Proceedings of the 2015 International Conference on Operations Excellence and Service Engineering
Orlando, Florida, USA, September 10-11, 2015

Characterization of Alzheimer's disease: An Operations Research Approach

Yazeli E. Cruz-Rivera
The Applied Optimization Group/Department of Industrial Engineering
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681
Yazeli.cruz@upr.edu

Yaritza M. Santiago
The Applied Optimization Group/Department of Industrial Engineering
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681
Yaritza.santiago@upr.edu

Valerie M. Gonzalez
The Applied Optimization Group/Department of Industrial Engineering
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681
Valerie.gonzalez9@upr.edu

Clara E. Isaza
The Applied Optimization Group/Department of Pharmacology
Ponce Health Sciences University
Ponce, Puerto Rico 00716
cisaza@psm.edu

Mauricio Cabrera-Rios
The Applied Optimization Group/Department of Industrial Engineering
University of Puerto Rico, Mayagüez Campus
Mayagüez, Puerto Rico 00681
Mauricio.cabrera@gmail.com

Abstract

To search for potential Alzheimer's disease biomarkers using microarray experiments often entails identifying genes with significant changes in their relative expression between control and illness-ridden tissues. These genes are deemed potentially important to characterize the illness. It is also important, however, to establish how these genes might interact with each other to contribute to the evolution of Alzheimer's disease. To this end, this work proposes that finding a path of maximal correlation among all genes of interest would lead to a structure capable to guide biological exploration with enhanced precision. Specifically, it proposes to find the path of maximal correlation among the relative changes of expression of important genes through mathematical optimization. The approach makes use of a mathematical network representation and its associated optimization problem known in the literature as the Travelling Salesman Problem (TSP). The analysis of this optimal path is aimed to determine if biological structure and function follows from this mathematical representation and solution. Furthermore, different correlation schemes for Alzheimer's disease biomarkers using microarray and microRNA experiments using network configurations are assessed to characterize their effects on the resulting paths.

Keywords

Optimization, Traveling salesman problem, Maximum Spanning Tree, correlation, bio-informatics

Introduction

The importance of an Alzheimer 's Disease biomarker stems from its role in early detection, diagnosis, prognosis and recurrence prediction of Alzheimer's Disease. A biomarker is defined as a substance that can be biologically measured and it is related with an increased risk of a disease". A good biomarking candidate is characterized by its distinct behavior in different states (Pérez-Morales Et al, 2014). When taken to a genetic level of relative expression in the presence of the illness. High throughput biological experiments like microarrays have been used to detect potential Alzheimer's Disease biomarkers under the difficulty of having to deal with large amounts of data. A microarray can measure relative expression levels for tens of thousands of genes simultaneously.

This work discusses the issue of finding the potential signaling path among a list of potential Alzheimer's Disease genetic biomarkers. The list of biomarkers is known a priori through the application of multiple criteria optimization, a strategy proposed by our research group. The novelty in this study is to attempt the characterization of the signaling path thought the well-known Travelling Salesman Problem (TSP) combinatorial optimization formulation (Watts Et al,2012).

Methodology

The analysis will be done with the use of a microarray database related to Alzheimer's Disease by finding coordinated behavior among the expression levels of different genes. This behavior can be measured as a statistical correlation. The statistical correlation is computed as linear as a first approach. The final amount of genes to be worked is was found through the application of multiple criteria optimization represented in figure 3. The database is represented in Figure 1 and 2.

Genes	Normal tissue			Alzheimer's Tissue			
	Control A	Control B	Control C	Patient A	Patient B	Patient C	Patient D
Gene 1	7.6233	7.7765	7.8555	7.8353	7.8143	7.9001	7.9381
Gene 2	9.7077	10.0280	9.6452	9.7888	9.9902	9.7609	10.0855
Gene 3	4.9772	4.8139	4.7596	4.7172	5.0581	4.8991	4.8584
Gene 4	9.3293	9.1734	9.3732	9.1128	8.9205	9.0751	9.1584
Gene 5	6.2785	6.1994	6.1902	6.2631	6.1811	6.1167	6.5137
Gene 6	6.8664	6.4720	6.2269	6.2911	6.1366	6.1712	6.4694
Gene 7	7.9830	8.7247	8.4095	8.4209	8.4822	8.6390	8.2859
Gene 8	5.0064	4.8204	4.7976	5.0792	4.9712	4.9199	4.8853
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Gene n	4.5892	4.5359	4.6324	4.6411	4.6598	4.6723	4.6493

Figure 1: Representation of the microarray database GSD2795 (Dunckley Et Al,2006)

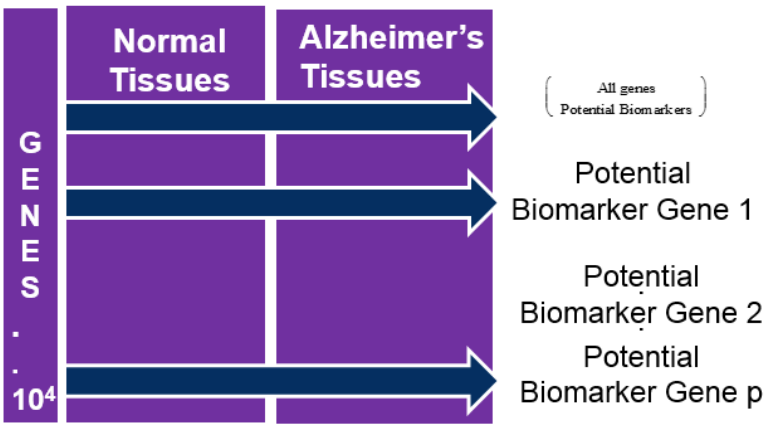

Figure 2: Another representation of the microarray database GSD2795 (Dunckley Et al, 2006)

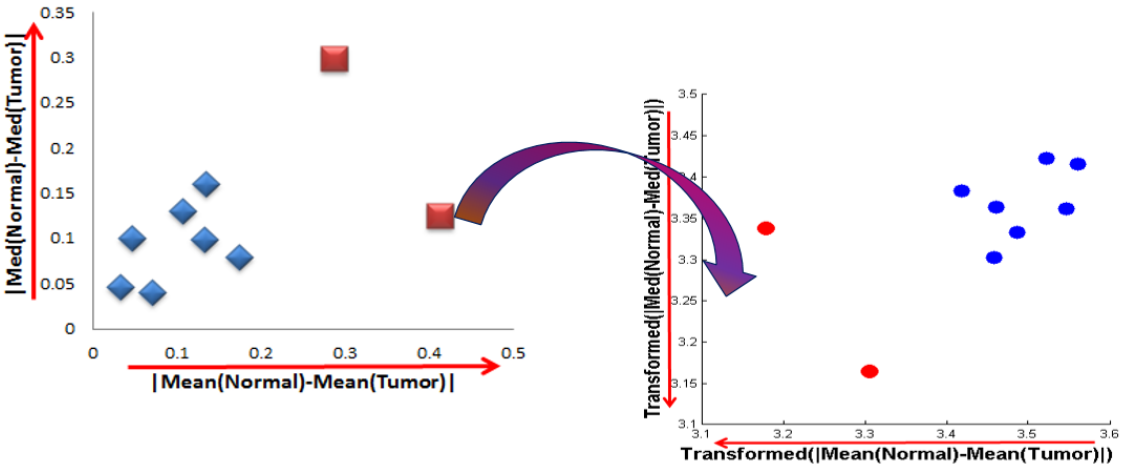

Figure 3: General representation of multiple criteria optimization problem using microarray with two performance measures transformed to a series of minimization cases through a linear transformation

There was a total of 14 genes (Table 1) identified with multiple criteria optimization. The computation of the correlations is necessarily carried out in a pairwise manner among all 14 genes. The linear correlation values found are proxies for suppressed or stimulated behavior in the expression levels of the two genes under analysis. Because these values range from -1 to 1, their absolute values indicate how strong the correlations are. Thus two genes will be strongly correlated if the absolute of their correlation value is close to 1. If each gene is represented through a node in a graph, then the undirected arc joining a pair of genes can hold their absolute correlation value. This leads naturally to the TSP formulation, where the idea is to find the most correlated complete tour. An illustration of how the resulting graph would look like is shown in Figure 4. In our effort, we solve the TSP to optimality capitalizing in the shortlist provided by the first part of the analysis. A Matlab code aided by the branch-and-bound method was used to this end.

Table 1: List of genes

Position	Position in Microarray	Identifier	Name
947	1553551_s_at	ND2	NADH dehydrogenase subuni
976	1553588_at	SH3KBP1	SH3-domain kinase binding protein 1
10037	200095_x_at	RSP10	radial spoke protein 10 tentative
10941	201492_s_at	RPL41	ribosomal protein L41
12988	203540_at	GFAP	glial fibrillary acidic protein
20926	211600_at	PTPRO	protein tyrosine phosphatase, receptor type, O
22094	212788_x_at	FTL	ferritin, light polypeptide
27093	217807_s_at	GLTSCR2	Glioma tumor suppressor candidate region gene 2
33040	224373_s_at	DCAF6	DDB1 and CUL4 associated factor 6
308608	229353_s_at	NUCKS1	casein kinase and cyclin-dependent kinase substrate 1
44327	235077_at	MEG3	maternally expressed 3 (non-protein coding)
47449	238199_x_at	COX3	Cytochrome c OXidase
54622	AFFX-CreX-5_at	AFFX-CreX-5_at	AFFX-CreX-5_at
54668	AFFX-r2-P1-cre-3_at	AFFX-r2-P1-cre-3_at	AFFX-r2-P1-cre-3_at

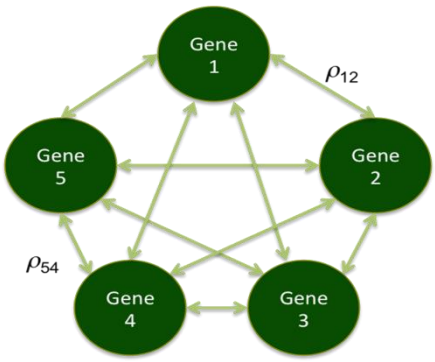

Figure 4: TSP representation

TSP Solution

TSP consists of finding the most correlated tour that visits each potential biomarker gene in a given list exactly. Coordinated behavior can be measured as a statistical correlation. The statistical correlation was computed as linear as a first approach. The computations of the correlations were carried out in a pairwise manner. Their absolute values indicate how strong the correlations are. Two genes will be strongly correlated if the absolute of their correlation value is close to 1. Through this method the following route between genes is represented in figure 5:

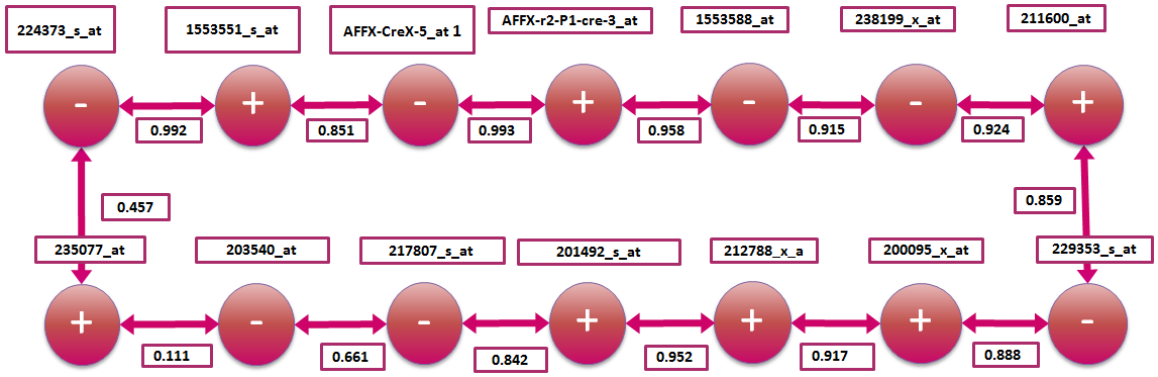

Figure 5: Gene coordinated behavior path determined by correlation

Correlation vs 1- P - Value

Another statistical measure used to find a route is the P-value. P-value is the probability of finding a difference at least as big as the observed by pure chance. Maximization of the sum of 1- P-values used to match the objective of the Traveling Salesman Problem. The solution is the path with the strongest signal. The route between genes found with this measurement is represented in figure 6:

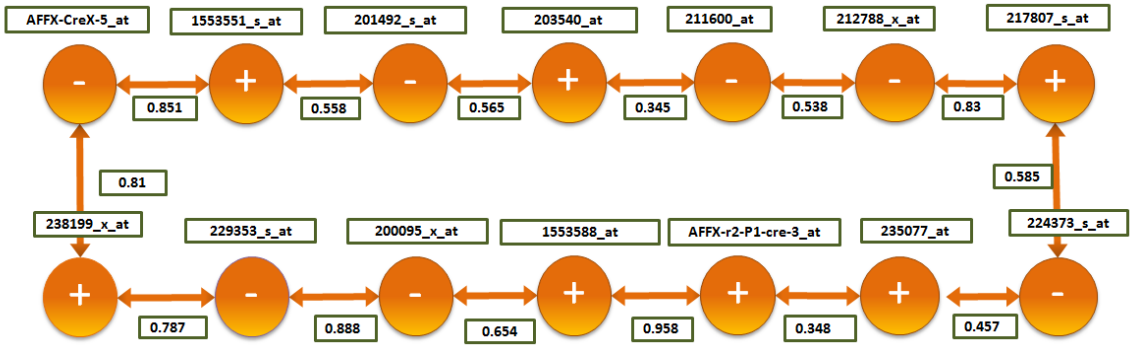

Figure 6: Gene coordinated behavior path determined by 1 - P-value

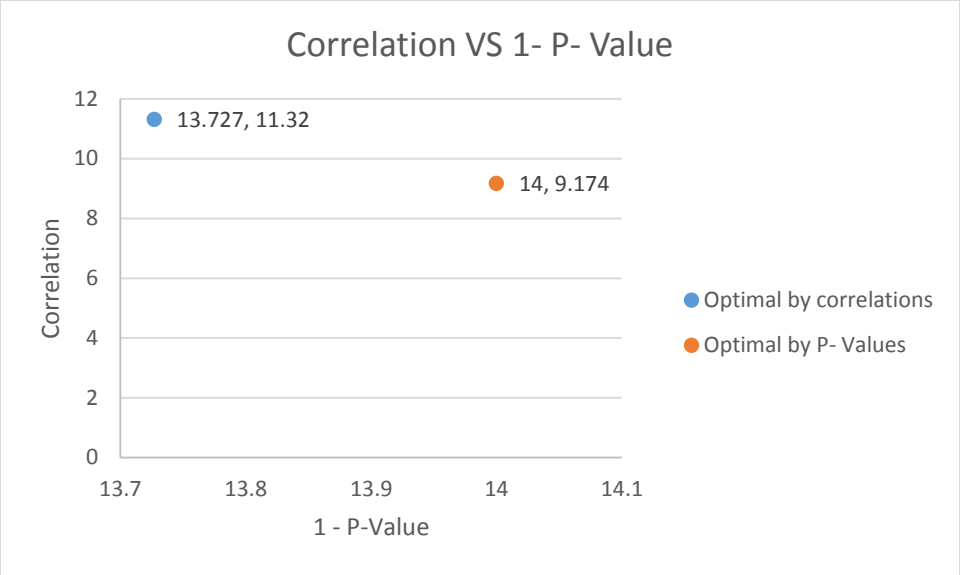

Figure 7: Correlation vs 1- P-value comparison

Figure 7, on the other hand, shows how the optimal solutions of both approaches {correlation values, p-values} compare to each other. They do not coincide, as there is a tradeoff between them. This study points to the need to further explore the different pieces of information that can be used to measure genetic co-expression. Ultimately, when biological exploration of these solutions is carried out in the future, a better assessment will be available to discern among the competing representation strategies. It is, however, important to point out that comparing optimal configurations provide a very efficient means to compare these strategies.

Biological Interpretation

Here we discuss the biological evidence of the 14 genes identified with multiple criteria optimization. According to scientific literature, ND2 has a relationship with Alzheimer due a mutation that causes a disorder in the mitochondria (Lin Et al, 1992). This is because between the entire hypothesis that can trigger Alzheimer stands the dysfunction of the mitochondria, because its charge of the central function of the cell (Santos R., Et al, 2012). The brain is the organ that is the most dependent of mitochondrial energy, representing 2% of the body's weight but consuming 20% of the oxygen (Coskun P. Et al, 2011). On the other hand SH3KBP1 has been identified to potentially play a role in Late-Onset. This is because the accumulation of β-amyloid (Aß) senile plaques in the brain as another factor that is capable of initiating Alzheimer's (Rosenthal S. Et al, 2012). In this case the paper recognizes that further study is necessary in the subject. Another important gene is the GFAP; this gene is particularly interesting because it is mostly expressed in the brain. The mutation in this gene is responsible for Alexander disease (a rare disorder of the central nervous system), to leukodystrophy and Alzheimer's disease. The disease causes the destruction of myelin (Chen Y., Et al, 2011). Another gene that stood out in our results was FTL, this gene is associated with neurodegenerative disorder associated with iron accumulation in the brain, primarily in the basal ganglia (Maciel P, Et al, 2005). The expression the NUCKS1 gene has a link with mood disorders, though little is known (Savitz J. Et al. 2013). Also this gene has a high expression in the brain like COX3.

However other of the selected genes have not a direct link - RSP10, PTPRO, DCAF6 - to Alzheimer's Disease but for some of them a link can be found through their roles in processes that can have an effect on Alzheimer's Disease. GLTSCR2 is a tumor suppressor (Kim YJ Et al, 2008), MEG3 a potential long non coding RNA tumor-suppressing gene (Balik V et al 2013) that has been associated with Huntington's disease (Johnson R. Et al, 2011) RPL41, it is also associated with ATF4 degradation (Wang A Et al, 2011) that in turn mediates neurodegeneration in Alzheimer's Disease and transmission of a neurodegenerative signal through some brain regions (Fayaz SM Et al, 2014) (Baleriola J Et al, 2013). By using a MCO point of view important differential express genes become apparent while a single criteria approach could fail to uncover all of them.

Conclusion

There is proof-of-concept on the mathematical part of this idea, biological validation will ensue. If biological structure and function follows, then a powerful discovery methodology is at hand. The need for models created with high-throughput biological experiments is clear in the fight against Alzheimer's disease. Its characterization has been very elusive. It is believed that optimization models, such as the TSP can support this effort in a transparent and effective way. Future work will consist on gathering more biological evidence to evaluate the resulting configurations elucidated in this work.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under Grant HRD 0833112 (CREST program), as well as the National Institutes of Health (NIH) MARC Grant 5T36GM095335-02 'Bioinformatics Programs at Minority Institutions'.

References

http://www.ncbi.nlm.nih.gov/pubmed/20127998 [June 25, 2012]

J Pérez-Morales, M Pérez-Santiago, E Jaramillo, CE Isaza, M Cabrera-Ríos, Data Envelopment Meta-analysis (DEM-A): an approach to identify potential cancer biomarkers from multiple incommensurable microarray experiments, Technical Report AOG_2014_June_1 DOI: 10.13140/2.1.3160.8009

E. Watts, M.L. Sanchez-Peña, C.E. Isaza, M. Cabrera-Ríos, Potential colon cancer biomarker search using more than two performance measures in a Multiple Criteria Optimization approach, Puerto Rico Health Sciences Journal 31:2 (2012) 59-63

Dunckley T, Beach TG, Ramsey KE, Grover A et al. Gene expression correlates of neurofibrillary tangles in Alzheimer's disease. *Neurobiol Aging* 2006 Oct;27(10):1359-71. PMID: 16242812

Lin, F.-H., Lin, R., Wisniewski, H. M., Hwang, Y.-W., Grundke-Iqbal, I., Healy-Louie, G., Iqbal, K. Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenase subunit 2 in Alzheimer's brains. Biochem. Biophys. Res. Commun. 182: 238-246, 1992.

Santos R, Correia S, Wang X , Perry G, Smith M, Moreira P and Zhu X. Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. Am J Neurodegener Dis. 2012; 1(2): 191–198.

Coskun P, Wyrembak J, Schriner S, Chen H, Marciniack C, LaFerla F ,Wallace D. A Mitochondrial Etiology of Alzheimer and Parkinson Disease. Biochim Biophys Acta. 2012 May; 1820(5): 553–564.

Rosenthal S, Wang X, Demirci F, Barmada M, Ganguli M, Lopez O, and Kamboh M. Beta-amyloid toxicity modifier genes and the risk of Alzheimer's disease. Exp Cell Res. 2011 Oct 1; 317(16): 2252–2266.

Chen Y, Lim S, Chen M, Quinlan R, Perng M. Alexander disease causing mutations in the C-terminal domain of GFAP are deleterious both to assembly and network formation with the potential to both activate caspase 3 and decrease cell viability. /Neurology. 2005 Aug 23;65(4):603-5.

Maciel P, Cruz VT, Constante M, Iniesta I, Costa MC, Gallati S, Sousa N, Sequeiros J, Coutinho P, Santos MM. Neuroferritinopathy: missense mutation in FTL causing early-onset bilateral pallidal involvement. Neurology. 2005 Aug 23;65(4):603-5.

Savitz J, Frank MB, Victor T, Bebak M, Marino JH, Bellgowan PS, McKinney BA, Bodurka J, Kent Teague T, Drevets WC. Inflammation and neurological disease-related genes are differentially expressed in depressed patients with mood disorders and correlate with morphometric and functional imaging abnormalities. Brain Behav Immun. 2013 Jul;31:161-71. doi: 10.1016/j.bbi.2012.10.007. Epub 2012 Oct 12.

Kim YJ, Cho YE, Kim YW, Kim JY, Lee S, Park JH. J Pathol. Suppression of putative tumor suppressor gene GLTSCR2 expression in human glioblastomas. 2008 Oct;216 (2):218-24)

Balik V, Srovnal J, Sulla I, Kalita O, Foltanova T, Vaverka M, Hrabalek L, Hajduch M. J Neurooncol. MEG3: a novel long noncoding potentially tumour-suppressing RNA in meningiomas. 2013 Mar; 112(1):1-8.)

Johnson R. Long non-coding RNAs in Huntington's disease neurodegeneration. Neurobiol Dis. 2012 May;46(2):245-54. doi: 10.1016/j.nbd.2011.12.006. Epub 2011 Dec 14.

Wang A, Xu S, Zhang X, He J, Yan D, Yang Z, Xiao S., J Pathol. Ribosomal protein RPL41 induces rapid degradation of ATF4, a transcription factor critical for tumour cell survival in stress. 2011 Oct; 225(2):285-92.

Fayaz SM, Rajanikant GK. ATF4: the perpetrator in axonal-mediated neurodegeneration in Alzheimer's disease. CNS Neurol Disord Drug Targets 2014;13(9):1483-4.

Baleriola J, Walker CA, Jean YY, Crary JF, Troy CM, Nagy PL, Hengst U. Axonally synthesized ATF4 transmits a neurodegenerative signal across brain regions Cell. 2014 Aug 28;158(5):1159-72.).

Biography

Yazeli E. Cruz Rivera is currently a senior undergraduate student in the Department of Industrial Engineering at University of Puerto Rico at Mayagüez (UPRM). She has been an Undergraduate Research Assistant at the Applied Optimization Group since 2012 working in the intersection of Operations Research and the analysis of high throughput biological experiments related to cervix cancer and Alzheimer`s disease.

Yaritza M. Santiago obtained her BS degree in Industrial Engineering from UPRM in 2012. She currently is a graduate student in the MS program in the same academic unit. Her thesis on the works relates genes and microRNAs attempts an optimization-driven characterization of Alzheimer´s disease. She expects to defend her MS thesis on December 2015.

Valerie M. González is a senior undergraduate student in Microbiology at UPRM. She has been an Undergraduate Research Assistant at the Applied Optimization Group since 2013 providing biological interpretation to results from microarray and microRNA analysis in cervix and lung cancer, as well as Alzheimer´s disease.

Clara E. Isaza is currently a postdoctoral researcher in the Department of Pharmacology at Ponce Health Sciences University. She obtained her doctoral degree in Biophysics from The Ohio State University in 2005. She also co-leads The Applied Optimization Group at UPRM. Her work relates to cellular signaling pathways.

Mauricio Cabrera-Ríos obtained his doctoral degree in Industrial and Systems Engineering from The Ohio State University in 2002. He currently holds an appointment as Associate Professor in the Department of Industrial Engineering at University of Puerto Rico at Mayagüez. He also co-leads the Applied Optimization Group. https://www.facebook.com/AppliedOptUPRM