# Gang Recruitment and Growth: A Cellular Automata and Directed Graph Approach to the Statistics of Gang Sizes

*William I. Newman, UCLA**

**Outline**

- Motivation
- Cellular Automata and Percolation
- Statics vs. Dynamics: Self-Organized Criticality & Graphs
- Forest Fires and Characteristic Earthquakes
- Gang-Warfare and Social Networks

---

** In collaboration primarily with D.L. Turcotte (Cornell & UC Davis), A. Gabrielov (Purdue), V.I. Keilis-Borok (UCLA & Moscow)*

## Motivation

- *Why discrete models?*
  - Applied math. often synonymous with ODEs/PDEs
  - Continuous representations not always helpful
  - Discrete models can be useful (metaphors or proxies)
- *Possible benefits of discrete models*
  - Discrete models can be amenable to combinatoric and graph theoretic methods, as well as more complex geometric structures (e.g., non-Euclidean, fractal)
  - Discrete models can exhibit scalings, be amenable to renormalization approaches, and manifest cascade (direct and indirect), and hierarchical structures
  - *But* potential problems (e.g., log periodicity)

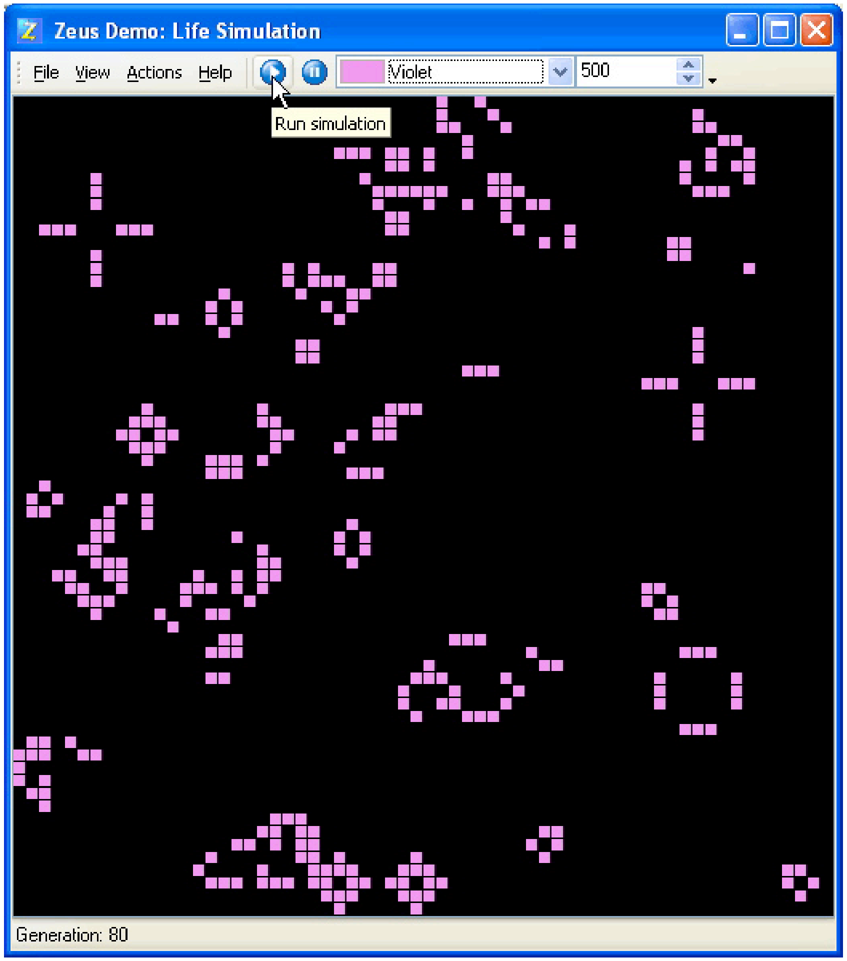
## **Motivation** (*cont.*)

- *Mathematics (pure and applied) and computation*
  - Problems in environmental and social sciences may be more readily described by discrete (*phenomenological*) models with quantitative treatments within grasp
  - Discrete models also validate need to develop new mathematical paradigms and tools
  - Discrete models also provide a testbed for new computational algorithms ($10^9$+ elements) and a path to new theorems (John von Neumann's view on computation)

## Cellular Automata and Percolation

- Cellular automata are generally associated with von Neumann's efforts in the 1940s to construct a "Turing machine," a hypothetical device using complex mathematical rules on a Cartesian grid that could build copies of itself
- Cambridge mathematician John Conway combined aspect of "Leech's problem" in group theory with von Neumann's self-replicating automata and devised the "Game of Life" (appearing in Oct. 1970 issue of *Scientific American*)
- Consists of a collection of "cells" which, based on a few mathematical rules and the initial conditions, can evolve complex patterns
- See following example: cellular automata's origins have a closer link to biology and forms foundation for "percolation theory"

# Cellular Automata and Percolation (*cont.*)

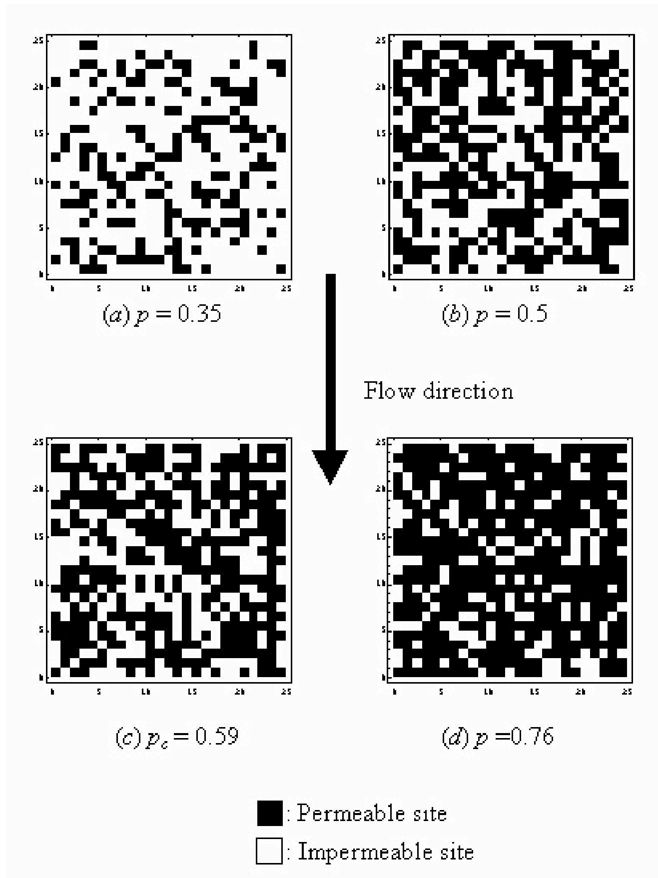



- Screen dump from *Game of Life*

## **Cellular Automata and Percolation** (*cont.*)

- Consider planting an orchard on a square grid
- Suppose that the trees are susceptible to a disease which can be transmitted to nearest-neighbor trees
- *What fraction of sites should have trees planted to avoid the possibility that the entire orchard succumbs to disease?*
- This question was first posed in the 1600's (R. Orbach).
- This corresponds to the "critical probability" $p_c \approx 0.592746$ for percolation across a (infinite) lattice.
- If the fraction (or probability) $p < p_c$, then the size of "clusters" of trees that could be claimed by disease remains finite.
- If $p = p_c$, then clusters can form that span entire lattice.
- Concepts relevant to conductivity and permeability.

## Cellular Automata and Percolation (*cont.*)




- Emergence of porosity/flow with increasing $p$ (through $p_c$)

## Statics vs. Dynamics: Self-Organized Criticality and Graphs

- To generate the previous picture, we chose a number $0 < p < 1$ for each panel
- Then, we regarded every site as being statistically independent and selected with uniform probability for each site a probability $0 \leq \mathcal{P} \leq \infty$
- If $\mathcal{P} \leq p$, we would say that the site was occupied; otherwise, we say that the site remained vacant.
- This is a time-independent or *static* means of determining the occupancy of sites
- We now want to develop a time-dependent scheme (in the spirit of the *ergodic hypothesis* in statistical physics)

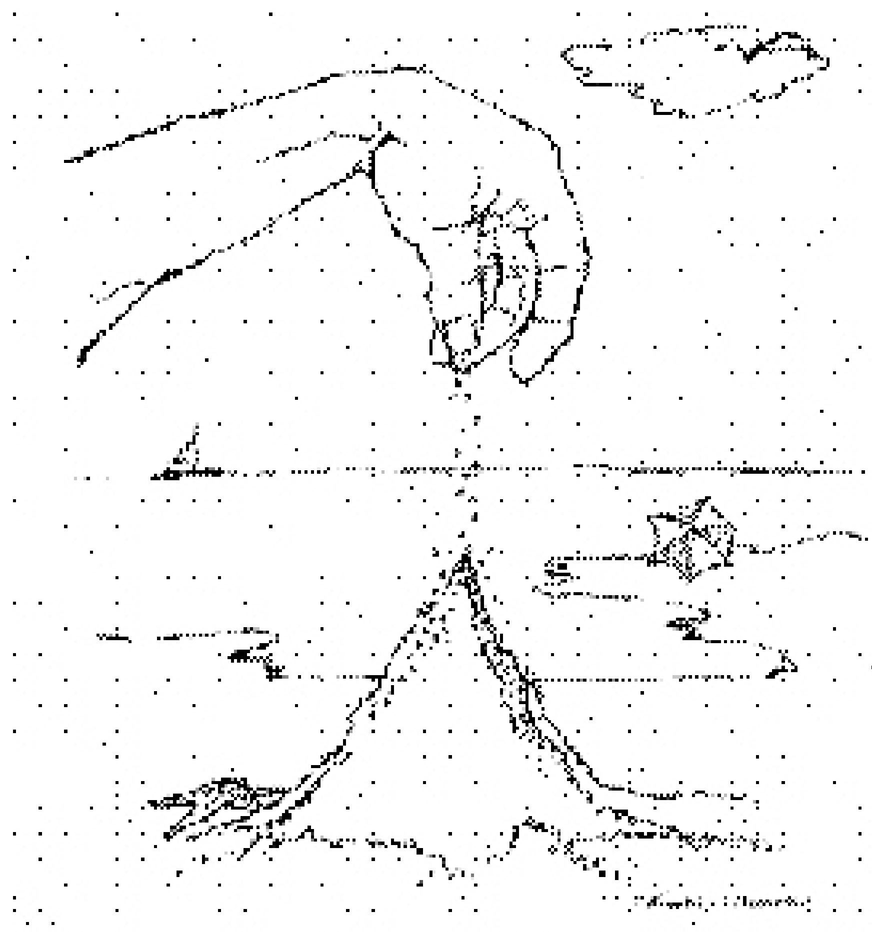
## **Statics vs. Dynamics** (*cont).*

- Let us assume that the grid is initially vacant
- Then, at each time step $n$, we select (with uniform probability) a site on the grid; if the site is vacant, then we occupy it but, if it is *not* vacant, then we do nothing
- We note that the total fraction $f$ of sites occupied grows as a Poissonian, namely $f = 1 - \exp\left[-n/N\right]$ where $N$ is the total number of sites in the grid
- It is easy to show, when $n \approx -N \ln\left[1 - p_c\right]$, a critical number of lattice sites will be occupied and percolation occurs
- This construction illustrates a simple device for evolving these systems (according to rules to be determined) in seeking to identify the nature of percolating or related phenomena

## Statics vs. Dynamics *(cont).*

- Return later to problem; how "renewal" processes can be used to model forest fires, earthquakes, and gang warfare
- We now briefly mention how this relates to self-organized criticality and the Bak-Tang-Wiesenfeld Sandpile Model
- Consider an empty 2-D array; at each time step, drop "grain of sand" onto randomly selected site, keeping track of number of grains that have accumulated on each site
- If a given site receives a $4^{\text{th}}$ grain, take all 4 grains and place 1 on each of 4 nearest-neighbor sites; if given site is on a boundary, then a transported grain falls off the edge
- If any of those nearest-neighbor sites has four grains, redistribute them onto their nearest neighbors, and so on
- "Avalanches" proxy for earthquakes?; Gutenburg-Richter law

## **Statics vs. Dynamics** *(cont).*



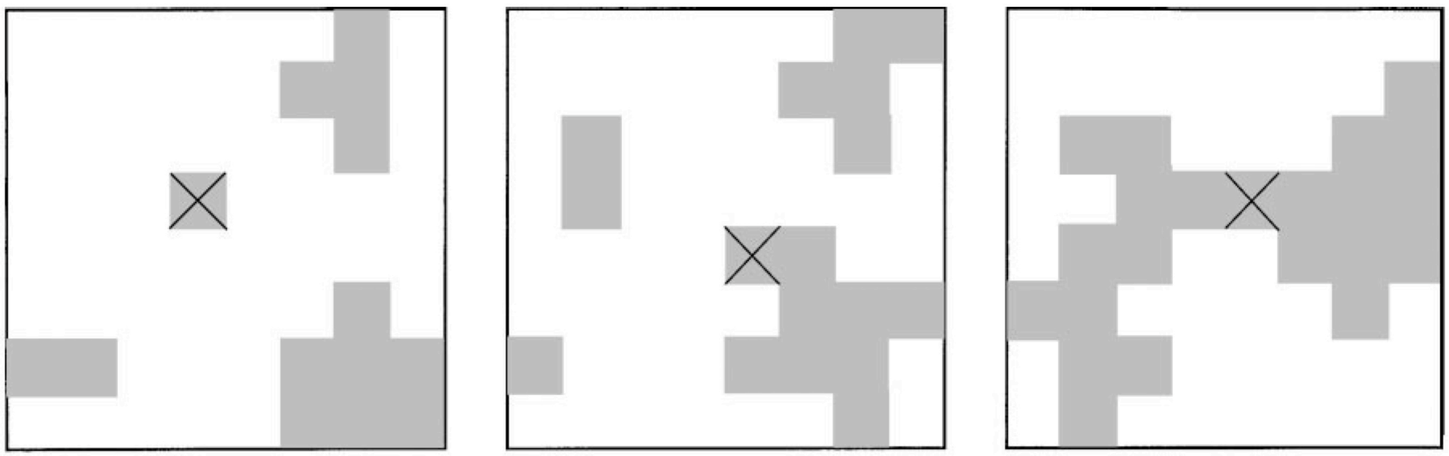
Bak-Tang-Wiesenfeld Sandpile Model

**Statics vs. Dynamics** *(cont).*

- At long time, the distribution of the number of grains at each site develops some intriguing behaviors, including fractal geometry, $1/f$ noise, power-law scaling in the Fourier spectrum
- The system is quite robust to parameter changes (e.g., system size); self-organization evident and many considered this a form of "critical behavior," hence self-organized criticality (SOC)
- Substantial debate has emerged whether this and related problems produce "critical behavior" or "complexity"
- SOC can produce avalanches, cascades, and other behavior reminiscent of more complicated (e.g., fluid) systems
- We turn now to the so-called "Forest Fire Model," originally claimed to be SOC by Bak and colleagues

## **Statics vs. Dynamics** *(cont).*

- Return to dynamic percolation model, i.e., sequential selection of unoccupied sites
- "Planting trees" (Forest Fire Model); gang recruitment?

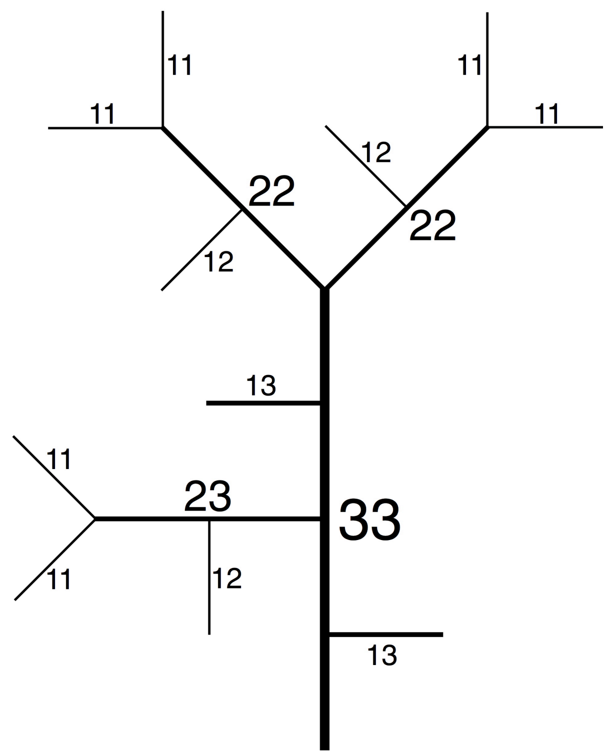


(a) (b) (c)

- Process results in formation of "clusters": (a) isolated clusters (with one element), (b) pre-existing clusters augmented by one tree, and (c) bridging of two pre-existing clusters

## **Statics vs. Dynamics** *(cont).*

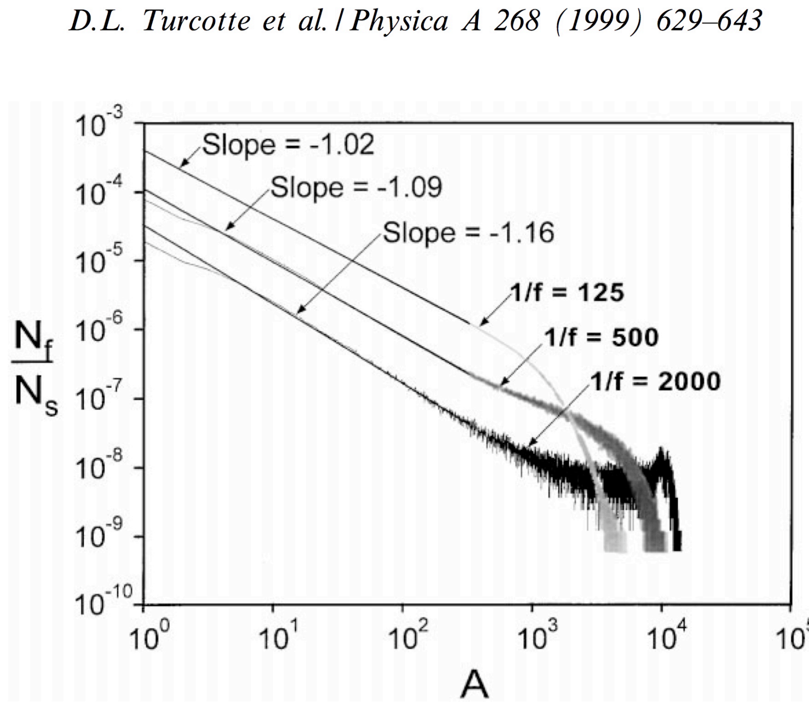
- Emergent clusters can be related to directed graphs; context of networks and classification schemes (applications to river systems and geomorphology)



## Forest Fires and Characteristic Earthquakes

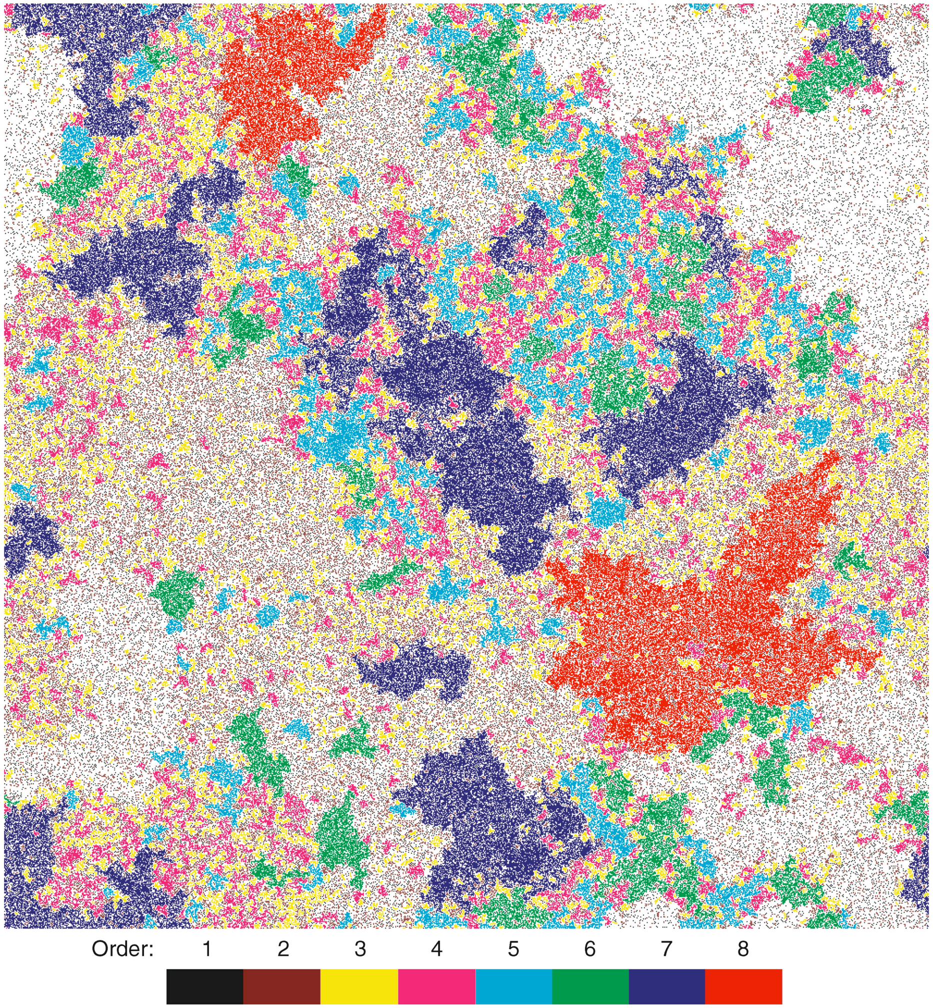
- Return to original percolation problem where we begin with an empty lattice and "plant" trees at randomly selected sites at each time step (if a "seed" lands on an occupied site, then nothing happens); relevance to real forests
- Suppose, now, you drop a match onto a randomly selected site on the grid with a *sparking frequency* $f$, i.e., one match is dropped for every $1/f$ trees planted
- This is the essence of the forest fire model, first introduced by Bak et al., which Bak claimed to be self-organized critical
- A magnitude-frequency figure (more correctly, a frequency-area burned plot) shows the emergence of power-laws that are independent of all parameters

# Forest Fires and Characteristic Earthquakes (*cont.*)




- Frequency-area statistics for Forest Fire model; observed power-law closely matches that seen in nature
- Graph-theoretic relationship for hierarchical clustering is also evident in the following figure

# Forest Fires and Characteristic Earthquakes (*cont.*)



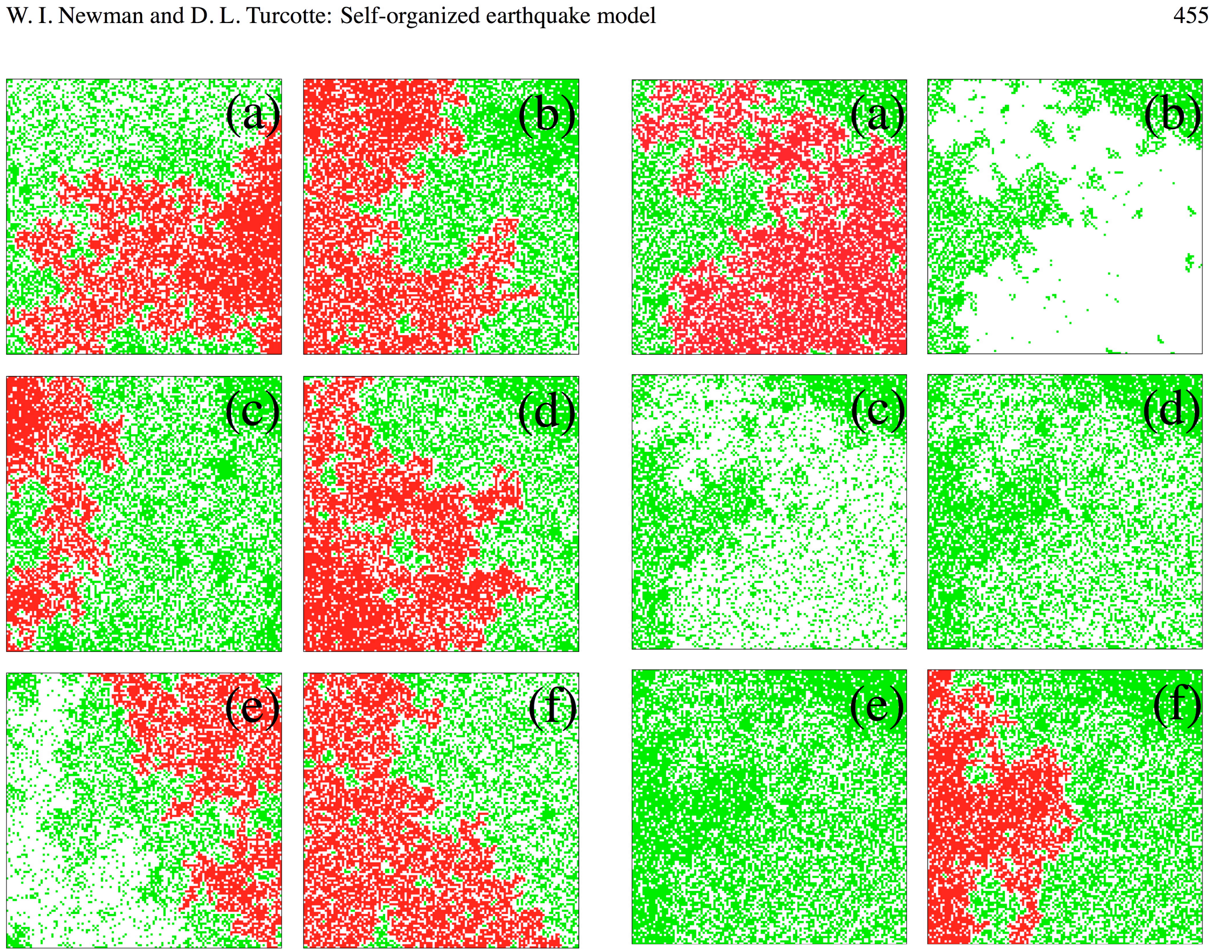
## Forest Fires and Characteristic Earthquakes (*cont.*)

- Gabrielov, Newman, and Turcotte (1999) *Phys. Rev. E* **60**, 5293. showed that SOC conjecture for forest fires is *not* correct and proved **theorem** showing emergence of power-laws
- Variations of model are relevant to other environmental or social science applications; they provide also a testbed for developing new mathematics, particularly relevant to combinatorics
- We wish to explore recurrence times between large earthquakes, so-called "characteristic earthquakes" that correspond to an entire region
- To model the latter, we consider the forest fire model modified so that "model fires" occur when percolation across the lattice occurs (not by triggering by a match)

**Forest Fires and Characteristic Earthquakes** (*cont.*)

- Newman and Turcotte (2002) *Nonl. Proc. Geophys.*, **9**, 453-461 explored this adaptation to recurrence of large-scale earthquakes
- Examples of percolating fires or characteristic earthquakes, as well as system between earthquakes
- Figures show role of memory in problem; statistics of events no longer expected to be Poissonian
- Power-law observed in cluster sizes
- Forecasting scheme developed from clustering hypothesis; possible applicability to real earthquakes

# Forest Fires and Characteristic Earthquakes (*cont.*)

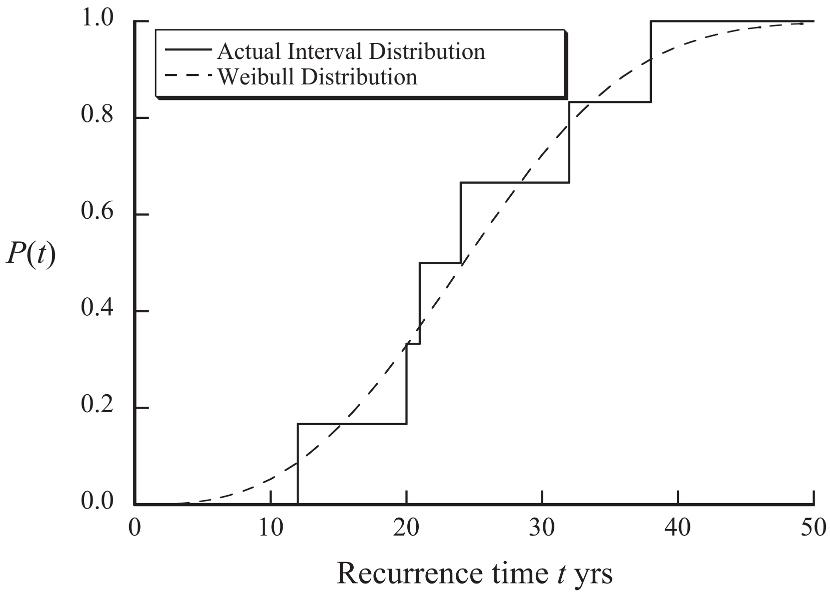


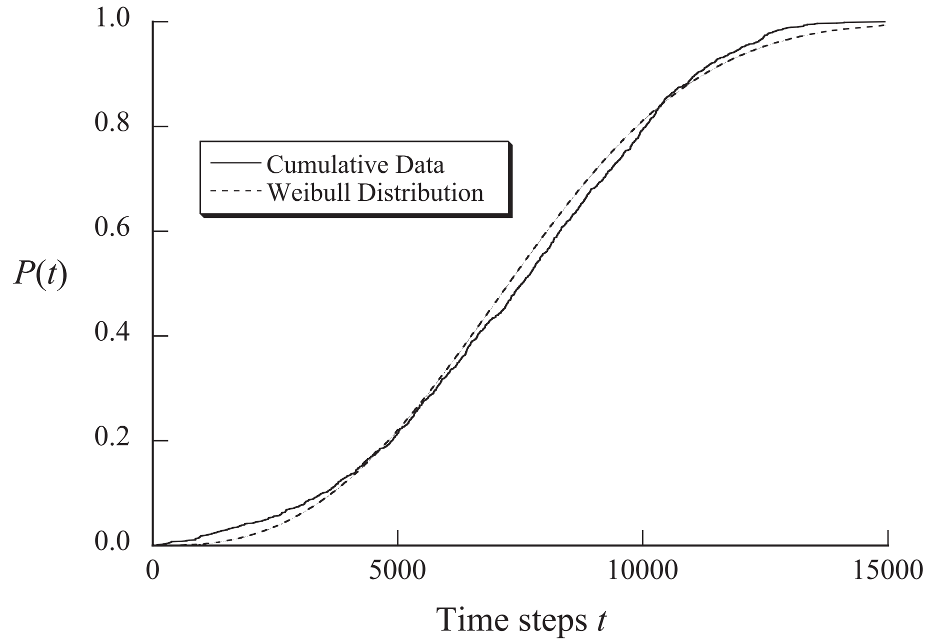

- Six realizations (left) and six stages of evolution (right)

## **Forest Fires and Characteristic Earthquakes** (*cont.*)

- *Apart from a power law frequency-magnitude relationship—akin to the empirical Gutenburg-Richter relation—does hybrid forest fire earthquake model reproduce any other observable?* What about the distribution of recurrence times between large earthquakes
- We show cumulative distribution functions for Parkfield earthquakes (near San Andreas north of LA) and our hybrid model
- Fits shown for Weibull dist. $P(t) = 1 - \exp\left[-(t/\tau)^{\beta}\right]$ where $\tau$ is related to the mean-recurrence time, and $\beta$ is 2.88 and 2.74 for the two respective cases
- Exploring relation between $\beta$ and criterion for earthquakes employed in model problem

# Forest Fires and Characteristic Earthquakes (*cont.*)

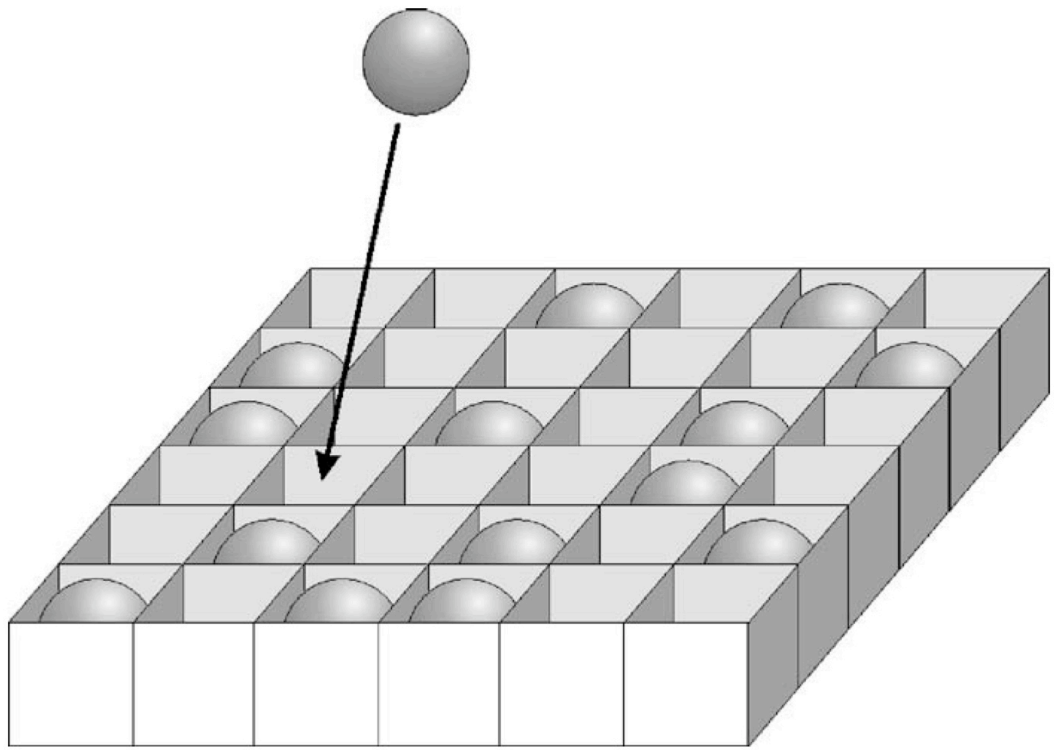



- Parkfield Earthquakes and Hybrid Forest Fire Model results
- Model inspired González, Gómez and Pacheco (2005) *Am. J. Phys.*, **73**, 946-952 to develop an "occupation of a box" as a toy model for seismicity

## Forest Fires and Characteristic Earthquakes (*cont.*)

- Unknown to them, their model is identical to the “coupon collectors problem” in combinatorics



## **Forest Fires and Characteristic Earthquakes** (*cont.*)

- It is *exactly* soluble (Newman et al., 2005) and, due to its linearity, has easily calculable asymptotics due to the Feller-Lindberg theorem (generalization of Central Limit Theorem) and yields normal distribution
- Other models have yielded other distributions, e.g., log-normal distribution
- A fundamental question remains: *Why Weibull?* This distills what is a question of fundamental importance in materials science (where Weibull is ubiquitous) as well as seismology (although not fully accepted)

## Gang-Warfare and Social Networks

- The previous discussion was largely devoid of any "physics"
- Could this imply that many environmental phenomena have a fundamental *geometric* underpinning?
- For example, the "$b$-value" problem of seismology, where the $b$ is the exponent in the observed Gutenberg-Richter power law which describes the frequency of seismic events as a function of their size—the exponent is essentially independent of the underlying geology, chemical makeup of underlying materials, geometry of faults and plate motion, etc.
- In our earlier discussion, we were depositing "stress" or "seeds" randomly on some field, awaiting their interaction, breakdown (earthquakes and fires), and renewal

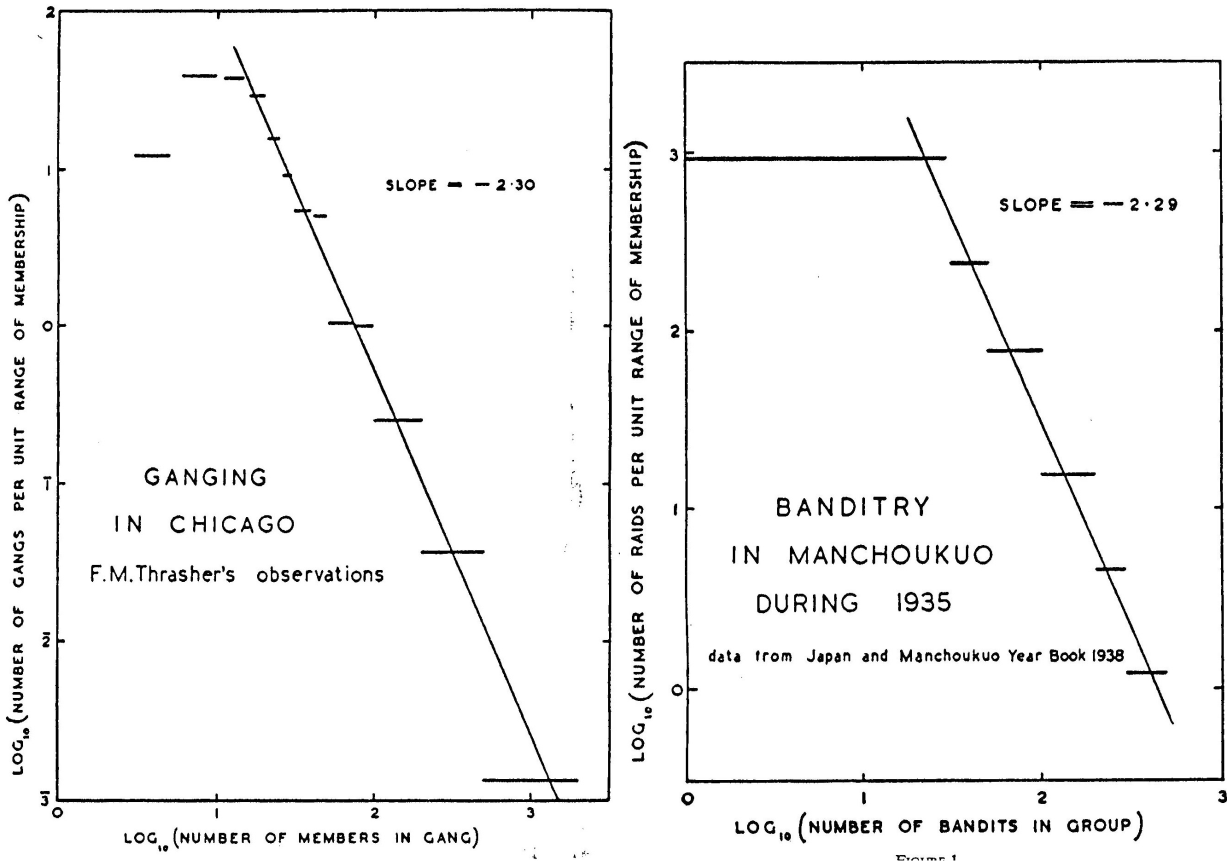
## **Gang-Warfare and Social Networks** (*cont.*)

- Could similar ideas apply to human interactions? Could gang warfare, as a lower-level description of class war, be modeled via cellular automata through a forest fire model? Could the role of "matches" be replaced by "police raids"?
- Finding data to explore on this topic is not simple; for example, reliable data for Los Angeles is non-existent or obfuscated to serve specific political agendae
- Lewis Fry Richardson (1881-1953) is best known as a hydrodynamicist and meteorologist; however, he is also well-known as a pacifist (as a Quaker, he volunteered to work in the Friends' Ambulance Service in the $16^{th}$ French Infantry Division) and a scientist who attempted to analyze the causes of war by mathematical means

## Gang-Warfare and Social Networks (*cont.*)

- While known, thanks to Benoit Mandelbrot, for quantifying the problem of identifying the coastline of Britain (owing to its fractal perimeter), Richardson began his investigation by exploring the probability of two nations going to war as a function of the length of their common border
- His work was published posthumously: L. F. Richardson (1960) *Statistics of Deadly Quarrels*, Pacific Grove, CA: Boxwood Press
- Richardson published data relating to gang warfare pertaining to Chicago gangland (Irish vs. Italian vs. Polish) & to Manchouko banditry (Japanese occupied Manchuria)
- For demographic as well as mathematical reasons, I believe that Los Angeles data (were it available) would be similar

# Gang-Warfare and Social Networks (*cont.*)




- The observed power laws are indistinguishable from those observed for real forest fires, and for the forest fire model

## Conclusions

- Cellular automata have enjoyed a rich history in pure mathematics and computer science, and is poised to make important progress in applications
- Discrete models permit the exploration of phenomena embracing scaling, hierarchical structures, cascades and avalanches, combinatorics, and graph theoretic structures that are largely inaccessible to ODE/PDE-derived methodologies
- Cellular automata will provide insights into many problems in the environmental and social sciences, as well as new mathematics, in the years to come