

Homework #2

This homework is complete and will not be changed. The homework does not require a lot of writing, but may require some thinking. It does not require a lot of processing power, but may require efficient programming. It accounts for 10% of the course grade. Any questions and comments regarding the homework should be directed to [Piazza](#).

Submission details

This homework is due on **April 1st** at 11:59pm, while late days expire on **April 4th** at 11:00am. The homework must be submitted through (1) [Gradescope](#) (entry code **6PDZD3**) and (2) [eUcilmica](#). (1) Submission to [Gradescope](#) should include answers to all questions, each on a separate page, which may also demand pseudocode, proofs, tables, plots, diagrams and other. (2) Submission to [eUcilmica](#) should include at least this cover sheet with signed honor code and all the programming code used to complete the exercises. The homework is considered submitted only when *both* parts have been submitted. Failing to include this honor code in the submission will result in **10% deduction**. Failing to submit all the developed code will result in **50% deduction**.

Honor code

Students are strongly encouraged to discuss the homework with other classmates and form study groups. Yet, each student must then solve the homework by itself without the help of others and should be able to redo the homework at a later time. In other words, students are encouraged to collaborate, but should not copy from one another. Referring to any solutions obtained from classmates, course books, previous years, found online or other material, is considered an honor code violation. Also, stating any part of the solutions in class or on [Piazza](#) is considered an honor code violation. Finally, failing to name the correct study group members, or filling out the wrong date or time of the submission, is also considered an honor code violation. Honor code violation will not be tolerated! Any student violating the honor code will be reported to **faculty disciplinary committee** and vice dean for education.

Name & SID: _____

Study group: _____

Date & time: _____

I acknowledge and accept the honor code.

Signature: _____

1 Where is SN100? (0.75 points)

Figure 1 shows social network of [bottlenose dolphins](#) famously studied by Lusseau [[LSB⁺03](#)]. After the disappearance of a particular dolphin named SN100 during the course of experiment [[LN04](#)], the rest split into two groups shown with different node colors. Your task is to analyze whether network analysis could be utilized to detect this obviously important role of SN100.

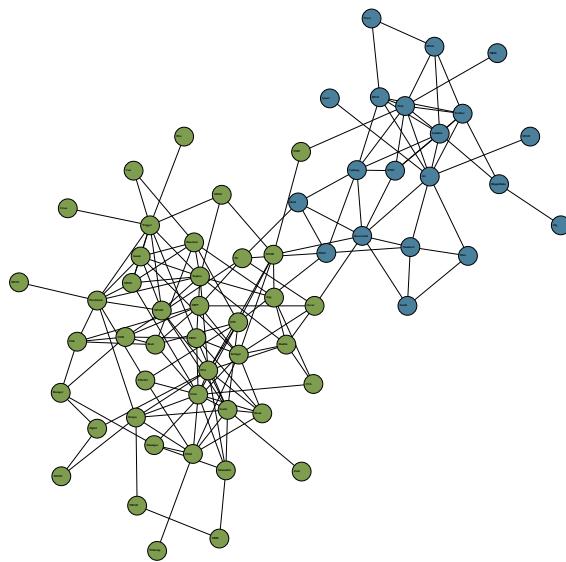


Figure 1: **Lusseau bottlenose dolphins network**

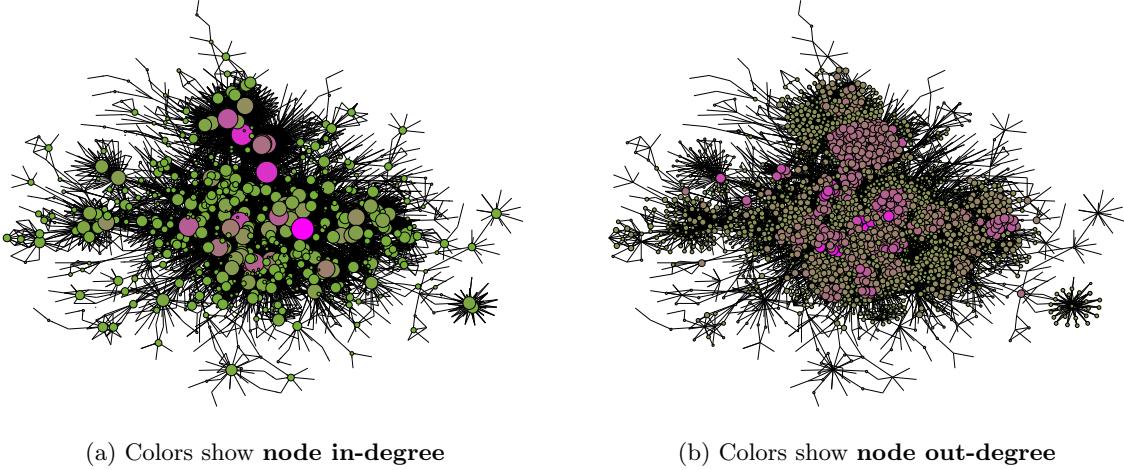
You can use any network analysis method, algorithm or technique. State how you measured the importance of SN100 and provide all necessary results.

What to submit?

State how you measured the importance of SN100 and briefly discuss the results (2×0.25 points). Give a printout of all the code used to solve the exercise (0.25 points).

2 Is software scale-free? (2 points)

Consider two software class dependency networks [[ŠB11](#)] representing [Java language](#) and [Lucene search engine](#) (Figure 2). These are directed networks where node i links to node j if software class represented by i depends on or *uses* class represented by j . Compute degree distributions p_k of both networks and plot each p_k on a separate log-log plot by representing each distinct (k, p_k) with a single dot, $k > 0$. (*Transformation to logarithmic axes should be done by your plotting software.*) Next, compute also in-degree distributions $p_{k^{in}}$ and out-degree distributions $p_{k^{out}}$ of both networks, and superimpose them on the same plots using dots of different color.

Figure 2: **Lucene class dependency network**

Compare all three degree distributions $p_{k^{\cdot}}$ of each network and highlight the differences among them. Do the distributions appear to be scale-free following a power-law $k^{-\gamma}$? For the distributions not appearing like a power-law, reason why. For the distributions that seem to follow a power-law, reason why and compute their power-law exponent γ using maximum likelihood formula [Bar16]

$$\gamma = 1 + n \cdot \left[\sum_{i=1}^n \ln \frac{k_i}{k_{min} - \frac{1}{2}} \delta(k_i \geq k_{min}) \right]^{-1},$$

where k_i is \cdot -degree of node i , $k_{min} \geq 1$ is some reasonable choice for the minimum degree cutoff and $n \leq n$ is the number of nodes thus considered.

What to submit?

Draw two plots with three distributions each and briefly discuss the results (3×0.25 points). Reason whether distributions are scale-free and why, and compute γ and k_{min} of seemingly scale-free distributions (4×0.25 points). Give a printout of all the code used to solve the exercise (0.25 points).

3 Errors and attacks on the Internet (2 points)

Consider [Internet overlay map](#) represented by an undirected graph. Routers, switches and hubs on the Internet fail from time to time. Your task is to study how this affects the Internet's ability to stay connected. Simulate such failure by random removal of nodes and measure the fraction of nodes in the largest connected component in the resulting network. Next, consider also a malicious individual attacking the Internet. Simulate such attack by removing most highly linked nodes or hubs and again measure the fraction of nodes in the largest connected component. For each of the two scenarios, plot the fraction of nodes in the largest connected component after removing 0%, 10%, 20%, 30%, 40% and 50% of randomly selected nodes or hubs. Finally, repeat the experiments also for a random graph [ER59] with the same number of nodes and links as the Internet map.

Briefly discuss the results by comparing robustness of the Internet and random graph. Would you claim that the Internet is robust to random failures? Would you claim that the Internet is robust to malicious attacks? What about random graphs? Can you explain your reasoning by some structural property of the Internet and/or random graphs.

What to submit?

Draw two plots showing the robustness of the Internet and random graph (3×0.25 points). Briefly comment on the results and give answers to all four questions (4×0.25 points). Give a printout of all the code used to solve the exercise (0.25 points).

4 HIV and network sampling (1.75 points)

When a patient is diagnosed with HIV, in most Western countries, she/he will be questioned about past sexual contacts. The authorities would then make an effort to track down those contacts and test them for HIV. The process is repeated with anyone who also tests positive, tracing her/his contacts as well, until all leads have been exhausted. This process is called contact tracing.

Notice that contact tracing gives a (biased) sample of the underlying social network [LF06]. Assuming that one gets HIV from a random sexual contact, contact tracing can be approximated by a simple random walk. Simulate a random walk on small [social network](#) until you sample 10% of the nodes and take an *induced* graph on the sampled nodes for your sampled network.

Is the original social network small-world [WS98] and/or seemingly scale-free [BA99]? Is the sampled network small-world and/or seemingly scale-free? Can you reason why? Support your answers with necessary computations.

What to submit?

Give brief answers to all three questions and provide necessary results to support your reasoning (3×0.5 points). Give a printout of all the code used to solve the exercise (0.25 points).

5 Who to vaccinate? (0.5 points)

Probability that an individual would spread the disease through her/his social network is proportional to k^2 [New10], where k is the degree of corresponding node. Consider two immunization schemes for preventing the spread of diseases. In the first scheme, you randomly select some number of individuals and vaccinate them. In the second scheme, you select the same individuals, but then rather vaccinate a random acquaintance of theirs.

Which of the two schemes do you expect to provide better immunization? Why?

What to submit?

Give brief answers to both questions (2×0.25 points).

References

- [BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [Bar16] A.-L. Barabási. *Network Science*. Cambridge University Press, Cambridge, 2016.
- [ER59] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.
- [LF06] Jure Leskovec and Christos Faloutsos. Sampling from large graphs. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 631–636, Philadelphia, PA, USA, 2006.
- [LN04] David Lusseau and M. E. J. Newman. Identifying the role that animals play in their social networks. *Proc. Biol. Sci.*, 271:S477–S481, 2004.
- [LSB⁺03] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? *Behav. Ecol. Sociobiol.*, 54(4):396–405, 2003.
- [New10] Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, 2010.
- [ŠB11] Lovro Šubelj and Marko Bajec. Community structure of complex software systems: Analysis and applications. *Physica A*, 390(16):2968–2975, 2011.
- [WS98] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.