Contagion

Complex Networks, CSYS/MATH 303, Spring, 2010

Prof. Peter Dodds

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont

Outline

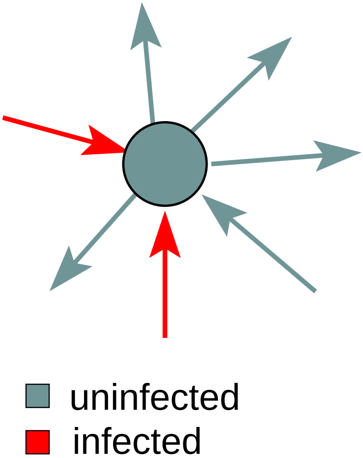
Contagion models

Some large questions concerning network contagion:

1. For a given spreading mechanism on a given network, what's the probability that there will be global spreading?
2. If spreading does take off, how far will it go?
3. How do the details of the network affect the outcome?
4. How do the details of the spreading mechanism affect the outcome?
5. What if the seed is one or many nodes?

- Next up: We'll look at some fundamental kinds of spreading on generalized random networks.

Spreading mechanisms

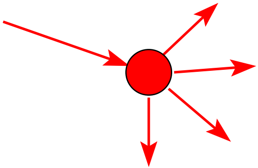


- General spreading mechanism:
  State of node *i* depends on history of *i* and *i*'s neighbors' states.
- Doses of entity may be stochastic and history-dependent.
- May have multiple, interacting entities spreading at once.

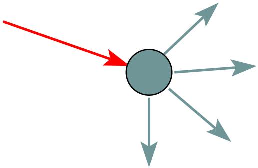
Spreading on Random Networks

- For random networks, we know local structure is pure branching.
- Successful spreading is $\therefore$ contingent on single edges infecting nodes.

Success



Failure:



- Focus on binary case with edges and nodes either infected or not.

Contagion condition

- We need to find:
  r = the average # of infected edges that one random infected edge brings about.
- Define β_k as the probability that a node of degree k is infected by a single infected edge.
- $$r = \sum_{k=0}^{\infty} \underbrace{\frac{kP_k}{\langle k \rangle}}_{\substack{\text{prob. of} \\ \text{connecting to} \\ \text{a degree } k \text{ node}}} \cdot \underbrace{\beta_k}_{\substack{\text{Prob. of} \\ \text{infection}}} \cdot \underbrace{(k-1)}_{\substack{\text{\# outgoing} \\ \text{infected} \\ \text{edges}}}$$
  $$+ \sum_{k=0}^{\infty} \overbrace{\frac{kP_k}{\langle k \rangle}} \cdot \underbrace{(1-\beta_k)}_{\substack{\text{Prob. of} \\ \text{no infection}}} \cdot \underbrace{0}_{\substack{\text{\# outgoing} \\ \text{infected} \\ \text{edges}}}$$

Contagion condition

- Our contagion condition is then:

$$\boxed{r = \sum_{k=0}^{\infty} \frac{(k-1)kP_k}{\langle k \rangle} \beta_k > 1.}$$

- Case 1: If $\beta_k = 1$ then

$$r = \frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1.$$

- Good: This is just our giant component condition again.

Contagion condition

- Case 2: If $\beta_k = \beta < 1$ then

$$r = \beta \frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1.$$

- A fraction (1-β) of edges do not transmit infection.
- Analogous phase transition to giant component case but critical value of $\langle k \rangle$ is increased.
- Aka bond percolation.
- Resulting degree distribution P'_k:

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

- We can show $F_{P'}(x) = F_P(\beta x + 1 - \beta)$.

Contagion condition

- Cases 3, 4, 5, ...: Now allow β_k to depend on k
- Asymmetry: Transmission along an edge depends on node's degree at other end.
- Possibility: β_k increases with k... unlikely.
- Possibility: β_k is not monotonic in k... unlikely.
- Possibility: β_k decreases with k... hmmm.
- $\beta_k \searrow$ is a plausible representation of a simple kind of social contagion.
- The story:
  More well connected people are harder to influence.

Contagion condition

- Example: $\beta_k = 1/k$.
- $$r = \sum_{k=1}^{\infty} \frac{(k-1)kP_k}{\langle k \rangle} \beta_k = \sum_{k=1}^{\infty} \frac{(k-1)kP_k}{\langle k \rangle k}$$
  $$= \sum_{k=1}^{\infty} \frac{(k-1)P_k}{\langle k \rangle} = \frac{\langle k \rangle - 1}{\langle k \rangle} = 1 - \frac{1}{\langle k \rangle}$$
- Since r is always less than 1, no spreading can occur for this mechanism.
- Decay of β_k is too fast.
- Result is independent of degree distribution.

Contagion condition

- Example: $\beta_k = H(\frac{1}{k} - \phi)$
  where $0 < \phi \leq 1$ is a threshold and H is the Heaviside function.
- Infection only occurs for nodes with low degree.
- Call these nodes vulnerables:
  they flip when only one of their friends flips.
-

$$r = \sum_{k=1}^{\infty} \frac{(k-1)kP_k}{\langle k \rangle} \beta_k = \sum_{k=1}^{\infty} \frac{(k-1)kP_k}{\langle k \rangle} H(\frac{1}{k} - \phi)$$

$$= \sum_{k=1}^{\lfloor \frac{1}{\phi} \rfloor} \frac{(k-1)kP_k}{\langle k \rangle} \quad \text{where } \lfloor \cdot \rfloor \text{ means floor.}$$

Contagion condition

- The contagion condition:

$$r = \sum_{k=1}^{\lfloor \frac{1}{\phi} \rfloor} \frac{(k-1)kP_k}{\langle k \rangle} > 1.$$

- As $\phi \rightarrow 1$, all nodes become resilient and $r \rightarrow 0$.
- As $\phi \rightarrow 0$, all nodes become vulnerable and the contagion condition matches up with the giant component condition.
- Key: If we fix ϕ and then vary $\langle k \rangle$, we may see two phase transitions.
- Added to our standard giant component transition, we will see a cut off in spreading as nodes become more connected.

Social Contagion

Some important models (recap from CSYS 300)

- Tipping models—Schelling (1971) [8, 9, 10]
  - Simulation on checker boards.
  - Idea of thresholds.
- Threshold models—Granovetter (1978) [7]
- Herding models—Bikhchandani et al. (1992) [1, 2]
  - Social learning theory, Informational cascades,...

Threshold model on a network

Original work:

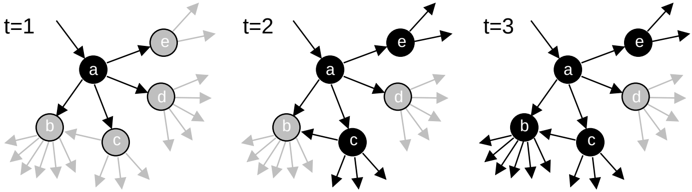
"A simple model of global cascades on random networks"
D. J. Watts. Proc. Natl. Acad. Sci., 2002 [12]

- Mean field Granovetter model → network model
- Individuals now have a limited view of the world

Threshold model on a network

- Interactions between individuals now represented by a network
- Network is sparse
- Individual i has k_i contacts
- Influence on each link is reciprocal and of unit weight
- Each individual i has a fixed threshold ϕ_i
- Individuals repeatedly poll contacts on network
- Synchronous, discrete time updating
- Individual i becomes active when fraction of active contacts $a_i \geq \phi_i k_i$
- Activation is permanent (SI)

Threshold model on a network



- All nodes have threshold $\phi = 0.2$.

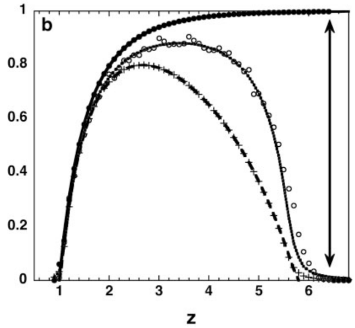
The most gullible

Vulnerables:

- Recall definition: individuals who can be activated by just one contact being active are vulnerables.
- The vulnerability condition for node i: $1/k_i \geq \phi_i$.
- Means # contacts $k_i \leq \lfloor 1/\phi_i \rfloor$.
- Key: For global cascades on random networks, must have a *global component of vulnerables* [12]
- For a uniform threshold ϕ, our contagion condition tells us when such a component exists:

$$r = \sum_{k=1}^{\lfloor \frac{1}{\phi} \rfloor} \frac{(k-1)kP_k}{\langle k \rangle} > 1.$$

Cascades on random networks

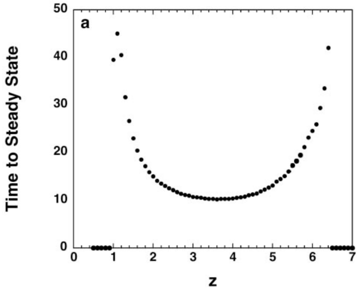


(n.b., $z = \langle k \rangle$)

- Top curve: final fraction infected if successful.
- Middle curve: chance of starting a global spreading event (cascade).
- Bottom curve: fractional size of vulnerable subcomponent. [12]

- Cascades occur only if size of vulnerable subcomponent > 0.
- System is robust-yet-fragile just below upper boundary [3, 4, 11]
- 'Ignorance' facilitates spreading.

Cascades on random networks

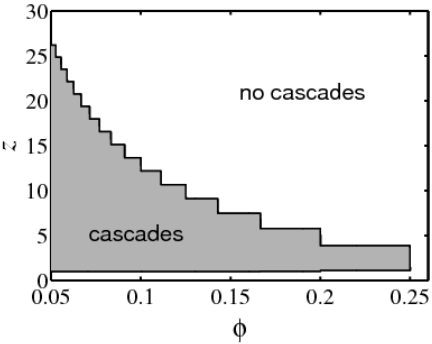


- Time taken for cascade to spread through network. [12]
- Two phase transitions.

(n.b., $z = \langle k \rangle$)

- Largest vulnerable component = critical mass.
- Now have endogenous mechanism for spreading from an individual to the critical mass and then beyond.

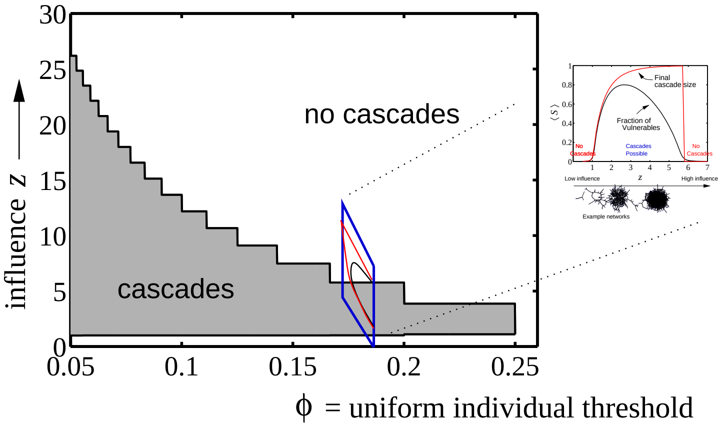
Cascade window for random networks



(n.b., $z = \langle k \rangle$)

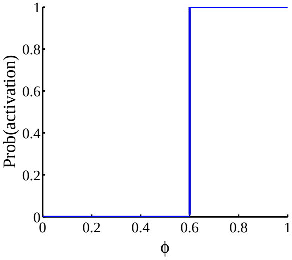
- Outline of cascade window for random networks.

Cascade window for random networks



Social Contagion

Granovetter's Threshold model—recap



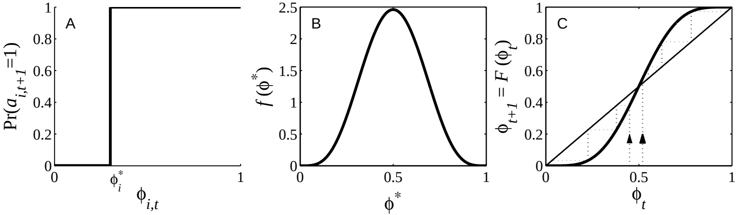
- Assumes deterministic response functions
- ϕ_* = threshold of an individual.
- $f(\phi_*)$ = distribution of thresholds in a population.
- $F(\phi_*)$ = cumulative distribution = $\int_{\phi'_*=0}^{\phi_*} f(\phi'_*)\mathrm{d}\phi'_*$
- ϕ_t = fraction of people 'rioting' at time step t.

Social Sciences—Threshold models

- At time $t+1$, fraction rioting = fraction with $\phi_* \le \phi_t$.
- $$\phi_{t+1} = \int_0^{\phi_t} f(\phi_*)\mathrm{d}\phi_* = F(\phi_*)|_0^{\phi_t} = F(\phi_t)$$
- $\Rightarrow$ Iterative maps of the unit interval $[0, 1]$.

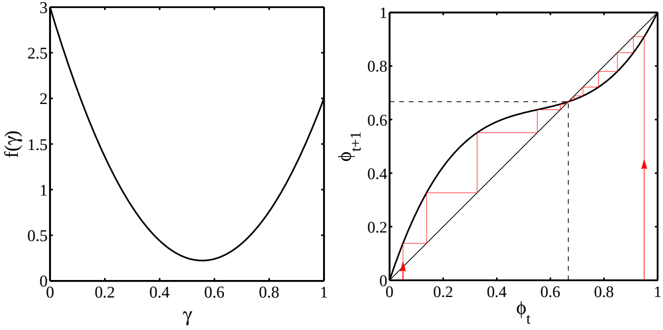
Social Sciences—Threshold models

Action based on perceived behavior of others.



- Two states: S and I
- Recover now possible (SIS)
- ϕ = fraction of contacts 'on' (e.g., rioting)
- Discrete time, synchronous update (strong assumption!)
- This is a Critical mass model

Social Sciences—Threshold models



- Example of single stable state model

Social Sciences—Threshold models

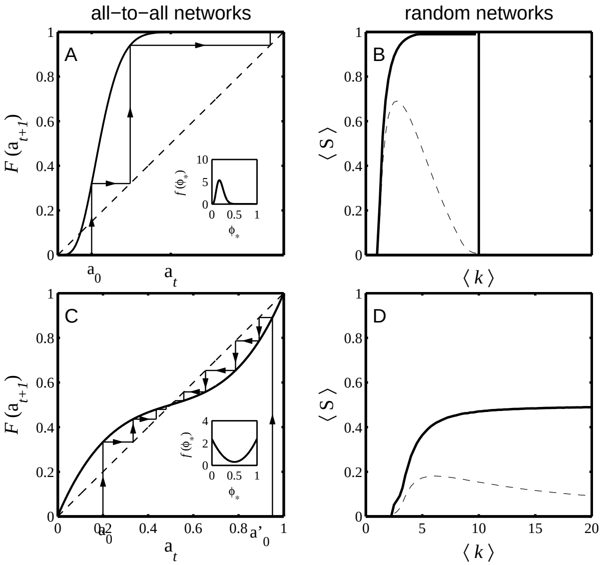
Implications for collective action theory:

1. Collective uniformity $\not\Rightarrow$ individual uniformity
2. Small individual changes $\Rightarrow$ large global changes

Next:

- Connect mean-field model to network model.
- Single seed for network model: $1/N \rightarrow 0$.
- Comparison between network and mean-field model sensible for vanishing seed size for the latter.

All-to-all versus random networks



Threshold contagion on random networks

Three key pieces to describe analytically:

1. The fractional size of the largest subcomponent of vulnerable nodes, S_{vuln}.
2. The chance of starting a global spreading event, $P_{\text{trig}} = S_{\text{trig}}$.
3. The expected final size of any successful spread, S.
    - n.b., the distribution of S is almost always bimodal.

Threshold contagion on random networks

- First goal: Find the largest component of vulnerable nodes.
- Recall that for finding the giant component's size, we had to solve:

$$F_\pi(x) = xF_P\left(F_\rho(x)\right) \text{ and } F_\rho(x) = xF_R\left(F_\rho(x)\right)$$

- We'll find a similar result for the subset of nodes that are vulnerable.
- This is a node-based percolation problem.
- For a general monotonic threshold distribution $f(\phi)$, a degree k node is vulnerable with probability

$$\beta_k = \int_0^{1/k} f(\phi)\mathrm{d}\phi\,.$$

Threshold contagion on random networks

- Everything now revolves around the modified generating function:

$$F_P^{(\mathrm{vuln})}(x) = \sum_{k=0}^{\infty} \beta_k P_k x^k.$$

- Generating function for friends-of-friends distribution is related in same way as before:

$$F_R^{(\mathrm{vuln})}(x) = \frac{\frac{\mathrm{d}}{\mathrm{d}x} F_P^{(\mathrm{vuln})}(x)}{\frac{\mathrm{d}}{\mathrm{d}x} F_P^{(\mathrm{vuln})}(x)|_{x=1}}.$$

Threshold contagion on random networks

- Functional relations for component size g.f.'s are almost the same...

$$F_{\pi}^{(\mathrm{vuln})}(x) = \underbrace{1 - F_{P}^{(\mathrm{vuln})}(1)}_{\substack{\text{central node}\\ \text{is not}\\ \text{vulnerable}}} + xF_{P}^{(\mathrm{vuln})}\left(F_{\rho}^{(\mathrm{vuln})}(x)\right)$$

$$F_{\rho}^{(\mathrm{vuln})}(x) = \underbrace{1 - F_{R}^{(\mathrm{vuln})}(1)}_{\substack{\text{first node}\\ \text{is not}\\ \text{vulnerable}}} + xF_{R}^{(\mathrm{vuln})}\left(F_{\rho}^{(\mathrm{vuln})}(x)\right)$$

- Can now solve as before to find $S_{\mathrm{vuln}} = 1 - F_{\pi}^{(\mathrm{vuln})}(1)$.

Threshold contagion on random networks

- Second goal: Find probability of triggering largest vulnerable component.
- Assumption is first node is randomly chosen.
- Same set up as for vulnerable component except now we don't care if the initial node is vulnerable or not:

$$F_{\pi}^{(\mathrm{trig})}(x) = x F_P \left(F_{\rho}^{(\mathrm{vuln})}(x) \right)$$

$$F_{\rho}^{(\mathrm{vuln})}(x) = 1 - F_R^{v)}(1) + x F_R^{(\mathrm{vuln})} \left(F_{\rho}^{(\mathrm{vuln})}(x) \right)$$

- Solve as before to find $P_{\mathrm{trig}} = S_{\mathrm{trig}} = 1 - F_{\pi}^{(\mathrm{trig})}(1)$.

Threshold contagion on random networks

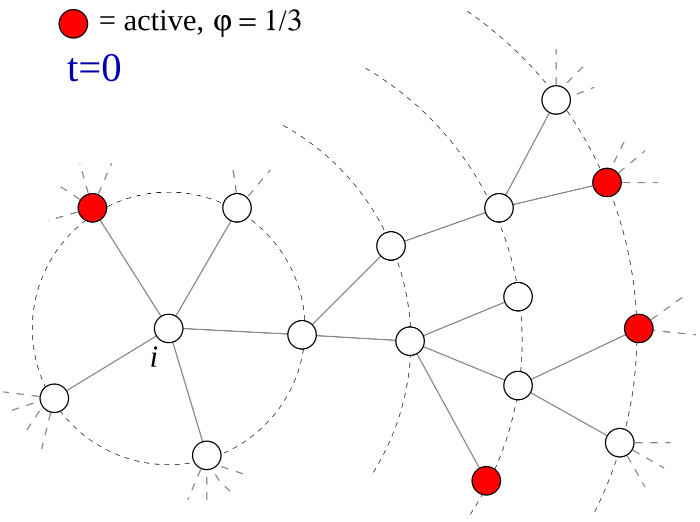
- Third goal: Find expected fractional size of spread.
- Not obvious even for uniform threshold problem.
- Difficulty is in figuring out if and when nodes that need ≥ 2 hits switch on.
- Problem solved for infinite seed case by Gleeson and Cahalane:
  "Seed size strongly affects cascades on random networks," Phys. Rev. E, 2007. [6]
- Developed further by Gleeson in "Cascades on correlated and modular random networks," Phys. Rev. E, 2008. [5]

Expected size of spread

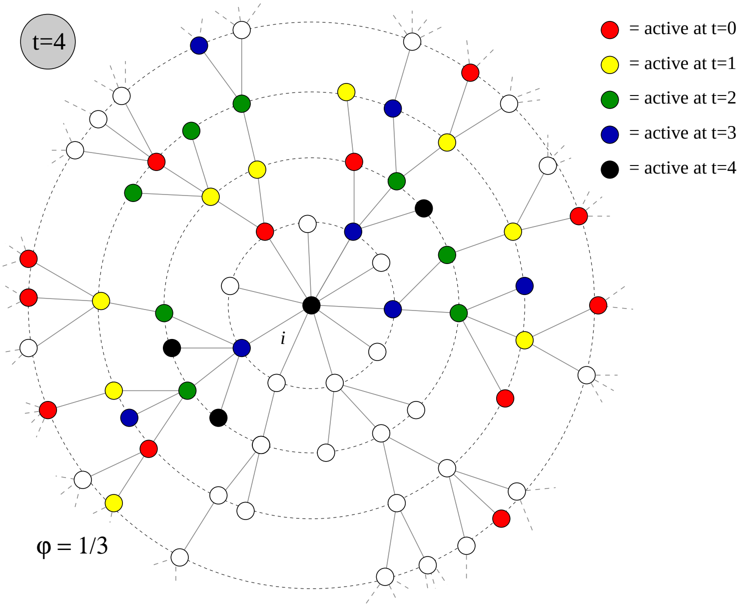
Idea:

- Randomly turn on a fraction ϕ_0 of nodes at time $t = 0$
- Capitalize on local branching network structure of random networks (again)
- Now think about what must happen for a specific node *i* to become active at time *t*:
- $t = 0$: *i* is one of the seeds (prob = ϕ_0)
- $t = 1$: *i* was not a seed but enough of *i*'s friends switched on at time $t = 0$ so that *i*'s threshold is now exceeded.
- $t = 2$: enough of *i*'s friends and friends-of-friends switched on at time $t = 0$ so that *i*'s threshold is now exceeded.
- $t = n$: enough nodes within *n* hops of *i* switched on at $t = 0$ and their effects have propagated to reach *i*.

Expected size of spread



Expected size of spread



Expected size of spread

Notes:

- Calculations are possible nodes do not become inactive.
- Not just for threshold model—works for a wide range of contagion processes.
- We can analytically determine the entire time evolution, not just the final size.
- We can in fact determine **Pr**(node of degree k switches on at time t).
- Asynchronous updating can be handled too.

Expected size of spread

Pleasantness:

- Taking off from a single seed story is about expansion away from a node.
- Extent of spreading story is about contraction at a node.

Expected size of spread

- Notation: **Pr**(node i becomes active at time t) $= \phi_{i,t}$.
- Notation: $\beta_{kj} =$ **Pr** (a degree k node becomes active if j neighbors are active).
- Our starting point: $\phi_{i,0} = \phi_0$.
- $\binom{k_i}{j}\phi_0^j(1-\phi_0)^{k_i-j}$ = **Pr** (j of node i's k_i neighbors were seeded at time $t=0$).
- Probability node i was a seed at $t=0$ is ϕ_0 (as above).
- Probability node i was not a seed at $t=0$ is $(1-\phi_0)$.
- Combining everything, we have:

$$\phi_{i,1} = \phi_0 + (1-\phi_0)\sum_{j=0}^{k_i}\binom{k_i}{j}\phi_0^j(1-\phi_0)^{k_i-j}\beta_{k_ij}.$$

Expected size of spread

- For general t, we need to know the probability an edge coming into node i at time t is active.
- Notation: call this probability θ_t.
- We already know $\theta_0 = \phi_0$.
- Story analogous to $t = 1$ case:

$$\phi_{i,t+1} = \phi_0 + (1 - \phi_0) \sum_{j=0}^{k_i} \binom{k_i}{j} \theta_t^j (1 - \theta_t)^{k_i - j} \beta_{k_i j}.$$

- Average over all nodes to obtain expression for ϕ_{t+1}:

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

- So we need to compute θ_t... massive excitement...

Expected size of spread

First connect θ_0 to θ_1:

- $\theta_1 = \phi_0 +$

$$(1-\phi_0)\sum_{k=1}^{\infty}\frac{kP_k}{\langle k\rangle}\sum_{j=0}^{k-1}\binom{k-1}{j}\theta_0^{\,j}(1-\theta_0)^{k-1-j}\beta_{kj}$$

- $\frac{kP_k}{\langle k\rangle} = R_k$ = **Pr** (edge connects to a degree k node).
- $\sum_{j=0}^{k-1}$ piece gives **Pr**(degree node k activates) of its neighbors $k-1$ incoming neighbors are active.
- ϕ_0 and $(1-\phi_0)$ terms account for state of node at time $t = 0$.
- See this all generalizes to give θ_{t+1} in terms of θ_t...

Expected size of spread

Two pieces:

1. $\theta_{t+1} = \phi_0 +$

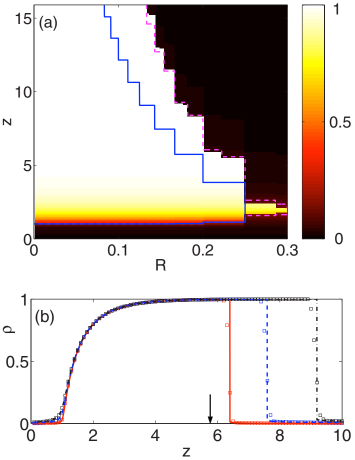
$$(1-\phi_0)\sum_{k=1}^{\infty}\frac{kP_k}{\langle k\rangle}\sum_{j=0}^{k-1}\binom{k-1}{j}\theta_t^{\,j}(1-\theta_t)^{k-1-j}\beta_{kj}$$

   with $\theta_0 = \phi_0$.

2. $\phi_{t+1} = \phi_0 +$

$$(1-\phi_0)\sum_{k=0}^{\infty}P_k\sum_{j=0}^{k}\binom{k}{j}\theta_t^{\,j}(1-\theta_t)^{k-j}\beta_{kj}.$$

Comparison between theory and simulations



From Gleeson and Cahalane [6]

- Pure random networks with simple threshold responses
- R = uniform threshold (our ϕ_*); z = average degree; $\rho = \phi$; $q = \theta$; $N = 10^5$.
- $\phi_0 = 10^{-3}$, 0.5×10^{-2}, and 10^{-2}.
- Cascade window is for $\phi = 10^{-2}$ case.
- Sensible expansion of cascade window as ϕ_0 increases.

Notes:

- Retrieve cascade condition for spreading from a single seed in limit $\phi_0 \rightarrow 0$.
- Depends on map $\theta_{t+1} = G(\theta_t; \phi_0)$.
- First: if self-starters are present, some activation is assured:

$$G(0;\phi_0) = \sum_{k=1}^{\infty} \frac{kP_k}{\langle k \rangle} \beta_{k0} > 0.$$

  meaning $\beta_{k0} > 0$ for at least one value of $k \geq 1$.
- If $\theta = 0$ is a fixed point of G (i.e., $G(0;\phi_0) = 0$) then spreading occurs if

$$G'(0;\phi_0) = \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} (k-1)kP_k \beta_{k1} > 1.$$

Insert question from assignment 6 (⊞)

Notes:

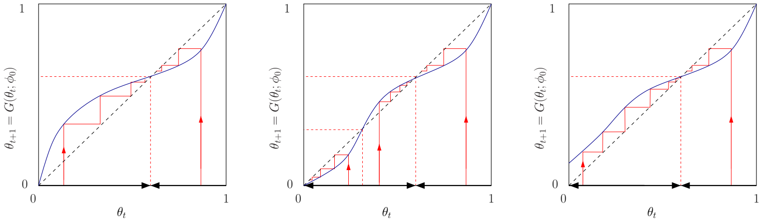
In words:

- If $G(0;\phi_0) > 0$, spreading must occur because some nodes turn on for free.
- If G has an unstable fixed point at $\theta = 0$, then cascades are also always possible.

Non-vanishing seed case:

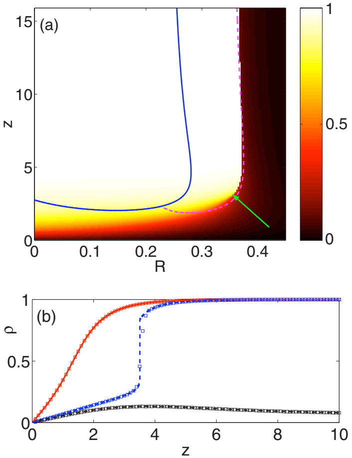
- Cascade condition is more complicated for $\phi_0 > 0$.
- If G has a stable fixed point at $\theta = 0$, and an unstable fixed point for some $0 < \theta_* < 1$, then for $\theta_0 > \theta_*$, spreading takes off.
- Tricky point: G depends on ϕ_0, so as we change ϕ_0, we also change G.

General fixed point story:



- Given $\theta_0 (= \phi_0)$, θ_∞ will be the nearest stable fixed point, either above or below.
- n.b., adjacent fixed points must have opposite stability types.
- Important: Actual form of G depends on ϕ_0.
- So choice of ϕ_0 dictates both G and starting point—can't start anywhere for a given G.

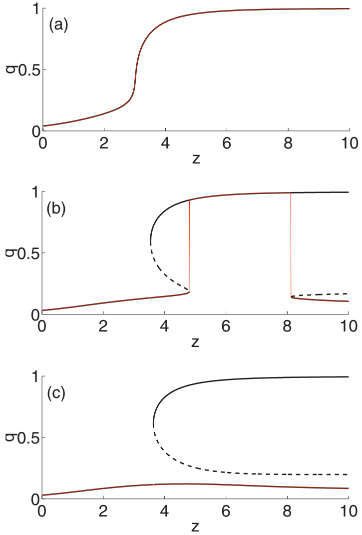
Comparison between theory and simulations



From Gleeson and Cahalane [6]

- Now allow thresholds to be distributed according to a Gaussian with mean R.
- R = 0.2, 0.362, and 0.38; $\sigma = 0.2$.
- $\phi_0 = 0$ but some nodes have thresholds ≤ 0 so effectively $\phi_0 > 0$.
- Now see a (nasty) discontinuous phase transition for low $\langle k \rangle$.

Comparison between theory and simulations



From Gleeson and Cahalane [6]

- Plots of stability points for $\theta_{t+1} = G(\theta_t; \phi_0)$.
- n.b.: 0 is not a fixed point here: $\theta_0 = 0$ always takes off.
- Top to bottom: R = 0.35, 0.371, and 0.375.
- n.b.: higher values of θ_0 for (b) and (c) lead to higher fixed points of G.
- Saddle node bifurcations appear and merge (b and c).

Spreadarama

Bridging to single seed case:

- Consider largest vulnerable component as initial set of seeds.
- Not quite right as spreading must move through vulnerables.
- But we can usefully think of the vulnerable component as activating at time $t = 0$ because order doesn't matter.
- Rebuild ϕ_t and θ_t expressions...

Spreadarama

Two pieces modified for single seed:

1. $\theta_{t+1} = \theta_{\text{vuln}} +$

$$(1 - \theta_{\text{vuln}}) \sum_{k=1}^{\infty} \frac{kP_k}{\langle k \rangle} \sum_{j=0}^{k-1} \binom{k-1}{j} \theta_t^{\,j} (1-\theta_t)^{k-1-j} \beta_{kj}$$

   with $\theta_0 = \theta_{\text{vuln}}$ = **Pr** an edge leads to the giant vulnerable component (if it exists).

2. $\phi_{t+1} = S_{\text{vuln}} +$

$$(1 - S_{\text{vuln}}) \sum_{k=0}^{\infty} P_k \sum_{j=0}^{k} \binom{k}{j} \theta_t^{\,j} (1-\theta_t)^{k-j} \beta_{kj}.$$

Time-dependent solutions

Synchronous update

- Done: Evolution of ϕ_t and θ_t given exactly by the maps we have derived.

Asynchronous updates

- Update nodes with probability α.
- As $\alpha \to 0$, updates become effectively independent.
- Now can talk about $\phi(t)$ and $\theta(t)$.
- More on this later...

References I

[1] S. Bikhchandani, D. Hirshleifer, and I. Welch.
A theory of fads, fashion, custom, and cultural change as informational cascades.
J. Polit. Econ., 100:992–1026, 1992.

[2] S. Bikhchandani, D. Hirshleifer, and I. Welch.
Learning from the behavior of others: Conformity, fads, and informational cascades.
J. Econ. Perspect., 12(3):151–170, 1998. pdf (⊞)

[3] J. Carlson and J. Doyle.
Highly optimized tolerance: A mechanism for power laws in design systems.
Phys. Rev. E, 60(2):1412–1427, 1999. pdf (⊞)

References II

[4] J. Carlson and J. Doyle.
Highly optimized tolerance: Robustness and design in complex systems.
Phys. Rev. Lett., 84(11):2529–2532, 2000. pdf (⊞)

[5] J. P. Gleeson.
Cascades on correlated and modular random networks.
Phys. Rev. E, 77:046117, 2008. pdf (⊞)

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

[7] M. Granovetter.
Threshold models of collective behavior.
Am. J. Sociol., 83(6):1420–1443, 1978. pdf (⊞)

References III

[8] T. Schelling.
Dynamic models of segregation.
J. Math. Sociol., 1:143–186, 1971.

[9] T. C. Schelling.
Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities.
J. Conflict Resolut., 17:381–428, 1973. pdf (⊞)

[10] T. C. Schelling.
Micromotives and Macrobehavior.
Norton, New York, 1978.

[11] D. Sornette.
Critical Phenomena in Natural Sciences.
Springer-Verlag, Berlin, 2nd edition, 2003.

References IV

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