PoCS What's The Story?

Principles of Complex Systems, Vols. 1 and 2
CSYS/MATH 6701, 6713
University of Vermont, Fall 2025
"Hey, could we be judged individually?"
Assignment 86

It's Always Sunny in Philadelphia: The Gang Goes to Hell, Part I and Part II, S2E12
Episode links: IMDB, Fandom, TV Tropes.

Due: Never
https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/assignments/86/
Some useful reminders:
Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)
Office: The Ether and/or Innovation, fourth floor
Office hours: See Teams calendar
Course website: https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse
Overleaf: LaTeX templates and settings for all assignments are available at
https://www.overleaf.com/read/tsxfwwmwdgxj.

Some guidelines:

1. Each student should submit their own assignment.
2. All parts are worth 3 points unless marked otherwise.
3. Please show all your work/workings/workingses clearly and list the names of others with whom you ~~conspired~~ collaborated.
4. We recommend that you write up your assignments in LaTeX (using the Overleaf template). However, if you are new to LaTeX or it is all proving too much, you may submit handwritten versions. Whatever you do, please only submit single PDFs.
5. For coding, we recommend you improve your skills with Python. And it's going to be a no for the catachrestic Excel. Please do not use any kind of AI thing unless directed. The (evil) Deliverator uses (evil) Matlab.
6. There is no need to include your code but you can if you are feeling especially proud.

Assignment submission:

Via **Brightspace** (which is not to be confused with the death vortex of the same name, just a weird coincidence). Again: One PDF document per assignment only.

The questions you don't have to do!

Some are open ended madnesses.

1. (9 points overall)

   Examine current weightlifting world records for the snatch, clean and jerk, and the total for scaling with body mass (three regressions).

   Do so separately for both women and men's current world records.

   This makes for six regressions.

   For weight classes, take the upper limit for the mass of the lifter.[1]

   For the open category (for women and for men), take the mass of the lifter with the world record. If unknown, omit this one data point.

   Wikipedia is an excellent source.

   (a) (3 points) Plot each set of data with the best fit regression line along with a adjacent line indicating the 2/3 fit.

   You can do this with just two plots, one for women and one for men, by plotting the three competitions on each axis.

   You can offset the 2/3 line vertically for clarity.

   (b) (3 points) How well does 2/3 scaling hold up?

   Is 2/3 scaling over, under, or thereabouts?

   Optional: support your observation by making mention of the errors reported by the linear regression method you used.

   (c) (3 points) Normalized by the scaling you determine, who holds the overall, rescaled world record?

   Again, there are six of world records.

   Normalization here means relative:

$$100 \times \left(\frac{M_{\textrm{world record}}}{c M^{\beta}_{\textrm{weight class}}} - 1 \right),$$

   where c and β are the parameters determined from a linear fit.

[1] In general, and beyond weightlifting, athletes will try to be at the upper weight limit of their sport.

2. Plot time series for the rank of the following baby names in the US over all years in the census data.

   Do so for raw ranks and $\log_{10}$ ranks.

   - Shirley.
   - Desmond.
   - Madison.
   - Aiden.
   - A name of your choice.

   Note that if you plotted relative frequency rather than rank, you would need to know (or estimate) the overall number of babies born. Ranks are both easy simple to work with and easy to understand.

3. **The complex geographies of fairness, greed, belief.**

   Let's start connecting people to places.

   Now: Source census population data as a function of location with corresponding map shape files.

   Goal: We will want to be able to connect density of people in regions with density of specific facilities.

   So the shape files should be as usefully fine in scale as possible. For the census, we have block, block groups, and tracts.

   Please do this collectively by discussing and sharing links/data in the assignments channel on Teams.

   Depending on the software you use, much of this data may be well curated.

4. From lectures on Supply Networks:

   Show that for large V and $0 < \epsilon < 1/2$

   $$\min V_{\text{net}} \propto \int_{\Omega_{d,D}(V)} \rho \, ||\vec{x}||^{1-2\epsilon} \, \mathsf{d}\vec{x} \sim \rho V^{1+\gamma_{\max}(1-2\epsilon)}$$

   Reminders: we defined $L_i = c_i^{-1} V^{\gamma_i}$ where $\gamma_1 + \gamma_2 + \ldots + \gamma_d = 1$, $\gamma_1 = \gamma_{\max} \geq \gamma_2 \geq \ldots \geq \gamma_d$., and $c = \prod_i c_i \leq 1$ is a shape factor.

   Assume the first k lengths scale in the same way with $\gamma_1 = \ldots = \gamma_k = \gamma_{\max}$, and write $||\vec{x}|| = (x_1^2 + x_2^2 + \ldots + x_d^2)^{1/2}$.

5. (3 + 3 points) **Supply networks and allometry:**

   This question's calculation is a specific, exactly-solvable case of the general result that you may attack (with optional relish and other condiments) in a nearby question.

   Consider a set of rectangular areas with side lengths L_1 and L_2 such that $L_1 \propto A^{\gamma_1}$ and $L_2 \propto A^{\gamma_2}$ where A is area and $\gamma_1 + \gamma_2 = 1$. Assume $\gamma_1 > \gamma_2$ and that $\epsilon = 0$.

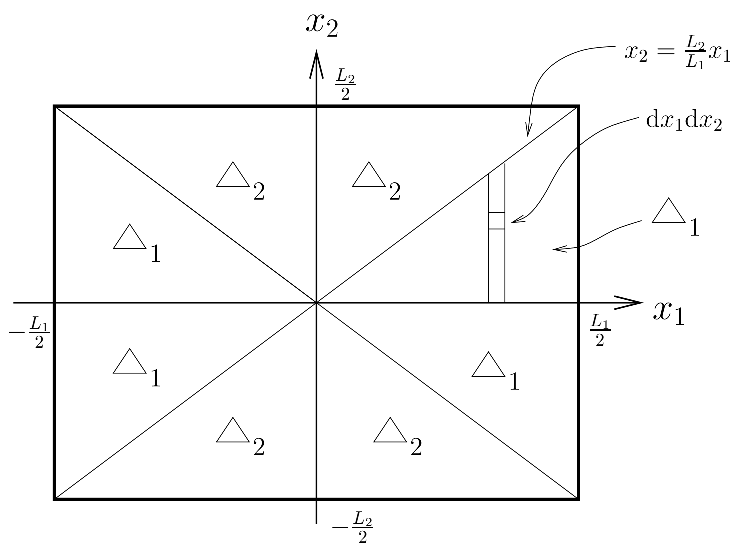
   Now imagine that material has to be distributed from a central source in each of these areas to sinks distributed with density $\rho(A)$, and that these sinks draw the same amount of material per unit time independent of L_1 and L_2.

   (a) Find an exact form for how the volume of the most efficient distribution network scales with overall area $A = L_1 L_2$. (Hint: you will have to set up a double integration over the rectangle.)

   (b) If network volume must remain a constant fraction of overall area, determine the maximal scaling of sink density ρ with A.

   Extra hints:

   - Integrate over triangles as follows.
   - You need to only perform calculations for one triangle.



6. Open:

   Derive a scaling law for the number of side branches that doesn't use stream ordering.

   How many parameters do we need? 3?

7. Come up with a microscopic description of branching river networks that builds from the outlet of the basin rather than the smallest streams.

   For bodies, move from aorta to capillaries.

8. (3 + 3 + 3)

   Estimating the rare:

   Google's raw data is for word frequency $k \geq 200$ so let's deal with that issue now.

   From Assignment 2, we had for word frequency in the range $200 \leq k \leq 10^7$, a fit for the CCDF of

   $$N_{\geq k} \sim 3.46 \times 10^8 k^{-0.661},$$

   ignoring errors.

   (a) Using the above fit, create a complete hypothetical N_k by expanding N_k back for $k = 1$ to $k = 199$, and plot the result in double-log space (meaning log-log space).

   (b) Compute the mean and variance of this reconstructed distribution.

   (c) Estimate:

      i. The hypothetical total number and fraction of unique words in Google's data set (think at the species or type level now),

      ii. The hypothetical fraction of words that appear once out of all words (think of words as organisms or tokens here),

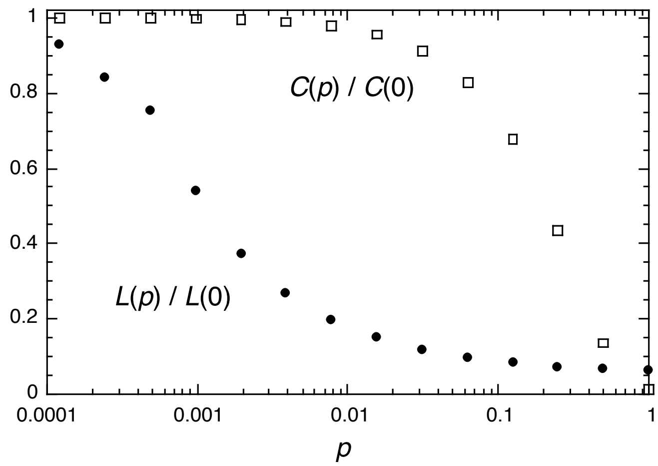
      iii. And what fraction of total words are left out of the Google data set by providing only those with counts $k \geq 200$ (back to words as organisms or tokens).

9. Simulate the small-world model and reproduce Fig. 2 from the 1998 Watts-Strogatz paper showing how clustering and average shortest path behave with rewiring probability p [1].

   Please find and use any suitable code online, and feel free to share with each other via Slack.

   Use $N = 1000$ nodes and $k = 10$ for average degree, and vary p from 0.0001 to 1, evenly spaced on a logarithmic scale (there are only 14 values used in the paper).

   Here's the figure you're aiming for:



10. (3 + 3 + 3 + 3 + 3 + 3 pts) **Generalized entropy and diversity:**

For a probability distribution of $i = 1, \ldots, n$ entities with the ith entity having probability of being observed p_i, Shannon's entropy is defined as [2]: $H = -\sum_{i=1}^{n} p_i \ln p_i$. There are other kinds of entropies and we'll explore some aspects of them here.

Let's use the setting of words in a text (another meaningful framing is abundance of species in an ecology). So we have word i appearing with probability p_i and there are n words.

Now, a useful quantity associated with any kind of entropy is diversity, D [3]. Given a text T with entropy H, we define D to be the number of words in another hypothetical text T' which (1) has the same entropy, and (2) where all words appear with equal frequency $1/D$. In text T', we have $p_i = 1/D$ for $i = 1, \ldots, D$.

Diversity is thus a number, and behaves in number-like ways that are more intuitive to grasp than entropy. (Entropy is still the primary thing here.)

Determine the diversity D in terms of the probabilities $\{p_i\}$ for the following:

(a) Simpson concentration:

$$S = \sum_{i=1}^{n} p_i^2.$$

(b) Gini index:

$$G \equiv 1 - S = 1 - \sum_{i=1}^{n} p_i^2.$$

Please note any connections between diversity for the Simpson and Gini indices.

(c) Shannon's entropy:

$$H = -\sum_{i=1}^{n} p_i \ln p_i.$$

(d) Renyi entropy:

$$H_q^{(\mathrm{R})} = \frac{1}{q-1}\left(-\ln \sum_{i=1}^{n} p_i^q\right),$$

where $q \neq 1$.

(e) The generalized Tsallis entropy:

$$H_q^{(\mathrm{T})} = \frac{1}{q-1}\left(1 - \sum_{i=1}^{n} p_i^q\right),$$

where $q \neq 1$.

Please note any connections between diversity for Renyi and Tsallis.

(f) Show that in the limit $q \to 1$, the diversity for the Tsallis entropy matches up with that of Shannon's entropy.

11. Determine the average value of samples with value $k \geq \min k_{\max}$ to find how the expected value of $k_{\max}$ (i.e., $\langle k_{\max}\rangle$) scales with N.

12. (3 + 3)

Allotaxonometry.

Rank-turbulence divergence (RTD) is defined as:

$$\begin{aligned} D_\alpha^{\mathrm{R}}(R_1 \,\|\, R_2) &= \sum_{\tau \in R_{1,2;\alpha}} \delta D_{\alpha,\tau}^{\mathrm{R}}(R_1 \,\|\, R_2) \\ &= \frac{1}{\mathcal{N}_{1,2;\alpha}} \frac{\alpha+1}{\alpha} \sum_{\tau \in R_{1,2;\alpha}} \left| \frac{1}{\left[r_{\tau,1}\right]^\alpha} - \frac{1}{\left[r_{\tau,2}\right]^\alpha} \right|^{1/(\alpha+1)}. \end{aligned} \tag{1}$$

Find the limits of RTD for:

(a) $\alpha \to 0$.

(b) $\alpha \to \infty$.

Leave $\frac{1}{\mathcal{N}_{1,2;\alpha}}$ as a constant.

13. For finite cutoffs a and b with $a \ll b$, which cutoff dominates the expression for the nth moment as a function of γ and n?

Note: both cutoffs may be involved to some degree.

14. *"Any good idea can be stated in fifty words or less."—Stanisław Ulam.*[2]

    Things have sped up since Ulam made his claim.

    The top of the narrative hierarchy:

    Read through Anderson's seminal paper "More is different" [4] and generate three descriptions of complexification with exactly the following lengths:

    (a) 1–3 words,

    (b) 4–6 words,

    (c) and 7–12 words.

    The 1–3 words one: Try to improve on "More is different".

15. For class discussion, read "Will a large complex system be stable?" by Robert May [5].

    Put together three comments and/or questions.

16. (3 + 3 + 3 + 3) This question is all about pure finite and infinite random networks

    We'll define a finite random network as follows. Take N labelled nodes and add links between each pair of nodes with probability p.

    (a) i. For a random node i, determine the probability distribution for its number of friends k, $P_k(p, N)$.

        ii. What kind of distribution is this?

        iii. What does this distribution tend toward in the limit of large N, if p is fixed?
        (No need to do calculations here; just invoke the right Rule of the Universe.)

    (b) Using $P_k(p, N)$, determine the average degree. Does your answer seem right intuitively?

    (c) Show that in the limit of $N \to \infty$ but with mean held constant, we obtain a Poisson degree distribution.

        Hint: to keep the mean constant, you will need to change p.

    (d) i. Compute the clustering coefficients C_1 and C_2 for standard finite random networks (N nodes).

        ii. Explain how your answers make sense.

        iii. What happens in the limit of an infinite random network with finite mean?

[2] At the very least, Ulam's claim is self-consistent.

17. (3 + 3)

    Determine the clustering coefficient for toy model small-world networks [1] as a function of the rewiring probability p. Find C_1, the average local clustering coefficient:

$$C_1(p) = \left\langle \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2} \right\rangle_i = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j_1 j_2 \in \mathcal{N}_i} a_{j_1 j_2}}{k_i(k_i-1)/2}$$

    where N is the number of nodes, $a_{ij} = 1$ if nodes i and j are connected, and $\mathcal{N}_i$ indicates the neighborhood of i.

    As per the original model, assume a ring network with each node connected to a fixed, even number m local neighbors ($m/2$ on each side). Take the number of nodes to be $N \gg m$.

    Start by finding $C_1(0)$ and argue for a $(1-p)^3$ correction factor to find an approximation of $C_1(p)$.

    Hint 1: you can think of finding C_1 as averaging over the possibilities for a single node.

    Hint 2: assume that the degree of individual nodes does not change with rewiring but rather stays fixed at m. In other words, take the average degree of individuals as the degree of a randomly selected individual.

    For what value of p is $C_1(p)/C_1(0) \simeq 1/2$?

    Does this seem reasonable given your simulation?

    (3 points for set up, 3 for solving.)

18. (3 + 3):

    Consider a modified version of the Barabàsi-Albert (BA) model [6] where two possible mechanisms are now in play. As in the original model, start with m_0 nodes at time $t = 0$. Let's make these initial guys connected such that each has degree 1. The two mechanisms are:

    M1: With probability p, a new node of degree 1 is added to the network. At time $t+1$, a node connects to an existing node j with probability

$$P(\text{connect to node } j) = \frac{k_j}{\sum_{i=1}^{N(t)} k_i} \qquad (2)$$

    where k_j is the degree of node j and $N(t)$ is the number of nodes in the system at time t.

M2: With probability $q = 1 - p$, a randomly chosen node adds a new edge, connecting to node j with the same preferential attachment probability as above.

Note that in the limit $q = 0$, we retrieve the original BA model (with the difference that we are adding one link at a time rather than m here).

In the long time limit $t \to \infty$, what is the expected form of the degree distribution P_k?

Do we move out of the original model's universality class?

Different analytic approaches are possible including a modification of the BA paper, or a Simon-like one (see also Krapivsky and Redner [7]).

Hint: You can attempt to solve the problem exactly and you'll find an integrating factor story.

Another hint, moment of mercy: Approximate the differential equation by considering large t (this will simplify the denominators).

(3 points for set up, 3 for solving.)

19. (3 + 3)

Using Gleeson and Calahane's iterative equations below, derive the contagion condition for a vanishing seed by taking the limit $\phi_0 \to 0$ and $t \to \infty$. In lectures, we derived the discrete evolution equations for the fraction of infected nodes ϕ_t and the fraction of infected edges θ_t as follows:

$$\phi_{t+1} = \phi_0 + (1 - \phi_0) \sum_{k=0}^{\infty} P_k \sum_{j=0}^{k} \binom{k}{j} \theta_t^j (1 - \theta_t)^{k-j} B_{kj},$$

$$\theta_{t+1} = G(\theta_t; \phi_0) = \phi_0 + (1 - \phi_0) \sum_{k=1}^{\infty} \frac{k P_k}{\langle k \rangle} \sum_{j=0}^{k-1} \binom{k-1}{j} {\theta_t}^j (1 - \theta_t)^{k-1-j} B_{kj},$$

where $\theta_0 = \phi_0$, and B_{kj} is the probability that a degree k node becomes active when j of its neighbors are active.

Recall that by contagion condition, we mean the requirements of a random network for macroscopic spreading to occur.

To connect the paper's model and notation to those of our lectures, given a specific response function F and a threshold model, the B_{kj} are given by $B_{kj} = F(j/k)$.

Allow B_{k0} to be arbitrary (i.e., not necessarily 0 as for simple threshold functions).

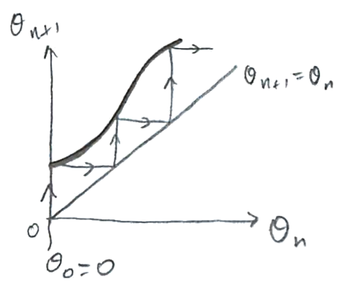
We really only need to understand how θ_t behaves. Write the corresponding equation as $\theta_{t+1} = G(\theta_t; \phi_0)$ and determine when

(a) $G(0;0) > 0$ (spreading is for free).

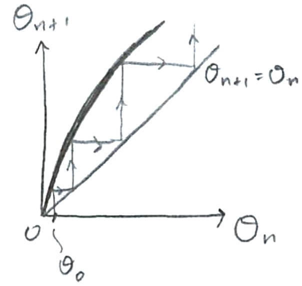
(b) $G(0;0) = 0$ and $G'(0;\phi_0) > 1$ meaning $\phi = 0$ is a unstable fixed point.

Here's a graphical hint for the three cases you need to consider as $\theta_0 \to 0$:

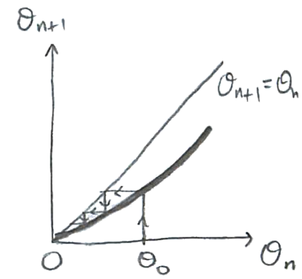
Success:



Sucesss:



Fail:



20. (3 + 3 + 3) Optional:

Solve Krapivsky-Redner's model for the pure linear attachment kernel $A_k = k$.

Starting point:

$$n_k = \frac{1}{2}(k-1)n_{k-1} - \frac{1}{2}kn_k + \delta_{k1}$$

with $n_0 = 0$.

(a) Determine n_1.

(b) Find a recursion relation for n_k in terms of n_{k-1}.

(c) Now find

$$n_k = \frac{4}{k(k+1)(k+2)}$$

for all k and hence determine γ.

21. (3 + 3) Optional:

From lectures:

(a) Starting from the recursion relation

$$n_k = \frac{A_{k-1}}{\mu + A_k} n_{k-1},$$

and $n_1 = \mu/(\mu + A_1)$, show that the expression for n_k for the Krapivsky-Redner model with an asymptotically linear attachment kernel A_k is:

$$\frac{\mu}{A_k} \prod_{j=1}^{k} \frac{1}{1 + \frac{\mu}{A_j}}.$$

(b) Now show that if $A_k \to k$ for $k \to \infty$ (or for large k), we obtain $n_k \to k^{-\mu-1}$.

22. (3 + 3 + 3)

From lectures, complete the analysis for the Krapivsky-Redner model with attachment kernel:

$$A_1 = \alpha \text{ and } A_k = k \text{ for } k \geq 2.$$

Find the scaling exponent $\gamma = \mu + 1$ by finding μ. From lectures, we assumed a linear growth in the sum of the attachment kernel weights $\mu t = \sum_{k=1}^{\infty} N_k(t) A_k$, with $\mu = 2$ for the standard kernel $A_k = k$.

We arrived at this expression for μ which you can use as your starting point:

$$1 = \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{1}{1 + \frac{\mu}{A_j}}$$

(a) Show that the above expression leads to

$$\frac{\mu}{\alpha} = \sum_{k=2}^{\infty} \frac{\Gamma(k+1)\Gamma(2+\mu)}{\Gamma(k+\mu+1)}$$

Hint: you'll want to separate out the $j = 1$ case for which $A_j = \alpha$.

(b) Now use result that [7]

$$\sum_{k=2}^{\infty} \frac{\Gamma(a+k)}{\Gamma(b+k)} = \frac{\Gamma(a+2)}{(b-a-1)\Gamma(b+1)}$$

to find the connection

$$\mu(\mu - 1) = 2\alpha,$$

and show this leads to

$$\mu = \frac{1 + \sqrt{1 + 8\alpha}}{2}.$$

(c) Interpret how varying α affects the exponent γ, explaining why $\alpha < 1$ and $\alpha > 1$ lead to the particular values of γ that they do.

23. Yes, even more on power law size distributions. It's good for you.

For the probability distribution $P(x) = cx^{-\gamma}$, $0 < a \leq x \leq b$, compute the mean absolute displacement (MAD), which is given by $\langle |X - \langle X \rangle| \rangle$ where $\langle \cdot \rangle$ represents expected value. As always, simplify your expression as much as possible.

MAD is a more reasonable estimate for the width of a distribution, but we like variance σ^2 because the calculations are much prettier. Really.

24. In the limit of $b \to \infty$, show that MAD asymptotically behave as:

$$\langle |X - \langle X \rangle| \rangle = \frac{2(\gamma-2)^{(\gamma-3)}}{(\gamma-1)^{(\gamma-2)}} a.$$

How does this compare with the behavior of the variance? (See the last question of Assignment todo???.)

25. *Simon's model II:*

A missing piece from the lectures: Obtain γ in terms of ρ by expanding Eq. **??** in terms of $1/k$. In the end, you will need to express n_k/n_{k-1} as $(1-1/k)^{\theta}$; from here, you will be able to identify γ. Taylor expansions and Procrustean truncations will be in order.

This (dirty) method avoids finding the exact form for n_k.

26. A spectacularly optional extra.

Warning:

- Only attempt if using registered safety equipment including welding goggles and a lead apron.
- Make sure to back up your brain in at least two geographically distant places beforehand (e.g., on different planets).

Dangerous feature:

- If you make it out, you will be very happy.

In lectures on lognormals and other heavy-tailed distributions, we came across a super fun and interesting integral when considering organization size distributions arising from growth processes with variable lifespans.

Show that

$$P(x) = \int_{t=0}^{\infty} \lambda e^{-\lambda t} \frac{1}{x\sqrt{2\pi t}} \exp\left(-\frac{(\ln \frac{x}{m})^2}{2t}\right) \mathrm{d}t$$

leads to:

$$P(x) \propto x^{-1} e^{-\sqrt{2\lambda(\ln \frac{x}{m})^2}},$$

and therefore, surprisingly, two different scaling regimes. Enjoyable suffering may be involved. Really enjoyable suffering. But many monks have found a way so you should follow their path laid out below.

Hints and steps:

- Make the substitution $t = u^2$ to find an integral of the form (excluding a constant of proportionality)

$$I_1(a, b) = \int_0^\infty \exp\left(-au^2 - b/u^2\right) \mathrm{d}u$$

  where in our case $a = \lambda$ and $b = (\ln \frac{x}{m})^2/2$.

- Substitute $au^2 = t^2$ into the above to find

$$I_1(a, b) = \frac{1}{\sqrt{a}} \int_0^\infty \exp\left(-t^2 - ab/t^2\right) \mathrm{d}t$$

- Now work on this integral:

$$I_2(r) = \int_0^\infty \exp\left(-t^2 - r/t^2\right) \mathrm{d}t$$

  where $r = ab$.

- Differentiate I_2 with respect to r to create a simple differential equation for I_2. You will need to use the substitution $u = \sqrt{r}/t$ and your differential equation should be of the (very simple) form

$$\frac{\mathrm{d}I_2(r)}{\mathrm{d}r} = -(\text{something}) I_2(r).$$

- Solve the differential equation you find. To find the constant of integration, you can evaluate $I_2(0)$ separately:

$$I_2(0) = \int_0^\infty \exp(-t^2) \mathrm{d}t,$$

  where our friend $\Gamma(frac12)$ comes into play.

A collection of questions from earlier seasons of PoCS, Vol 2 (also variously known as CoNKs, CocoNuTs, and Complex Networks).

This is all a big soup and some questions may be poorly constructed or repeated.

- The first series of questions will explore real networks by performing some key measurements introduced in Principles of Complex Systems, Vol. 1.

- For general coherence with other humans, you are encouraged to use Python. Also very good: Unix command line tools, R, Julia, Matlab. But you can of course use whatever system you like.

- Data is available in two compressed formats:
  - Matlab + text (tgz): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.tgz
  - Matlab + text (zip): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.zip

  and can also be found on the course website (helpfully) under data.
- The main Matlab file containing everything is networkdata_combined.mat.
- For directed networks, the ijth entry of the adjacency matrix represents the weight of the link from node i to node j. Adjacency matrices for undirected networks are symmetric.
- For all questions below, treat each network as undirected unless otherwise instructed.
- For this assignment, convert all weights on links to 1, if the network is weighted.
- You do not have to use Matlab for your basic analyses. Python would be a preferred route for many.
- The supplied text versions may be of use for visualization using gml.
- The Matlab command spy will give you a quick plot of a sparse adjacency matrix.
- Real data sets used here are taken from Mark Newman's compilation (and linked-to sites) at http://www-personal.umich.edu/~mejn/netdata/.

1. Record in a table the following basic characteristics:
   - N, the number of nodes;
   - m, the total number of links;
   - Whether the network is undirected or directed based on the symmetry of the adjacency matrix;
   - $\langle k \rangle$, the average degree ($\langle k_{\rm in} \rangle$ and $\langle k_{\rm out} \rangle$ if the network is directed);
   - The maximum degree $k^{\rm max}$ (for both out-degree and in-degree if the network is directed);
   - The minimum degree $k^{\rm min}$ (for both out-degree and in-degree if the network is directed).
2. (3+3)

(a) Plot the degree distribution P_k as a function of k. In the case that P_k versus k is uninformative, also produce plots that are clarifying. For example, $\log_{10} P_k$ versus $\log_{10} k$.

(Note: Always use base 10.)

(b) See if you can characterize the distributions you find (e.g., exponential, power law, etc.).

3. Measure the clustering coefficient C_2 where

$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}.$$

For directed networks, transform them into undirected ones first.

One approach is to compute C_2 as

$$C_2 = \frac{3 \times \frac{1}{6}\mathrm{Tr}A^3}{\frac{1}{2}\left(\sum_{ij}[A^2]_{ij} - \mathrm{Tr}A^2\right)}.$$

Note: avoiding computing A^3 is important and can be done.

4. For each of our main six networks, compute and present distributions of the shortest path length between all pairs of nodes. Notation: $d_{i,j}$ is the shortest distance between i and j.

Also compute the average shortest path length, $\langle d \rangle$.

5. Generate ensembles of random networks of the same 'size' as the six networks. Process 1 random network and then scale up as computing power/time/sanity permits. 1000 random networks would be good.

Size here means having the same number of nodes and the same number of edges.

As for the real networks, compute the shortest path lengths for these random networks and present frequency distributions.

6. Determine how well/poorly random networks produce the shortest path distributions of real world networks.

Using whatever tests you like, show how well both the average shortest path length and the full distributions compare between the real network and their random counterparts.

7. Given N labelled nodes and allowing for all possible number of edges m, what's the total number of undirected, unweighted networks we can construct?

How does this number scale with N?

8. Given N labelled nodes and a variable number of m edges, for what value of m do we obtain the largest diversity of networks? And for this m, how does the number of networks scale with N?

9. We've seen that large random networks have essentially no clustering, meaning that locally, random networks are pure branching networks. Nevertheless, a finite, non-zero number of triangles will be present.

   For pure random networks, with connection probability $p = \langle k \rangle/(N-1)$, what is the expected total number of triangles as $N \to \infty$?

10. Repeat the preceding calculation for cycles of length 4 and 5 (triangles are cycles of length 3).

11. Show that the second moment of the Poisson distribution is

$$\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle.$$

   and hence that the variance is $\sigma^2 = \langle k \rangle$.

12. We've figured out in class that for large enough N (and $\langle k \rangle$ fixed), a random network always has a Poisson degree distribution:

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

   where $\lambda = \langle k \rangle$. And as we've discussed, we don't find these networks in the real world (they don't arise due to simple mechanisms). Let's investigate this oddness a little further.

   Compute the expected size of the largest degree in an infinite random network given $\langle k \rangle$ and as a function of increasing sample size N. In other words, in selecting (with replacement) N degrees from a pure Poisson distribution with mean $\langle k \rangle$, what's the expected minimum value of the largest degree $\min k_{\max}$?

   A good way to compute $k_{\max}$ is to equate it to the value for which we expect $1/N$ of our random selections to exceed. (We had a question in 300 along these lines for power-law size distributions.)

   Hint—Of course we'll be using Stirling's Approximation.:
   http://www.youtube.com/watch?v=uK5yakuX59M

13. Generating functions and giant components: In this question, you will use generating functions to obtain a number of results we found in class for standard random networks.

(a) For an infinite standard random network (Erdös-Rényi/ER network) with average degree $\langle k \rangle$, compute the generating function F_P for the degree distribution P_k.

(Recall the degree distribution is Poisson: $P_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$, $k \geq 0$.)

(b) Show that $F_P'(1) = \langle k \rangle$ (as it should).

(c) Using the joyous properties of generating functions, show that the second moment of the degree distribution is $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$.

14. (a) Continuing on from Q1 for infinite standard random networks, find the generating function $F_R(x)$ for the $\{R_k\}$, where R_k is the probability that a node arrived at by following a random direction on a randomly chosen edge has k outgoing edges.

(b) Now, using $F_R(x)$ determine the average number of outgoing edges from a randomly-arrived-at-along-a-random-edge node.

(c) Given your findings above and the condition for a giant component existing in terms of generating functions, what is the condition on $\langle k \rangle$ for a standard random network to have a giant component?

15. (a) Find the generating function for the degree distribution P_k of a finite random network with N nodes and an edge probability of p.

(b) Show that the generating function for the finite ER network tends to the generating function for the infinite one. Do this by taking the limit $N \to \infty$ and $p \to 0$ such that $p(N-1) = \langle k \rangle$ remains constant.

16. (a) Prove that if random variables U and V are distributed over the non-negative integers then the generating function for the random variable $W = U + V$ is

$$F_W(x) = F_U(x) F_V(x).$$

Denote the specific distributions by $\mathbf{Pr}(U = i) = U_i$, $\mathbf{Pr}(V = i) = V_i$, and $\mathbf{Pr}(W = i) = W_i$.

(b) Using the your result in part (a), argue that if

$$W = \sum_{j=1}^{U} V^{(j)}$$

where $V^{(j)} \stackrel{d}{=} V$ then

$$F_W(x) = F_U(F_V(x)).$$

Hint: write down the generating function of probability distribution of $\sum_{j=1}^{k} V^{(j)}$ in terms of $F_V(x)$.

17. (a) Again, given

$$W = \sum_{i=1}^{U} V^{(i)} \text{ with each } V^{(i)} \stackrel{d}{=} V$$

where we know that

$$F_W(x) = F_U\left(F_V(x)\right),$$

determine the mean of W in terms of the means of U and V.

(b) For $W = U + V$, similarly find the mean of W in terms of U and V via generating functions. Your answer should make rather good sense.

18. Consider the family of generalized random networks with

$$P_k = a\delta_{k1} + (1-a)\delta_{k3}$$

where $0 \le a \le 1$.

General note: We worked through the $a = 1/2$ case in class so those notes should be rather helpful.

Determine the following (3 points each for a–d):

(a) i. The distribution of other friends for a node arrived along a randomly chosen direction of a randomly chosen edge, R_k.

ii. The generating function $F_P(x)$.

iii. The generating function $F_R(x)$, both directly from R_k and via $F_R(x) = F_P'(x)/F_P'(1)$.

(b) For which values of a a giant component exists, noting the critical value a_c if any phase transition is present.

(c) i. The generating function $F_\rho(x)$. Note: Do not expand the form you find.

ii. The probability that a random edge leads to a subcomponent of finite size, $F_\rho(1)$.

(d) i. The generating function $F_\pi(x)$.

ii. The fractional size of the largest component $S_1 = 1 - F_\pi(1)$ as a function of a.

19. By expanding $F_\rho(x)$ as a formal power series, find the probabilities that a random edge leads to components of finite size 1, 2, 3, 4, and 5, all as a function of a.

20. Using Python's NetworkX (or similar package in any language), simulate random networks with $N = 10^4$ nodes and determine the fractional size of the giant component as a function of a.

Plot the simulation's output against your theoretical curve determined in the first question.

21. (3 + 3 + 3 + 3)

    Generalize the theory for the previous questions and solve for the same quantities and features in Q1a–Q1d for random networks with:

    $$P_k = a\delta_{k1} + (1-a)\delta_{kk'}$$

    for fixed $k' \geq 2$ with $0 \leq a \leq 1$.

    Modifications:

    You will be able to do Q1a and Q1b exactly.

    Important: Please minimally set up and then solve Q1c and Q1d numerically (only) for $k' = 3, \ldots, 10$.

    Put everything on the same plot.

    (a) i. The distribution of other friends for a node arrived along a randomly chosen direction of a randomly chosen edge, R_k.
        ii. The generating function $F_P(x)$.
        iii. The generating function $F_R(x)$, both directly from R_k and via $F_R(x) = F'_P(x)/F'_P(1)$.

    (b) For which values of a a giant component exists, noting the critical value a_c if any phase transition is present.

    (c) i. The generating function $F_\rho(x)$. Note: Do not expand the form you find.
        ii. The probability that a random edge leads to a subcomponent of finite size, $F_\rho(1)$.

    (d) i. The generating function $F_\pi(x)$.
        ii. The fractional size of the largest component $S_1 = 1 - F_\pi(1)$ as a function of a.

22. Plan: Work through some random bipartite calculations reproducing a few results from the classic Newman *et al.* paper [8]. Our stories are their stars, and our tropes are their movies.

    Please note that we use a different convention for defining certain distributions, not just notation. It's a bit confusing. Okay, it's very confusing.

    Here's a key to help:

Feature:	Our notation:	Newman *et al.* [8]:
First node type, symbol	stories, $\faFilm$	movies, 0
Second node type, symbol	tropes, $\faLightbulbO$	actors, 1
Number of type 1 nodes	$N_{\faFilm}$	M
Number of type 2 nodes	$N_{\faLightbulbO}$	N
Average affiliations of type 1 nodes	$\langle k \rangle_{\faFilm}$	ν
Average affiliations of type 2 nodes	$\langle k \rangle_{\faLightbulbO}$	μ
Affiliation distribution for type 1 nodes	$P_k^{(\faFilm)}$	q_k
Affiliation distribution for type 1 nodes	$P_k^{(\faLightbulbO)}$	p_k
P Generating function for type 1 nodes	$F_{P^{(\faFilm)}}$	g_0
P generating function for type 2 nodes	$F_{P^{(\faLightbulbO)}}$	f_0
R generating function for type 1 nodes	$F_{P^{(\faFilm)}}$	g_1
R generating function for type 2 nodes	$F_{P^{(\faLightbulbO)}}$	f_1
Induced P generating function for type 1 nodes	$F_{P^{(\faFilm)}_{\text{ind}}}$	F_0
Induced P generating function for type 2 nodes	$F_{P^{(\faLightbulbO)}_{\text{ind}}}$	G_0
Induced R generating function for type 1 nodes	$F_{R^{(\faFilm)}_{\text{ind}}}$	F_1
Induced R generating function for type 2 nodes	$F_{R^{(\faLightbulbO)}_{\text{ind}}}$	G_1

Note: You can of course use something simple like a and b instead of the film and lightbulb glyphs. Nevertheless, for notation happiness, feel free to use font awesome and the following structures:

```
\usepackage{fontawesome}

%% random biparite networks

\newcommand{\rbone}{\textnormal{\faFilm}}
\newcommand{\rbtwo}{\textnormal{\faLightbulbO}}

\newcommand{\rboneng}{N_{\rbone}}
\newcommand{\rbtwong}{N_{\rbtwo}}

\newcommand{\Prboneind}{P^{(\rbone)}_{\textnormal{ind},k}}
\newcommand{\Prbtwoind}{P^{(\rbtwo)}_{\textnormal{ind},k}}

\newcommand{\Rrboneind}{R^{(\rbone)}_{\textnormal{ind},k}}
\newcommand{\Rrbtwoind}{R^{(\rbtwo)}_{\textnormal{ind},k}}

\newcommand{\Prboneindplain}{P^{(\rbone)}_{\textnormal{ind}}}
```

```
\newcommand{\Prbtwoindplain}{P^{(\rbtwo)}_{\textnormal{ind}}}

\newcommand{\Rrboneindplain}{R^{(\rbone)}_{\textnormal{ind}}}
\newcommand{\Rrbtwoindplain}{R^{(\rbtwo)}_{\textnormal{ind}}}
```

Show that the triple-triangle clustering coefficient for the induced networks produced by an arbitrary random bipartite affiliation graph are

$$C_2^{(\text{🎞})} = \frac{N_{\text{💡}}}{N_{\text{🎞}}} \frac{F'''_{P^{(\text{💡})}}(1)}{F''_{P^{(\text{🎞})}_{\text{ind}}}(1)}$$

and

$$C_2^{(\text{💡})} = \frac{N_{\text{🎞}}}{N_{\text{💡}}} \frac{F'''_{P^{(\text{🎞})}}(1)}{F''_{P^{(\text{💡})}_{\text{ind}}}(1)}$$

23. (6 + 6 + 6) Consider the following bipartite affiliation graph degree distributions.

    (a) Fixed degree and fixed degree: $k_{\text{🎞}}$ and $k_{\text{💡}}$, both at least 1.

    (b) Poisson (mean $\langle k \rangle_{\text{🎞}}$) and fixed degree ($k_{\text{💡}}$):

    (c) Poisson and Poisson with mean degrees $\langle k \rangle_{\text{🎞}}$ and $\langle k \rangle_{\text{💡}}$.

    For each case, determine these generating functions: $F_{P^{(\text{🎞})}}(x)$, $F_{P^{(\text{💡})}}(x)$, $F_{R^{(\text{🎞})}}(x)$, $F_{R^{(\text{💡})}}(x)$, $F_{P^{(\text{🎞})}_{\text{ind}}}(x)$, $F_{R^{(\text{🎞})}_{\text{ind}}}(x)$, $F_{P^{(\text{💡})}_{\text{ind}}}(x)$, and $F_{R^{(\text{💡})}_{\text{ind}}}(x)$.

24. For the three bipartite graphs given above, determine the condition for a giant component in both induced networks, i.e.,

$$\langle k \rangle_{R,\text{🎞},\text{ind}} \equiv \langle k \rangle_{R,\text{💡},\text{ind}} > 1$$

    where

$$\begin{aligned}\langle k \rangle_{R,\text{🎞},\text{ind}} = \langle k \rangle_{R,\text{💡},\text{ind}} &= \frac{F''_{P^{(\text{💡})}}(1)}{F'_{P^{(\text{💡})}}(1)} \frac{F''_{P^{(\text{🎞})}}(1)}{F'_{P^{(\text{🎞})}}(1)} \\ &= \frac{\langle k(k-1) \rangle_{\text{🎞}}}{\langle k \rangle_{\text{🎞}}} \frac{\langle k(k-1) \rangle_{\text{💡}}}{\langle k \rangle_{\text{💡}}}.\end{aligned}$$

25. Using whatever network package you like, construct random bipartite affiliation networks to reproduce Fig. 7 from [8]:

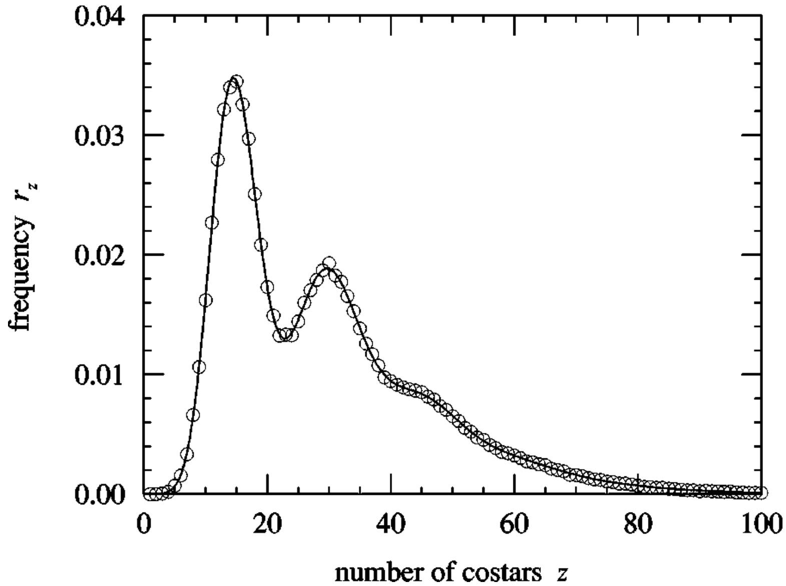


FIG. 7. The frequency distribution of numbers of co-stars of an actor in a bipartite graph with $\mu = 1.5$ and $\nu = 15$. The points are simulation results for $M = 10\,000$ and $N = 100\,000$. The line is the exact solution, Eqs. (89) and (90). The error bars on the numerical results are smaller than the points.

Consider this to be $P^{(\text{💡})}_{\text{ind},k}$, the probability a trope shares appears alongside k other tropes in stories.

Parameters: $N_{\text{🎞}} = 10^4$, $N_{\text{💡}} = 10^5$, $\langle k \rangle_{\text{🎞}} = 15$, and $\langle k \rangle_{\text{💡}} = 1.5$.

26. Plot the induced distribution $P^{(\text{🎞})}_{\text{ind},k}$, the probability a story is connected to k other stories through shared tropes.

27. (optional)

    Derive equation 89 in [8] for the degree distribution:

$$P^{(\text{🎞})}_{\text{ind},k} = \frac{(\langle k \rangle_{\text{💡}})^k}{k!} e^{\langle k \rangle_{\text{🎞}} (e^{-\langle k \rangle_{\text{💡}}} - 1)} \sum_{i=1}^{k} \left\{ {k \atop i} \right\} \left[\langle k \rangle_{\text{🎞}} \, e^{-\langle k \rangle_{\text{💡}}} \right]^i ,$$

    where

$$\left\{ {k \atop i} \right\} = \sum_{j=1}^{i} \frac{(-1)^{i-j}}{j!(i-j)!} j^k$$

    is the Stirling number of the second kind.

28. (optional) Add the theoretical curve obtained above to the plot you generated before that.

29. Data snaring and wrangling:

    Find two (2) interesting, large network data sets online. The networks may be weighted or not, directed or undirected.

    Transform each network's representation into row, column, and weight vectors as per the first assignment. The row vector contains the node at the start of an edge, the column vector the ends, and the weights, well, the weight of the edge.

    Include a one line description for each network along with a link to the data source.

    This time round, if you haven't already, please give NetworkX a shot too.

    Please submit your data via email with the subject heading "CocoNuTS: Network submission for "Hey, could we be judged individually?"".

    In the next assignment, we'll examine all submitted networks. Possibly.

    For questions 30–35:

    Consider the simple spreading mechanism on generalized random networks for which each link has a probability $\beta \leq 1$ of successfully transmitting a disease.

    We assume that this transmission probability is tested only once: either a link will or will not be able to send an infection one way or the other (this is a bond percolation problem). We'll call these edges 'active.'

    Denote the degree distribution of the network as P_k and the corresponding generating function as F_P. In class, we wrote down the probability that a node has k active edges as

    $$\tilde{P}_k = \beta^k \sum_{i=k}^{\infty} \binom{i}{k} (1-\beta)^{i-k} P_i.$$

30. Given a random network with degree distribution P_k, find $F_{\tilde{P}}$, the generating function for $\tilde{P}_k$, in terms of F_P.

31. Find the generating function for $\tilde{R}_k$, the analogous version of R_k, the probability that a random friend has k other friends.

32. For standard random (ER) networks, use your results from the preceding questions to find the critical value of $\langle k \rangle$ above which global spreading occurs.

33. Find an expression connecting the three quantities β, the average degree $\langle k \rangle$, and the size of the giant component $\tilde{S}_1$.

34. What is the slope of the $\tilde{S}_1$ curve near the critical point for ER networks?

35. Using whichever method you find most exciting, plot how $\tilde{S}_1$ depends on $\langle k \rangle$ for $\beta = 1$, $\beta = 0.8$, and $\beta = 0.5$.

36. Using either generating function methods (original) or the physical approach (better) from slides on contagion, reproduce the following pieces from Watts's 2002 paper [9] on global cascades on random networks:

    (a) The cascade windows diagram in Fig. 1.

    (b) The vulnerable and triggering component curves in Fig. 2b.

    Note 1: Only the vulnerable component was determined theoretically in [9]. The slides go further and determine the triggering component's size.

    Note 2: This question is all theory but you will need to solve the second and third problems numerically.

37. Using Gleeson and Calahane's iterative equations below, derive the contagion condition for a vanishing seed by taking the limit $\phi_0 \to 0$ and $t \to \infty$.

$$\phi_{t+1} = \phi_0 + (1-\phi_0) \sum_{k=0}^{\infty} P_k \sum_{j=0}^{k} \binom{k}{j} \theta_t^j (1-\theta_t)^{k-j} B_{kj},$$

$$\theta_{t+1} = \phi_0 + (1-\phi_0) \sum_{k=1}^{\infty} \frac{kP_k}{\langle k \rangle} \sum_{j=0}^{k-1} \binom{k-1}{j} {\theta_t}^j (1-\theta_t)^{k-1-j} B_{kj},$$

where $\theta_0 = \phi_0$, and B_{kj} is the probability that a degree k node becomes active when j of its neighbors are active.

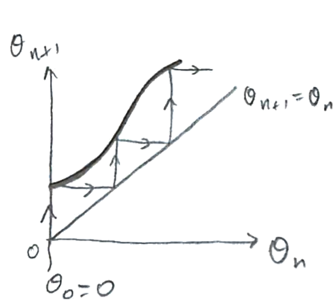
Recall that by contagion condition, we mean the requirements of a random network for macroscopic spreading to occur.

To connect the paper's model and notation to those of our lectures, given a specific response function F and a threshold model, the B_{kj} are given by $B_{kj} = F(j/k)$.

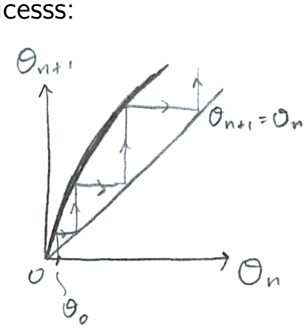
Allow B_{k0} to be arbitrary (i.e., not necessarily 0 as for simple threshold functions).

Here's a graphical hint for the three cases you need to consider as $\theta_0 \to 0$:

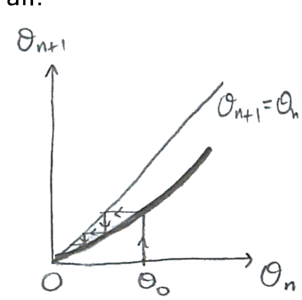
Success:



Sucesss:



Fail:

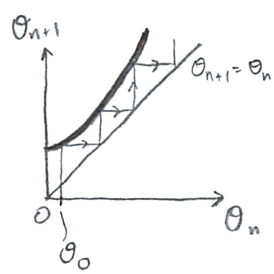


38. Derive equation 4 in Gleeson and Cahalane (2007) [10]:

$$C_\ell = \sum_{k=\ell+1}^{\infty} \sum_{j=0}^{\ell} \binom{k-1}{\ell} \binom{\ell}{j} (-1)^{\ell+j} \frac{kP_k}{\langle k \rangle} F\left(\frac{j}{k}\right).$$

39. (9 pts)

(a) Derive equation 6 in Gleeson and Cahalane (2007), which is a second order approximation to the cascade condition for vanishing seeds.
Here's an example of how this must work:

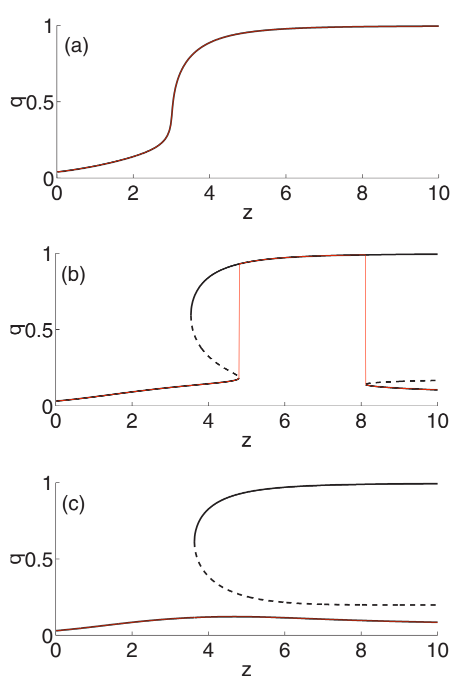


(b) Hence reproduce the dashed analytic curve shown in Figure 1 of their paper.

(c) Explain why there are jumps in the cascade window outline that do not occur at reciprocals of the integers.

40. (6 pts)

(a) By numerically finding the fixed points of $\theta_{t+1} = G(\theta_t; 0)$, reproduce Figure 3 in Gleeson and Cahalane (2007):



(b) Also plot $G(\theta_t; 0)$ for an average threshold $\phi_*(= R)$ of 0.371 for $\langle k \rangle (= z) = 1, 2, 3, \ldots, 10$.

Add the cobweb diagram for a $\phi_0 = 0$ seed.

Do this by creating a recursive plotting script in matlab, for example.

You can use the following Matlab scripts/data as a basis, and most of the work is done. You'll need to improve the plots with some labels, and interpret them properly. The first function calls the other two.

https://pdodds.w3.uvm.edu//share/matlab/Gfunction.m
https://pdodds.w3.uvm.edu//share/matlab/gleeson_{f}ig3_{0}2.mat
https://pdodds.w3.uvm.edu//share/matlab/cobweb3.m

(c) Discuss how the stable points move with $\langle k \rangle$.

Note: $\phi_* = 0.371$ matches plot (b) in Figure 3 of [10].

41. We've figured out in class that for large enough N (and $\langle k \rangle$ fixed), a random network always has a Poisson degree distribution:

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where $\lambda = \langle k \rangle$. And as we've discussed, we don't find these networks in the real world (they don't arise due to simple mechanisms). Let's investigate this oddness a little further.

Compute the expected size of the largest degree in an infinite random network given $\langle k \rangle$ and as a function of increasing sample size N. In other words, in selecting (with replacement) N degrees from a pure Poisson distribution with mean $\langle k \rangle$, what's the expected minimum value of the largest degree $\min k_{\max}$?

A good way to compute $k_{\max}$ is to equate it to the value for which we expect $1/N$ of our random selections to exceed. (We had a question in 300 along these lines for power-law size distributions.)

Hint—Of course we'll be using Stirling's Approximation.:
http://www.youtube.com/watch?v=uK5yakuX59M

42. In 1-d, consider a population density $\rho(x) = cx^{-\gamma}$ for $x \geq 1$ and $\gamma > 2$ (note that $c = \gamma - 1$).

Find the ideal distribution for N sources where N is large.

Hint: draw yourself a clear picture of what's going on.

Hint: guess the form of the locations of the centers and work from there.

Also: Feel free to do some numerics to see how things work.

43. Repeat the above treatment for $\rho(x) = \lambda e^{-\lambda x}$ for $x \geq 0$.

44. Yes, even more on power law size distributions. It's good for you.

    For the probability distribution $P(x) = cx^{-\gamma}$, $0 < a \leq x \leq b$, compute the mean absolute displacement (MAD), which is given by $\langle |X - \langle X \rangle| \rangle$ where $\langle \cdot \rangle$ represents expected value. As always, simplify your expression as much as possible.

    MAD is a more reasonable estimate for the width of a distribution, but we like variance σ^2 because the calculations are much prettier. Really.

45. In the limit of $b \to \infty$, show that MAD asymptotically behave as:

$$\langle |X - \langle X \rangle| \rangle = \frac{2(\gamma - 2)^{(\gamma-3)}}{(\gamma - 1)^{(\gamma-2)}} a.$$

    How does this compare with the behavior of the variance? (See the last question of Assignment todo???.)

46. Using the CCDF and standard linear regression, measure the exponent $\gamma - 1$ as a function of the upper limit of the scaling window, with a fixed lower limit of $k_{\min} = 200$.

    Please plot γ as a function of $k_{\max}$, including 95% confidence intervals.

    Note that the break in scaling should mess things up but we're interested here in how stable the estimate of γ is up until the break point.

    Comment on the stability of γ over variable window sizes.

    Pro Tip: your upper limit values should be distributed evenly in log space.

47. (3 + 3 + 3)

    Estimating the rare:

    Google's raw data is for word frequency $k \geq 200$ so let's deal with that issue now.

    From Assignment 2, we had for word frequency in the range $200 \leq k \leq 10^7$, a fit for the CCDF of

$$N_{\geq k} \sim 3.46 \times 10^8 k^{-0.661},$$

    ignoring errors.

    (a) Using the above fit, create a complete hypothetical N_k by expanding N_k back for $k = 1$ to $k = 199$, and plot the result in double-log space (meaning log-log space).

    (b) Compute the mean and variance of this reconstructed distribution.

    (c) Estimate:

i. the hypothetical fraction of words that appear once out of all words (think of words as organisms here),

ii. the hypothetical total number and fraction of unique words in Google's data set (think at the species level now),

iii. and what fraction of total words are left out of the Google data set by providing only those with counts $k \geq 200$ (back to words as organisms).

48. Starting from here: http://mskcc.convio.net/pdf/cycle_{f}or_{s}urvival/cfs_{c}ancer_{f}act_{s}heet1.pdf, explore the "rare are legion" aspect of heavy-tailed distributions for cancer.

49. Explain the scaling of RPM for engines.

50. **Zombies!**

    (Optional. But taking practical precautions for your survival in the event of a global zombie attack is not optional.)

    Network version of the SZR model:

    Based on the work of Munz *et al.* [11], we will model Zombie attacks on generalized random networks (the paper is here).

    There are three states: S, susceptible, Z, zombie, and, R, removed.

    For the random mixing model studied by Munz et al., the differential equations are

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \theta - \beta S Z - \delta S,$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta S Z + \zeta R - \alpha S Z,$$

$$\text{and } \frac{\mathrm{d}R}{\mathrm{d}t} = \delta S + \alpha S Z - \zeta R,$$

    where

    θ is the birth rate of new susceptibles;

    β is the rate at which susceptibles who bump into zombies become zombies

    δ is the background, non-zombie related death rate for susceptibles;

    ζ is the rate at which the dead (removed) are resurrected as zombies;

    and α is the rate at which susceptibles defeat zombies (through traditional methods shown in movies).

    For our purposes, consider a random network with degree distribution P_k containing completely susceptible individuals and discrete time updates. We'll now think of the parameters above as probabilities, and ignore birth and death processes ($\theta = \delta = 0$).

We'll further assume that if a susceptible takes out a zombie, the latter cannot resurrect. So this means there's a fourth category, let's call it D for definitely dead.

Assume that in each time step, all edges convey interactions, meaning each individual interacts with each of their neighbors.

Under what conditions (P_k and spreading parameters) will local zombification be guaranteed to take off (i.e., grow exponentially, at least in the short term), given one randomly chosen individual becomes the first zombie?

(The long term dynamics will likely be complicated so we will focus on the initial dynamics.)

See http://www.wired.com/wiredscience/2009/08/zombies/ for more information/enjoyment.

51. (12 pts) Consider a family of undirected random networks with degree distribution

$$P_k = c\delta_{k1} + (1-c)\delta_{k3},$$

where δ_{ij} is the Kronecker delta function, and where c is a constant to be determined below. Also allow nodes to be correlated according to the following node-node mixing probabilities.

Conditional probability version, $P(k\,|k'\,)$:

$$P(1\,|\,1) = \frac{1}{2}(1+r),$$
$$P(3\,|\,1) = \frac{1}{2}(1-r),$$
$$P(1\,|\,3) = \frac{1}{2}(1-r),$$
$$\text{and } P(3\,|\,3) = \frac{1}{2}(1+r).$$

where $-1 \le r \le 1$ is the family's tunable parameter.

Newman's correlation probability version:

$$E = [e_{ij}] = \left[\begin{array}{cc} e_{00} & e_{02} \\ e_{20} & e_{22} \end{array}\right] = \frac{1}{4}\left[\begin{array}{cc} (1+r) & (1-r) \\ (1-r) & (1+r) \end{array}\right]$$

where e_{ij} is the probability that a randomly chosen edge connects a node of degree $i+1$ an a node of degree $j+1$, and only the non-zero values of E are shown.

For the following questions, you can use either formulation.

(a) Determine c so that purely disassortative networks are achievable if r is tuned to -1.

(b) Determine which networks in this family have a giant component. In other words, find the values of r for which a giant component exists.

Note which value (or values) of r mark a phase transition.

(c) Analytically determine the size of the giant component as a function of r.

(d) Determine the size of the largest component containing only degree 3 nodes as a function of r.

Hint: allow degree 3 nodes to be always vulnerable ($\beta_{3i} = 1$ for $i = 0$, 1, 2, and 3) and degree 1 nodes to be never vulnerable ($\beta_{1i} = 0$ for $i = 0$ and 1).

52. Spreading on assortative networks: Starting from

$$\theta_{j,t+1} = G_j(\vec{\theta}_t) = \phi_0 + (1 - \phi_0)\times$$

$$\sum_{k=1}^{\infty} \frac{e_{j-1,k-1}}{R_{j-1}} \sum_{i=0}^{k-1} \binom{k-1}{i} \theta_{k,t}^{i} (1 - \theta_{k,t})^{k-1-i} B_{ki}.$$

show the matrix for which we must have the largest eigenvalue greater than 1 for spreading to occur is

$$\frac{\partial G_j(\vec{0})}{\partial \theta_{k,t}} = \frac{e_{j-1,k-1}}{R_{j-1}} (k-1)(\beta_{k1} - \beta_{k0}).$$

53. Show that for uncorrelated networks, i.e, when $e_{jk} = R_j R_k$, the above condition collapses to the standard condition

$$\sum_{k=1}^{\infty} (k-1) \frac{k P_k}{\langle k \rangle} (\beta_{k1} - \beta_{k0}) > 1.$$

54. (3 + 3 + 3) Optional:

Solve Krapivsky-Redner's model for the pure linear attachment kernel $A_k = k$.

Starting point:

$$n_k = \frac{1}{2}(k-1)n_{k-1} - \frac{1}{2}k n_k + \delta_{k1}$$

with $n_0 = 0$.

(a) Determine n_1.

(b) Find a recursion relation for n_k in terms of n_{k-1}.

(c) Now find

$$n_k = \frac{4}{k(k+1)(k+2)}$$

for all k and hence determine γ.

55. (3 + 3) Optional:

From lectures:

(a) Starting from the recursion relation

$$n_k = \frac{A_{k-1}}{\mu + A_k} n_{k-1},$$

and $n_1 = \mu/(\mu + A_1)$, show that the expression for n_k for the Krapivsky-Redner model with an asymptotically linear attachment kernel A_k is:

$$\frac{\mu}{A_k} \prod_{j=1}^{k} \frac{1}{1 + \frac{\mu}{A_j}}.$$

(b) Now show that if $A_k \to k$ for $k \to \infty$ (or for large k), we obtain $n_k \to k^{-\mu-1}$.

56. (3 + 3 + 3) Optional:

From lectures, complete the analysis for the Krapivsky-Redner model with attachment kernel:

$$A_1 = \alpha \text{ and } A_k = k \text{ for } k \geq 2.$$

Find the scaling exponent $\gamma = \mu + 1$ by finding μ. From lectures, we assumed a linear growth in the sum of the attachment kernel weights $\mu t = \sum_{k=1}^{\infty} N_k(t) A_k$, with $\mu = 2$ for the standard kernel $A_k = k$.

We arrived at this expression for μ which you can use as your starting point:

$$1 = \sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{1}{1 + \frac{\mu}{A_j}}$$

(a) Show that the above expression leads to

$$\frac{\mu}{\alpha} = \sum_{k=2}^{\infty} \frac{\Gamma(k+1)\Gamma(2+\mu)}{\Gamma(k+\mu+1)}$$

Hint: you'll want to separate out the $j = 1$ case for which $A_j = \alpha$.

(b) Now use result that [7]

$$\sum_{k=2}^{\infty} \frac{\Gamma(a+k)}{\Gamma(b+k)} = \frac{\Gamma(a+2)}{(b-a-1)\Gamma(b+1)}$$

to find the connection

$$\mu(\mu-1) = 2\alpha,$$

and show this leads to

$$\mu = \frac{1+\sqrt{1+8\alpha}}{2}.$$

(c) Interpret how varying α affects the exponent γ, explaining why $\alpha < 1$ and $\alpha > 1$ lead to the particular values of γ that they do.

57. (10 pts) What is the clustering coefficient C for a standard random network with degree distribution P_k? Compute C for the following two cases:

(a) N is finite and links between nodes exist with probability p.

(b) The random network is infinite with mean degree $\langle k \rangle = z$.

Use the definition $C = \frac{3\#\text{triangles}}{\#\text{triples}}$, or equivalently, that C is the probability that if a is connected to b and c, then b and c are connected.

(c) What's the interpretation for the local structure of infinite random networks given your answer to (b)?

58. (25 pts) Generating functions and giant components. In this question, you will use generating functions to obtain a number of results we found in class for standard random networks.

(a) For an infinite standard random network with average degree $\langle k \rangle = z$, compute the generating function for the degree distribution P_k.

(Recall the degree distribution is Poisson: $P_k = e^{-z} z^k / k!$, $k \geq 0$.)

(b) Using your answer to (a) and the joyous properties of generating functions, show that $\langle k \rangle = z$ and that the degree variance is $\langle k^2 \rangle = z^2 + z$.

(c) Find the generating function for the $\{\tilde{q}_k\}$, where q_k is the probability that a node arrived at by following a random direction on a randomly chosen edge has k outgoing edges.

(d) Using your result for (c), determine the average number of outgoing edges from a randomly-arrived-at-along-a-random-edge node.

(e) Based on (d), what is the condition on z for a standard random network to have a giant component?

(Hint: you need to find for what values of z, a randomly chosen neighbor will, on average, have at least one other neighbor.)

59. (15 pts) In Krapivsky and Redner's treatment of growing random network for linear attachment kernels, they assumed $\sum_{k=1}^{\infty} n_k A_k = \mu t$ and found that μ must be such that

$$1 = \sum_{k=1}^{\infty} \prod_{j=1}^{k} \left(1 + \frac{\mu}{A_j}\right)^{-1}.$$

(a) Show that when the attachment kernel is purely linear, $A_j = j$, and when $\mu = 2$, the above equation above is satisfied.

(b) Bonus question territory: Krapivsky and Redner also looked at the specific attachment kernel $A_1 = \alpha$ and $A_k = k$ for $k > 1$, where $\alpha > 0$. They determined that the resulting degree distribution has a power-law tale obeying $k^{-\gamma}$ where $\gamma = (3 + \sqrt{1 + 8\alpha})/2$.

Using this modified linear attachment kernel, show that $\mu(\mu - 1) = 2\alpha$.

60. (20 pts)

Aspects of Kleinberg's search problem in one dimension:

Consider N nodes connected in a 1-d line graph (i.e., a sequence of N nodes lying on a line, with adjacent nodes connected), labelled $i = 1$ to N.

Take our starting node to be at one end of the line, say $i = 1$, and the target node to be at the other end, $i = N$.

Let node $i = 1$ have exactly one long distance link (i.e., a shortcut link).

In attempting to construct a searchable network, we add a link from our start node $i = 1$ to another node $j = 2, \ldots, N$ with probability $cr^{-\alpha}$, where c is a constant of proportionality and $r = j - i$ is the distance between i and j. (Normally, we add links to all nodes but for this question, we're only interested in what happens with the first step from $i = 1$.)

(a) Compute the constant of proportionality c (Hint: the sum over the probabilities of attaching to all other nodes must be unity; use an integral approximation again.)

(b) Show that for $\alpha = 1$, the chance of the link from node $i = 1$ reaching a node at position $j \geq N/2$ is on the order of $1/\ln N$. (This effectively means that by moving along the line starting at i, we should find a shortcut to the other half of the line within a factor of $\ln N$ steps. This is pretty good.)

(c) For $\alpha > 1$, show that the probability of i having a shortcut to the other half of the line decays as an inverse power of N. (This means that our shortcut is likely too close to i and won't help us jump to the other half of the line.)

(d) If $\alpha < 1$, our shortcut will link to the other half of the line with a finite, constant probability, independent of N for large N. So what's the drawback here?

61. More of a note:

- Newman[12]:
$$C_3 = \frac{6 \times \#\text{triangles}}{\#\text{ordered pairs}}$$
- Now count each triple twice
- Same as C_2 but interpretation is different
- Probability that a friend of i's friend is also i's friend.

- C_1 = probability that two friends of a randomly chosen node are connected
- C_2 = probability that two nodes are connected given they have a friend in common.
- $C_3 (= C_2)$ = probability that a node's friend of a friend is also a friend of that node.

- For sparse networks, C_1 tends to discount highly connected nodes.
- While C_1 is a measure of clustering, it doesn't quite as simple interpretation as C_2.
- Some variability in which measure is used in the literature.
- Not always clear which one is being used...

62. Generating functions and giant components: In this question, you will use generating functions to obtain a number of results we found in class for standard random networks.

(a) For an infinite standard random network (Erdös-Rényi/ER network) with average degree $\langle k \rangle$, compute the generating function F_P for the degree distribution P_k.

(Recall the degree distribution is Poisson: $P_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$, $k \geq 0$.)

(b) Show that $F'_P(1) = \langle k \rangle$ (as it should).

(c) Using the joyous properties of generating functions, show that the second moment of the degree distribution is $\langle k^2 \rangle = \langle k \rangle^2 + \langle k \rangle$.

(d) Find the generating function for the degree distribution P_k of a finite random network with N nodes and an edge probability of p.

(e) Show that the generating function for the finite ER network tends to the generating function for the infinite one. Do this by taking the limit $N \to \infty$ and $p \to 0$ such that $p(N-1) = \langle k \rangle$ remains constant.

63. (a) Continuing on from Q1 for infinite standard random networks, find the generating function $F_R(x)$ for the $\{R_k\}$, where R_k is the probability that a node arrived at by following a random direction on a randomly chosen edge has k outgoing edges.

    (b) Now determine the average number of outgoing edges from a randomly-arrived-at-along-a-random-edge node.

    (c) Given your findings above, what is the condition on $\langle k \rangle$ for a standard random network to have a giant component?

    (Hint: you need to find for what values of $\langle k \rangle$, a randomly chosen neighbor will, on average, have at least one other neighbor.)

64. Consider the simple spreading mechanism on generalized random networks for which each link has a probability $\beta \leq 1$ of successfully transmitting a disease.

    We assume that this transmission probability is tested only once: either a link will or will not be able to send an infection one way or the other (this is a bond percolation problem). We'll call these edges 'active.'

    Denote the degree distribution of the network as P_k and the corresponding generating function as F_P. In class, we wrote down the probability that a node has k active edges as

$$P'_k = \beta^k \sum_{i=k}^{\infty} \binom{i}{k} (1-\beta)^{i-k} P_i.$$

    (a) Given a random network with degree distribution P_k, find $F_{P'}$, the generating function for P'_k, in terms of F_P.

    (b) Find the generating function for R'_k, the analogous version of R_k, the probability that a random friend has k other friends.

65. (a) For standard random networks, use your results for Q3 to find an expression connecting β, the average degree $\langle k \rangle$, and the size of the giant component S'_1.

    (b) What is slope of the S'_1 curve near the critical point for ER networks?

    (c) Using whichever method you find most exciting, plot how S'_1 depends on $\langle k \rangle$ for $\beta = 1$, $\beta = 0.8$, and $\beta = 0.5$.

66. Consider a network with a degree distribution that obeys a power law and is otherwise random.

    Assume that the network is drawn from an ensemble of networks which have N nodes whose degrees are drawn from the probability distribution $P_k = ck^{-\gamma}$ where $k \geq 1$ and $2 < \gamma < 3$.

    (a) Estimate $\min k_{\max}$, the approximate minimum of the largest degree in the network, finding how it depends on N. (Hint: we expect on the order of 1 of the N nodes to have a degree of $\min k_{\max}$ or greater.)

    (b) Determine the average degree of nodes with degree $k \geq \min k_{\max}$ to find how the expected value of $k_{\max}$ scales with N.

Repeats:

I. Supply networks and allometry:

Consider a set of rectangular areas with side lengths L_1 and L_2 such that $L_1 \propto A^{\gamma_1}$ $L_2 \propto A^{\gamma_2}$ where A is area and $\gamma_1 + \gamma_2 = 1$. Assume $\gamma_1 > \gamma_2$.

Now imagine that material has to be distributed from a central source in each of these areas to sinks distributed with density $\rho(A)$, and that these sinks draw the same amount of material per unit time independent of L_1 and L_2.

1. Find an exact form for how the volume of the most efficient distribution network scales with overall area $A = L_1 L_2$. (Hint: you will have to set up a double integration over the rectangle.)

2. If network volume must remain a constant fraction of overall area, determine the maximal scaling of sink density ρ with A.

II. Size-density law:

In two dimensions, the size-density law for distributed source density $D(\vec{x})$ given a sink density $\rho(\vec{x})$ states that $D \propto \rho^{2/3}$. We showed in class that an approximate argument that minimizes the average distance between sinks and nearest sources gives the 2/3 exponent.

1. Repeat this argument for the d-dimensional case and find the general form of the exponent β in $D \propto \rho^{\beta}$.

- We will explore real networks throughout the course performing some key measurements introduced in Principles of Complex Systems.
- you are encouraged to use Python (along with, for example, NetworkX or graph-tools).
- Data is available in two compressed formats:
  - Matlab + text (tgz): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.tgz
  - Matlab + text (zip): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.zip

  and can also be found on the course website (helpfully) under data.
- The main Matlab file containing everything is networkdata_combined.mat.
- For directed networks, the ijth entry of the adjacency matrix represents the weight of the link from node i to node j. Adjacency matrices for undirected networks are symmetric.
- For all questions below, treat each network as undirected unless otherwise instructed.
- For this assignment, convert all weights on links to 1, if the network is weighted.
- You do not have to use Matlab for your basic analyses. Python would be a preferred route for many.
- The supplied text versions may be of use for visualization using gml.
- The Matlab command spy will give you a quick plot of a sparse adjacency matrix.
- Real data sets used here are taken from Mark Newman's compilation (and linked-to sites) at http://www-personal.umich.edu/~mejn/netdata/.

1. Record in a table the following basic characteristics:
   - N, the number of nodes;
   - m, the total number of links;
   - Whether the network is undirected or directed based on the symmetry of the adjacency matrix;
   - $\langle k \rangle$, the average degree ($\langle k_{\rm in} \rangle$ and $\langle k_{\rm out} \rangle$ if the network is directed);

- The maximum degree $k^{\max}$ (for both out-degree and in-degree if the network is directed);
- The minimum degree $k^{\min}$ (for both out-degree and in-degree if the network is directed).

2. (3+3)

   (a) Plot the degree distribution P_k as a function of k. In the case that P_k versus k is uninformative, also produce plots that are clarifying. For example, $\log_{10} P_k$ versus $\log_{10} k$.
   (Note: Always use base 10.)

   (b) See if you can characterize the distributions you find (e.g., exponential, power law, etc.).

3. Measure the clustering coefficient C_2 where

$$C_2 = \frac{3 \times \#\text{triangles}}{\#\text{triples}}.$$

   For directed networks, transform them into undirected ones first.

   One approach is to compute C_2 as

$$C_2 = \frac{3 \times \frac{1}{6}\mathrm{Tr}A^3}{\frac{1}{2}\left(\sum_{ij}[A^2]_{ij} - \mathrm{Tr}A^2\right)}.$$

   Note: avoiding computing A^3 is important and can be done.

- We will explore real networks throughout the course performing some key measurements introduced in Principles of Complex Systems.
- you are encouraged to use Python (along with, for example, NetworkX or graph-tools).
- Data is available in two compressed formats:
  - Matlab + text (tgz): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.tgz
  - Matlab + text (zip): https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/data/303complexnetworks-data-package.zip

  and can also be found on the course website (helpfully) under data.
- The main Matlab file containing everything is networkdata_combined.mat.

- For directed networks, the ijth entry of the adjacency matrix represents the weight of the link from node i to node j. Adjacency matrices for undirected networks are symmetric.
- For all questions below, treat each network as undirected unless otherwise instructed.
- For this assignment, convert all weights on links to 1, if the network is weighted.
- You do not have to use Matlab for your basic analyses. Python would be a preferred route for many.
- The supplied text versions may be of use for visualization using gml.
- The Matlab command spy will give you a quick plot of a sparse adjacency matrix.
- Real data sets used here are taken from Mark Newman's compilation (and linked-to sites) at http://www-personal.umich.edu/~mejn/netdata/.

1. Okay, let's get back to the 6 networks we explored in the first assignment. Questions 2 through 4 will focus on them.

   Measure the degree-degree assortativity. This is the standard Pearson correlation coefficient and the focus is on links, and then the nodes at the end of each link.

   For undirected networks, we need to think about how we choose the ordering of an edge's two degrees when we perform the correlation. Which degree goes first? Or should we include both orderings? How about randomly choosing the ordering? Does it matter?

   For directed networks, various correlations are possible (in-in, in-out, etc.). For this question, measure the correlation of the in-degree of the source node and the out-degree of the destination node for each link.

2. Produce plots of the adjacency matrices.

3. Using a network visualization tool of your choice, produce plots of the networks (if possible, depending on size).

   For the smaller ones, please label the nodes numerically.

4. (3 + 3)

   Consider a modified version of the Barabàsi-Albert (BA) model [6] where two possible mechanisms are now in play. As in the original model, start with m_0 nodes at time $t = 0$. Let's make these initial guys connected such that each has degree 1. The two mechanisms are:

M1: With probability p, a new node of degree 1 is added to the network. At time $t+1$, a node connects to an existing node j with probability

$$P(\text{connect to node } j) = \frac{k_j}{\sum_{i=1}^{N(t)} k_i} \tag{3}$$

where k_j is the degree of node j and $N(t)$ is the number of nodes in the system at time t.

M2: With probability $q = 1 - p$, a randomly chosen node adds a new edge, connecting to node j with the same preferential attachment probability as above.

Note that in the limit $q = 0$, we retrieve the original BA model (with the difference that we are adding one link at a time rather than m here).

In the long time limit $t \to \infty$, what is the expected form of the degree distribution P_k?

Do we move out of the original model's universality class?

Different analytic approaches are possible including a modification of the BA paper, or a Simon-like one (see also Krapivsky and Redner [7]).

Hint: You can attempt to solve the problem exactly and you'll find an integrating factor story.

Another hint, moment of mercy: Approximate the differential equation by considering large t (this will simplify the denominators).

(3 points for set up, 3 for solving.)

5. Optional:

   Watch "Remedial Chaos Theory."

   Community, S3E04.

   https://en.wikipedia.org/wiki/Remedial_Chaos_Theory

6. Tokunaga's law is statistical but we can consider a rigid version. Take $T_1 = 2$ and $R_T = 2$ and draw an example network of order $\Omega = 4$ with these parameters.

   Please take some effort to make your network look somewhat like a river network.

7. Show $R_s = R_\ell$. In other words show that Horton's law of stream segments matches that of main stream lengths, and do this by showing they imply each other.

8. Tokunaga's law implies Horton's laws:

   In lectures, we established the following:

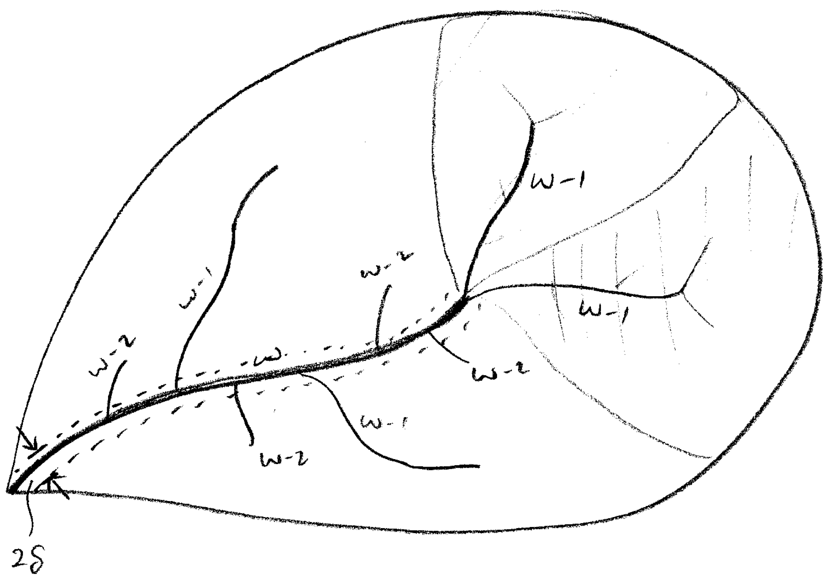
$$n_\omega = \underbrace{2\, n_{\omega+1}}_{\text{generation}} + \sum_{\omega'=\omega+1}^{\Omega} \underbrace{T_{\omega'-\omega}\, n_{\omega'}}_{\text{absorption}}$$

   From here, derive Horton's law for stream numbers: $n_\omega / n_{\omega+1} = R_n$, where $R_n > 1$ and is independent of ω, and find R_n in terms of Tokunaga's two parameters T_1 and R_T.

9. Show $R_n = R_a$ by using Tokunaga's law to find the average area of an order ω basin, $\langle a \rangle_\omega$, in terms of the average area of basins of order 1 to $\omega - 1$.

   (In lectures, we use Horton's laws to roughly demonstrate this result.)

   Here's the set up:



   Using the Tokunaga picture, we see a basin of order ω can be broken down into non-overlapping sub-basins.

   Connect $\langle a \rangle_\omega$ to the average areas of basins of lower orders as follows:

$$\langle a \rangle_\omega = 2 \langle a \rangle_{\omega-1} + \sum_{\omega'=1}^{\omega-1} T_{\omega,\omega'} \langle a \rangle_{\omega'} + 2\delta \langle s \rangle_\omega .$$

   The first term on the right hand side corresponds to the two 'generating' streams of order $\omega - 1$. The second term (the sum) accounts for side streams entering the sole order ω stream segment in the basin. And the last term gives the contribution of 'overland flow,' i.e., flow that does not arrive in the main stream segment through a stream. The length scale δ is the typical distance from stream to ridge.

10. For river networks, basin areas are distributed according to $P(a) \propto a^{-\tau}$.

    Determine the exponent τ in terms of the Horton ratios R_n and R_s.

    Guide:

    Follow the same procedure shown in lectures for $P(\ell) \propto \ell^{-\gamma}$.

    In class, we derived $P(\ell) \propto \ell^{-\gamma}$ starting from Horton's laws (see the section of scaling relations in the slides on Branching Networks II. In doing so, we started with the following observation:

    $$P_{>}(\ell_\omega) = \frac{N_{>}(\ell_\omega; \Delta)}{N_{>}(0; \Delta)}$$

    where $N_{>}(\ell_\omega; \Delta)$ was the number of sites with main stream length $> \ell_\omega$.

    Now, we can equally well use the right hand side to count the number of sites with drainage area exceeding a_ω. So,

    $$P_{>}(a_\omega) \propto \left(\frac{R_n}{R_s}\right)^{-\omega} = e^{-\omega \ln(R_n/R_s)}.$$

    Our task is now to wrangle the right hand side so that we see it in terms of a_ω.

11. Derive Murray's law.

    Per lectures, find the minimum rate of energy expenditure working from the assertion that:

    $$P = P_{\mathrm{drag}} + P_{\mathrm{met}} = \Phi^2 \frac{8\eta\ell}{\pi r^4} + c_{\mathrm{met}} r^2 \ell,$$

    where met stands for metabolic.

    We are interested in how P varies with the tube radius r.

    Per lectures, we defined the 'parent' branch's radius as r_{parent}, and the 'offspring' branches as having radii $r_{\mathrm{offspring1}}$ and $r_{\mathrm{offspring2}}$ (which need not be the same).

    Show that minimizing energy expenditure leads to $r^3_{\mathrm{parent}} = r^3_{\mathrm{offspring1}} + r^3_{\mathrm{offspring2}}$.

    Note that in the LaTeX settings for assignments, various derivative notations are included.

    Here, you will want to use partial derivatives, and here's a start.

    Note the code-like formatting as expounded on here. Far easier to create, edit, debug, read.

$$\boxed{\frac{\partial P}{\partial r} = \frac{\partial}{\partial r}\left(\Phi^2 \frac{8\eta\ell}{\pi r^4} + c_{\text{met}} r^2 \ell\right)}$$

```
$
\partialdiff{P}{r}
=
\partialdiff{}{r}
\left(
\Phi^{2}
\frac{
  8 \eta \ell
}{
  \pi r^{4}
}
+
c_{\textnormal{met}}
r^{2}
\ell
\right)
$
```

12. Derive the equivalent of Murray's law for branching networks where material moves by diffusion. Perhaps surprisingly, this connects the inner workings of insects, electrical networks, and search on networks.

    For diffusion, the impedance of a vessel is now $Z = c_{\text{diff}}\ell r^{-2}$ where c_{diff} is a constant, ℓ is vessel length, and r is vessel radius.

    In terms of general impedance, the expression for the rate of energy expenditure is:

    $$P = P_{\text{drag}} + P_{\text{met}} = \Phi^2 Z + c_{\text{met}} r^2 \ell.$$

13. Now derive the generalized version of Murray's law for a generalized impedance $Z = c_{\text{imp}}\ell r^{-2\alpha}$, where c_{imp} is a general impedance constant, ℓ is vessel length, and r is vessel radius.

    We can assume $\alpha > 0$ as impedance should decrease with wider vessels.

    We choose $r^{-2\alpha}$ because cross sectional area πr^2 can be considered the essential parameter here, and because we skipped to the end of the book and decided to rewrite the start.

14. Murray's law for real data.

    See if you can track down a data set for real branching networks where Murray's law might reasonably apply, and then test how well Murray's law holds up.

    Could be blood vessels, trees, ... [13, 14, 15].

    As always, you are welcome to collaborate. Feel free to share data sets on Teams.

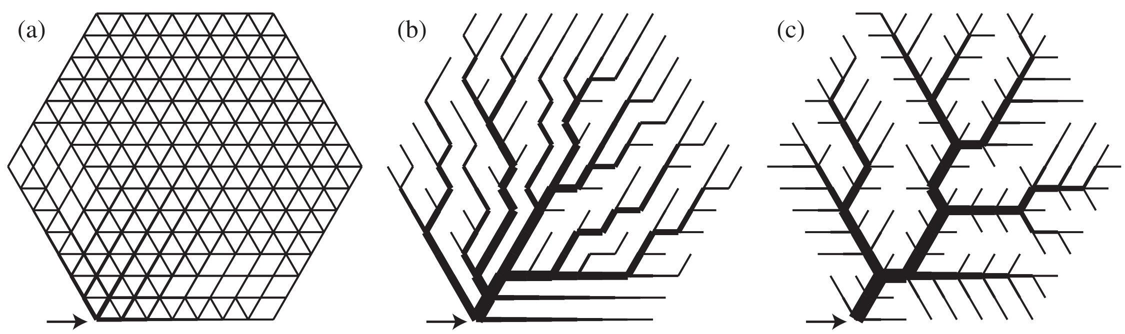
15. (3 + 3)

    Let's start on trying to reproduce reproduce Bohn and Magnasco's Figs. 2a and 2b in [16].

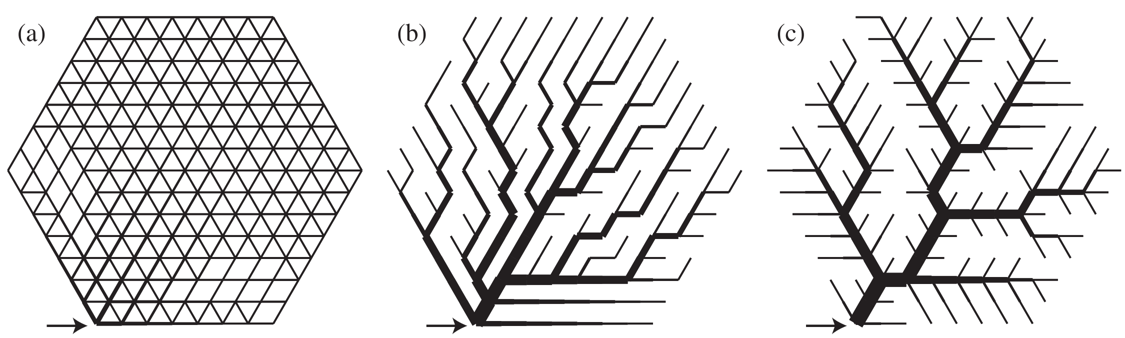
    A profound physical result. For movement of stuff, when should networks exist?

    Preliminary work:

    - Construct an adjacency matrix for the underlying hexagonal lattice where the side number of nodes is a variable n.
    - Plot the $n = 8$ version to match with the grids underlying the figures below.



16. (3 + 3) Reproduce Bohn and Magnasco's Figs. 2a and 2b in [16]:



    Steps are given below but please read through the paper to understand how they set things up.

    The full team is encouraged to work together on Teams.

    (a) Done (previous assignment): Construct an adjacency matrix $\mathbf{A}$ representing the hexagonal lattice used in [16]. Plot this adjacency matrix.

    (b) Run a minimization procedure to construct Figs. 2a and 2b which correspond to $\gamma = 2$ and $\gamma = 1/2$. Steps:

i. Set each link's length to 1 (the d_{kl}). The goal then reduces to minimizing the cost
$$C = \sum_{k,l} |I_{kl}|^{\Gamma}$$
where I_{kl} is the current on link kl and $\Gamma = 2\gamma/(\gamma+1)$.

ii. Place a current source of nominal size i_0 at one node (as indicated in Fig. 2 above).

iii. All other nodes are sinks, drawing a current of
$$i_k = -\frac{i_0}{N_{\text{nodes}-1}}.$$

iv. Suggest setting $i_0 = 1000$ (arbitrary but useful value given the size of the network).

v. Generate an initial set of random conductances for each link, the $\{\kappa_{kl}\}$. From the paper, these must sum to some global constraint as
$$K = \left(\sum_{k,l} \kappa_{kl}^{\gamma}\right)^{1/\gamma}.$$
This constraint is meant to represent a limitation on the amount of material that can be used to build the network.
Note: There seems to be no reason not to set $K = 1$. However, taking the initial value of K determined by the initial set of random conductances would work.
To our notational peril, we now have a lot of k types on deck.

vi. Solve the following to determine the potential U at each node, and hence the current on each link using:
$$i_k = \sum_{l} \kappa_{kl}(U_k - U_l),$$
and then
$$I_{kl} = \kappa_{kl}(U_l - U_k).$$
Note: the paper erroneously has $I_{kl} = R_{kl}(U_l - U_k)$ below equation 4; there are a few other instances of similar miswritings of R_{kl} instead of κ_{kl}.

vii. Now, use scaling in equation (10) to compute a new set of $\{\kappa_{kl}\}$ from the I_{kl}. Everything boils down to
$$\kappa_{kl} \propto |I_{kl}|^{-(\Gamma-2)},$$
where the constant of proportionality is determined by again making sure $K^{\gamma} = \sum_{k,l} \kappa_{kl}^{\gamma}$.

Some help—Let's sort out the key equation:

$$i_k = \sum_l \kappa_{kl}(U_k - U_l).$$

Each time we loop around through this equation, we know the i_k and the κ_{kl} and must determine the U_k. In matrixology, we love $A\vec{x} = \vec{b}$ problems so let's see if we can fashion one:

$$\begin{aligned}
i_k &= \sum_l \kappa_{kl}(U_k - U_l) \\
&= \sum_l \kappa_{kl} U_k - \sum_l \kappa_{kl} U_l \\
&= U_k \sum_l \kappa_{kl} - \sum_l \mathbf{K}_{kl} U_l \\
&= \lambda_k U_k - [\mathbf{K}\vec{U}]_k
\end{aligned}$$

where we have set $\lambda_k = \sum_l \kappa_{kl}$, the sum of the kth row of the matrix $\mathbf{K}$. We now construct a diagonal matrix Λ with the λ_k on the diagonal, and obtain:

$$\vec{i} = (\Lambda - \mathbf{K})\,\vec{U}.$$

The above is in the form $A\vec{x} = \vec{b}$ so we can solve for $\vec{U}$ using standard features of R, Matlab, Python, ... (hopefully).

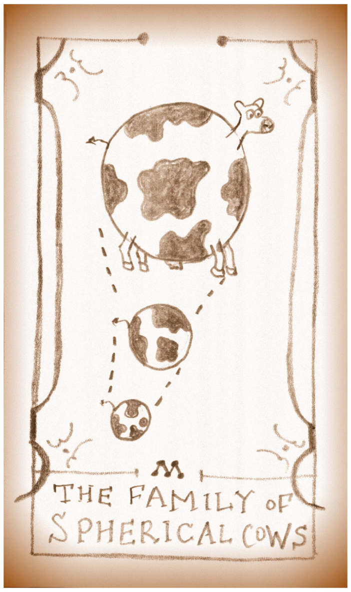
17. Surface area of allometrically growing ~~Love~~Minecraftian organisms:

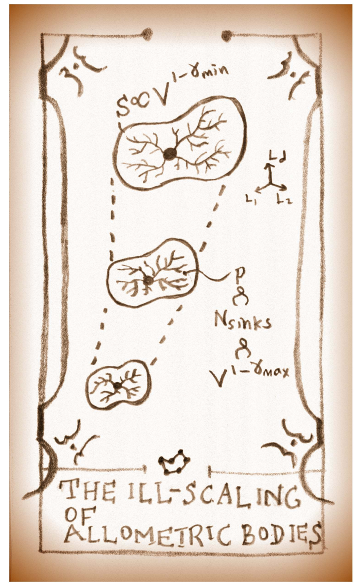
    Let's consider animals as parallelepipeds (e.g., the well known box cow), with dimensions L_1, L_2, and L_3 and volume $V = L_1 \times L_2 \times L_3$.

    Let's assume length L_i scales with volume as $L_i = c_i^{-1} V^{\gamma_i}$ where the exponents satisfy $\gamma_1 + \gamma_2 + \gamma_3 = 1$ and the c_i are prefactors such that $c_1 \times c_2 \times c_3 = 1$. Let's also arrange our organisms so that $\gamma_1 \geq \gamma_2 \geq \gamma_3$.

    (a) Show that the scalings $L_i = c_i^{-1} V^{\gamma_i}$ mean that indeed $L_1 \times L_2 \times L_3 = V$.
    (b) Write down the γ_i corresponding to isometric scaling.
    (c) Calculate the surface area S of our imaginary blockular beings for general allometric scaling of the sides.
    (d) Show how S behaves as V becomes large (i.e., which term(s) dominate).
    (e) Which sets of γ_i give the fastest and slowest possible scaling of S as a function of V?

Relevant tarot cards, for your consideration:





References

[1] D. J. Watts and S. J. Strogatz. Collective dynamics of ‘small-world’ networks. Nature, 393:440–442, 1998. pdf

[2] C. E. Shannon. A mathematical theory of communication. The Bell System Tech. J., 27:379–423,623–656, 1948. pdf

[3] L. Jost. Entropy and diversity. Oikos, 113:363–375, 2006. pdf

[4] P. W. Anderson. More is different. Science, 177(4047):393–396, 1972. pdf

[5] R. M. May. Will a large complex system be stable? Nature, 238:413–414, 1972. pdf

[6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 286:509–511, 1999. pdf

[7] P. L. Krapivsky and S. Redner. Organization of growing random networks. Phys. Rev. E, 63:066123, 2001. pdf

[8] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E, 64:026118, 2001. pdf

[9] D. J. Watts. A simple model of global cascades on random networks. Proc. Natl. Acad. Sci., 99(9):5766–5771, 2002. pdf

[10] J. P. Gleeson and D. J. Cahalane. Seed size strongly affects cascades on random networks. Phys. Rev. E, 75:056103, 2007. pdf

[11] P. Munz, I. Hudea, J. Imad, and R. J. Smith? When zombies attack!: Mathematical modelling of an outbreak of zombie infection. In J. M. Tchuenche and C. Chiyaka, editors, Infectious Disease Modelling Research Progress, pages 133–150. Nova Science Publishers, Inc., 2009. pdf

[12] M. E. J. Newman. The structure and function of complex networks. SIAM Rev., 45(2):167–256, 2003. pdf

[13] T. F. Sherman. On connecting large vessels to small. The meaning of Murray's law. The Journal of general physiology, 78(4):431–453, 1981. pdf

[14] K. A. McCulloh, J. S. Sperry, and F. R. Adler. Water transport in plants obeys Murray's law. Nature, 421:939–942, 2003. pdf

[15] K. A. McCulloh, J. S. Sperry, and F. R. Adler. Murray's law and the hydraulic vs mechanical functioning of wood. Functional Ecology, 18:931–938, 2004. pdf

[16] S. Bohn and M. O. Magnasco. Structure, scaling, and phase transition in the optimal transport network. Phys. Rev. Lett., 98:088702, 2007. pdf