Lecture 13 – More to Know About Power Law Non-Normal Distributions and SIR Network Epidemic Models

A Three Roads to Power Laws

1)Road 1 is to generalize **Central Limit theorem (**iid errors with mean/variance, then as N → infinity distribution is NORMAL) – to distributions with infinite variance. Stable distributions where CLT is special case.
Have property that linear combination of two independently drawn copies of the variable has the same distribution.

Three closed form representatives of stable distribution: Normal is bell-shape with mean and variance. **Cauchy** is symmetric with such a thick tail and no mean or variance; **Levy is non-negative variable** – not symmetric. Being STABLE distributions, all three are "attractors" – if lots of random "stuff" happens end up with this distribution.

A stable distribution has four parameters: – key is *stability* parameter $\alpha \in (0, 2]$ (*skewness* parameter; s*cale* parameter; *location* parameter) Stable distributions with infinite variances likely to show jumps, which fits "many time series appear to exhibit "discontinuities(e.g., large jumps)" Evidence + Generalized Central Limit Theorem justifies stable models in finance & economics, where data poorly described by Gaussian model, but well described by a stable distribution eg **stock prices** (Journal of Business & Economic Statistics,(Apr., 1990)).

2) Road 2 positive feedback loops to random shocks that push distribution toward power law in two ways. (Exam type question: "Positive feedback loops always lead to discontinuous change/big jumps." T, F and explain

STRUCTURE A: Stochastic/proportionate growth **plus a barrier/bound (**% growth + bounds) generate power law (associated with Herb Simon in debate with Mandelbrot).

Without barriers/bounds stochastic growth gives log-normal: **random ln/% growth –> log-normal** with var $\sigma^{2.}$ Rate of growth independent of initial size and variance of growth that is the same for all units – Gibrat's law in economics that the proportional rate of growth of a firm is independent of its absolute size , which produces log-normal. -- yields equation for growth of firm (http://docentes.fe.unl.pt/~jmata/gibrat.pdf):

% change in SIZE = σ so that SIZE (t) = (1+ σ) SIZE (t-1)–> ln SIZE (t) = ln SIZE (t-1) + ln (1+σ)

Need something to fatten tails to go beyond log-normal. Some lower bound/friction. Gibrat + lower bound–> Zipf. This is the STEADY STATE distribution. The bound produces "Reflected Brownian motion" – originally shown by Champernowne for income distribution. **Lower barrier frees "extra mass" to add to distribution's tail.**

Gabaix model for cities: Cities of different sizes have same growth rate with a constant variance. The position of cities can change, but the distribution replicates itself. LA surpasses Chicago as number 2 in US but number 2 city is still proportionate (½ of the largest city) in Zipf's law with coefficient 1.

How the barrier works: You follow the average growth rate + random component unless you are very small. **If you are very small you grow at 0 or at some positive value that depends on average growth and random shock.** Moving density from the bottom pushes the distribution toward fatter tails. All but smallest have same growth rate with constant variance (presumably if "policies" matter this will no be true).

Alternative way to see mechanism, consider **fixed total population that distributes itself among cities**. With same % growth larger cities have greater absolute growth. **Must have more small cities to maintain the fixed** population. City with 4M growing at 25% would add 1M so must have lots of small cities to give up 1M; and conversely if city with 4M declines by 25% … must have lots of small cities for people to move to.

Consider world in which cities either double or halve every period, where P % double every period; (1-P) = % of cities that halve. This shows how the fixed rule produces distribution (but does not allow for the variance in growth rates) Scale the fixed population at 1 so 2(p) + ½ (1-p) = 1. Solving we get p = 1/3 → cities with size 2 make up 1/3rd; cities with size ½ made up 2/3rd of population. **There are twice as many small cities as large cities.**

What about next period, with the same process? Some cities get half the population and others double. City A of size 2 becomes 4 and city B of size 2 becomes 1, etc

Rank of city by size	size	Frequency	
1	4	1/9	
2	1	2/9	
3	1	2/9	
4	1/4	4/9	And next period and so on

A **stable distribution** by size classes after a long period of doubling/halving needs same absolute changes, which holds only if size classes have the same total population and fits Zipf with bins:

SIZE	# CITIES	POPULATION IN CLASS	
small	40	1	40
larger	20	2	40
big	10	4	40
biggest	5	8	40

STRUCTURE B – Preferential attachment http://ccl.northwestern.edu/netlogo/models/PreferentialAttachment)

The power law story for web pages is that new sites more likely to attach to (older) larger sites. Small # of larger sites will grow more rapidly than smaller/newer sites → power law. Can also explain why citations fit power law. https://en.wikipedia.org/wiki/Preferential_attachment .

But also need to model entry and exit. A new site/paper enters and gathers some followers/citations according to attachment while some sites die off. The power law is presumed to hold for internet with mixture of older and younger firms. But the lifetime of sites at a moment in time is often **exponential** with a few long-lived sites and many short-lives sites with a difference < power law. **Mixture of exponential and log-normal gives power law.**

Other ways to get power law: as the inverse of a function that follows a power law; as combinations of exponential; random walk distribution of lifetime → lots of short lives, few older ones. The different variants of preferential attachment suggest different processes that direct attention at different ways to affect power law distribution if, say, society viewed it as "too weighted" at tail for some reason – Billionaires/1% vs rest of us.

Road 3) Through OPTIMIZING behavior that brings system to "brink" of large changes

LOCAL INTERACTIONS AND OPTIMIZATION --> SOC self organized criticality

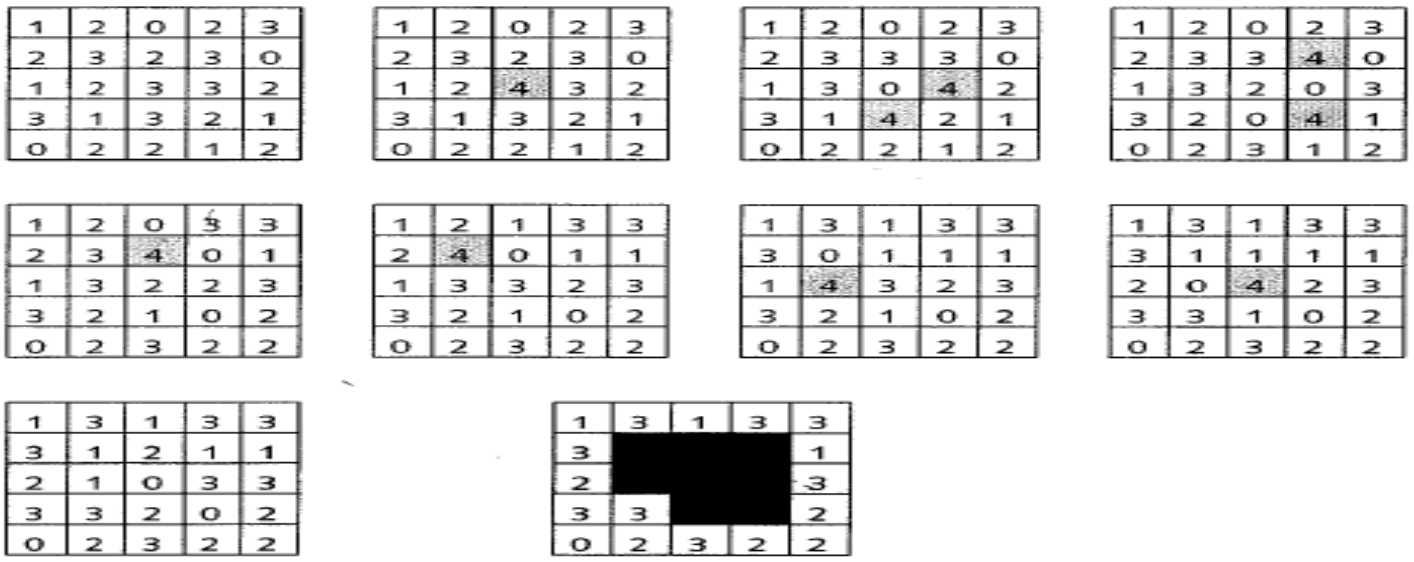
System has birth/death process that moves it to border area where it is subject to risk of major disruptions, producing power law. This is the P. Bak "sandpile theory" explanation: systems naturally move to a point where they generate "avalanche" events. But if then "we must also abandon any idea of detailed long-term determinism or predictability". There will be Silicon Valley or an economic collapse, but you cannot predict where it occurs or when. In a short period, you get few BIG EVENTS, but you never know when. There will be an epidemic or big car crash or traffic jam as cars follow given route and pack roads until some random event → stoppage.

Bak: "Large fluctuations ... in economics indicate an economy operating at the SOC state, in which minor shocks can lead to avalanches of all sizes ... there is no way one can stabilize the economy... eventually something different and quite unexpected will upset ... balance ... and there will be a major avalanche somewhere else" (p 191).

Al Gore Paean to the Sandpile Model: "The sandpile theory – self-organized criticality – is irresistible as a metaphor; one can begin by applying it to the developmental stages of human life. The formation of identity is akin to a formation of the sandpile, with each person being unique and thus affected by events differently. A personality reaches the critical state once the basic contours of its distinctive shape are revealed; then the impact of each new experience reverberates throughout the whole person, both directly, at the time it occurs, and indirectly, by setting the stage for future change. Having reached this mature configuration, a person continues to pile up grains of experience, building on the existing base. But sometimes, at midlife, the grains start to stack up as if the entire pile is still pushing upward, still searching for its mature shape. The unstable configuration that results makes one vulnerable to a cascade of change."

The model is a **cellular automata** that uses nearest neighbor interactions to produce an avalanche-- lots of places changing. Think of debts/ bankruptcy. Add an extra debt: Owe--> Owe + 1. If you hit a debt limit, you go bankrupt, pushing your debts to neighbors by lowering their assets. In the diagram, numbers measure debts. You drop a new debt onto the model -- 4 in the second diagram. That person can't pay loans to neighbors, which adds to their debt. They go under. And so on. The avalanche is defined as the number of sites that hit 0 -- go bankrupt.

Figure 12. Illustration of toppling avalanche in a small sandpile. A grain falling at the site with height 3 at the center of the grid leads to an avalanche composed of nine toppling events, with a duration of seven update steps. The avalanche has a size $s = 9$. The black squares indicate the eight sites that toppled. One site toppled twice.



This is a LOCAL INTERACTION MODEL in which several sites get close to an avalanche, so that "the next straw breaks the camels' back". The model has power law: # of Avalanches of Size in period = (Size)$^{-1.1}$

Avalanche Size	# of Avalanches
2	.47
10	.08
100	.006

HOT (highly optimized tolerance) develops power laws from optimization. Systems are optimized along some dimensions and robust but risk failure from cascade of shocks in other dimensions. If optimize return to investment hy going highly leveraged you leave open the door to catastrophe –> heavy tail of financial avalanche as in Wall Street implosion.

Example is **Forest fire model.** Consider the forest with trees on a grid. Random lightning bolts cause a fire. If a bolt lands on an empty space, no fire; if it lands on tree, it burns the tree, which spreads to all neighbors. The chance of a bolt/ fire at any tree is p. #of trees that burn is inversely related to # of fires per time period in a power law: many small fires in which few trees burn and a few large fires. The key metric is $R(p)$, the probability that there is a path across the space called a spanning cluster so that whole forest burns. $R(p)$ makes a dramatic transition from low to high values---a phase transition---at a critical value when the density of trees is about 0.59.

Neat Result: Large fires more likely when few sparks. Why? Because with few small fires, the forest gets denser –> one big fire. Lots of sparks → space between trees. Carlson and Doyle explain this with model in which forester plants trees to optimize amount of lumber, subject to fires that burn trees. The optimizing strategy is to plant trees in blocks with narrow fire breaks between them to prevent fire from spreading. Smaller blocks in regions of likely fire; larger blocks in regions where fire is unlikely. The structure is designed to respond optimally to small shocks.

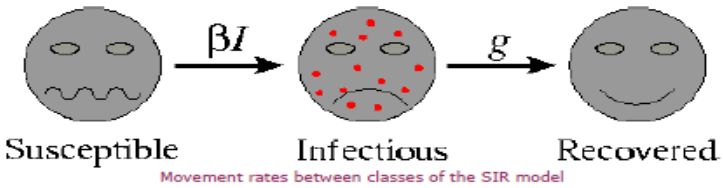
But optimal performance in normal times risks ruinous collapse if get unlikely shock. In an area with not many sparks you put up few firebreaks. The trees are close together. Unlikely spark event → forest burns down. In area with lots of sparks, you put up lots of firebreaks. Unlikely spark event happens – you are safe. What is analogy to financial leveraging?

Optimization to common perturbations leads to good properties with respect to those shocks but are "fragile" to rare events, unanticipated changes in the environment, and flaws in the design

SIR Models of Epidemics

SIR – Classic Susceptible, Infected, Recovered model – differential equation model. Modified by network geometry and mutation evolution of pathogen.

Infections are transmitted from nodes in a network map to neighbors – local "geometry" with no long distance links. Population of N consists of S+I+R = N. At outset, everyone is S but as time proceeds people will shift, first to I and then to R. Scale N to 1 so get shares. Model has two transitions: Susceptible → Infected and Infected → recover. Each infected person generates **bs(t)** new infected individuals per period and a fraction k of the infected group recover per period. If average duration of infection is three periods, 1/3d of infected recovers each period



Movement rates between classes of the SIR model

s(t) = S(t)/N, the susceptible fraction of the population,

i(t) = I(t)/N, the infected fraction of the population, and

r(t) = R(t)/N, the recovered fraction of the population.

Thus, ds/dt = -b s(t) i(t), where b is rate at which infected ==> turn susceptible into infects
di/dt = bs(t)i(t) – ki(t) bcs where k is the rate at which infected recover

Also, ds/dt + di/dt + dr/dt =0. Why – bcs shares of population sum to 1

This generates:
1)The disease always dies out. Diagram is uni-directional. S-->I → R so if rate of infection is high, everyone gets infected and recovers.

Figure 1 contains solutions of the SIR system simulating a highly virulent ($R_e = 3.5$) flu epidemic in a town of 50, 000 people.

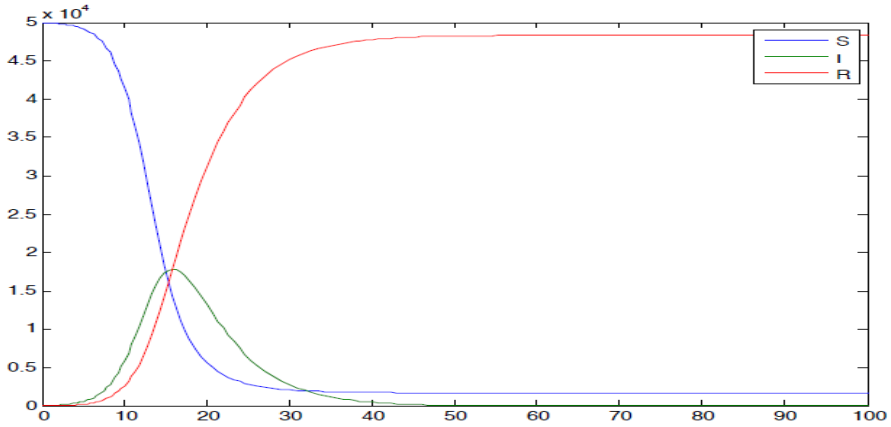


Figure 1: Solutions of SIR system of ODEs with $\beta = 0.7/50000, \nu = 1/5, S(0) = 49955, I(0) = 5, R(0) = 0$

2)But there is another solution in which disease dies out. This is when b < k – infection is slower than recovery.
Epidemic threshold theorem – stated in terms of R = b/v known as epidemiological parameter – get rid of disease without everyone getting it.

The value of R_0 for some well-known diseases	
Disease	**R_0**
AIDS	2 to 5
Smallpox	3 to 5
Measles	16 to 18
Malaria	> 100

So public health policy is to reduce R... lower b and raise v.
1. Reduce the contact rate by self-isolation of susceptible individuals--> on line education
2. Find vaccines to create new state for susceptibles so disease cannot invade you
3. Reduce the transmissibility by encouraging frequent hand washing and face masks.
4. Reduce the duration of infection D with antiviral drugs

Globalization and Small World → Transportation Networks

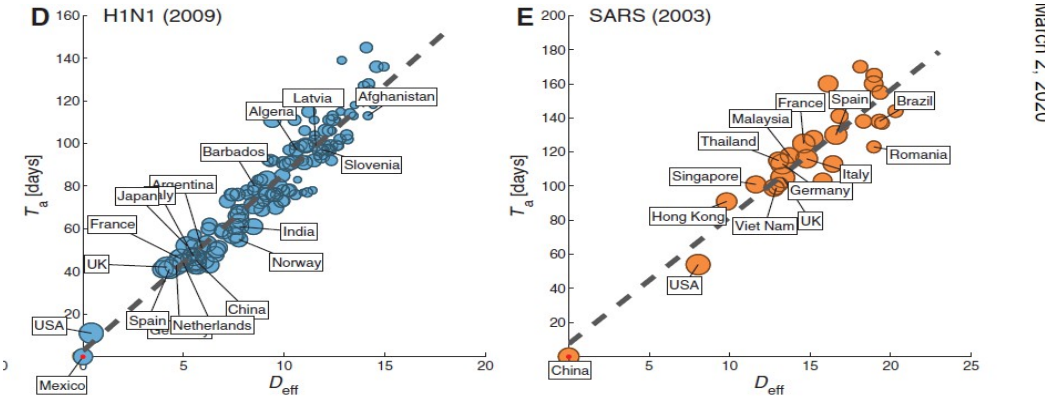
If small world, not just local neighbors – someone far away may connect to you. Send you email, your computer may get infected but not you. But if someone from infected area comes to your block from far-away with disease, your neighborhood can get infected.

So linked analysis to real networks of travel – flight patterns and once infected arrives in destination can study where they go. Infected is a basketball fan – yikes for sports events. Infected gives talk at Harvard – shut down any seminars with outside speakers. Venice fighting Black Death by quarantining ships for 40 days before passengers and crew could go ashore during Black Death plague. Need not know if person is infected or not. just stop all contact.

Transportation network differs from standard measure of networks as static nodes and edges because it introduces time and dynamic changes into the model. You and I have a connection and we both have lots of other connections but not all nodes are concurrently active. Map shows possible roads but some may be slower than others at some times and faster than others. Individuals in a social network do not interact simultaneously with all of their acquaintances etc. So need some sort of dynamic network/measure of mobility patterns to capture reality.

One model --The Hidden Geometry of Complex,Network-Driven Contagion Phenomena Dirk Brockmann and Dirk Helbing Science 13 DECEMBER 2013 VOL 342 **replaces conventional geographic distance by a measure of effective distance derived from the underlying mobility network.** In epidemic analysis this says, your neighbors may be "far away" in distance but close if there is a lot of air traffic. Also could depend on size of populations. Boston would be closer to NY than to Martha's Vineyard because more people come to Boston from NY than from the island. You "rewrite" your network to reflect transportation parameters. Many infected from planes to Boston more likely to create problem than small number on ferry.

"In addition to the local dynamics, individuals travel between nodes according to a transportation equation that determines **effective distance** from a node n to a connected node m … This concept of effective distance reflects the idea that a small fraction of traffic n→m is effectively equivalent to a large distance, and vice versa. The complexity of the spatiotemporal pattern is largely determined by the structure of the mobility component and not by the nonlinearities or the disease-specific, epidemiological rate parameters of the model.



Linear relationship between effective distance and arrival time for the 2009 H1N1 pandemic (D) and the 2003 SARS epidemic (E). The arrival time data are the same as in Fig. 1, D and E. The effective distance was computed from the projected global mobility network between countries. Strong correlation between arrival time and effective distance.

Another Model Traffic-driven epidemic spreading in finite-size scale-free networks Sandro Melonia, Alex Arenasb,c, and Yamir Morenoc PNAS October 6, 2009 vol. 106 no. 40 16897–16902

The value of the epidemic threshold in scale-free networks depends directly on flow conditions, in particular on the first and second moments of the betweenness distribution given a routing protocol. Bounded delivery provokes congestion, slowing down the spreading of the disease and setting a limit for the epidemic incidence. **Nodes do not interact at all times t, but only when they exchange – ie have a flight – epidemic can spread between nodes every time an interaction takes place.** Contagion is driven by traffic or interaction flow. If "congestion" arises, the number of contacts between the system elements decreases, leading to a less-efficient spreading of the disease and therefore to a significant reduction of the average number of infected individua The disease-propagation process has two dynamical components:one intrinsic to the disease itself (β) and the other to the underlying traffic dynamics (the flow).

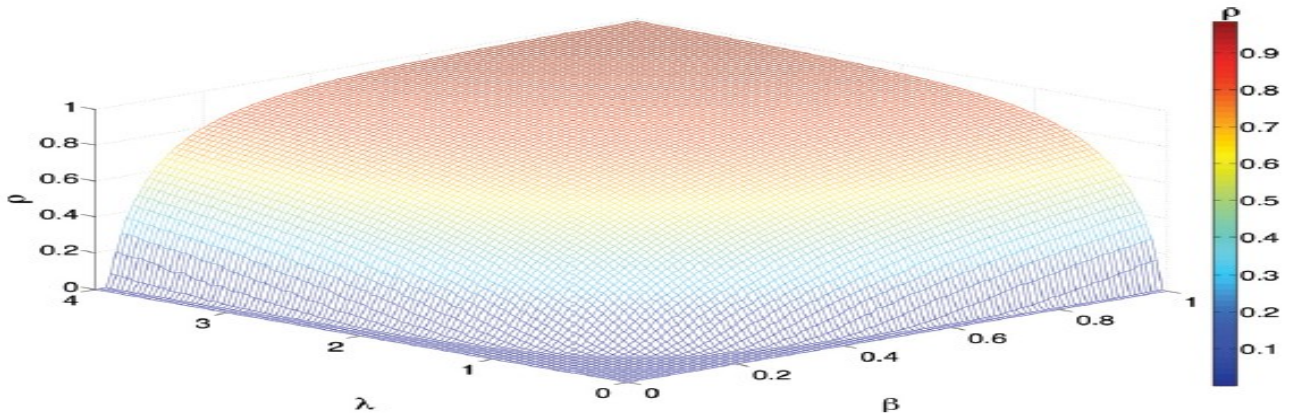


Fig. 1. Dependence of epidemic incidence on traffic conditions for unbounded delivery rate. The density of infected nodes, ρ, is shown as a function of the spreading rate β and the intensity of flow λ in SF networks. Flow conditions (controlled by λ) determine both the prevalence level and the values of the epidemic thresholds. Increasing the number of packets traveling through the system has a malicious effect: The epidemic threshold decreases as the flow increases. Each curve is an average of 100 simulations starting from an initial density of infected nodes $\rho_0 = 0.05$. The results correspond to the greedy routing scheme and the network is made up of 10^3 nodes by using the model in ref. (20). The remaining parameters are $\alpha = 2$, $\gamma = 2.6$ and $\langle k \rangle = 3$.

Evolution of Pathogens/Ideas in Network

But there is another problem – pathogen is not constant. It evolves. 60% of the (approximately) 400 emerging infectious diseases identified since 1940 were initially poorly adapted, poorly replicated, and inefficiently transmitted (S. S. Morse et al., Prediction and prevention of the next pandemic zoonosis. Lancet (2012) and K. E. Jones et al., Global trends in emerging infectious diseases. Nature (2008). So you need another dimension to understand/predict/find policies to deal with SIR type world.

Pathogens often evolve in response to changing environments and medical interventions.

Information is often modified by individuals before being forwarded. Evolutionary adaptations can impact the threshold, probability, and final size of epidemics. More elements lead to more complex model, with more interactions and parameters in which structural properties of the network – transportation/mobility – and the evolutionary adaptations of the spreading process move simple SIR to a simulation model – agent-based economics.

One model: Risk factors for the evolutionary emergence of pathogens H. K. Alexander* and T. Day J. R. Soc. Interface (2010) "When first introduced to a population, a pathogen is often poorly adapted to its new host and must evolve in order to escape extinction. Theoretical arguments and empirical studies have suggested various factors to explain why some pathogens emerge and others do not, including host contact structure, pathogen adaptive pathways and mutation rates. **Using a multi-type branching process, we model the spread of an introduced pathogen evolving through several strains**...(with) a network-based approach to separate host contact patterns from pathogen transmissibility. We also allow for arbitrary adaptive pathways. These generalizations lead to novel predictions regarding the impact of hypothesized risk factors.

Pathogen fitness depends on the host population in which it circulates, and the 'riskiest' contact distribution and adaptive pathway depend on initial transmissibility.

Emergence probability is sensitive to mutation probabilities and number of adaptive steps required, with the possibility of **large adaptive steps (e.g. simultaneous point mutations or recombination) having a dramatic effect.** In most situations, increasing overall mutation probability increases the risk of emergence; however, notable exceptions arise when deleterious mutations are available.

Economics of New Products/Ideas and Flow of Information in Social Network

Models of pathogens can be used to understand Economics of new products or ideas: new product/idea invades market. What is correlate to infection? But here want to consider optimizing. Pathogen mutation can be viewed as "dumb optimizing" – natural selection survival.

SEE YOU ON ZOOM OR VIDEO AFTER THE BREAK: From China one of my colleagues writes "I have been teaching online (live streaming) for a few weeks now. It turns out that students are more willing to interact online than face-to-face. I guess this virus will have long-lasting influence on how people work---wide acceptance for work from home, more flexible work time, and possibly more gender equality with the flexible work time."