001&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA
- Niere, F., Namjoshi, S., Song, E., Dilly, G. A., Schoenhard, G., Zemelman, B. V., Raab-Graham, K. F. (2016). Analysis of Proteins That Rapidly Change Upon Mechanistic/Mammalian Target of Rapamycin Complex 1 (mTORC1) Repression Identifies Parkinson Protein 7 (PARK7) as a Novel Protein Aberrantly Expressed in Tuberous Sclerosis Complex (TSC). *Molecular & Cellular Proteomics : MCP*, *15*(2), 426-444. doi:10.1074/mcp.M115.055079
- O'Brien, R. J., & Wong, P. C. (2011). Amyloid Precursor Protein Processing and Alzheimer's Disease. *Annual Review of Neuroscience*, *34*, 185-204. doi:10.1146/annurev-neuro-061010-113613

- Pei, J., & Hugon, J. (2008). mTOR-dependent signalling in Alzheimer's disease. *Journal of Cellular and Molecular Medicine*, 12(6b), 2525-2532. doi:10.1111/j.1582-4934.2008.00509.x
- RAB28 - Ras-related protein Rab-28 - Homo sapiens (Human) | UniProtKB | UniProt. Retrieved from <https://www.uniprot.org/uniprotkb/P51157/entry>
- Rodrigues, F. A., Costa, L. d. F., & Barbieri, A. L. (2011). Resilience of protein–protein interaction networks as determined by their large-scale topological features. *Molecular BioSystems*, 7(4), 1263-1269. doi:10.1039/C0MB00256A
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations - ScienceDirect. *NeuroImage*, 52(3), 1059-1069. Retrieved from <https://www.sciencedirect.com/science/article/pii/S105381190901074X?via%3Dihub>
- Saxton, R. A., & Sabatini, D. M. (2017). mTOR Signaling in Growth, Metabolism, and Disease. *Cell*, 168(6), 960-976. doi:10.1016/j.cell.2017.02.004
- Sporns, O., Honey, C. J., & Kötter, R. (2007). Identification and Classification of Hubs in Brain Networks. *PLoS One*, 2(10), e1049. doi:10.1371/journal.pone.0001049
- Steen, M., Hayasaka, S., Joyce, K., & Laurienti, P. (2011). Assessing the consistency of community structure in complex networks. *Physical Review E*, 84(1), 016111. doi:10.1103/PhysRevE.84.016111
- Swiech, L., Perycz, M., Malik, A., & Jaworski, J. (2008). Role of mTOR in physiology and pathology of the nervous system. *Biochimica et Biophysica Acta*, 1784(1), 116-132. doi:10.1016/j.bbapap.2007.08.015
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Mering, C. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(Database issue), D607-D613. doi:10.1093/nar/gky1131
- Telesford, Q. K., Simpson, S. L., Burdette, J. H., Hayasaka, S., & Laurienti, P. J. (2011). The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain. <https://Home.Liebertpub.Com/Brain>, doi:10.1089/brain.2011.0055
- Tomkins, J. E., & Manzoni, C. (2021). Advances in protein-protein interaction network analysis for Parkinson's disease. *Neurobiology of Disease*, 155, 105395. doi:10.1016/j.nbd.2021.105395
- Uneri, A., Niere, F., Macauley, S. L., Ma, T., Keene, C. D., Craft, S., & Raab-Graham, K. F. (2021). mRNA-Binding Protein DJ-1 as a pivotal protein in AD pathology. *Alzheimer's & Dementia*, 17(S2), e058602. doi:10.1002/alz.058602

van Wijk, B. C. M., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS One*, 5(10), e13701. doi:10.1371/journal.pone.0013701

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440-442. doi:10.1038/30918

Wood, K. C., & Gutkind, J. S. (2022). Challenges and Emerging Opportunities for Targeting mTOR in Cancer. *Cancer Research*, 82(21), 3884-3887. doi:10.1158/0008-5472.CAN-22-0602

Zotenko, E., Mestre, J., O’Leary, D. P., & Przytycka, T. M. (2008). Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. *PLoS Computational Biology*, 4(8), e1000140. doi:10.1371/journal.pcbi.1000140

CURRICULUM VITAE

