Final project

In the final project, you will read a scientific article and utilize the new tools learned during the course to show what happens between the lines.

Learning outcomes

After finishing the course project, you

- have improved your skills in reading scientific text,
- have experience in recognizing the most important parts of a scientific article,
- have seen how to apply the tools and techniques learned during the course on a real-life research problem,
- have gained more self-confidence for working on network mathematics and reading articles about it, and
- can write an alone-standing, well-structured report on a scientific article.

Instructions

In order to pass the course project, you should:

- Read the article and its supplementary information. While reading, concentrate in particular on the mathematical tools and methods used to obtain the results;
- Determine which parts of the article are the most important ones: what is needed for understanding the results of the article and for evaluating the validity and meaning of the results.
  - Depending on the article you choose, there might be a large number of different ideas and results, or the article might simply focus on a specific result.
  - Keep in mind that quality is more important than quantity. It is better to cover a few results very well than to glance over all of the results.
- Go over the mathematics of the results that you chose. You can think of this as checking the validity of the equations in the article.
  - Explain how the formulas are derived and why they are used.
  - List all assumptions, also those that have been left implicit in the original article.
  - The level of detail should be about the same as in the exercises.
  - If you are not able to offer derivation for some formulas that you find important, mark this to your report (*e.g.* by writing "Unfortunately, I was not able to derive this formula").

- Write a report in which you cover the most central mathematics of the article. In the report, you should:
  - Write a short introduction (1-2 paragraphs) that tells what is the topic of the article and why it is important.
  - Describe what are the most important results/calculations of the article and why.
  - Show the intermediate steps of the calculations. You can write, for example, "The authors show this equation ... for the number of links in the network. Let us derive it in detail..."
- We have a strict policy for not allowing any plagiarism. Make sure that you're not copying from the original article. Clearly state what parts are from the article or some other source; everything else should be your own contribution.
  - Figures can be adapted, but there needs to be explicit mention of this together with a citation. Only citing is not enough.
  - Never start by copy-pasting sentences from the original sources and editing them later. Sentences where some of the words have just been changed are still sentences of the original author. They are considered to be plagiarism even though they might evade some automatic plagiarism detectors. If you need to include something directly from the original article, you can do so by using quotation marks "..." and citing properly.
  - Copying from other sources is also forbidden. For example, you are not allowed to copy from the projects completed in previous years. If you have returned the project in a previous year but have not passed the course, please contact the course staff to ask about the possibility of reusing your project this year.
  - If we find any plagiarism, we can fail the project without a chance for a boomerang.
  - Rules for handling plagiarism and other academic dishonesty can be found here.

The amount of work on the project should be around the same as for a single exercise round. This means that you might have time to go through in detail a few (or even one) key derivations depending on how long they are. Plan your work and justify your choices.

You might need to follow the citations made in the source article to find out some details that are omitted in the original article. Note that some articles might not be open access, but in network science, it is a common practice to post published articles in arXiv (the layout of the articles might be different from the journal version, but the content is usually exactly the same). Remember always to cite your sources properly.

NOTE: your report should be able to stand alone, *i.e.* be understandable for a reader who has not read the original article. So it is important to introduce the reader to the topic. Also, ensure that your report is well-structured and easy to read.

The deadline for handing in your solutions is **3rd March 2024 23:55**. Return your solutions (one `.pdf` file) in MyCourses (Assignments tab).

Grading

The project reports will be graded as failed, pass, or pass with distinction.

In order to pass, you should follow the instructions given above. It should be visible from your report that you have read and understood the article and been able to pick the most important parts of it. You should be able to derive most of the important mathematics; however, a couple of missing derivations or intermediate steps do not automatically lead to failing, especially if you have marked them properly (see above).

Passing the project should not require a lot more work than an exercise round. If you end up spending more than 20 hours on the project, you are probably doing something more than required – some students may be able to finish the project in a significantly shorter time.

In order to pass with distinction, your report should be excellent in all aspects. Students who pass with distinction will have their course grades increased by one. However, passing with distinction is rather an exception than a rule and requires special effort.

Here are some features the pass-with-distinction projects might have:

- The presentation style and the language that is used are particularly good
- There are no (or very few) mistakes in the mathematics
- The chosen mathematical parts of the article are particularly challenging and often omitted by the other students
- The project contains more work and details than is expected. (Note that in grading, quality is still valued over quantity, and lots of bad quality content is worse than smaller amounts of good quality content.)
- Content outside of the article: for example, alternative ways of reaching similar conclusions or formulas.

Failing the project can be caused by several reasons: you will fail if you either omit some important contents of the article or are not able to explain them in sufficient detail. A serious lack in the structure of the report can also lead to failure. In case of failure, you will get a possibility to improve your report and resubmit (the so-called *bumerang* system).

Articles

Herd Immunity and Vaccination Homophily

Herd immunity is a state of a population such that it is collectively immune to the invasion of an infectious disease. This state can be reached when enough people in the population have recovered from a disease or when enough people have been vaccinated against a disease and have developed protective antibodies against future infection. Finding the herd immunity threshold is not a straightforward problem due to different levels of heterogeneities in the contact network and health behaviors. In this paper, they study the effect of homophily with respect to vaccine adoption on the critical vaccine coverage for herd immunity and on the size of a major epidemic. They find that the critical coverage is higher for larger homophily. The epidemic size also increases as a function of homophily when the vaccine is highly effective, while it is maximized at an intermediate level of homophily when the efficacy is limited. Such results highlight the importance of the correlation between vaccine adoption and social network structure, which is not captured by the traditional assumption of homogeneous mixing. This work is available here: `https://journals.aps.org/pre/abstract/10.1103/PhysRevE.105.L052301` or `https://arxiv.org/abs/2112.07538`

Epidemic tracing and mitigation with apps

With the hit of new pandemic threats, scientific frameworks are needed to understand the unfolding of the epidemic. At the mitigation stage of the epidemics in which several countries are now, the use of mobile apps that are able to trace contacts is of utmost importance in order to control new infected cases and contain further propagation. Here, the authors present a theoretical approach using both percolation and message-passing techniques to the role of contact tracing in mitigating an epidemic wave. This work is available at `https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.3.L012014` or `https://arxiv.org/abs/2007.05277`

Random directed graphs and spreading phenomena

Most models of epidemic spread, including many designed specifically for COVID-19, implicitly assume that social networks are undirected, i.e., that the infection is equally likely to spread in either direction whenever contact occurs. In particular, this assumption implies that the individuals most likely to spread the disease are also the most likely to receive it from others. Here, the authors review results from the theory of random directed graphs, which show that many important quantities, including the reproductive number and the epidemic size, depend sensitively on the joint distribution of in- and out-degrees ("risk" and "spread"), including their heterogeneity and the correlation between them. They also show that directed networks have a forward and backward version of the classic "friendship paradox" – forward links tend to lead to individuals with high risk, while backward links lead to individuals with high spread – such that a combination of both forward and backward contact tracing is necessary to find superspreading events and prevent future cascades of infection. This work is available at `https://arxiv.org/abs/2005.11283`

Dynamical random graphs

Random networks are usually considered to be static so that they do not change in time. In the article "Random graph models for dynamic networks," Eur. Phys. J. B **90**, 200 (2017) (available at `https://arxiv.org/abs/1607.07570`) X. Zhang et. al. extend the usual random graph models (ER-graphs, configuration model and the degree corrected block model) by including rates at which links appear and disappear. They show how one can use snapshots of a random network, for example, to estimate the time scales of the evolution of the network or to infer community structure.

Maximum entropy in dynamic complex networks

This paper explores how representing diverse systems as complex networks and using network randomization can reveal common characteristics. By applying the maximum entropy principle and the maximum caliber approach, it constructs network evolutions based on specific constraints, analyzing both static and dynamic properties. The findings validate the approach against simulations, discuss its relationship with other entropy-based methods, and suggest future research directions. The paper is available from here `https://arxiv.org/abs/2401.15090`.

Stochastic Block Models

If you are interested in stochastic block models, you can build on your knowledge from week 5 and read "Parsimonious module inference in large networks" by T. Peixoto, Phys. Rev. Lett. **110**, 148701 (2013) (available at `https://arxiv.org/abs/1212.4794`). This article builds on an earlier article by the same author: `https://arxiv.org/abs/1112.6028`. For slightly more recent work, you might want to have a look at `https://arxiv.org/abs/1310.4377`, but you probably want to limit your project to the earlier articles due to time limitations.

Message passing and percolation on sparse networks

In this work, the authors reformulate percolation as a message-passing process and demonstrate how the resulting equations can be used to calculate, among other things, the size of the percolating cluster and the average cluster size. The calculations are exact for sparse networks when the number of short loops in the network is small, but even on networks with many short loops, we find them to be highly accurate when compared with direct numerical simulations. By considering the fixed points of the message-passing process, they also show that the percolation threshold on a network with few loops is given by the inverse of the leading eigenvalue of the so-called non-backtracking matrix. The paper is available from here `https://arxiv.org/abs/1405.0483`.

Spread of epidemic disease on networks

The study of social networks, and in particular the spread of disease on networks, has attracted considerable recent attention in the physics community. In this paper, Newman shows that a large class of standard epidemiological models, the so-called susceptible/infective/removed (SIR) models can be solved exactly on a wide variety of networks. In addition to the standard but unrealistic case of fixed infectiveness time and fixed and uncorrelated probability of transmission between all pairs of individuals, he solve cases in which times and probabilities are nonuniform and correlated. He also considers one simple case of an epidemic in a structured population, that of a sexually transmitted disease in a population divided into men and women. He confirms the correctness of his exact solutions with numerical simulations of SIR epidemics on networks. The paper is available from here `https://journals.aps.org/pre/abstract/10.1103/PhysRevE.66.016128`.

Article to submit if you have missed a homework submission

You can choose this article where the calculations are more simple and closer to the exercises and submit it instead of one problem set.

Random graphs with clustering

Throughout the whole course we often made the simplifying assumption that the networks are locally tree-like. Many real networks (e.g. social networks), however, have high clustering and many small loops. M.E.J. Newman proposes a solution to this problem in "Random Graphs with Clustering", Phys. Rev. Lett. **103**, 058701 (2009) (available at `https://arxiv.org/abs/0903.4009`) by considering an extension to the configuration model. In the extension triangles are considered as link-type objects that connect three verices instead of two. The calculations in this article are close to those on week 3.

Note that there is a typo in the arXiv version of the article in Equation 25. The version published in the journal is correct (Equation 24 in the journal). The Equation 25 should be: $v^2 = 1 - 2\phi(1-\phi)^2[1 - g_r(u, v^2)] - \phi^2(3 - 2\phi)[1 - g_r^2(u, v^2)]$.

Feedback

Give us feedback about the project and the last lecture at `https://forms.gle/FwryTertgmntJzGw6` and gain one bonus point!