Applications of Linear Algebra applied to Big Data Analytics

Dr. Rajendran Swamidurai, Alabama State University

Dr. Rajendran Swamidurai is an Associate Professor of Computer Science at Alabama State University. He received his BE in 1992 and ME in 1998 from the University of Madras, and PhD in Computer Science and Software Engineering from Auburn University in 2009. He is an IEEE senior Member.

Dr. Cadavious M Jones,

Dr. Cadavious M. Jones is an Associate Professor of Mathematics at Alabama State University. He received his BS in 2006 and MS in 2008 from Alabama State University, and PhD in Mathematics from Auburn University in 2014. He is a contributor to the Australian Maths Trust, and member of the MASAMU international research group for mathematics.

Dr. Carl Pettis

Carl S. Pettis, Ph.D. Professor of Mathematics Department of Mathematics and Computer Science Alabama State University

Administrative role:

Interim Associate Provost Office of Academic Affairs Alabama State University

Dr. Uma Kannan

Dr. Uma Kannan is Assistant Professor of Computer Information Systems in the College of Business Administration at Alabama State University, where she has taught since 2017. She received her Ph.D. degree in Cybersecurity from Auburn University in 2017. She specialized in Cybersecurity, particularly on the prediction and modelling of insidious cyber-attack patterns on host network layers. She also actively involved in core computing courses teaching and project development since 1992 in universities and companies.

<u>Applications of Linear Algebra applied to Big Data Analytics</u>

1. Introduction

The digital universe (the data we create and copy annually) is doubling every two years and will reach 44 zettabytes (44 trillion gigabytes) in 2020 [1]. The stored digital data volume has grown exponentially over the past few years [2, 3]. In 1986, only three exabytes of data existed and in 2011 it went up to 300 exabytes [3], and at the end of 2020 it might reach 44,000 exabytes [1]. Moreover, the International Data Corporation (IDC) forecasts that in 2025 we will be producing 165 zettabytes of data per year [4]. There is an increasing demand for doing everything online whether it is in our work or private life and this increase is responsible for this explosion of data being created. In conjunction, we also employ smart devices that are continuously connected to the internet and produce constant streams of real-time data about things that range from our heartrate to our current location. It is estimated that today (in 2019) more than four billion people are online and in 2020 every person will produce 1.7 megabytes of data every second [4], and millions of enterprises are becoming more and more web based. In addition, there are millions of sensors and communicating devices transmitting data over the internet adding to the digital universe [1]. It is expected that in 2020, over 10 billion mobile devices will be in use [5] and this will make the digital universe ever larger. This exponential data growth is observed almost in all sectors, including government, healthcare, banking, manufacturing, retail, transportation, and education [6]. For example, in 2019, there were 2.3 billion active Facebook users [4] and more than 30 petabytes of data handled by Facebook [5]. Twitter users send over 230 million tweets every day [5] at an average half-a-million tweets per minute [4]. There are over 1 billion Google searches every day [5], Walmart records around 267 million transactions (4 petabytes) per day [7], Netflix streams over 1 billion hours of TV shows and movies per month [1], and YouTube users upload 5 hours of video per second [8]. In addition to this, the modern telescopes are data driven and they produce enormous amounts of data, for example, the Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope streams data at a rate of 2.8 Gigabytes per second [1], and the proposed Large Synoptic Survey Telescope (LSST) will record 15 Terabytes of data every night [9]. Similarly, the Large Hadron Collider (LHC), a particle accelerator [10], helps us to understand the workings of the Universe and will generate 60 terabytes of data per day [7].

The sudden increase in the digital universe (Big Data) opened doors for new types of data analytics called big data analytics and new job opportunities [11]. In 2012, only 23% of organizations had an enterprise-wide Big Data strategy [5, 12], whereas today 97.2% of organizations are investing in Big Data [4]. A recent Harvard Business Review [13] survey of senior Fortune 500 and federal agency business and technology leaders report that 70% of the respondents plan to hire data scientists. The U.S. Bureau of Labor Statistics (BLS), Occupational Outlook Handbook 2018 [11] projects that there will be a 34% increase in data analytics jobs from 2016 to 2026. A McKinsey Global Institute research report [14] indicates that the demand for big data analytical talent could be very high and will produce 50 to 60 percent more data analytic jobs. Similarly, a Forbes report indicates that there will be 2.7 million data science and analytic job openings by 2020 [4].

A recent survey from Harvard Business Review [13] indicates that 85% of the organizations that they surveyed revealed that they planned to fill 91% of their data science jobs with new graduates. Though the private sector asks at least a master's degree in mathematics or statistics for data analytics jobs, the government sector requires only a bachelor's degree [15]. Moreover, it is impractical to fill this huge demand for big data analytics through only from graduate degree holders in mathematics-related fields. The Harvard Business Review [13] report also indicates that 70% of the organizations that they surveyed report that finding big data talent as challenging or impossible. The hiring scale for big data jobs is 73%; this high score indicates the amount of difficulty in finding skillful candidates for the job [16]. In order to address this serious problem, Alabama State University with the support of Auburn University employed a unique technique called infusing big data analytics in various undergraduate mathematics and statistics courses. Our big data course modules walked students through producing working solutions by having them perform a series of hands-on big data exercises developed specifically to apply cutting-edge industry techniques with each mathematics and statistics course module. We strongly believe that equipping students with such skills greatly improves their employability. Linear algebra concepts such as feature extraction, clustering, and classification involving the manipulation of large matrices are extensively used in big data analytics; therefore, this is a natural course to start introducing students to big data analytics. This paper presents our four years' experience in adapting and integrating big data concepts into undergraduate linear algebra courses.

2. Linear Algebra and Big Data

Linear algebra topics such as linear equations, eigenvalue problems, principal component analysis, singular value decomposition, quadratic forms, linear inequalities, linear programming, optimization, linear differential equations, modeling and prediction, and data mining algorithms (frequent pattern analysis, classification, clustering, and outlier detection) are frequently used in practice by Big Data Analytical applications. [17] Particularly, matrix algorithms constitute the core of modern Big Data Analysis. Because, matrices provide a convenient mathematical structure for modeling a wide range of applications' data. For example, information about 'N' objects with 'D' features can be easily described/encoded by an 'NxD' matrix. [18] Manipulations of large matrices are used in feature extraction, clustering, and classification. Matrix decomposition is used in principal component analysis for dimension reduction. Similarly, application of eigenvectors used in Google's PageRank method. [19]

The versatility of graphs can be seen from their ability to illustrate important aspects of modern computer science, the intricacies geography, linguistic complexities, and the consistency of chemical structures. Incorporating linear algebra allows for representation of these graphs as matrices, this completes the pertinent task of enhancing their computational aspects. [20] Linked data is usually represented by a graph in Big Data applications [19]. We define a graph G as an ordered pair (V (G),E(G)) consisting of a set V (G) of vertices and a set E(G), disjoint from V (G), of edges, together with an incidence function ψ_G that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G. Using this definition the vertices of a graph can represent webpages, genes, image pixel, or interacting users, and edges represent relations or links between the vertices [21]. Notions such as centrality, shortest path, and reachability can be derived from the graph using graph analytics. A widely used practical application of large graph analytics is the internet search engine. [19] Some widely used methods in Big Data Analytics that incorporate the utilization of graphs is to visualize big data as graphs (e.g. the World Wide Web), computation for strongly connected large graphs (e.g. PageRank for strongly connected graphs), and finding matchings in bipartite graphs (e.g. internet advertising) are [19]. Many practical applications of large graph analytics exist, including internet search engines. One obstacle that presents itself with graphs is the humongous sizes that are involved with the numerous millions of vertices that could exist. Low rank approximation of the adjacent matrices or graph Laplacians in relation to the graphs compared in the analysis of and interpretation of the data. [21]

Practical Big Data applications that use linear algebra include, but are not limited to: 1) Google's Page Rank Algorithm, 2) Recommender Systems (e.g., Netflix, Pandora, Spotify), 3) Topic Modeling (e.g., Wikipedia, Genome Sequence Analysis), 4) Social Network Analysis (e.g., Facebook, MySpace, LiveJournal, YouTube), 5) Internet Search, 6) Complex System Analysis (e.g., Biological Networks), 7) Image Segmentation, 8) Graph Clustering, 9) Link Prediction, and 10) Cellular Networks. In Big Data Analytics, linear equations along with matrices are widely used in large network analysis, Leontief economic models, a model for the economics of a whole country/region in which consumption equals production [22], and ranking of sports teams [23]. [17] Eigenvalues and eigenvectors are used in Google's PageRank algorithm, networks clustering, and weather system modeling [17] and spectral decomposition, a matrix approximation technique which uses eigenvectors, are used in spectral clustering [24], link prediction in social networks [25], recommender systems with side-information [26], densest k-subgraph problem [27], and graph matchings [28]. [29] Principal component analysis and Singular Value Decomposition techniques are used to compare the structure of folded proteins and in Dimension Reduction techniques such as image compression, face recognition, and El Nino [17]. Optimization, a minimization of a quadratic expression, and *linear programming* are widely used in the stable marriage problem, production planning, portfolio selection [23], transportation problem, minimization of production costs, minimization of environmental damage, and maximization of profits. [17] Practical applications like face recognition, fingerprint recognition, plagiarism finding, and Netflix movie ratings are using similar items and frequent patterns concepts [17, 30].

3. Infusing Big Data Analytics in UG Linear Algebra Course

To facilitate the Big Data infusion and active learning in the linear algebra course, we employed a two-part module. The first part focused on theoretical and conceptual ideas behind the methods under discussion and the second part had hands-on experimentation using real-world data. The students are advised to use both R and Python general-purpose programming languages to complete their projects. The students can also use MATLAB programming to perform their project as well as MS Excel.

The initial set of topics in which we integrated big data analysis methods were chosen using two criteria: suitability of material for pedagogical integration of big data methods and impact on all computing and Mathematics majors. Instructors may eventually choose to expand the integration of Big Data concepts to other computing and Mathematics courses in the future. The following big data lectures and lab modules were infused to the existing linear algebra course:

Lecture: To begin, the students were provided with a pretest to gauge their understanding of Big Data Analytics and how linear algebra can be applied. This data was later paired with a posttest that was given as the last component of the module. The instructor presented the class with a concept of "Big Data" that best suits the linear algebraic viewpoint. In linear algebraic terms we define big data as data that can be represented as an $m \times n$ array with large m and large n. The goal of the lecture was to reinforce topics already outlined in the course syllabus while only presenting additional information, if it was absolutely necessary for students to understand aspects of the modules. Some of the topics already incorporated into the course curriculum include linear equations and matrices, eigenvalues, eigenvectors, and singular value decomposition. The lecture focused on methods for gathering data and representing such data in the form of matrices and the utilization of basic applications of linear algebra on said matrices. The primary source for such data was www.data.gov and similar sites. In addition, students were presented with the PageRank algorithm and a scenario utilizing it. Lastly, the lecture introduced the topic of the Leslie Matrix and population change. Examples were kept as simple as possible for students to understand the complexities of certain algorithms or unfamiliar methods.

Hands-on activities: Students were asked to:
1) Classify data sets into categories that describe the shape of the data distribution. For this lab activity students were encouraged to use the practical big data techniques explained during the lecture when considering linear equations relating to business problems, tax problems, economic planning models, problems for the input-output matrix for an economy producing transportation, and interpret data analytic problems. The data sets for this portion of the activity were prearranged in order to allow for certain controlled outcomes and provide key discussion points.
2) Investigate real data using eigenvalues, and eigenvectors to decipher information from the data set. For this lab, students were encouraged to solve practical problems such as economic development problems, analysis of situations as diverse as land problems, applications in structural engineering, control theory problem, vibration analysis problem, electric circuits problem, and advanced dynamic problem and so on. Of the previous topics stated, instructors were given the freedom to select from this group areas to focus on as part of the second portion of the activity. However, data was once again provided from sources such as www.data.gov.

MATLAB was used as the primary computing tool when calculations would exceed those gained from simple introductory examples. Given that all students were not previously familiar with MATLAB, step-by-step "cheat sheets" were used when working through examples and as a reference for using MATLAB.

Students were then provided with an out-of-class assignment to further their research, as well as reinforce their understanding of how linear algebra could be applied to Big Data Analytics.

Assignment: The assignment focused on a complexity analysis of the PageRank Algorithm and was titled "The Mathematics of Google Search." Given that this was an undergraduate course models were constructed from real world data, but altered as not to produce an unnecessary number of iterations that would distract from the purpose of the assignment.

Upon their return to class students were asked to engage in a classroom discussion based on questions contained in their assignment called "Questions for Class Discussion." These questions were meant to understand, from the students' point of view, the value of the module and their feelings toward applying class material in such a manner.

In order to understand as completely as possible the student's competency, certain standards were considered. There were several standards addressed to some degree by this project. The standards are: Students will be able to: collect data, display data in a graphical manner, interpret data as a matrix, apply techniques already contained within the curriculum; develop models; determine levels of accuracy needed; organize materials; interpret the data and draw a conclusion from the data; explain their thought process.

The criteria which identified indicators of good performance on the task and in class discussions were:

- accuracy of calculations
- accuracy of models and graphs
- usage of algorithms
- organization of calculations
- clear explanations

As mentioned, the culmination of the module came in the form of a posttest designed to, among other things, show if the students had a better understanding of the topic than when they began.

4. Results

During spring 2016, fall 2016, and spring 2017 semesters, Alabama State University faculty developed Big Data modules to infuse into the existing Introduction to Linear Algebra and Discrete Mathematics courses. After the beta test between spring 2016 to spring 2017, these Big Data modules went through various updates – some based on student feedback and some due to the change in computer hardware and software from fall 2017 to fall 2019. These modules were evaluated for their effectiveness through pre- and post-tests. In addition, students in all offered classes were asked to complete a survey pertaining to their coursework, confidence in using big data modules in their classes, and strategies they use to learn in their math classes.

4.1. Student Knowledge

Students in each class completed pre- and post-tests to examine changes over the duration of the module implementation. In each class, there were students that failed to complete the pre, post, or both tests. Overall, scores on the pre-tests averaged just 36.63% while averaging 80.69% on the post-tests. The box plot and paired t-test results are shown in figure 1 and figure 2 respectively. The two-tailed P value for the 95% confidence interval less than 0.0001, by conventional criteria, this difference is extremely statistically significant.

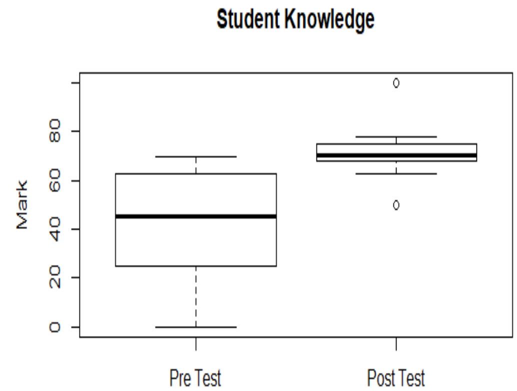

Figure-1: Box Plot (Student Knowledge)

P value and statistical significance:
The two-tailed P value is less than 0.0001
By conventional criteria, this difference is considered to be extremely statistically significant.

Confidence interval:
The mean of PreTest minus PostTest equals -44.0588
95% confidence interval of this difference: From -52.2025 to -35.9152

Intermediate values used in calculations:
t = 10.8667
df = 50
standard error of difference = 4.054

Review of the data:

Group	**PreTest**	**PostTest**
Mean	36.6275	80.6863
SD	23.3169	15.8600
SEM	3.2650	2.2208
N	51	51

Figure-2: Paired *t* test results

4.2. Matched Pre-Post Student Knowledge

To better examine gains made by students after using these modules, the analysis was limited to those students with complete pre- and post-test data. A total of 44 students had completed both the pre- and post-test. Scores for this matched sample increased from pre-test (M=35.14, SD=23.5) to post-test (M=83.61, SD=14.75). Using a paired-samples t-test, changes from pre-test to post-test were statistically significant (t=14.09, p<0.0001). These results are summarized in figure-3 (boxplot) and figure-4 (paired t-test).

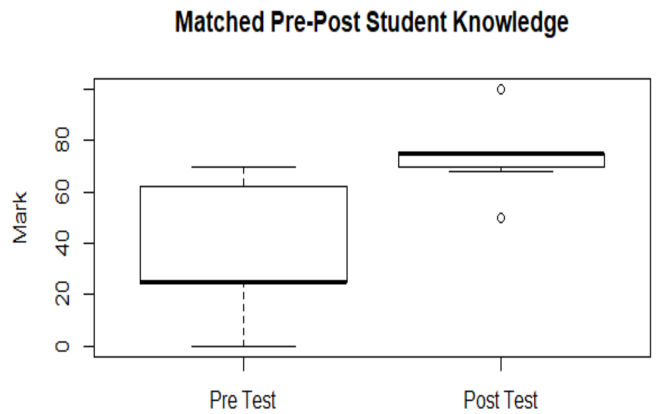

Figure-3: Box Plot (Matched Student Knowledge)

P value and statistical significance:
The two-tailed P value is less than 0.0001
By conventional criteria, this difference is considered to be extremely statistically significant.

Confidence interval:
The mean of PreTest minus PostTest equals -48.4773
95% confidence interval of this difference: From -55.4153 to -41.5393

Intermediate values used in calculations:
t = 14.0910
df = 43
standard error of difference = 3.440

Review of the data:

Group	**PreTest**	**PostTest**
Mean	35.1364	83.6136
SD	23.4963	14.7463
SEM	3.5422	2.2231
N	44	44

Figure-4: Paired *t* test results

4.3. Confidence in using Big Data Modules in Class

In spring 2016, nearly 80% of the overall survey respondents were either juniors or seniors and nearly 30% were enrolled as computer science majors. The sample was balanced in terms of gender (52.9% female), but offered little diversity in terms of race, ethnicity or disability. In Fall 2016, nearly 95% of the overall survey respondents were either juniors or seniors and over 38% were enrolled as computer science majors. The sample offered little diversity in terms of race, ethnicity or disability and over 32% were female. In spring 2017, nearly 95% of the overall survey respondents were either juniors or seniors and nearly 28% were enrolled as computer science majors. The sample had a larger number of males (53.1%), with majority of participants identifying as Black (87.5%) and primarily not identifying with Hispanic or Latino ethnicity

(90.6%). Using a 5-point scale (1=little of no confidence…5=A great deal of confidence), students were asked to respond to 31 different potential big data modules/applications. These responses were requested prior to the implementation of modules in math coursework. In spring 2016, only 8 out of 26 modules (30.8%) received an average response of 3 or above, in fall 2016, only 2 out of 26 modules (6.5%) received an average response of 3 or above, and in fall 2017, 30 out of 31 modules (96.8%) received an average response of 3 or above.

4.4. Student Academic Efficacy, Motivation and Learning Strategies in Math Courses
Finally, students were asked to respond to survey items pertaining to their level of academic efficacy, motivation and goals in learning math, and strategies that they use and prefer to learn math.

Academic Efficacy: Students were asked to respond to five items related to their academic efficacy as it pertains to the math class in which they were enrolled. Overall, students reported a great deal of confidence in their academic abilities with the average for each term above 4 (on a 5-point scale). Students believed that they would learn if they tried, worked hard, and did not give up. They also believed that they could master the skills and figure out the most difficult class work.

Goals in Math: While all goals were important to them, students believed that getting a good grade was most important. They also wanted to meet requirements for their degree, improve their ability to communicate math ideas to others, learn new ways of thinking and specific procedures for solving math problems.

Preferred Learning Environments: When asked to indicate their perceptions of statements describing different learning environments, students reported the greatest agreement with “the instructor explains the solutions to problems” and “the assignments are similar to the examples considered in class.” Students also indicated situations in which they compared their math knowledge to other students, studied their notes, explained ideas to others, worked in small groups, and got frequent feedback on their mathematical thinking. They were less supportive of having the class critique their solutions, exams that prove their skills and group presentations.

General Learning Strategies used by Students: In general, students reported using a variety of strategies in their math classes and not giving up when they get stuck. They most frequently reported finding their own ways of thinking and understanding and reviewing their work for mistakes or misconceptions. They also reported checking their understanding of what a problem is asking, studying on their own and using their intuition about what an answer should be.

Motivation to learn Math - Task Value: Students reported high levels of task value, indicating their belief in the importance and utility of course content in their math classes. Their understanding of math is extremely important to them and their motivation to learn math is strong.

Learning Strategy – Critical Thinking: In terms of learning math, students reported many strategies that require critical thinking. They reported developing their own ideas based on course content and evaluating the evidence before accepting a theory or conclusion. They also reported questioning what they read or heard in class and thinking of possible alternatives.

Learning Strategy – Self- Regulation: Students reported using many effective self-regulation strategies in their math classes. In particular, they pay careful attention to concepts that they find confusing and focus on studying and reviewing these, so they learn them.

Learning Strategy – Time and Study Environment Management: Another positive strategy reported by students related to the management of their time and study environment. They reported attending class regularly, finding a place to study and keeping up with the weekly readings and assignments.

The reliability of these scales was generally supportive, with internal consistency estimates ranging from 0.491 to 0.926, with a median of 0.867. Perceptions were also very positive as overall scale means exceeded the scale midpoints.

5. Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No.1436871.

6. Conclusions

We have created many one-week linear algebra big data modules and infused them into existing core undergraduate mathematics courses over a period of four years. The modules were taught using examples that were worked through interactively during class. The students then worked on assignments that incorporated the new big data instructional concepts. We have evaluated the big data modules effectiveness through pre- and post-tests, and surveys. The paired-samples t-test results show that matched pre-post student knowledge is statistically significant. Regarding confidence in using big data modules in class, we had mixed results. Students' perception was very positive as overall scale means exceeded the scale midpoints. We feel the courses were a success but indicated there was room for improvement.

References

1. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, EMC Digital Universe with Research & Analysis by IDC, April 2014, https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm
2. Sara Royster, "Working with big data," Occupational Outlook Quarterly, 57, 3, 2-10, 2013
3. Nicolaus Henke, Jacques Bughin, Michael Chui, James Manyika, Tamim Saleh, Bill Wiseman and Guru Sethupathy, THE AGE OF ANALYTICS: COMPETING IN A DATA-DRIVEN WORLD, December 2016, McKinsey & Company.
4. Christo Petrov, "Big Data Statistics 2020," Tech Jury, March 2019, https://techjury.net/stats-about/big-data-statistics/#gref
5. Quick Facts and Stats on Big Data, IBM Big Data & Analytics Hub, Last Accessed 1/21/2020 6:58PM, https://www.ibmbigdatahub.com/gallery/quick-facts-and-stats-big-data

6. Ralph Jacobson, "2.5 quintillion bytes of data created every day. How does CPG & Retail manage it?", IBM, April 24, 2013, https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/
7. Bryant R. E., Katz R. H., & Lazowska E. D. (2008). Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. http://cra.org/ccc/resources/ccc-led-whitepapers/
8. Bernard Marr, "Big Data: 20 Mind-Boggling Facts Everyone Must Read", Forbes, September 30, 2015, https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#322cdf0017b1
9. Legacy Survey of Space and Time: Opening a Window of Discovery on the Dynamic Universe, https://www.lsst.org/
10. The Large Hadron Collider, https://home.cern/science/accelerators/large-hadron-collider
11. Bureau of Labor Statistics, U.S. Department of Labor, Occupational Outlook Handbook, Mathematicians and Statisticians, on the Internet at https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm (visited January 30, 2018)
12. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, December 2012, https://www.emc.com/leadership/digital-universe/2012iview/index.htm
13. Paul Barth and Randy Bean, "There's No Panacea for the Big Data Talent Gap", Harvard Business Review, November 29, 2012, https://hbr.org/2012/11/the-big-data-talent-gap-no-pan
14. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, May 2011
15. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
16. Louis Columbus, "Where Big Data Jobs Will Be In 2016", Forbes, NOV 16, 2015, https://www.forbes.com/sites/louiscolumbus/2015/11/16/where-big-data-jobs-will-be-in-2016
17. Eli Tziperman, Applied Mathematics 120: Applied linear algebra and big data, https://canvas.harvard.edu/courses/4766 (Spring 2016), Last updated: May 19, 2016.
18. Jiyan Yang, Randomized Linear Algebra for Large-Scale Data Applications, August 2016, Stanford University, http://purl.stanford.edu/wr092fb7484
19. Carl Pettis, Rajendran Swamidurai, Ash Abebe, and David Shannon, "Infusion of Big Data Concepts Across the Undergraduate Computer Science Mathematics and Statistics Curriculum," 2018 ASEE Annual Conference & Exposition, June 24-27, 2018, Salt Lake City, UT, USA
20. Dylan Johnson, GRAPH THEORY AND LINEAR ALGEBRA, May 2017, http://www.math.utah.edu/~gustafso/s2017/2270/projects-2017/dylanJohnson/Dylan%20Johnson%20Graph%20Theory%20and%20Linear%20Algebra.pdf
21. Berkant Savas and Inderjit S. Dhillon, "Clustered low rank approximation of graphs in information science applications," Proceedings of the 2011 SIAM International Conference on Data Mining. 2011, 164-175, https://doi.org/10.1137/1.9781611972818.15
22. Gerald Höhn, Application to economics: Leontief Model, https://www.math.ksu.edu/~gerald/leontief.pdf
23. Gilbert Strang, "Linear Algebra and Its Applications," Cengage Learning; 4th edition, 2006
24. A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: analysis and an algorithm. In NIPS, pages 849–856, 2001
25. D. Shin, S. Si, and I. S. Dhillon. Multi-scale link prediction. In CIKM, pages 215–224, 2012
26. N. Natarajan and I. S. Dhillon. Inductive matrix completion for predicting gene-disease associations. Bioinformatics, 30(12):i60–i68, 2014
27. D. Papailiopoulos, I. Mitliagkas, A. Dimakis, and C. Caramanis. Finding dense subgraphs via low-rank bilinear optimization. In ICML, pages 1890–1898, 2014
28. R. Patro and C. Kingsford. Global network alignment using multiscale spectral signatures. Bioinformatics, 28(23):3105–3114, 2012
29. Si Si, Donghyuk Shin, Inderjit S. Dhillon and Beresford N. Parlett, Multi-Scale Spectral Decomposition of Massive Graphs, Neural Information Processing Systems Conference, 2014, https://papers.nips.cc/paper/5519-multi-scale-spectral-decomposition-of-massive-graphs.
30. Jure Leskovec, Anand Rajaraman and Jeffrey D. Ullman, "Mining of Massive Datasets," 2019, http://www.mmds.org