MA 4404 (4-0) Structure and Analysis of Complex Networks.

Contact Information: Prof. Ralucca Gera
Office 260 Spanagel, Naval Postgraduate School
Email rgera@nps.edu and Phone 831-656-2230
website: https://nps.edu/web/faculty-rgera/ma4404

Course Description This course focuses on the emerging science of complex networks and their applications, through an introduction to techniques and models for understanding network's behavior. The topics discussed will be building mainly on the concepts of graph-theory, and they will address the mathematics of networks, their applications and use in research. Students will be exposed to the building blocks of network analysis, learn about the ongoing research in the field through presentations and research articles, and apply their knowledge to the analysis of networks through the individual *Network Profile Summary* and the *Research Project*.

Many real-life networks are considered complex networks, and have received increasing attention in recent years from those seeking to understand emerging phenomena in technology and society. Some examples of complex networks are on-line social networks, the Internet, the World Wide Web, neural networks, food-webs, metabolic networks, power grids, airline networks, national highway networks, the brain, etc. We will study models created for these real networks, beginning with the Erdős–Rényi random networks and followed by more sophisticated models of network formation: Milgram's 1967 experiment and the Watts–Strogatz small-world networks, the Barabási–Albert preferential attachment growing model and its variants, the Malloy-Reed configuration model, the random geometric model, and community preserving models. We conclude with the study of their properties, degree distributions, centralities, shortest paths, clustering, robustness, community detection, followed by ideas in new areas of current research.

Course Learning Outcomes and Objectives: Your goal for this class is to develop the mathematical sophistication needed to understand what properties to search for in these complex networks, and what they tell you about the network. In doing this, you will

1. analyze new networks using the main concepts of complex network analysis:
   - identify network models and explain their structures;
   - choose between several methodologies in analyzing networks;
   - be able to grasp the meaning of a new research paper in complex networks;
2. evaluate:
   - contrast networks and synthetic models;
   - synthesize the new research papers;
   - critique peer's research;
3. create new network research:

- design new network models building on the existing ones and on the collected data;
- design experiments to test hypotheses;
- generate new theory by expanding on the designed experiments.

The learning outcomes above are achieved through building and analyzing the *Network Profile Summaries*, reading scientific papers and writing the *Research Project* for possible publication in network science conferences. They are exactly the puzzle pieces for the learning objectives of the course: understanding the concepts, models and methodologies needed to identify how to use knowledge on complex networks to either produce a research article or apply the knowledge in a real world situation. These learning objectives provide you the mathematical sophistication and confidence to use gained experience as situations arise.

Generative AI and Academic Integrity While NPS policy encourages you to explore generative AI tools, such as ChatGPT 4.0, Dall-E 2, and Copilot, among many others, you must use them ethically and responsibly. Your coursework must accurately represent your underlying work and ideas, not generative AI's capabilities. Review the NPS Interim Guiding Principles for Use of Generative Artificial Intelligence (AI) Tools, the DKL's Citation LibGuide advice, learn more on NPS's generative AI resource hub, and ask me if you have questions. Do not input into generative AI tools any classified information or CUI, unless the platform has been explicitly approved by DOD or DONCIO. If you need reasonable accommodations for a scheduled absence, please see me early in the quarter.

Course Content: The course materials are available at https://nps.edu/web/faculty-rgera/ma4404. The topics of the course are the following:

- Types of networks, such as technological, social, information and biological.
- Synthetic network models: the Erdős–Rényi random networks, the Watts–Strogatz small-world networks, the Barabási–Albert preferential attachment growing model and its variants, the Malloy-Reed configuration model, the random geometric model, community preserving models.
- Network properties: degree, clustering coefficient, diameter, density, shortest paths, node similarity, homophily.
- Centralities: degree, closeness, betweeness, eigenvector, Katz, PageRank, hubs and authorities.

Textbooks & References There is a fragmented literature, with inconsistent terminology and frequent reinvention of concepts and methodologies of network science due to the mix of the backgrounds of the researchers in this field. The class builds on several manuscripts and conference presentations. Presentation slides are available for each interactive lecture at https://nps.edu/web/faculty-rgera/ma4404

The main article used as the main reference for the class is The Structure and Function of complex networks by Newman (2003), the most cited paper in a SIAM journal. The expanded version of this article appears in the textbook "Networks: An Introduction" by Mark Newman. A supplemental paper is the review Characterization of complex networks: A survey of measurements by Costa et. al (2007). A free and interactive book in network science is available on Barabási's site at http://barabasi.com/networksciencebook/.

Primary Analysis Software:

- NetworkX (https://networkx.github.io/) and igraph (http://igraph.org/redirect.html), two Python libraries developed for the study of graphs and networks. Python (https://www.python.org/) is an open source programming language with wide interoperability and many, varied tools.

- Packages of the R programming language as well: igraph (http://kateto.net/networks-r-igraph) or Statnet (http://www.statnet.org/ for which a tutorial can be found here: https://statnet.org/trac/raw-attachment/wiki/Resources/introToSNAinR_sunbelt_2012_tutorial.pdf).

Primary Visualization Software: Gephi (http://gephi.org/) which is open source graph visualization and manipulation software. Other software choices are available and welcome to use, here is a list: http://www.gmw.rug.nl/∼huisman/sna/software.html.

Course Format and Schedule: Our class meets Mondays-Thursdays, following this format:

Mondays	Tuesdays	Wednesdays	Thursdays
Asynchronous guided learning	Interactive lectures	Research project on open research topics	Network Profile Summary presentations

The following activities will support your learning:

Activity	Points (of 300 total)
In-class participation	70
Multilayer Networks Group Activity	30
Network Profile Summary	100
Research Team Project	100

Table 1: The breakdown of the points for the final grade

In class participation (25 points):
I have a very interactive teaching style that expects everyone to participate in classroom discussions.

Assessments of learning: Participation is measured by evidence of class preparation, interactions during class (asking relevant questions, showing the ability to express critical thinking) and making connections even if they are not correct. You are encouraged to be engaged in these discussions, but give everyone else a chance to participate as well. The in class conversations allows you to modify and improve your existing perceptions about the network science topics.

Multilayer Networks Group Activity (25 points):
Based on the description and tasks of the Multilayer Network Activity, the multilayer activity presentation is due on the 3rd Thursday of the quarter. Each team will have 15 minutes to present their reasoning for the multilayer network and the team formation (presentation could be using a report document, the board, presentation slides or other creative ideas).

Assessments of learning: The assessment will be driven by the accuracy and creativity of the model and its solution. The following should shape the presentations as well: helpful visual aids, clear, complete and organized presentation (labeled pictures/tables).

Research Team Project (100 points):
The goal is to explore a novel research topic and learn the process of turning an exploration into a research paper. To begin this process, each student will review one paper that will be used as one of the references for the project, and turn in one to two paragraphs (or up to a page) synthesizing the paper. This allows you to critically look at research, having examples of papers to follow as you work on your project.

Researchers learn a great deal from both success and failures, so this is your chance to think creatively and make educated hypothesis. Mistakes are inevitable. Maintain the perspective that you are making

progress; missteps spark creativity and deepen your understanding. We learn as much from the mistakes as we do from the correct answers. The final goals for the team project are:

- a short, $5-10$ minute update of your research idea and findings each Tuesday,
- a final 20 minute presentation during the last week of classes, and
- a research paper (abut $8-12$ pages), due the week of finals.

Assessments of learning: The weekly updates serve as formative assessments to prepare you for the summative ones. I strongly suggest that you build your presentation slides and research paper weekly, so that I can provide feedback. The rubric on the last page of this syllabus will be used for assessing your team's research. The skeleton below will guide the research project that will emerge as a paper.

- 25 points for final presentation slides
- 25 points for self-assessed contribution to research team paper
- 50 points for research team paper

This paper's quality is given by knowledge integration:

1. Accuracy and Vision: The modeling assumptions need to be appropriate and the model needs to be checked against true network(s). The publication needs to give insight beyond a restatement of existing work and the exposition of the raw analysis of data; however, it should be related to existing work so that it has a place the current field of research.
2. Critical reasoning and exposition of relevant course material: Contrast your new methodology/parameter to existing ones. Present arguments for your methodology and test metrics demonstrating competence with the content of complex networks. Explain connections to the real world and the observations of your results.
3. Clarity: Communicate the problem and questions you are addressing, your methodology and approach in addressing them, and your insights, solutions, and remaining open questions. Make your explanations concise by eliminating unnecessary verbiage.
4. Rigor and precision: Your paper must be mathematically precise (using proofs if such results are presented), and logical in its reasoning throughout. Any methodology used should be justified, and limitations or assumptions should be clarified.

The description of the team project will be shared during the 3rd week of classes. The main two parts of the project are: (1) Theory development: Propose a theoretical direction and present reasons for your new methodology. (2) Validation of the Methodology: Run your theory on data set(s).

Network profile Summary (100 points):
By the end of the first week, each of you must choose a network that interests you. For the remaining of the quarter, you will be analyzing this network and present results about it. You are welcome to use your thesis data, collect data, find an existing network/data, download a data set from https://www.kaggle.com/ or download a network from http://faculty.nps.edu/rgera/MA4404/NetworkProfileSummaryResources.html.

This project transitions from knowledge exposure to practice (without mimicking the instructor). During and outside the class you will have the opportunity to experiment on your network, exploring the variety of topics as they are presented and as your intellectual curiosity inspires you.

By the end of each week, you need to create one (or two slides if absolutely needed) with your analysis of the topics presented that week, applied to your network. Each Thursday four students will present their

findings, followed by a class discussion of all the networks' analysis (similarities and dissimilarities). You will receive feedback on your presentation; use it to update the presentation slide for the final slide deck. Each of these three presentations is worth 20 points; the final presentation slides is worth 40 points. The final presentation slides are due on your presentation day of the last week of classes. Your presentation must be convincing, coherent, and cohesive, including the suggestions you have received throughout the quarter.

Assessment Rubric for Weekly Presentations (and final presentation slides): This knowledge, skill and ability assessment is both formative and summative assessment, allowing you to incorporate weekly feedback to refine the final project. A refined comprehensive presentation slide deck (including the weekly updated slide and a conclusion slide) is due on the finals' week.

Criteria	Task	Detailed Step (0-10 points)
Content	An Analysis is Performed	Correct analysis synthesizing learned concepts
		A (9-10 points) Relevant & clearly explained findings, Insightful contextualization of findings, and thoughtful synthesis & interpretation of metrics.
		B (8 points) Minor errors
		C (7 points) Significant errors
		D (6 points) Major conceptual errors
		F (0-5 points) Little to no work of merit
Presentation	Results are presented	Clarity and style of graphics (Could they be presented in a more significant way)
		A (9-10 points) **Slide Deck** clear & succinct slides, correct spelling and mathematical notation figures and tables are labeled and have captions consistent in tense and active voice references are provided **Conveying the information** clear verbal explanation correct use of terminology while explaining clear and loud speaking
		B (8 points) Minor errors
		C (7 points) Significant errors
		D (6 points) Incoherent presentation
		F (0-5 points) Little to no work of merit

Table 2: Assessment rubric for slides and all the presentations