

Cascades and Social Influence on Networks

Mason A. Porter

Mathematical Institute, University of Oxford

Collaborators: Davide Cellai, James Gleeson, Adam Hackett, **Sergey Melnik**, Peter Mucha, JP Onnela, Felix Reed-Tsochas, Jonathan Ward

Outline

- Introduction
- Modeling Social Influence
- Multi-Stage Complex Contagions
- Interlude: Locally Tree-Like Approximations and Mean-Field Theories
- Analytics for Multi-Stage Complex Contagions
- Briefly: Dynamics on Modular Networks with Heterogeneous Correlations
- Briefly: A Model for Popularity Cascades on Facebook
- Conclusions

Familiar: ODEs and PDEs

- Example: a toy model for a biological epidemic in a well-mixed population (SIR)

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS \\ \frac{dI}{dt} &= \beta IS - \nu I \\ \frac{dR}{dt} &= \nu I\end{aligned}$$

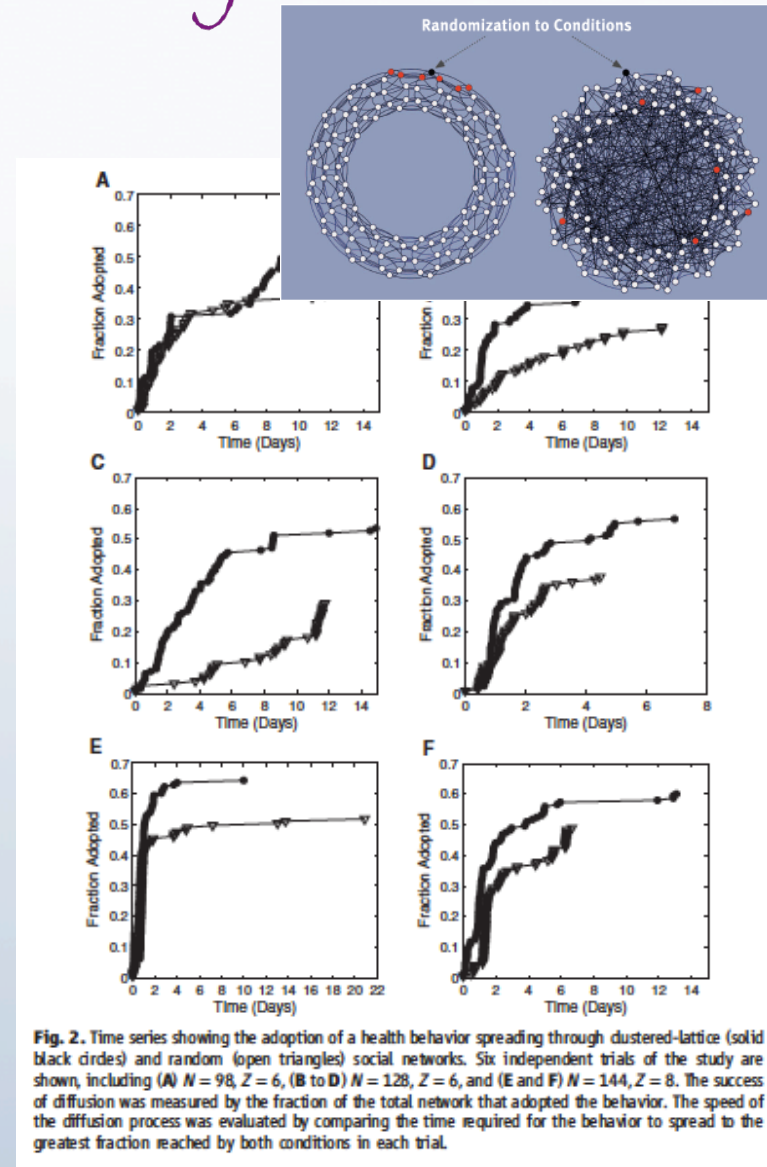
- Can incorporate space by adding diffusion terms (Δ).

Dynamics on Networks

- Incorporate which people (nodes) interact with which other people via their ties (edges).
- This yields a dynamical system on a network.
- **Basic question:** How does network structure affect dynamics (and vice versa)?

Cascades on Networks: Why Bother?

- Math
 - Provide interesting and tractable examples to study effects of network structure on dynamical systems
- Real World
 - "It's a nice place to visit."
 - "Complex contagions" versus "simple contagions"
 - Social reinforcement
 - E.g. Centola, *Science*, 2010
 - Social influence
 - E.g. mass movements
 - E.g. memes on Facebook and Twitter
 - E.g. Adoption of Facebook apps (Onnela and Reed-Tsochas, PNAS, 2010)
 - Epidemics on networks



Challenge in Modelling Social Influence

- Given a possible “social contagion” (e.g. the spread of obesity studied by Fowler & Christakis) based on empirical data, how does one distinguish among the following:
 - 1. Genuine spread via social influence
 - Note also: social learning
 - 2. Homophily
 - Agents do the same thing due to common traits
 - 3. Environment
 - Common external influence on agents
- Control strategies (e.g. legislation) depend on whether it's 1, 2, 3, or a mixture of them.
- Efforts date back at least to the 1970s: E.g., DeGroot, Friedkin, Granovetter, etc.

How to Gain Insights on Social Influence?

- 1. Well-controlled experiments and careful data analysis (and statistics)
- 2. Simple, tractable generative mathematical models
 - Our approach!
- Approach 1 and its relatives are far more common in this field. (I'd like to convince some of you to help with #2!)
- A general goal is to advance models of social influence.
 - Develop and study new generative models (as opposed to statistical models) and then ultimately combine them with data to estimate parameter values, improve models, etc.
 - Can one emulate the success of models of biological epidemics (with e.g. their empirically estimated basic reproductive numbers)?

Watts Threshold Model

- D. S. Watts, PNAS, 2002
- Each node j has a (frozen) threshold R_j drawn from some distribution and can be in one of two states (0 or 1)
- Choose a seed fraction $p(0)$ of nodes (e.g. uniformly at random) to initially be in state 1 ("infected", "active", etc.)
- Updating can be either:
 - Synchronous: discrete time; update all nodes at once
 - Asynchronous: "continuous" time; update some fraction of nodes in time step dt
- Update rule: Compare fraction of infected neighbors (m/k_j) to R_j . Node j becomes infected if $m/k_j \geq R_j$. Otherwise no change.
 - Variant (Centola-Macy): Compare number of active neighbors (m) rather than fraction of active neighbors
- Monotonicity: Nodes in state 1 stay there forever.

Watts Model: “Response Function”

- Nodes have binary states:

S_0 - *inactive* (don't influence neighbors)

S_1 - *active* (influence neighbors)

- At each time step, a node becomes S_1 -active if the fraction of its S_1 -active neighbors exceeds its threshold:

$$F(m, k) = \begin{cases} 1, & \text{if } m / k \geq R \\ 0, & \text{otherwise} \end{cases}$$

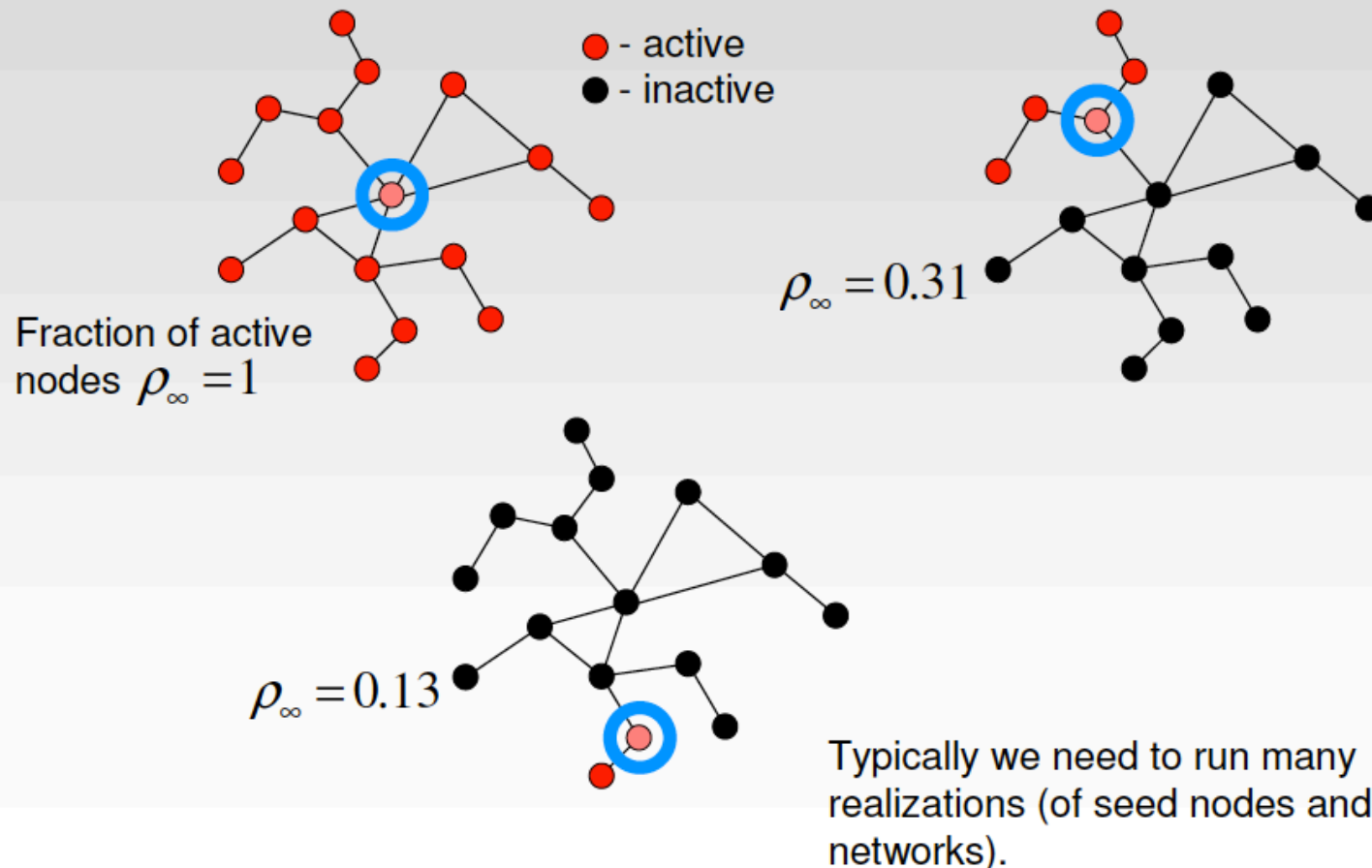
m - number of S_1 -active nbrs.
 k - total number of nbrs.
 R - threshold for switching to S_1

Response function $F(m, k)$ describes the probability that a k -degree node becomes S_1 -active when m of its neighbors are S_1 -active.

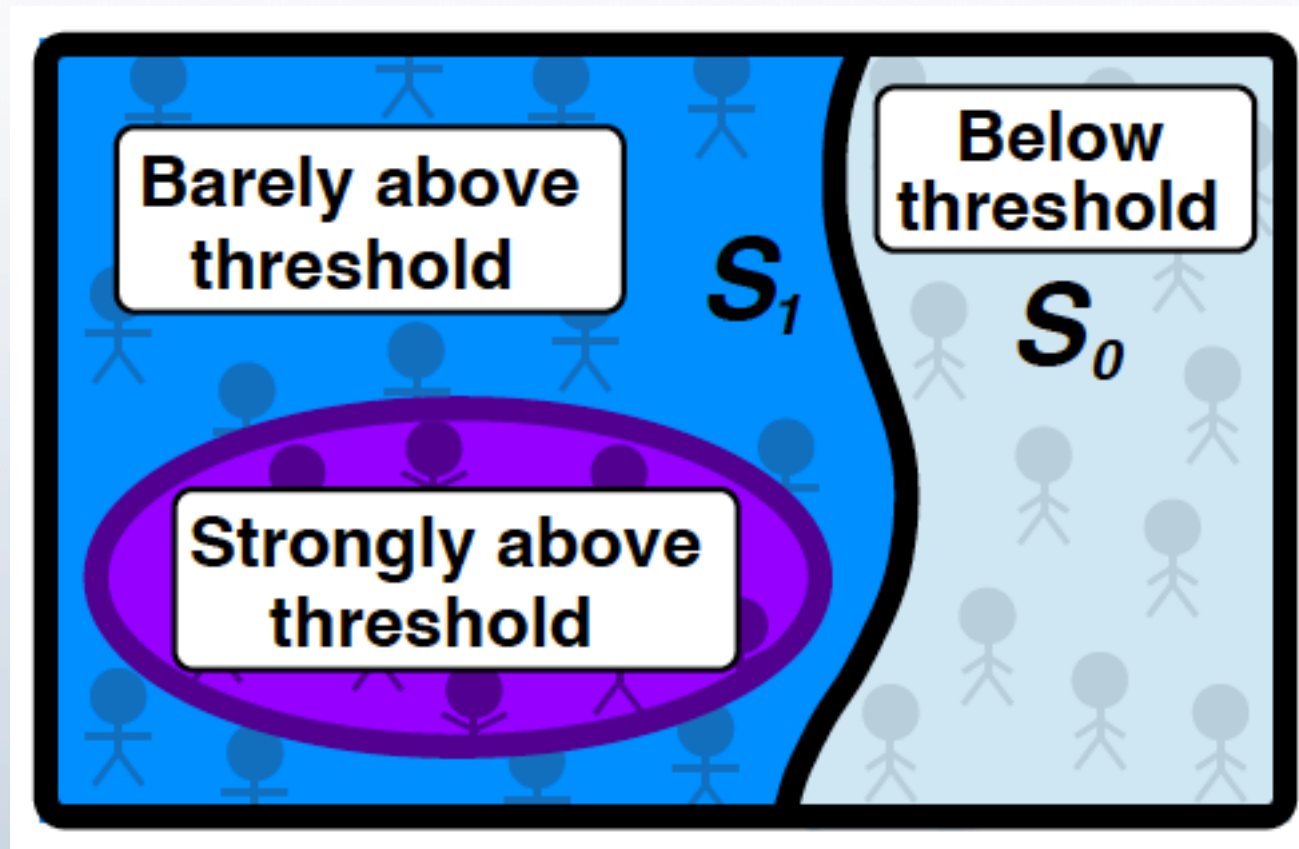
Equilibrium Levels of Adoption

Example: all nodes have threshold $R=0.3$

(A node having $\geq 30\%$ of its neighbors active will become active itself)

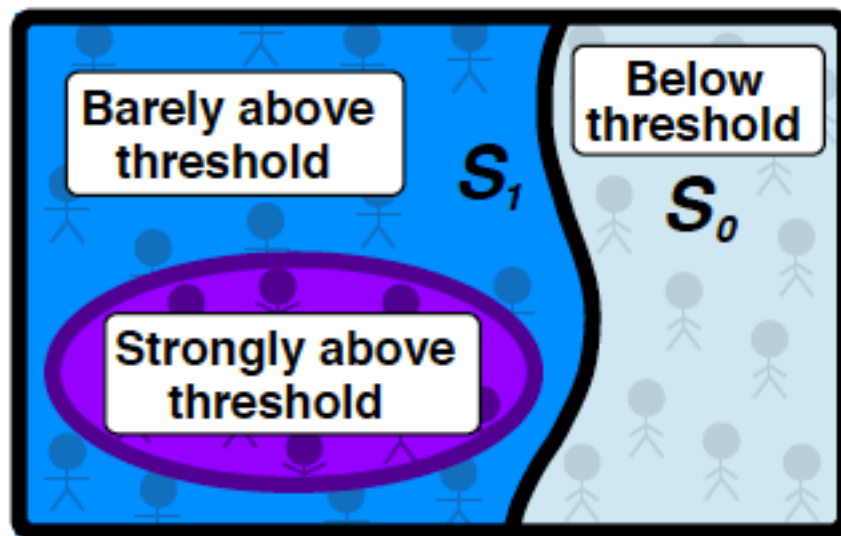


Schematic: Watts Model

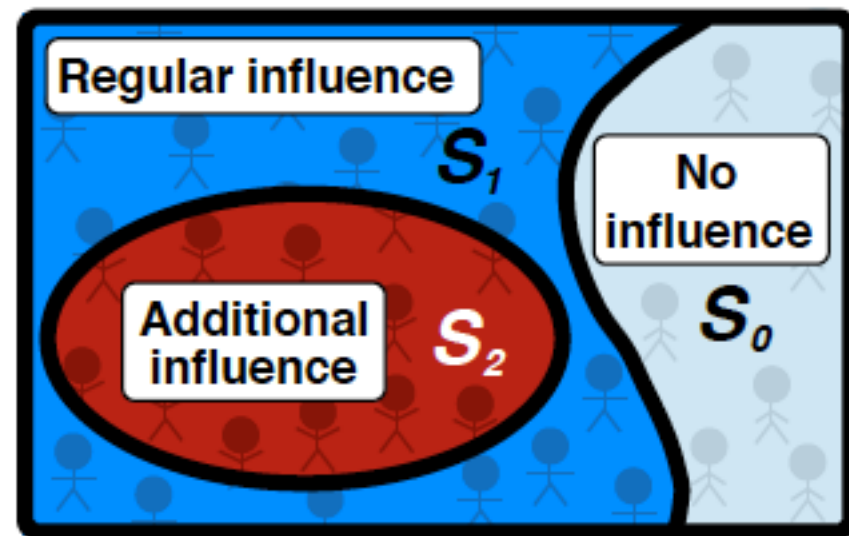


Single-Stage vs Multi-Stage

(a) Single-stage case



(b) Multi-stage case



S. Melnik, J. A. Ward, J. P. Gleeson, & MAP, "Multi-Stage Complex Contagions", *Chaos*, 23(1), 013124 (2013)

Types of Nodes

Passive (S_0)



No influence

Active (S_1)



Influences neighbors

Hyper-active (S_2)



Influences neighbors,
but with bonus influence
compared to Active
nodes

Note: $S_2 \subseteq S_1$ but $S_i \not\subseteq S_0$ for $i > 0$

Peer Pressure and Response Function

- Peer pressure = total influence experienced by a degree- k node
 - $P = (m_1 + \beta m_2)/k$
 - m_1 = number of neighbors in S_1
 - m_2 = number of neighbors in S_2
 - β = bonus influence ($\beta = 0 \rightarrow$ only S_0 and S_1 ; no S_2 state)
- Update step: Node j becomes S_i -active if $P_j \geq R_{j,i}$
 - 2 different thresholds for each node; chosen from some distributions
- If $R_{j,i} = R_i$ for all j (i.e. same threshold for all nodes), then the response function F_i is written:

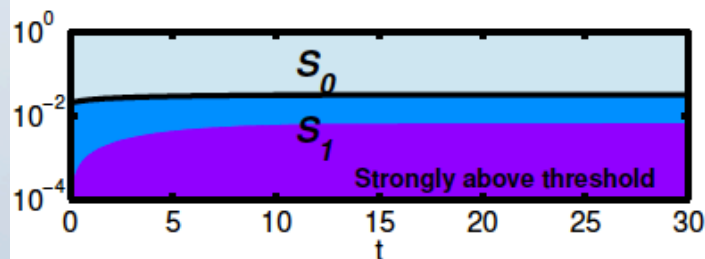
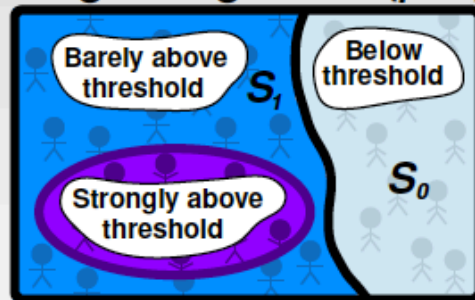
$$F_i(m_1, m_2, k) = \begin{cases} 1, & \text{if } (m_1 + \beta m_2)/k \geq R_i \\ 0, & \text{otherwise} \end{cases}$$

S_2 Cascades can Facilitate S_1 Cascades

Example: Facebook network of students at the University of Oklahoma.
All nodes have identical thresholds $R_1=0.15$, $R_2=0.3$, about 2% of nodes are initially S_1 -active.

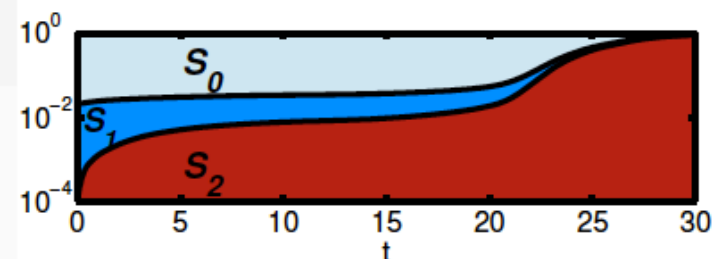
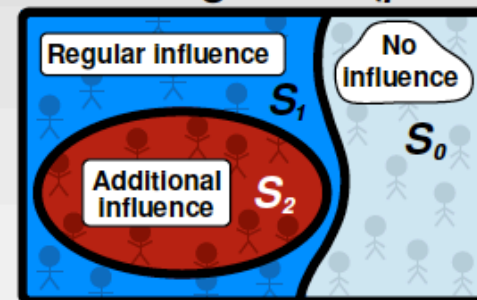
$$F_i(m_1, m_2, k) = \begin{cases} 1, & (m_1 + \beta m_2)/k \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad (i = \{1, 2\})$$

Single-stage case ($\beta=0$)



Activations do not spread

Multi-stage case ($\beta=0.5$)



All nodes eventually become S_1 - and S_2 -active

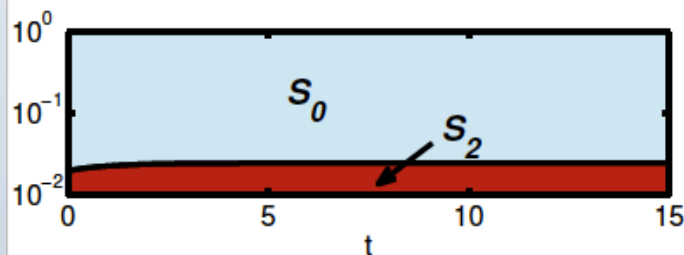
S_1 Cascades can Facilitate S_2 Cascades

Example: Facebook network of students at the University of Oklahoma.
 $\beta=0.3$, and about 2% of nodes are initially S_2 -active.

$$F_i(m_1, m_2, k) = \begin{cases} 1, & (m_1 + \beta m_2)/k \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad (i = \{1, 2\})$$

Single-stage case

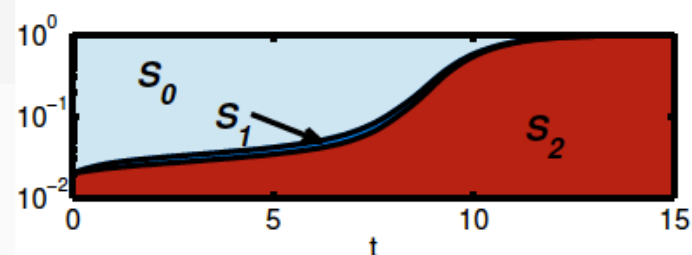
All nodes have thresholds $R_1=R_2=0.2$,
i.e. S_1 dynamics are slaved to S_2 dynamics.



Activations do not spread

Multi-stage case

All nodes have thresholds $R_1=0.15$, $R_2=0.2$



All nodes eventually become S_1 - and S_2 -active

(z_1, z_2) -Regular Random Graphs

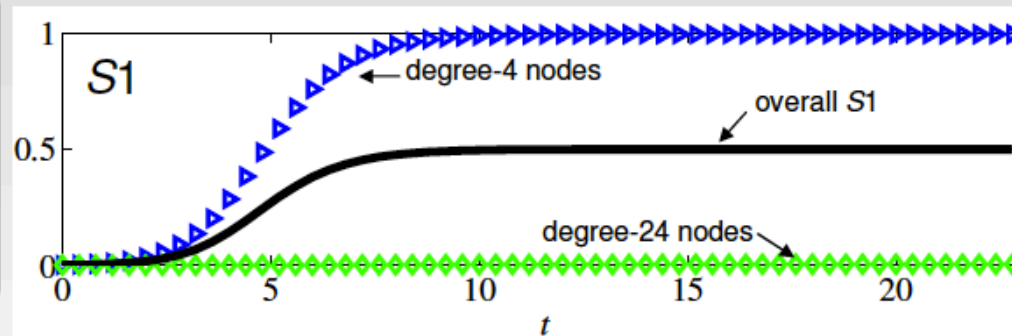
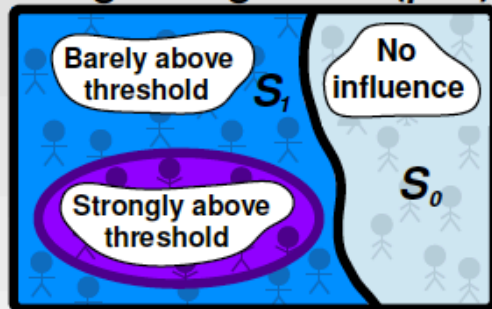
- A useful random graph ensemble to study the dynamics in more detail.
 - Precise knowledge of when nodes have state changes
- Fix degree distribution $P(k)$ and possibly also fix joint degree-degree distribution $P(k, k')$
 - Otherwise connect uniformly at random
- Example:
 - Half of nodes have degree $z_1 = 4$ and the other half have degree $z_2 = 24$
 - Ensemble in which each node has on average all but one neighbor from its own degree class
 - Assume all nodes have identical thresholds $R_1 = 0.2$ and $R_2 = 0.8$
 - Consider the case in which S_2 activations drive S_1 activations.

Dynamics on (z_1, z_2) -Regular Random Graphs

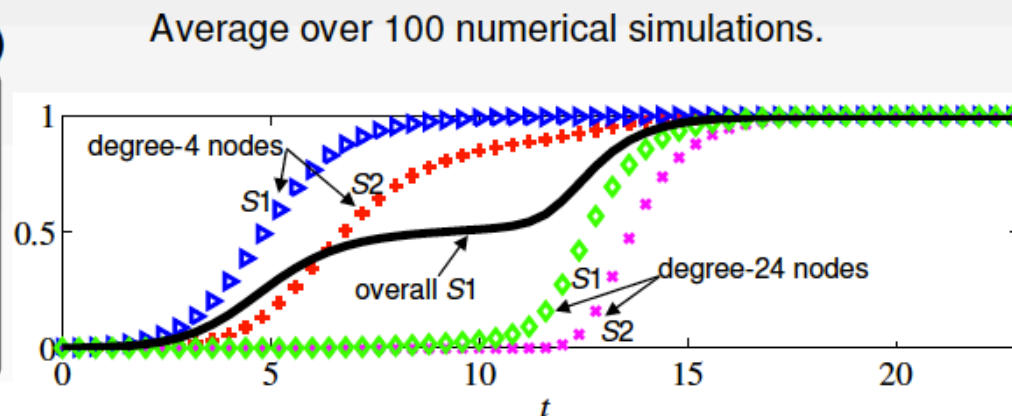
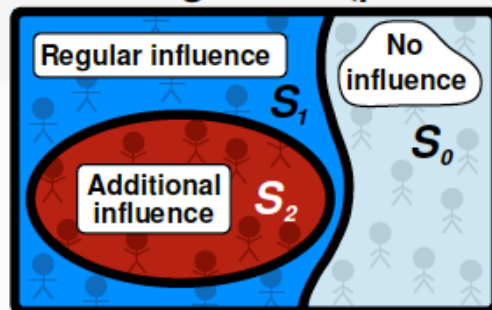
$$F_i(m_1, m_2, k) = \begin{cases} 1, & \text{if } (m_1 + \beta m_2)/k \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad R_1 = 0.2, R_2 = 0.8$$

$(i = \{1, 2\})$

Single-stage case ($\beta=0$)



Multi-stage case ($\beta=0.45$)



Some Analytics

- Consider dynamics on uncorrelated (z_1, z_2) -regular random graphs
 - Fix degree distribution $P(k)$ [but don't fix $P(k, k')$] and connect stubs uniformly at random (i.e. special case of configuration model)
- One can derive analytical expressions for final fraction of S_1 -active and S_2 -active nodes as well as temporal evolution of S_1 -active and S_2 -active nodes
- Analytical approach: a tree-based mean-field theory
 - Assumes network is “locally tree-like” and ignores both structural correlations and dynamical correlations
 - SM, A. Hackett, MAP, PJM, & JPG, *PRE* **83**(3), 036112 (2011)
 - JPG, SM, JAW, MAP, & PJM, *PRE* **85**(2), 026106 (2012)

Analytical approach

Generalization of results for single-stage model (Gleeson PRE'08, **77**, 046117) to two stages.

Advantages:

- intuitive
- readily generalizable to the study of other cascade-type problems on networks (e.g. site and bond percolation, k -cores)
- should work for any $F_2(m_1, m_2, k) \leq F_1(m_1, m_2, k)$, both non-decreasing with m_1 and m_2 .

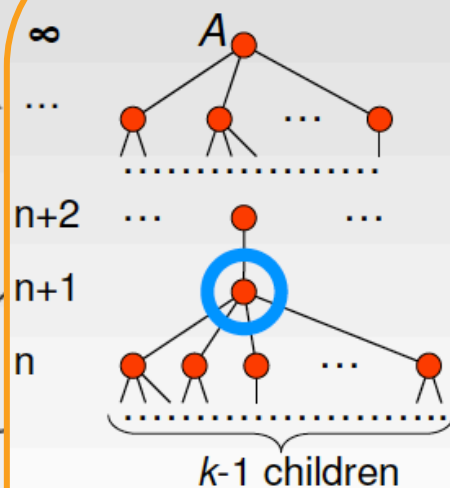
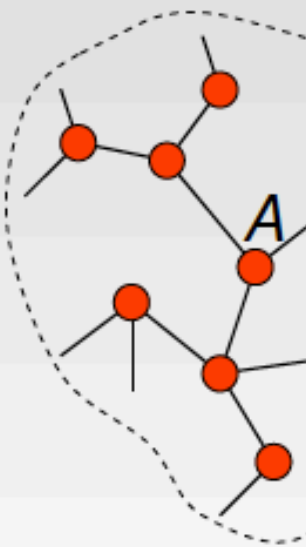
Limitations:

- works for monotone dynamics only (note: S_2 was defined as a subset of S_1)
- gives the expected active fraction, not the whole distribution

I will explain this approach using the single-stage case as an example...

Analytical Approach

Main idea: pick a node A at random and calculate its probability of becoming active. This will give the average fraction of active nodes.



q_n : probability that a node on level n is active, conditioned on its parent (on level $n+1$) being inactive.

- Start at level 0 (bottom of the tree)
- Examine the propagation of activity from level n to level $n+1$, proceeding one level at a time
- Non-seed nodes at level $n+1$ are inactive unless their children cause them to become active

$$q_{n+1} = \rho_0 + (1 - \rho_0) \times \sum_{k=1}^{\infty} \frac{k}{z} P_k \times \left(\sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1 - q_n)^{k-1-m} F(m, k) \right)$$

(seed: active)
(non-seed: initially inactive)
(has degree k ; $k-1$ children)

(m out of $k-1$ children active) (activated by m active neighbours)

Analytical approximation (single-stage case)

Degree distribution of the network: P_k

Fraction of initially active nodes: $q_0 = \rho_0$

Auxiliary variable: $q_{n+1} = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} \frac{k}{z} P_k \sum_{m=0}^{k-1} \binom{k-1}{m} q_n^m (1 - q_n)^{k-1-m} F(m, k)$

The expected fraction of active nodes at time step n :

$$\rho_n = \rho_0 + (1 - \rho_0) \sum_{k=1}^{\infty} P_k \sum_{m=0}^k \binom{k}{m} q_n^m (1 - q_n)^{k-m} F(m, k)$$

Note: different F can be used for other dynamical processes, for example:

Bond percolation: $F(m, k) = 1 - (1 - p)^m$

Site percolation: $F(m, k) = \begin{cases} 0, & \text{if } m = 0 \\ Q_k, & \text{if } m > 0 \end{cases}$

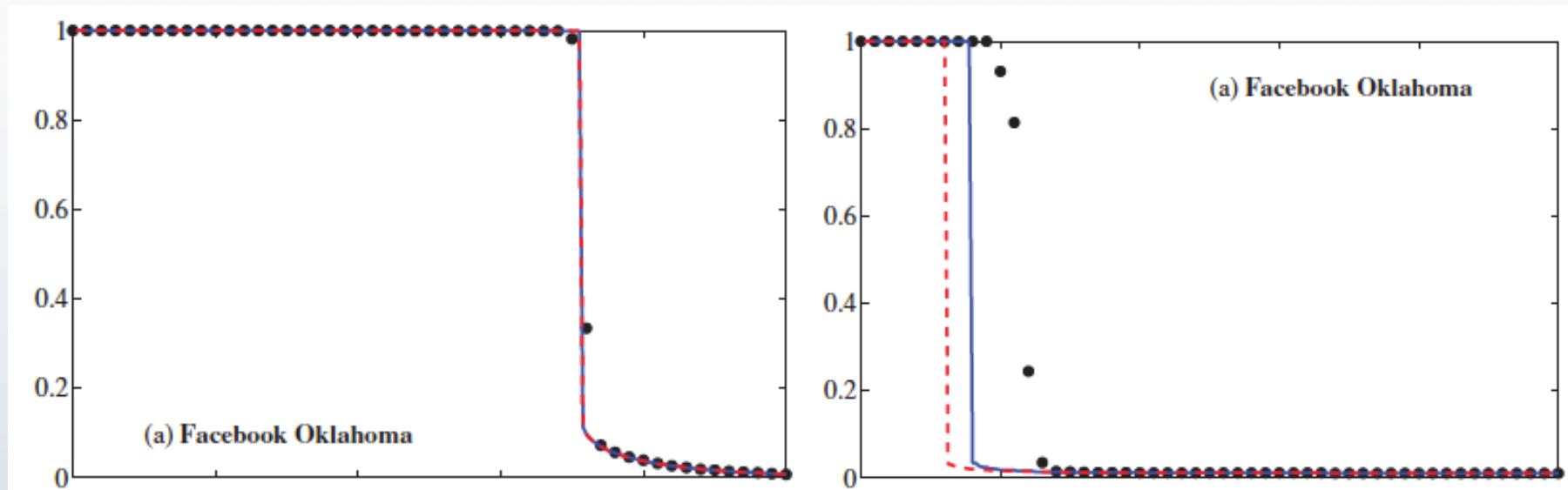
K -core sizes: $F(m, k) = \begin{cases} 0, & \text{if } m \leq k - K \\ 1, & \text{if } m > k - K \end{cases}$

Interlude...

Locally Tree-Like Approximation?

- SM, AH, MAP, PJM, & JPG, “The unreasonable effectiveness of tree-based theory for networks with clustering”, *PRE* 83(3), 036112 (2011)
 - Fun Fact: The use of “The” at the beginning of an article title is against official *Physical Review* policy (but we got away with it).
- **Main result:** the tree-based theory I’ve discussed can yield extremely accurate results even for many networks with very large clustering (high value of clustering coefficient). It tends to work well when the mean inter-node distance is sufficiently small (so if it’s a “sufficiently small” small world; not just a small world)
 - Results for quantities like effective fraction of infected nodes at equilibrium, size of largest connected component in percolation, etc.
- “Locally tree-like” means that cycles have zero measure as the number of nodes $N \rightarrow \infty$. (So you want the clustering coefficient $C \rightarrow 0$ as $N \rightarrow \infty$, and you want few small cycles for finite N .)

Example: Watts Model



Gaussian distributed threshold values with mean μ and standard deviation σ

Vertical axes: Equilibrium fraction of 'infected' nodes; horizontal axes: μ

Left: $\sigma = 0.04$; right: $\sigma = 0$ (i.e. uniform threshold)

Black dots: numerics; red dash: $P(k)$ theory; blue solid: $P(k, k')$ theory

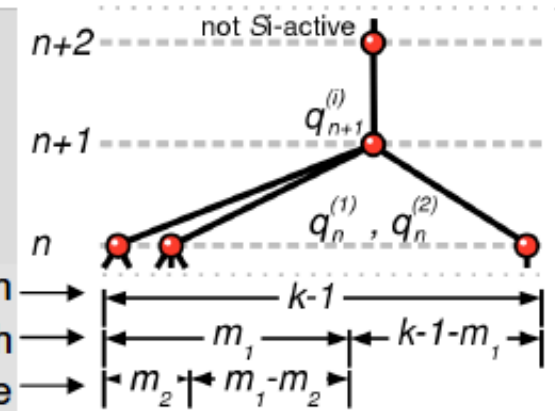
Mean-Field Theories

- JPG, SM, JAW, MAP, & PJM, “Accuracy of mean-field theory for dynamics on real-world networks”, PRE 85(2), 026106 (2012)
- Loosely, one can think of mean-field theories as (possibly sophisticated) ways of assuming some form of being well-mixed.
- Mean-field theories are typically derived under the following assumptions:
 - 1. Absence of local clustering (i.e. locally tree-like)
 - 2. Absence of modularity
 - → E.g. all degree- k nodes are well-described by the same equation (obtained by averaging over all degree- k nodes)
 - 3. Absence of dynamical correlations
 - I.e. the state of node i and those of its neighbors can be treated as independent when updating node i
 - Can be relaxed by considering pair approximations or moment closure at even higher moments
 - Contrast: *Structural correlations*, like degree-degree correlations in $P(k,k')$ theory

And now back to our regular program,
“Multi-Stage Complex Contagions”...

Analytical approximation for multi-stage contagion model

Degree distribution of the network: P_k
 Fraction of initially Si-active nodes: $q_0^{(i)} = \rho_0^{(i)}$



Auxiliary variables

$$\text{For S1: } q_{n+1}^{(1)} = \rho_0^{(1)} + (1 - \rho_0^{(1)}) \sum_{k=1}^{\infty} \frac{k}{Z} P_k \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}(q_n^{(1)}) \sum_{m_2=0}^{m_1} B_{m_2}^{m_1} \left(\frac{q_n^{(2)}}{q_n^{(1)}} \right) F_1(m_1, m_2, k)$$

$$\text{For S2: } q_{n+1}^{(2)} = \rho_0^{(2)} + (1 - \rho_0^{(2)}) \sum_{k=1}^{\infty} \frac{k}{Z} P_k \sum_{m_1=0}^{k-1} B_{m_1}^{k-1}(q_n^{(1)}) \sum_{m_2=0}^{m_1} B_{m_2}^{m_1} \left(\frac{q_n^{(2)}}{q_n^{(1)}} \right) \times$$

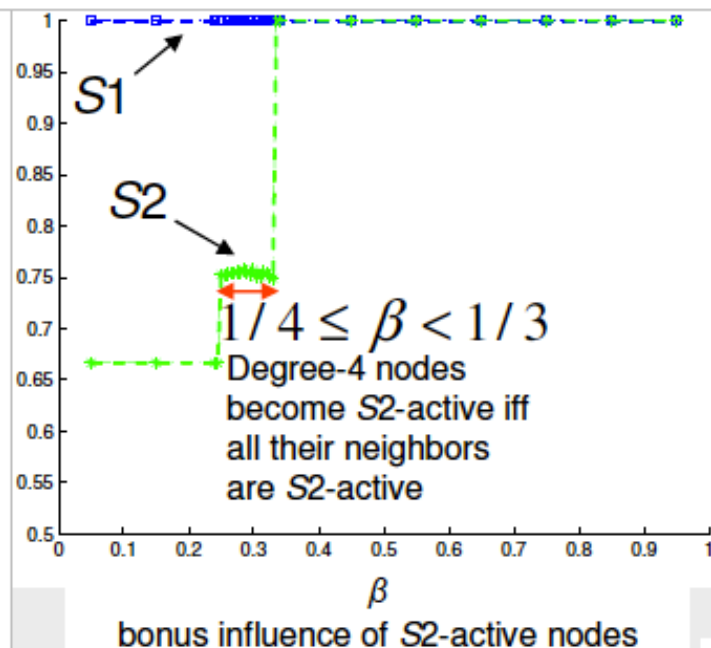
$$\left[(1 - q_n^{(1)}) F_2(m_1, m_2, k) + q_n^{(1)} F_2(m_1 + 1, m_2, k) \right] \quad \text{where } B_m^k(q) = \binom{k}{m} q^m (1 - q)^{k-m}$$

Fraction of Si-active nodes at time step n

$$\rho_n^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_{k=1}^{\infty} P_k \sum_{m_1=0}^k B_{m_1}^k(q_n^{(1)}) \sum_{m_2=0}^{m_1} B_{m_2}^{m_1} \left(\frac{q_n^{(2)}}{q_n^{(1)}} \right) F_i(m_1, m_2, k)$$

$(i = \{1, 2\})$

Time evolution



$$F_i(m_1, m_2) = \begin{cases} 1, & \text{if } m_1 + \beta m_2 \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad (i = \{1, 2\})$$

Network:

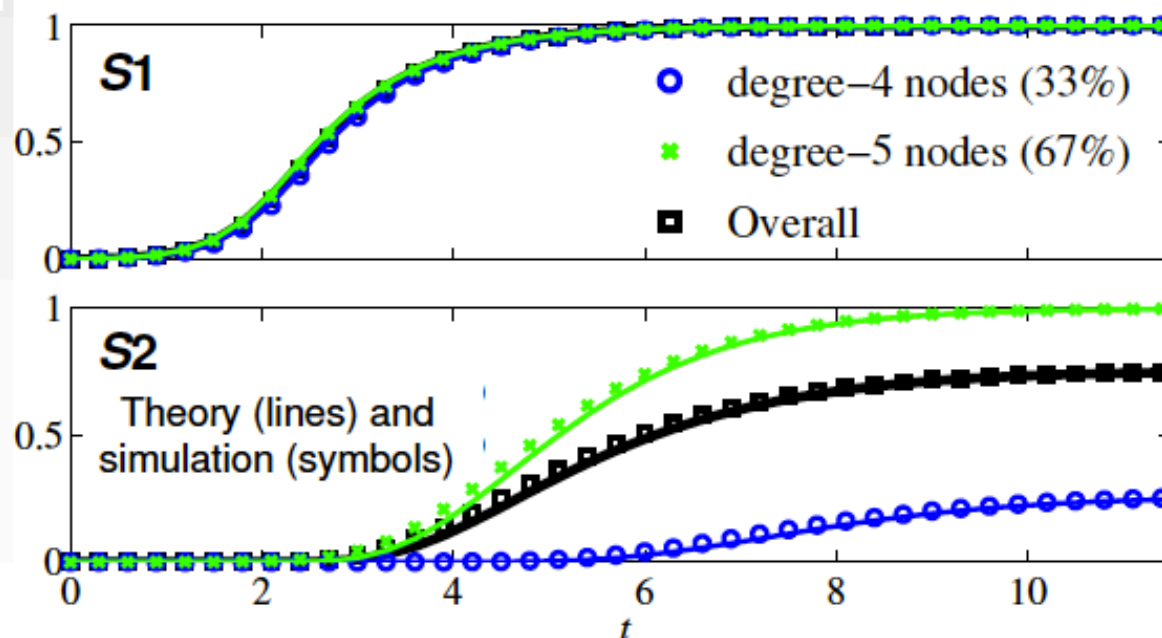
33% of nodes have degree 4

67% of nodes have degree 5

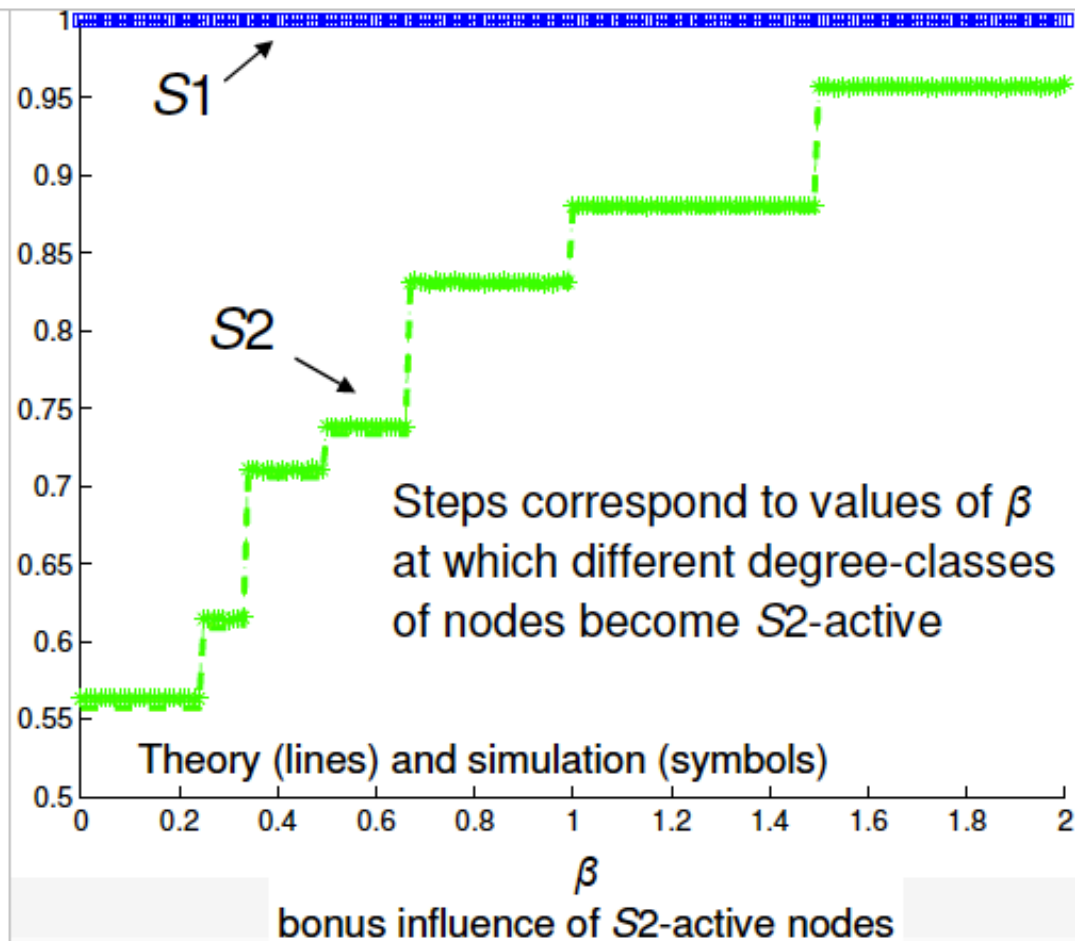
All nodes have identical thresholds $R_1=1, R_2=5$

(Plots normalized by total number of nodes)

Time evolution of S_1 and S_2 for $1/4 \leq \beta < 1/3$



Final fraction of active nodes

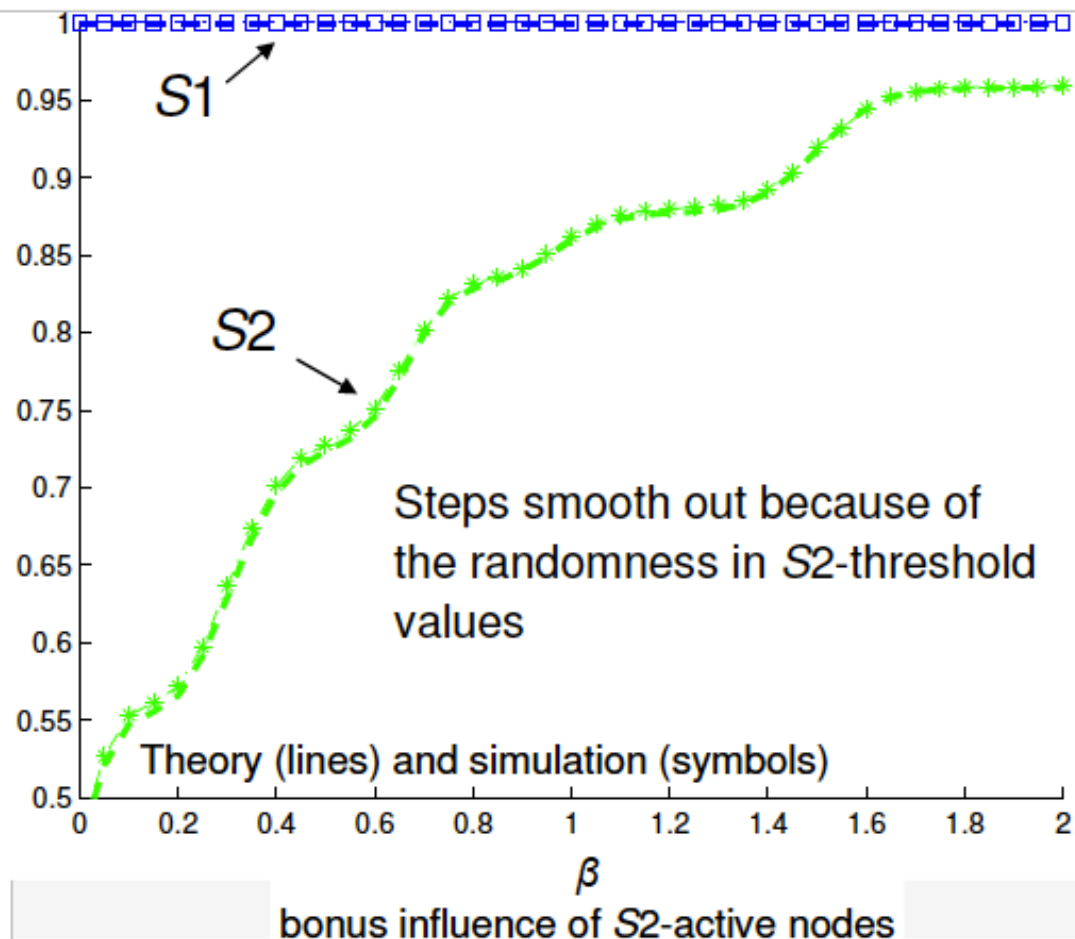


Erdős-Rényi network with mean degree $z = 5$

All nodes have identical thresholds $R_1=1, R_2=5$

$$F_i(m_1, m_2) = \begin{cases} 1, & \text{if } m_1 + \beta m_2 \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad (i = \{1, 2\})$$

Final fraction of active nodes



Erdős-Rényi network with mean degree $z = 5$

All nodes have identical threshold $R_1=1$ as before, but R_2 are drawn from normal distribution with mean=5 and std=0.2.

$$F_i(m_1, m_2) = \begin{cases} 1, & \text{if } m_1 + \beta m_2 \geq R_i \\ 0, & \text{otherwise} \end{cases} \quad (i = \{1, 2\})$$

Model Features that will Necessitate Generalization of our Analytical Approach

- Some areas where improved analytical methodology is needed...
 - We haven't yet developed this.
- "June Bug" effect
 - Believing one's own gossip: An S_1 -active node j influences neighboring S_0 nodes to become S_1 -active, which subsequently influence node j to S_2 -active
 - Dynamical correlation
 - Final fraction of S_2 -active nodes is correct, but the theory underestimates number of S_2 -active nodes in the temporal dynamics
- State Segregation ("You hang up first!" "No, you hang up first!")
 - A node j has a sufficiently high threshold R_2 to become S_2 -active that it needs all neighbors to be S_1 -active for this to happen, but one of those neighbors can't become S_1 -active unless j is S_2 -active.
 - Dynamical correlation
 - Final fractions again correct, but theory overestimates observed fractions of active nodes in temporal dynamics

Dynamics on Modular Networks with Heterogeneous Correlations

- SM, MAP, PJM, & JPG, “Dynamics on Modular Networks with Heterogeneous Correlations”, arXiv: 1207.1809
- We develop an ensemble of random graphs in which degree-degree correlations can be different in different modules.
 - E.g. coupled social networks from different social circles (or different social media), such that the degree homophily differs across them
- We examine binary monotonic dynamics (e.g. Watts model) on such networks.
- Similar type of theory as before, but the random graph ensemble is more complicated
 - $P(i,i';k,k')$ theory: joint distribution of a pair of nodes with degrees k and k' , and located in modules i and i'

Popularity Cascades on Facebook

- Data analysis in Onnela and Reed-Tsochas, PNAS, 2010
- A model for popularity cascades
 - JPG, D. Cellai, MAP, J.-P. Onnela, & F. Reed-Tsochas, "A Simple Generative Model of Collective Online Behaviour", arXiv:1305:7440
- We develop and analyze a generative model that has a good quantitative match to the time-dependent statistical characteristics of adoption of Facebook apps from Facebook's Cambrian era (a couple of months in 2007).
 - We find several models that fit the long-time statistics very well, but they can be distinguished with the temporal behavior.
 - Features: Local information (friends' behavior as opposed to global top-app lists) and long memory window.
 - Note: No network structure in this model

Conclusions

- Modeling social influence is an excellent playground for applied mathematics. I hope some of you will be interested in taking on this challenge!
- There are subtleties regarding when tree-based and mean-field theories are valid, and some results are valid even when the hypothesis used to derive them are violated fantastically.
 - Open problems: Derive the “deeper” theories that use different (and, for the most part, somewhat weaker hypothesis).
- I introduced an analytically tractable model of multi-stage complex contagions.
 - Much more work to do!
- Paper in preparation from my group on using ideas from algebraic topology to study complex contagions (D. Taylor, et al.)

2014 AMS 'Mathematics Research Community' in Network Science

- 24—30 June 2014, Snowbird Ski Resort, Utah
- For PhD students and early postdocs
- **Organizers:** Mason Porter, Aaron Clauset (CU Boulder), David Kempe (USC); with help from Dan Larremore (Harvard)
- **More information:**
<http://www.ams.org/programs/research-communities/mrc-14>
- Applications due 1 March