The Evolutionary Dynamics of Incubation Periods

Bertrand Ottino-Löffler
Advisor: Steve Strogatz, Co-Author: Jacob Scott

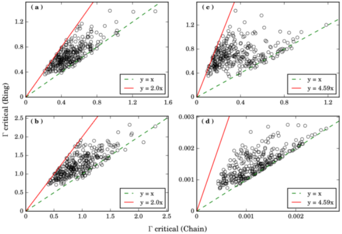
Cornell University

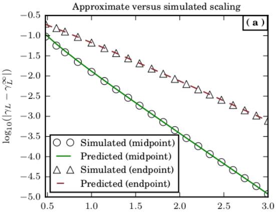
01/11/18

Things I have done

Previously:

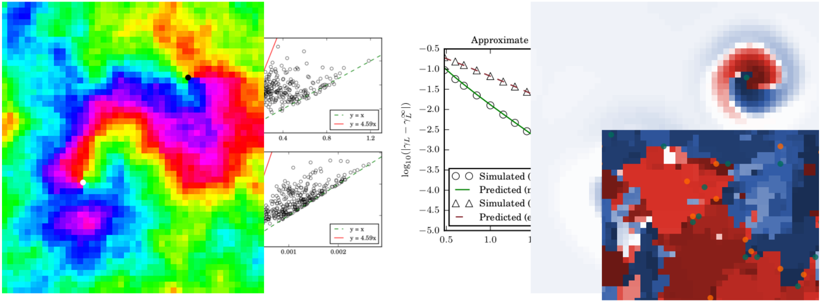
Previously:
Asymptotic Analysis!





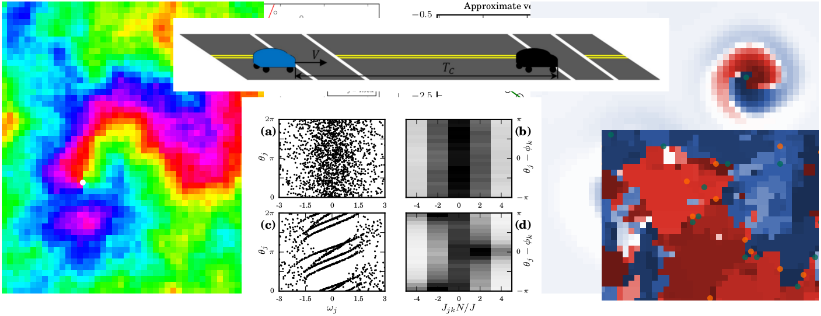
Things I have done

Previously:
Asymptotic Analysis! Coupled Oscillators!



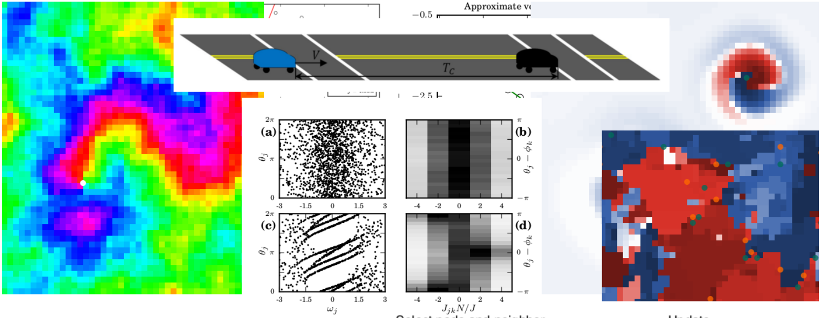
Things I have done

Previously:
Asymptotic Analysis! Coupled Oscillators! Dynamical Systems!

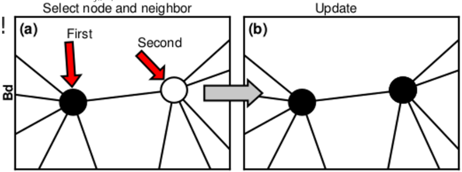


Things I have done

Previously:
Asymptotic Analysis! Coupled Oscillators! Dynamical Systems!



Today: Evolutionary dynamics!
Probability!
Stochastic Processes!
Typhoid!



This talk

eLIFE
elifesciences.org

RESEARCH ARTICLE

Evolutionary dynamics of incubation periods

Bertrand Ottino-Loffler[1], Jacob G Scott[2,3], Steven H Strogatz[1]*

[1]Center for Applied Mathematics, Cornell University, Ithaca, United States; [2]Department of Translational Hematology and Oncology Research , Cleveland Clinic, Cleveland, United States; [3]Department of Radiation Oncology, Cleveland Clinic, Cleveland, United States

Personal: people.cam.cornell.edu/∼bjo34/
Twitter: @OttinoLoffler

Outline

1. Incubation periods, Sartwell's law
2. Moran models, Bd and Db
3. Infinite r, complete graph and star graph
4. Lattices, critical dimensions
5. Resilience under relaxation
6. Summary and closing

The Incubation Period

Definition

The **Incubation Period** of a disease is defined to be the time between first exposure to a contagion and observation of first symptoms.

The Incubation Period

Then incubation period of a disease is important for...

- ... individual diagnosis.
- ... deciding quarantine policy.
- ... predicting secondary outbreaks of epidemics.

The Incubation Period

NINETY-THREE PERSONS INFECTED BY A TYPHOID CARRIER AT A PUBLIC DINNER

WILBUR A. SAWYER, M.D.
Director of the Hygienic Laboratory of the California State Board of Health
BERKELEY, CAL.

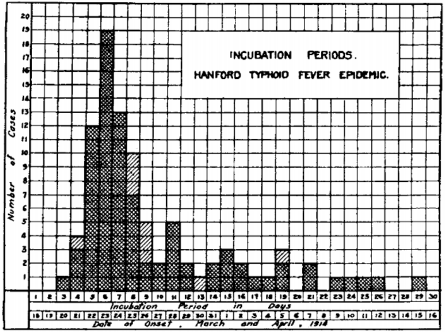


Chart of the cases in the Hanford typhoid fever epidemic, showing incubation periods and dates of onset. The heavily shaded areas represent definite cases of typhoid fever. The lightly shaded areas represent the doubtful cases.

The Incubation Period

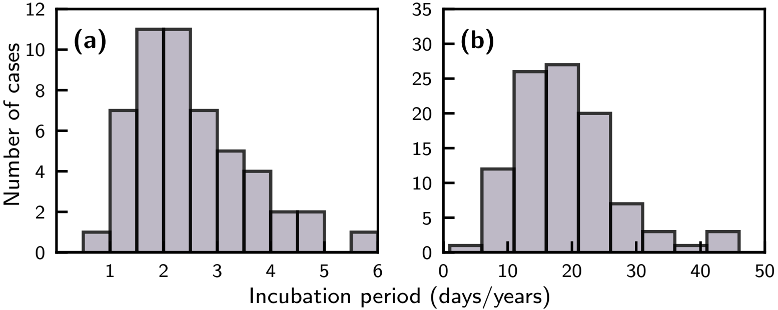
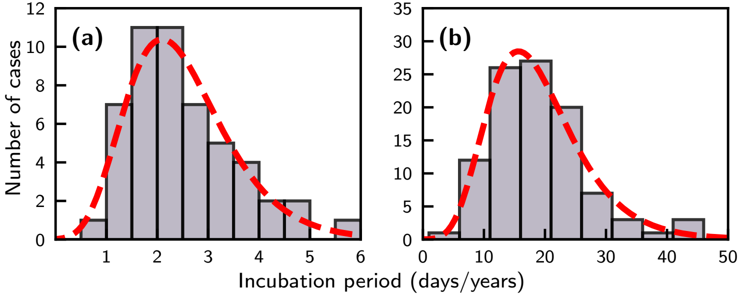


Figure: Frequency distributions of incubation periods for two diseases. Data redrawn from historic examples. (a) Data from an outbreak of food-borne streptococcal sore throat, reported in 1950 (Sartwell, 1950). Time is measured in units of days. (b) Data from a 1949 study of bladder tumors among workers following occupational exposure to a carcinogen in a dye plant (Goldblatt, 1949).

Sartwell's Law (1966)

Sartwell's Law

Incubation periods for diseases tend to be distributed as lognormals; more generally, they will be right-skewed.



Explanations?

Traditionally: **Heterogeneity** of ...

- ... The contagion's fitness.
- ... The host's immunosensitivity.
- ... The inoculum of contagion.

Q: Can Sartwell's Law arise from intrinsic chance alone?

An Evolutionary Graph Theory Approach

Many illnesses consist of spreading on a within-host network

The Illness ...	... Takes over the ...	... Which is a ...
Typhoid	well-mixed gut microbiome	Complete graph
Influenza	uncompromised tracheal cells	2D lattice
Leukemia	healthy bone marrow cells	3D lattice

Evolutionary Graph Theory

- We emulate the **Incubation Period** of a disease via the **The Takeover/Fixation Time** of an evolutionary takeover process on a network.

Definition

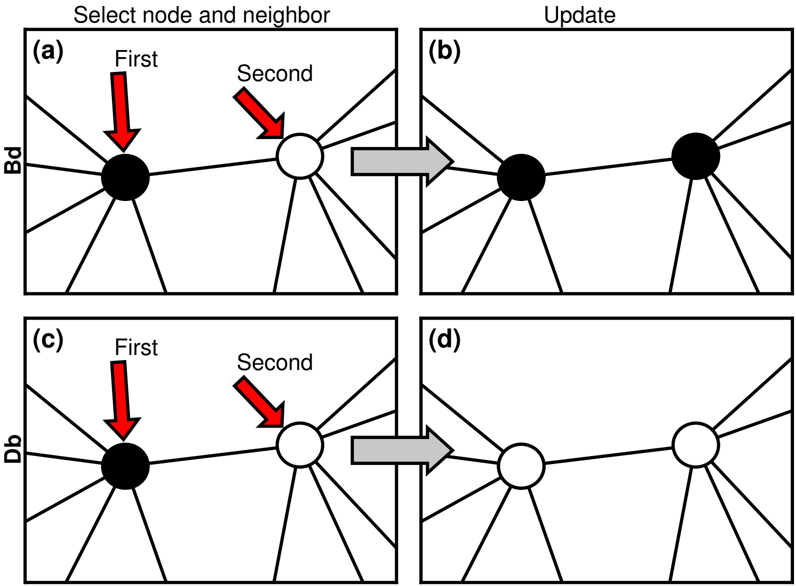
The **Fixation (or Takeover) Time** of a network evolutionary process is the time between the appearance of a single invader and 100% of the resident nodes being replaced by invaders. (The initial population of invaders and the final takeover threshold can both be adjusted, if desired.)

Evolutionary Graph Theory

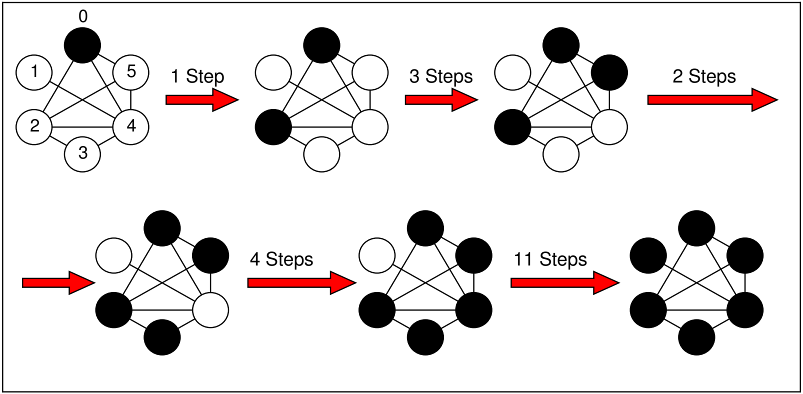
Definition

The **Moran Birth-death (Bd) model** consists of three steps:
1. With probability proportional to fitness (r), randomly select a node on the network to give birth.
2. Uniformly randomly select a neighbor of the first node to die.
3. The dying node takes on the type of the birthing node.

The Moran Model



A Path To Takeover: $r = \inf$



Complete Graph: Summary

Complete Graph: Simplified

Infinite r Bd on a complete graph (rephrased)

Each time step, a random node from a set of $N-1$ is selected uniformly with replacement. If we pick a healthy node, we relabel it and toss it back, and repeat until there are no healthy nodes left. What is the distribution of times T until all nodes are relabeled?

The Coupon Collector's Problem

The Coupon Collector's Problem

Each day, a kid gets one trading card, uniformly at random. Given that there are N distinct cards, what is the distribution of times T required to form a complete set?

The Complete Graph

The mean of the collection time is $\mu \approx N\log(N) + N\gamma$. Then we find

$$\frac{T-\mu}{N} \xrightarrow{d} \text{Gumbel}(-\gamma, 1). \tag{1}$$

Here $\gamma \approx 0.5772$ is the Euler-Mascheroni constant, $\xrightarrow{d}$ denotes convergence in distribution, and a Gumbel(α, β) random variable has a density given by

$$h(x) = \beta^{-1} e^{-(x-\alpha)/\beta} \exp\left(-e^{-(x-\alpha)/\beta}\right). \tag{2}$$

The Star Graph is Similar

Proposition 1: Agreement of Geometric and Exponential Variables

Proposition

Suppose we have a family of sequences $(p_m)_{m=1}^{M}$, with $0 \leq p_m \leq 1$ for all m and M, where p_m may depend on M. Define $Geo(p)$ to be a geometric random variable with distribution

$$P(Geo(p) = k) = (1 - p)^{k-1} p$$

for $k = 1, 2, \ldots$. Further, let $\mathcal{E}(p)$ be an exponential random variable with distribution

$$P(\mathcal{E}(p) = x)dx = pe^{-px} dx$$

for $x \geq 0$. Given some function $L := L(M)$ such that $\lim_{M\to\infty} L = \infty$ and $\lim_{M\to\infty} \sum_{m=1}^{M} \frac{1}{p_m L^2} = 0$, and given $T_G := \sum_{m=1}^{M} X(p_m)$, $T_E := \sum_{m=1}^{M} \mathcal{E}(p_m)$, and $\mu := \sum_{m=1}^{M} 1/p_m$, then

$$\frac{T_G - \mu}{L} \sim \frac{T_E - \mu}{L}.$$

The symbol "$\sim$" means the ratio of characteristic functions goes to 1 as N gets large. That is, the random variables on both sides converge to each other in distribution as M gets large.

Proposition 1 (Short Version)

Proposition

It's usually okay to replace sums of geometric random variables with a similar sum of exponential random variables.

Complete Graph and Star Graph

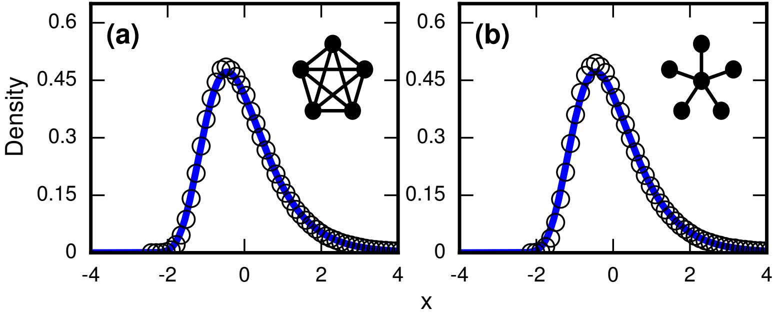
For the complete graph, we have

$$\frac{T - N(\log(N) + \gamma)}{N} \xrightarrow{d} \text{Gumbel}(-\gamma, 1). \quad (3)$$

By a similar (but more complicated) trajectory, we have that the corresponding fixation time for a star network with N spokes is

$$\frac{T - N^2(\log(N) + \gamma - 1)}{N^2} \xrightarrow{d} \text{Gumbel}(-\gamma, 1). \quad (4)$$

Complete Graph and Star Graph



Results for Lattices

Lattices: A Problem

- The derivation for the complete and star graphs relied on p_m (the probability of adding a new invader, given there are currently m invaders).
- **However, p_m as a concept isn't well defined for lattices** and other more complicated networks. The probability of a new invader being added depends on the *configuration* of the existing invaders, not just their number.

Lattices: Geometric Simplification

Lattices: Geometric Simplification 2

In many cases, it is possible to make an analogy to **first-passage percolation**. So these clusters are described by shape theorems, stating that these have a simple convex (but non-ball) limit shapes.

Lattices: Surface Area to Volume

- In a d dimensional lattice, a simple convex shape of volume V has a surface area proportional to V^η, where $\eta = 1 - 1/d$.
- Moreover, the probability of adding a new invader is proportional to the probability of selecting a node on the boundary of the cluster of invader nodes. So:

$$p_m \propto \frac{1}{m} \cdot \text{Surface area of of the invader cluster}$$
$$\propto q_m := \frac{\min(m, N-m)^\eta}{m}.$$

Therefore,

$$T \approx \sum_{m=1}^{N-1} \text{Geo}(q_m).$$

Lattices: Low Dimensions First

Proposition 2: A Condition for Normality

Proposition

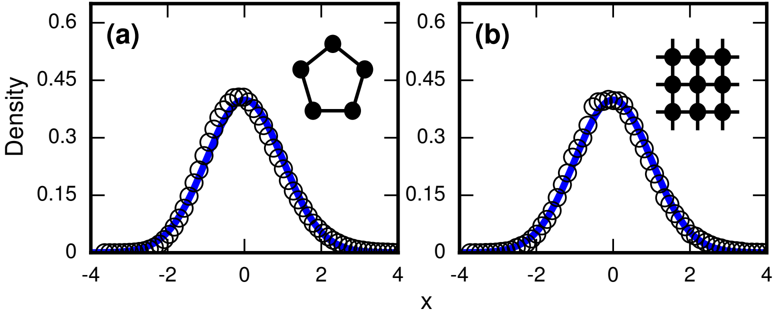
Let $T = \sum_{m=1}^{M} \mathcal{E}(p_m)$, *define* $\sigma^2 = Var(T) = \sum_{m}^{M} p_m^{-2}$, *and let* $\lim_{M\to\infty} p_m\sigma = \infty$. *If*

$$\lim_{M\to\infty} \sum_{m=1}^{M} \exp\left(-\epsilon p_m \sigma\right) = 0, \qquad (5)$$

then

$$\frac{T-\mu}{\sigma} \xrightarrow{d} Normal(0,1). \qquad (6)$$

Lattices: $d = 1$ and $d = 2$



Lattices: No Closed-Form for High Dimensions

Lattices: Skew Results

Lemma

If the independent random variables X_i have variances σ_i^2 and skews κ_i, then their sum has a skew of

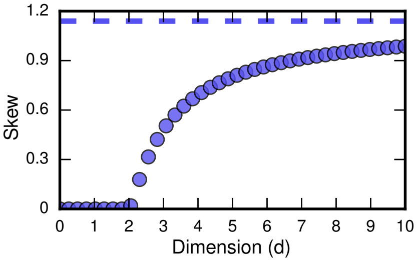
$$Skew\left(\sum_i X_i\right) = \frac{\sum_i \kappa_i \sigma_i^3}{\left(\sum_i \sigma_i^2\right)^{3/2}}$$

Lattice: $d \geq 3$.

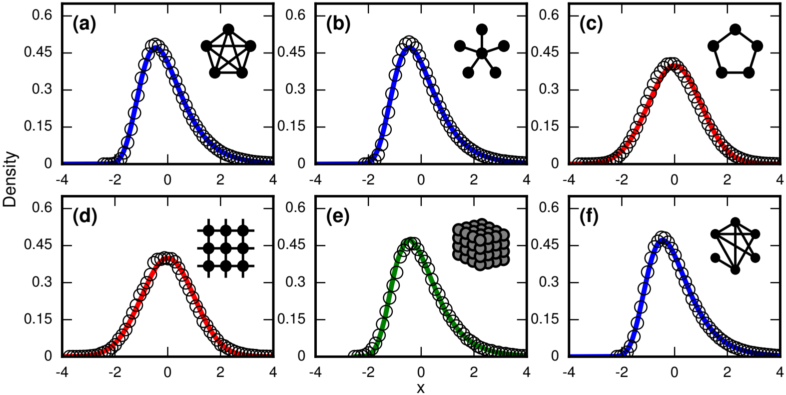
Theorem

Letting $\eta = 1 - 1/d$, the asymptotic skew of the takeover times for a $d > 2$ dimensional lattice is given by

$$Skew(d) = \frac{2\zeta(3\eta)}{\zeta(2\eta)^{3/2}}, \quad where\ \zeta(x) = \sum_{n=1}^{\infty} \frac{1}{n^x}.$$



Summary: Infinite r



Summary: Complete Graph, $r = 1$

Summary: Complete Graph, $r = 1$

By symmetry, $p_m^+ = p_m^- = \frac{m(N-m)}{N(N-1)}$. Therefore,

$$X_n := \text{The population level after } n \text{ changes} = \sum_{i=1}^{n} x_i,$$

where $x \in \{-1, +1\}$, each with probability $1/2$.

Summary: Complete Graph, $r = 1$

- Fact: We can only record an incubation period if someone *actually* gets sick.
- Therefore, we need to condition on the population X_n hitting N before ever hitting 0.

Summary: Complete Graph, $r = 1$

Proposition

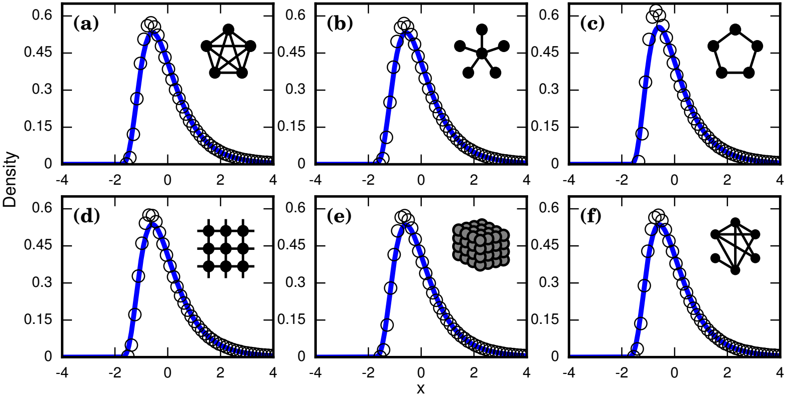
(Via optional stopping) An unbiased random walk which starts at 1 and hits N before hitting 0 has a high level of skew (≈ 1.807).

Summary: Complete Graph, $r = 1$

Proposition

Conditioning on the success of the invaders induces a high level of skew.

Summary: $r = 1$



Summary: Realism?

- Sartwell measured **Dispersion Factors**, the standard deviations of the logs of the data, across many diseases.
- He measured dispersion factors between **1.1 and 1.5** for real-world diseases.
- For our high fitness simulations, we measured factors between **1.1 and 1.4**.
- For neutrally fit invaders, we measured factors between **1.6 and 1.7**.

Summary: Heterogeneity

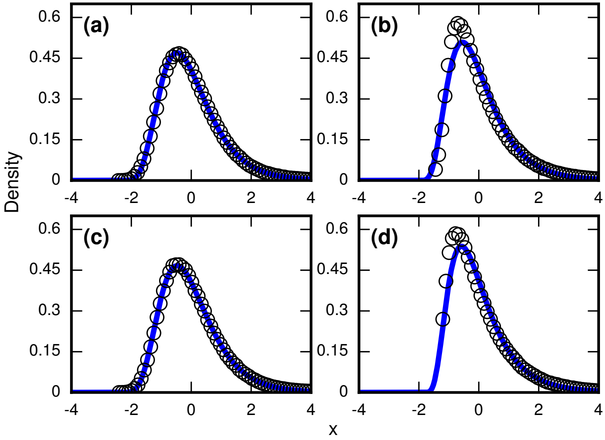


Figure: Complete graph with $r = 10$ under various forms of heterogeneity. (a) heterogeneity in invaders. (b) Heterogeneity of host sensitivity. (c) Heterogeneity of initial dosage. (d) Heterogeneity of all three.

Summary: Main Points

Important Facts:

- When the invader fitness is high, dynamics are dominated by **The Coupon Collector's Problem**, leading to right skewed distributions.
- When invader fitness is low, dynamics are dominated by a **Conditioned Random Walk**, leading to right skewed distributions.
- There is a **Critical Dimension** in the infinite fitness case, with higher dimensional topologies leading to more skewed distributions.
- While population-level heterogeneity can be tuned to cause right-skewed distributions, **Such Heterogeneity Isn't Necessary** for these distributions, and just accentuate the fundamental mechanisms we already observe.

Summary: Main Points

And Most Importantly...

- While lognormal-like distributions can be justified in any number of ways, **Evolutionary Network Dynamics** is the only phenomenon common to the diverse range of diseases shown to obey Sartwell's law.

Selected References

- B Ottino-Löffler, J G Scott, and S H Strogatz, Evolutionary Dynamics of Incubation Periods, `http://www.biorxiv.org/content/early/2017/05/30/144139`.
- B Ottino-Löffler, J G Scott, and S H Strogatz, Takeover Times for a Simple Model of Network Infection, Physical Review E, 96m 012313 (2017).
- W A Sawyer, Ninety-Three Persons Infected by a Typhoid Carrier at a Public Dinner, Journal of the American Medical Association, 63(18) (1914).
- P E Sartwell, The Distribution of Incubation Periods of Infectious Disease, Am. J. Hyg. 51, 310-318 (1950).
- M W Goldblatt, Vesical tumours induced by chemical compounds, Occupational and Environmental Medicine 6:6581 (1949).
- L Ryan, Meet the best 9-year-old Pokemon card player in Minnesota, Star Tribune, March 9 (2015).
- P A P Moran, The Effect of Selection in a Haploid Genetic Population, Proc. Camb. Phil. Soc., 54(8) (1958).
- E Lieverman, C Hauert, and M A Nowak, Evolutionary Dynamics on Graphs, Nature, 433(7023) (2005).
- P Erdős and A Rényi, On a Classical Problem of Probability Theory, Publ. Math. Inst. Hung. Acad. Sci 6 (1961).
- L E Baum and P Billingsley, Asymptotic Distributions for the Coupon Collector's Problem, The Annals of Mathematical Statistics, 36(6) (1965).
- A Auffinger, M Damron, and J Hanson, 50 years of First Passage Percolation, `https://arxiv.org/abs/1511.03262` (2016).

This research was supported by a Sloan Fellowship to Bertrand Ottino-Löffler, in the Center for Applied Mathematics in Cornell, as well as by NSF grants DMS-1513179 and CCF-1522054.

Questions?

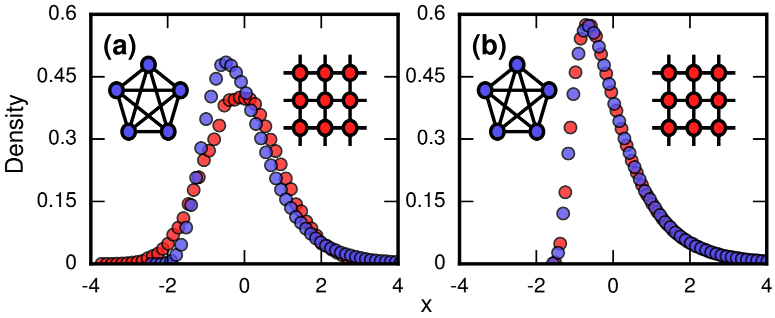
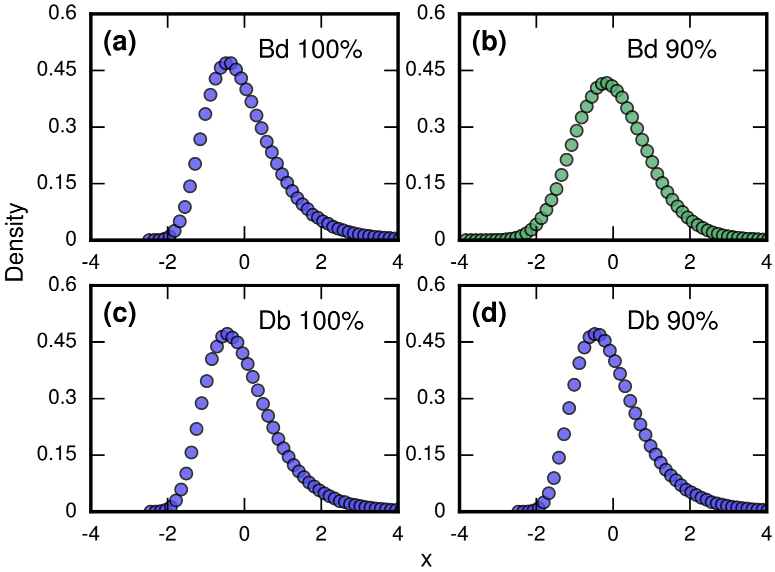


Figure: (a) $r = \infty$. (b) $r = 1$.

Personal: people.cam.cornell.edu/∼bjo34/
Twitter: @OttinoLoffler

Appendix: Additional info

Summary: Truncation



Summary: Complex Networks

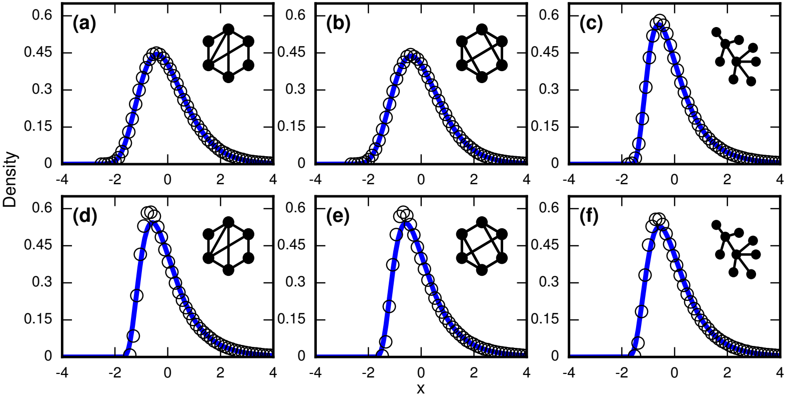


Figure: Top row: $r = \infty$. Bottom row: $r = 1$.

Complete Graph: Derivation Sketch, $r = \infty$

Complete Graph: Part 1

Define $p_m :=$ The probability of adding a new invader, given there are currently m invaders.

$$p_m := P(\text{Choose an invader}) \cdot P(\text{Neighbor is resident})$$
$$= \frac{mr}{mr + (N-m)} \cdot \frac{N-m}{N-1} \xrightarrow{r \to \infty} \frac{N-m}{N-1}$$

Complete Graph: Part 2

Probability of invader population $m \to m+1$ taking t steps $=$ $(1-p_m)^{t-1} p_m$.

$$P(\text{Geo}(p_m) = t) = (1-p_m)^{t-1} p_m; t \geq 1$$

Complete Graph: Part 3

$$\begin{aligned} T &= \text{The total fixation time} \\ &= \sum_{m=1}^{N-1} \text{Geo}\left(p_m\right) = \sum_{m=1}^{N-1} \text{Geo}\left(\frac{N-m}{N-1}\right). \end{aligned}$$

This is a simple sum of Geometric random variables with increasing means.

Summary: Complete Graph, $r = 1$

The important stopping time is the first hitting time of 0 or N, so use

$$S = \min\{S_0, S_N\}, \text{ where } S_m = \min\{n | X_n = m\}.$$

To find the appropriate moments of the conditioned fixation times, set up the moments

$$\mu_i := E(S^i | X_S = N).$$

Summary: Complete Graph, $r = 1$

By using martingales and the optional stopping theorem, we get

$$\mu_1 = \frac{N^2 - 1}{3}$$
$$\mu_2 = \frac{7N^4 - 20N^2 + 13}{45}$$
$$\mu_3 = \frac{31N^6 - 147N^4 + 189N^2 - 73}{315}$$

So in the large N limit, the skew of the conditioned random walk becomes

$$\text{Skew} = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^2}{(\mu_2 - \mu_1^2)^{3/2}} \approx 1.807 \gg 0.$$