

Influence Dynamics on Social Networks

Srinivasan Venkatramanan

Ph.D. advisor: Prof. Anurag Kumar

Department of ECE, IISc.

August 28, 2014



Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

Social Networks

- Networks are all-pervasive
 - **Natural:** protein interaction networks, food webs, neuronal networks
 - **Man-made:** transportation networks, power grids, Internet
- Social networks: individuals as nodes, and relationships as edges
- Edges can be directed, weighted
- Some examples: email-contact networks, co-authorship networks
- Online social networks: Massive active user base; Facebook (>1billion), Twitter (>250million)

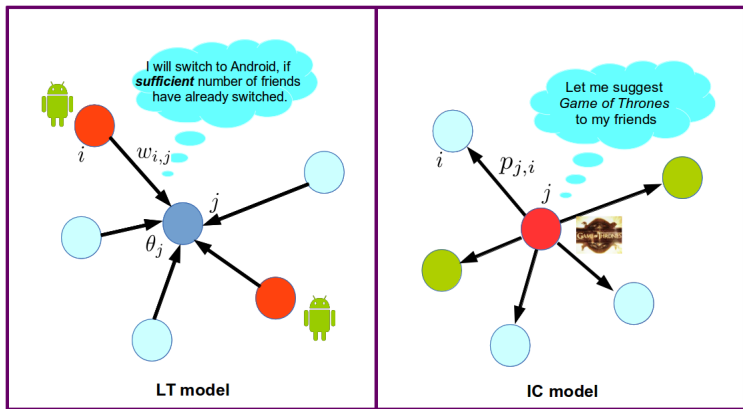


Social Network Analysis

- Several structural properties have been empirically observed
 - **small-world:** small average path length, *six degrees of separation*
 - **high clustering coefficient:** large number of triangles
 - **scale-free:** power law degree distribution
- Generative models: Preferential attachment, Watts-Strogatz model
- Network dynamics: evolution of edge structure, communities
- Focus of this thesis: **Analytical approach to dynamics on networks**
 - Characterizing cascades on networks
 - Spread of influence, information, epidemics, node-failures

Influence spread models

- Threshold model for collective behavior [Granovetter 1978]
- Maximizing influence spread (viral marketing) [Kempe et al. 2003]
- Two widely used models:
 - Linear Threshold (LT) model, Independent Cascade (IC) model



Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

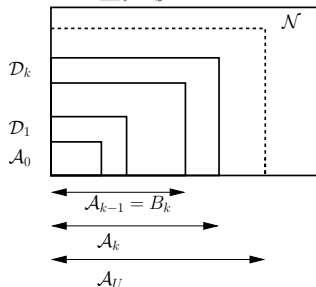
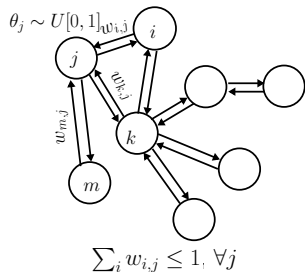
Linear Threshold Model

Introduced by Kempe et al. [2003]

- The social network is modeled by a directed weighted graph $\mathcal{N} = (V, E)$, with edge weights $w_{i,j}$, $\sum_i w_{i,j} \leq 1$
- At the beginning, each node j independently chooses a random threshold $\theta_j \sim U[0, 1]$
- Begin with initial set of *seed* nodes $\mathcal{A}_0$
- At step k , a node j gets activated if, it had been inactive until step $k - 1$ and

$$\sum_{i \in \mathcal{A}_{k-1}} w_{i,j} \geq \theta_j$$

- The process stops when a *terminal set* $\mathcal{A}_U$ is reached
- **Expected Influence:**
 $\sigma^{(\mathcal{N}, \mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N}, \mathcal{A}_0)}[|\mathcal{A}_U|]$



Influence Maximization: Greedy Algorithm

Picking the most influential set of K nodes

$$\max \sigma(\mathcal{N}, \mathcal{A}_0)$$

$$\text{s.t. } \mathcal{A}_0 \subset \mathcal{N}$$

$$|\mathcal{A}_0| = K$$

- The problem is NP-hard: equivalent to *Vertex cover*
- $\sigma(\mathcal{N}, \mathcal{A}_0)$ is monotone and submodular (i.e., diminishing returns) in $\mathcal{A}_0$
- Network heuristics (degree centrality, distance centrality, etc.) perform poorly
- Greedy hill-climbing algorithm: $(1 - \frac{1}{e})$ approximation guarantee, computationally expensive
- Other efficient algorithms/heuristics: CELF [Leskovec et al. 2007], LDAG [Chen et al. 2010]
- Our approach: **Leverage analytical insights from the LT model**

Activation probabilities

$g_j^{(\mathcal{N}, \mathcal{A}_0)}(k)$: probability that node j
gets activated in step k ,
starting with seed set $\mathcal{A}_0$ in network $\mathcal{N}$

$$\sigma^{(\mathcal{N}, \mathcal{A}_0)} = \sum_{j \in \mathcal{N}} g_j^{(\mathcal{N}, \mathcal{A}_0)}$$

Lemma

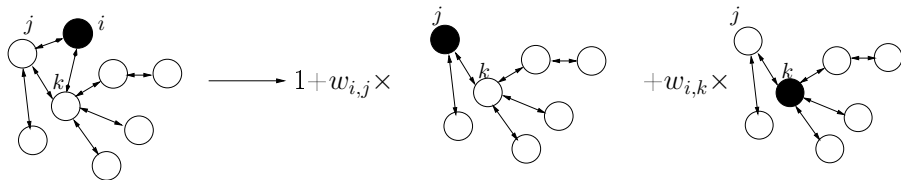
- 1 $j \in \mathcal{A}_0$,
 - 1 $g_j^{(\mathcal{N}, \mathcal{A}_0)}(0) = 1$
 - 2 $g_j^{(\mathcal{N}, \mathcal{A}_0)}(k) = 0$, for all $k > 0$
- 2 $j \notin \mathcal{A}_0$,
 - 1 $g_j^{(\mathcal{N}, \mathcal{A}_0)}(0) = 0$
 - 2 $g_j^{(\mathcal{N}, \mathcal{A}_0)}(k) = \sum_{l \in \mathcal{N} \setminus \{j\}} g_l^{(\mathcal{N} \setminus \{j\}, \mathcal{A}_0)}(k-1) w_{l,j}$, for all $k > 0$

Total Expected Activation due to $\{i\}$: $\sigma^{(\mathcal{N},i)}$

Theorem

Given a social network $\mathcal{N}$, with influence matrix $\mathbf{W}$, the total influence of any node i in the network under the LT model is given by

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{j \in \mathcal{N} \setminus \{i\}} w_{i,j} \sigma^{(\mathcal{N} \setminus \{i\}, j)} \quad (1)$$



Singleton Initial Set $\{i\}$

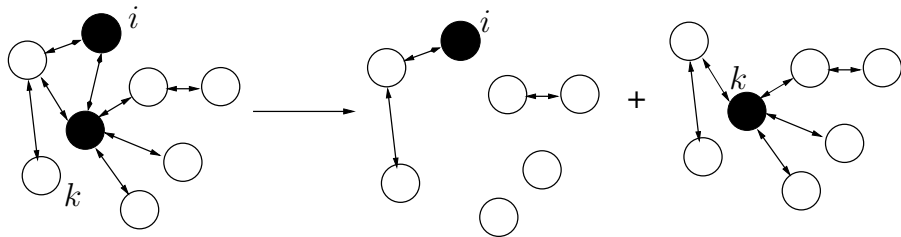
Total Expected Activation due to Set $\mathcal{A}_0$: $\sigma^{(\mathcal{N}, \mathcal{A}_0)}$

Theorem

Given a network $\mathcal{N}$ with influence matrix $\mathbf{W}$ and an initial set $\mathcal{A}_0$, for all $i \in \mathcal{A}_0$, define sub-networks $\mathcal{N}_i^{\mathcal{A}_0} = \{\mathcal{N} \setminus \mathcal{A}_0\} \cup \{i\}$.

Then the expected influence of the initial set $\mathcal{A}_0$ is given by,

$$\sigma^{(\mathcal{N}, \mathcal{A}_0)} = \sum_{i \in \mathcal{A}_0} \sigma^{(\mathcal{N}_i^{\mathcal{A}_0}, i)} \quad (2)$$

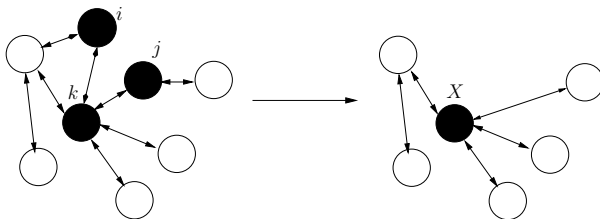


Initial Set $\mathcal{A}_0 = \{i, k\}$

Replacing an Initial Set with a “Supernode”

Theorem

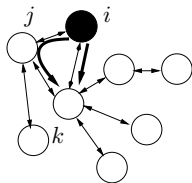
Given a network $\mathcal{N}$, with influence matrix $\mathbf{W}$, an initial set $\mathcal{A}_0$ can be replaced by a supernode X as shown below. Then, $\sigma^{(\mathcal{N}, \mathcal{A}_0)}$ is equal to the influence of X in the modified network, with X being counted as $|X|$ nodes instead of 1.



Replacing the initial set with a supernode

Equivalence to Markov chains: Insights into Pagerank

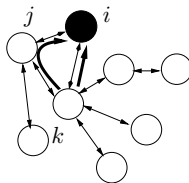
- Reverse the edges in the social network under LT model to yield a Markov chain
- Influence spread $\sim$ acyclic path probabilities (APP) to the initial set
- Provides insights into performance of Pagerank
 - Pagerank performs better in graphs with fewer cycles



$$g_k^{(\mathcal{N}, i)} = w_{i,k} + w_{i,j} \times w_{j,k}$$

Social Network

$$p_{i,j} = w_{j,i}$$



$$c(k \rightarrow i) = p_{k,i} + p_{k,j} \times p_{j,i}$$

Equivalent Markov Chain

Toy Networks

Star topology

- Edge weights: $\alpha (\leq 1)$ and $\beta (\leq \frac{1}{N-1})$
- $\exists \alpha^*$ such that for $\alpha < \alpha^*$, hub does not feature in the optimal initial set

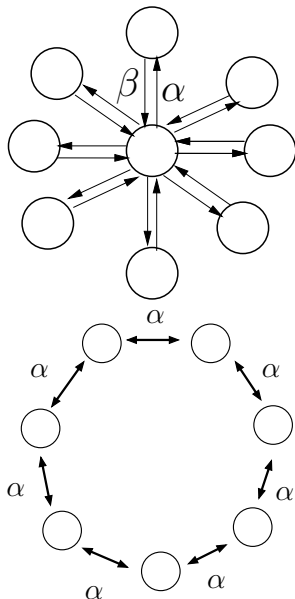
$$\alpha^* = \frac{K}{(N-K)\frac{1}{\beta} - K(N-K-1)}$$

Ring topology

- $\alpha \leq 0.5$: influence of any node on each of its neighbours

$$\sigma(\mathcal{N}, A(K)) = K + 2\frac{\alpha}{1-\alpha} \left(K - \sum_{i=1}^K \alpha^{l_i} \right)$$

- Optimal seed: Uniformly spaced K nodes
- With $\alpha = 0.5$, for large N and small K , the influence of $A(K)$ grows as $3K$

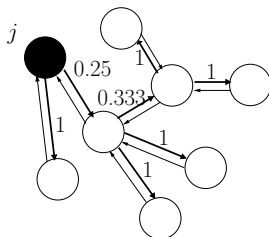


Degree Based Model: General Result on Influence

Theorem

Consider an acyclic undirected network $\mathcal{N}$. For $(i,j) \in E$, define $w_{i,j} = \frac{1}{d_j}$, where d_j is the degree of node j . Then, for any node $i \in \mathcal{N}$,

$$\sigma^{(\mathcal{N},i)} = d_i + 1$$

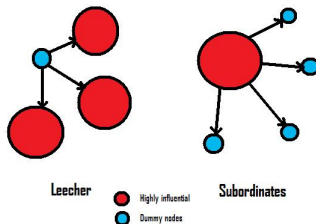


$$\sigma^{(\mathcal{N},j)} = 1 + 1 + 0.25(1 + 1 + 1 + 0.333(1 + 1 + 1)) = 3 = d_j + 1$$

- A node's local property (degree) governs its global property (influence on the entire network)

Influence Maximization: Improving the Greedy Algorithm

- Greedy algorithm: k th stage involves influence computation for $|\mathcal{N}| - k + 1$ sets of size k and this is repeated for K stages
- **G1**: Nodes listed in decreasing order of individual influences
- Top k nodes in **G1** not effective because it may contain *dummy* nodes
- Let $\mathcal{X}$ denote the set of nodes in **G1** which are above i and have been picked to be part of the initial set
- A node i is a **leecher** to set $\mathcal{X}$, if $\sigma^{(\mathcal{N}, i)} \approx \sum_{j \in \mathcal{X}} w_{ij} \sigma^{(\mathcal{N} \setminus i, j)}$
- A node i to be an α -**subordinate** of set $\mathcal{X}$ if $g_i^{(\mathcal{N}, \mathcal{X})} > \alpha$



G1-Sieving Algorithm

Evaluate $\sigma^{(\mathcal{N},i)}$ for all $i \in \mathcal{N}$

Sort nodes in decreasing order of $\sigma^{(\mathcal{N},i)}$ to get G1

$\mathcal{X} = G1(1)$, $G1 = G1 \setminus \mathcal{X}$

for $c = 2$ *to* K **do**

if $G1 = \emptyset$ **then**

 break

end

for $i \in G1$ **do**

if $g_i^{(\mathcal{N},\mathcal{X})} > \alpha \parallel \sigma^{(\mathcal{N} \setminus \mathcal{X},i)} < \epsilon$ **then**

$G1 = G1 \setminus \{i\}$

 continue

end

 Evaluate $\sigma^{(\mathcal{N} \setminus \mathcal{X},i)}$

end

$v = \arg \max_{i \in G1} \sigma^{(\mathcal{N} \setminus \mathcal{X},i)}$

$\mathcal{X} = \mathcal{X} \cup v$, $G1 = G1 \setminus \{v\}$

end

- $\mathcal{R}$: set of all papers (excluding single-authored)
- $\mathcal{N}$: union of all authors of the papers in $\mathcal{R}$
- $\forall r \in \mathcal{R}, n_r$: number of coauthors for paper r
- $\delta(i, r)$: Indicator for author i being a co-author of paper r
- Strength of collaboration between authors i and j

$$\tilde{w}_{i,j} = \sum_{r \in \mathcal{R}} \frac{\delta(i, r) \delta(j, r)}{n_r - 1}$$

- The influence matrix W is then obtained by normalizing $\tilde{W}$.
- The simulations in the following sections are carried out on a coauthorship network of NetScience community containing 1589 nodes.

Performance of G1-sieving

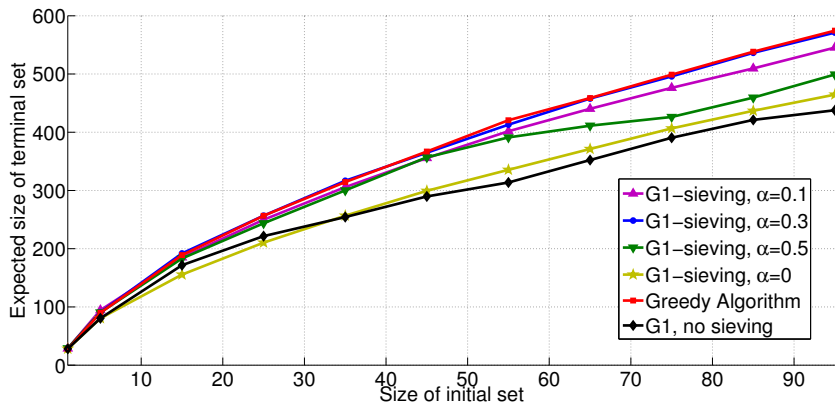


Figure: Comparison of various sieving thresholds with Greedy algorithm

Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model**
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

Technology Adoption

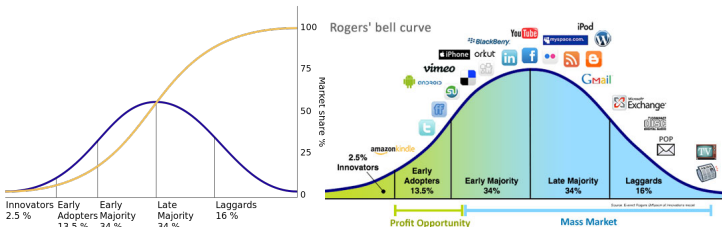


Figure: Everett Roger's bell curve shows adoption of an innovation over time

- bell-shaped curve: incremental change in adoption
- S-shaped curve: cumulative adoption
- Innovators, Early adopters ~ least threshold for adoption
- “bell curve” does not imply normally distributed threshold

General threshold distribution [Granovetter 1978]

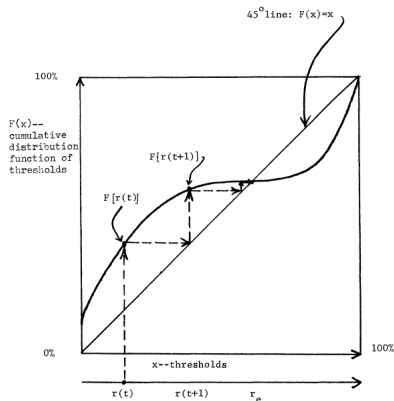
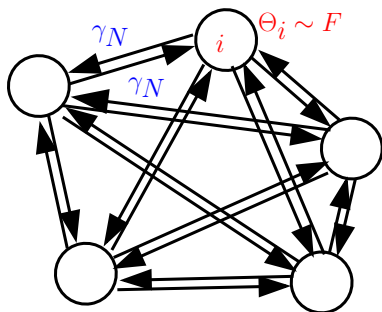


Figure: Granovetter's Fixed point dynamics $r(t+1) = F(r(t))$. Intersections indicate equilibrium points.

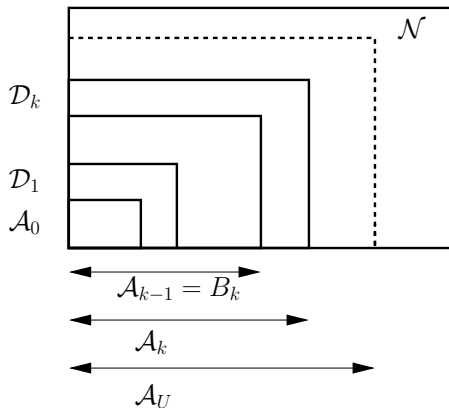
Issues with Granovetter's dynamics

- Thresholds are resampled: influence spread is not *progressive*
- Predicts no spread for uniform distribution

HILT model

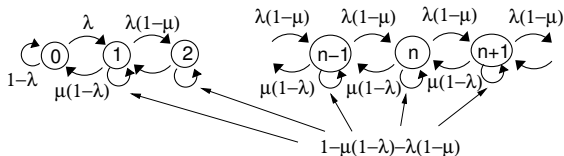


$$\gamma_N = \frac{\Gamma}{N-1}, \Gamma \leq 1$$

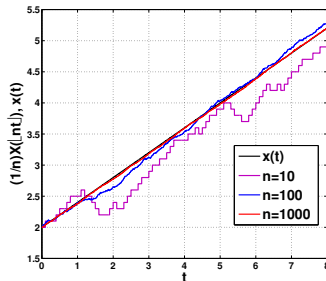
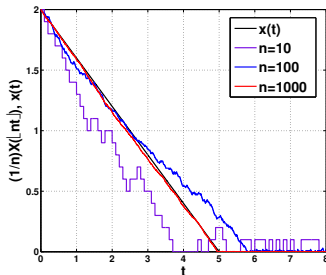


- Complete network on $\mathcal{N}$, edge weights $\gamma_N = \frac{\Gamma}{N-1}$
- Random threshold $\Theta_i \sim F, i \in \mathcal{N}$
- $A(k) = |\mathcal{A}(k)|$, $D(k) = |\mathcal{D}(k)|$ and $B(k) = |\mathcal{B}(k)|$
- $(B(k), D(k))$ is a Markov chain
- Kurtz theorem [1970]: Convergence of Markov chains to o.d.e.s

Fluid Limits - Kurtz Theorem

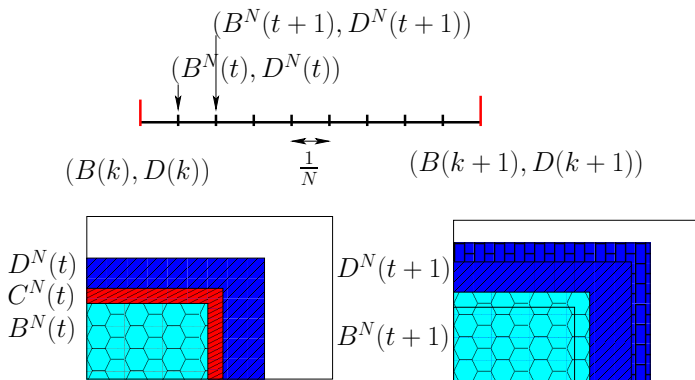


- Queue length process, $X(k), k \geq 0$, $\bar{X}^{(N)}(t) := \frac{1}{N} X(\lfloor Nt \rfloor)$
- ODE: $\dot{x}(t) = (\lambda - \mu)I_{\{x(t) > 0\}}$ with $x(0) = \bar{X}^{(N)}(0)$, for each N



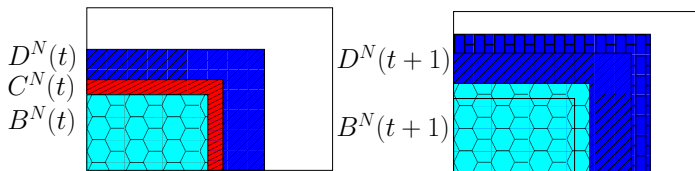
Scaling and Drift Equations

Scaled Markov process $(B^N(t), D^N(t))$ evolving over mini-slots



- Each infectious destination in $D^N(t)$ attempts to influence the relays with probability $\frac{1}{N}$
 - Success: Contributes its influence of $\frac{\Gamma}{N-1}$ and moves to $B^N(t+1)$
 - Failure: Stays in the $D^N(t+1)$

Scaling and Drift Equations



$$C^N(t) = \frac{D^N(t)}{N} + Z_b^N(t+1)$$

$$B^N(t+1) = B^N(t) + C^N(t)$$

$$D^N(t+1) = D^N(t) + \mathbb{E} \left[\frac{F(\gamma_N(B^N(t) + C^N(t))) - F(\gamma_N B^N(t))}{1 - F(\gamma_N B^N(t))} \right] \times \\ \left(N - B^N(t) - D^N(t) \right) - C^N(t) + Z_d^N(t+1)$$

Fluid Limit for Interest Evolution - HILT Model

- Defining $\tilde{B}^N(t)$, $\tilde{C}^N(t)$ and $\tilde{D}^N(t)$ as the fractional processes, we can then state the following theorem

Theorem

Given the interest evolution Markov process $(\tilde{B}^N(t), \tilde{D}^N(t))$, for the threshold distribution with density $f(\cdot)$, with bounded $f'(\cdot)$ and hazard function $h_F(x) = \frac{f(x)}{1-F(x)}$, we have for each $T > 0$ and each $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{0 \leq u \leq T} \|(\tilde{B}^N(\lfloor Nu \rfloor), \tilde{D}^N(\lfloor Nu \rfloor)) - (b(u), d(u))\| > \epsilon\right) \xrightarrow{N \rightarrow \infty} 0$$

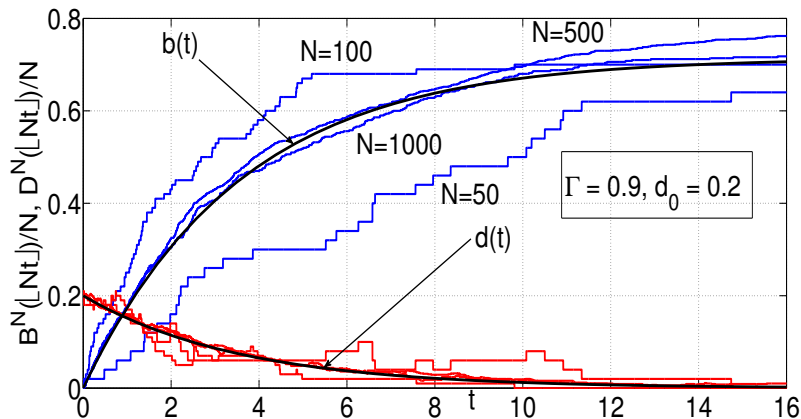
where $(b(u), d(u))$ is the unique solution to the o.d.e.,

$$\dot{b} = d \tag{3}$$

$$\dot{d} = h_F(\Gamma b) \Gamma d (1 - b - d) - d \tag{4}$$

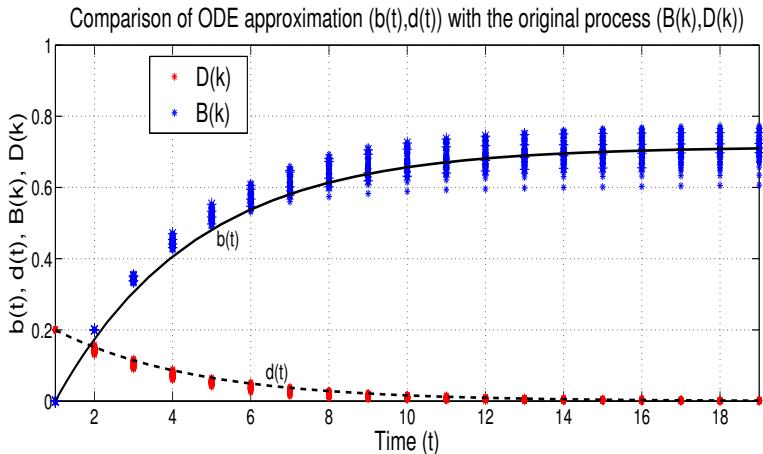
with initial conditions $(b(0) = 0, d(0) = d_0)$.

Convergence of the Scaled HILT Markov Chain to the ODE



- Influence threshold distribution is Uniform[0, 1]

Comparison of ODE with Unscaled HILT Markov Chain



- $N = 1000$
- We can conclude that the approximation works well for populations of a few 1000s

HILT Model: Effect of Threshold Distribution

- **Uniform threshold distribution:** $h_F(x) = \frac{1}{1-x}$. Let $r = 1 - \Gamma + \Gamma d_0$ then we have,

$$b(t) = \frac{d_0}{r} - \frac{d_0}{r} e^{-rt} \quad (5)$$

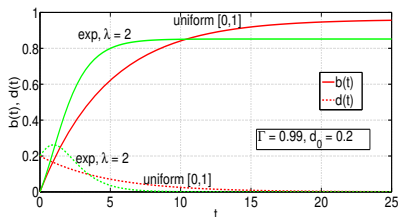
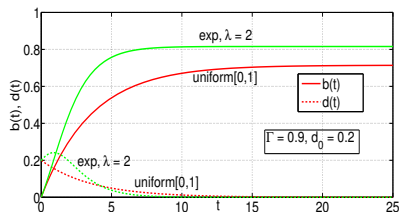
$$d(t) = d_0 e^{-rt} \quad (6)$$

- **Exponential distribution with parameter λ :** $h_F(x) = \lambda$ (memoryless). The fluid limit is the solution to the o.d.e.,

$$\begin{aligned} \dot{b} &= d \\ \dot{d} &= \lambda \Gamma d (1 - b - d) - d \end{aligned}$$

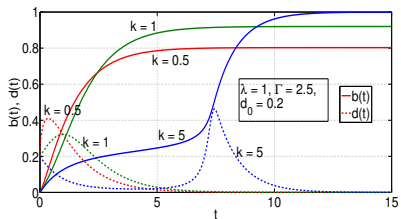
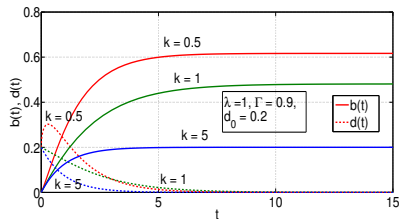
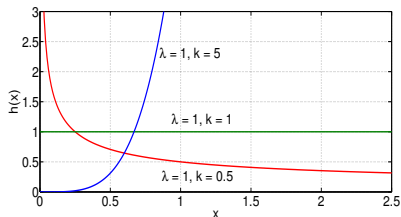
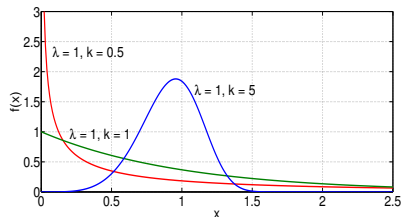
This is equivalent to the SIR epidemic model with infection rate $\lambda \Gamma$ and recovery rate of 1

Comparison: Uniform vs Exponential



- small Γ : exponential case yields a larger terminal influence spread
 - higher slope near zero; more nodes with threshold close to zero
- large Γ : uniform distribution yields a larger terminal influence spread
 - thresholds bounded above by 1

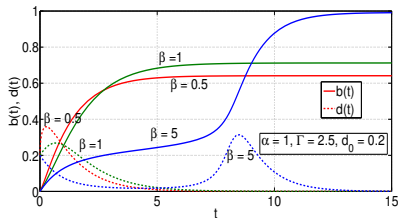
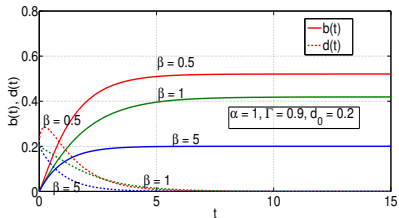
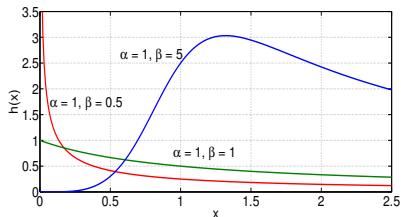
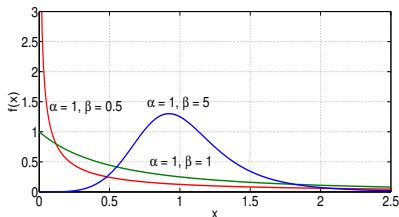
Weibull distribution



$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{(k-1)} e^{-\left(\frac{x}{\lambda}\right)^k}, \quad x \geq 0$$

$$h(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{(k-1)}, \quad x \geq 0$$

Loglogistic distribuion



$$f(x; \alpha, \beta) = \frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^{\beta}\right)^2}, \quad x \geq 0$$

$$h(x; \alpha, \beta) = \frac{\beta}{\alpha} \left[\frac{\left(\frac{x}{\alpha}\right)^{\beta-1}}{1 + \left(\frac{x}{\alpha}\right)^{\beta}} \right], \quad x \geq 0$$

Multiclass HILT model

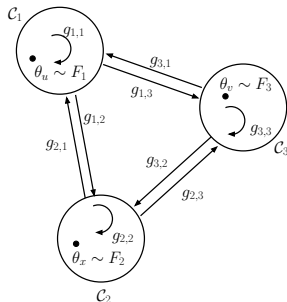


Figure: Heterogeneous network with three communities, shown with entries of the influence matrix $\mathcal{G}$. Nodes u , x and v belong to communities $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, and have their thresholds distributed according to $F_1(\cdot)$, $F_2(\cdot)$ and $F_3(\cdot)$ respectively.

$$\dot{b}_i = d_i$$

$$\dot{d}_i = -d_i + [\mathcal{G}^T \mathbf{d}]_i h_{F_i}([\mathcal{G}^T \mathbf{b}]_i)(n_i - b_i - d_i)$$

Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

Mobile Opportunistic Networks (MON)

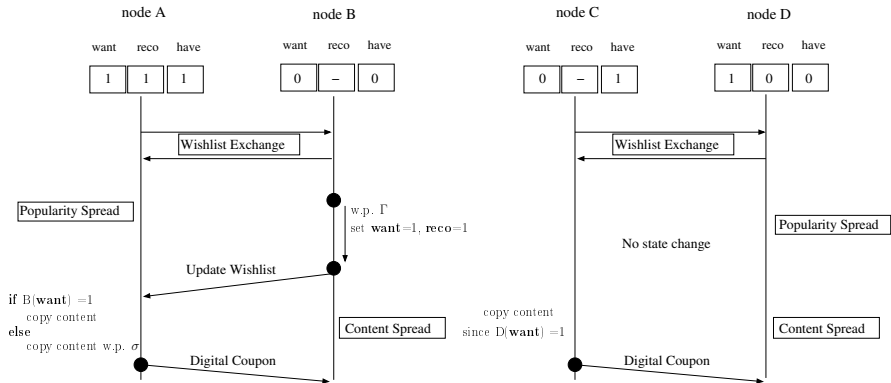
- Proliferation of smart mobile devices
 - Smartphones, tablets, wearable computing devices, etc.
- Mobile content delivery - a challenge
 - Direct delivery - not scalable
 - Alternative: p2p delivery (via Bluetooth, WiFi-Direct, etc.)
- A single item of content may be of **interest** to several co-located users
 - e.g., slides of a conference keynote/ course lecture
- The item can be forwarded between devices when their p2p radio interfaces make contact
 - Akin to **epidemic** spread
 - A multi-hop opportunistic mobile network
 - Provides an approach to *delay tolerant* networking

Mobile Coupon Delivery Ecosystem

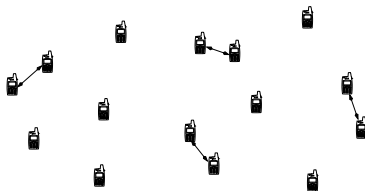
- **Scenario:** Pre-release promotions of a product (movie, book, etc.)
- **Content:** discount coupon for pre-ordering the product
- Mobile users can
 - maintain a wishlist of products of interest
 - share wishlists with peers
 - receive coupons from peers or central server
- wishlists + peer recommendation + coupon delivery



A Possible Application Framework



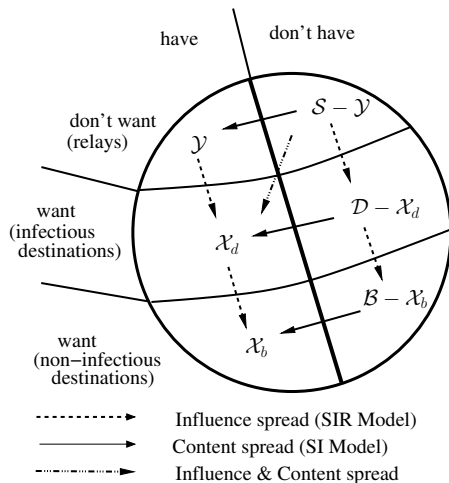
Content Popularity and Dissemination: SIR-SI model



- Population size: N (fixed)
- Pairwise meetings at points of independent Poisson processes
- The coupon needs to reach certain *destination* nodes
- Some destination nodes are given the coupon initially
- The set of destination nodes grows, as more nodes express their interest
- *Do-not-want* nodes could help in forwarding (relays)

Objective: Quickly spread content to a large fraction of destinations nodes while minimizing the residual number of relays that have the content

SIR-SI States and Evolution



- λ_N : Poisson meeting rate for each pair of nodes; $\lambda_N = \frac{\lambda}{N}$
- β_N : Recovery rate of infectious destinations; $\beta_N = \beta$
- Γ : Influence probability
- σ : Copying probability to a relay (control)

SIR-SI CTMC Markov Chain: Transitions at Meeting Epochs

- Let $k = 0, 1, 2, \dots$ index the meeting/recovery epochs at times t_k
- System state: $Z(k) = (B(k), D(k), X_b(k), X_d(k), Y(k))$

Epoch type	Rate	State update δ_k
$\mathcal{D} - \mathcal{X}_d$ recovers	$\beta_N(D(k) - X_d(k))$	$(1, -1, 0, 0, 0)$
$\mathcal{X}_d$ recovers	$\beta_N X_d(k)$	$(1, -1, 1, -1, 0)$
$\mathcal{B} - \mathcal{X}_b$ meets $\mathcal{X} + \mathcal{Y}$	$\lambda_N(B(k) - X_b(k))(X(k) + Y(k))$	$(0, 0, 1, 0, 0)$
$\mathcal{D} - \mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N(D(k) - X_d(k))Y(k)$	$(0, 0, 0, 1, 0)$ + $(0, 1, 0, 1, -1)$ w.p. Γ
$\mathcal{D} - \mathcal{X}_d$ meets $\mathcal{X}$	$\lambda_N(D(k) - X_d(k))X(k)$	$(0, 0, 0, 1, 0)$
$\mathcal{X}_d$ meets $\mathcal{Y}$	$\lambda_N X_d(k)Y(k)$	$(0, 1, 0, 1, -1)$ w.p. Γ
$\mathcal{S} - \mathcal{Y}$ meets $\mathcal{X}_b + \mathcal{Y}$	$\lambda_N(X_b(k) + Y(k))(S(k) - Y(k))$	$(0, 0, 0, 0, 1)$ w.p. σ
$\mathcal{S} - \mathcal{Y}$ meets $\mathcal{D} - \mathcal{X}_d$	$\lambda_N(D(k) - X_d(k))(S(k) - Y(k))$	$(0, 1, 0, 0, 0)$ w.p. Γ
$\mathcal{S} - \mathcal{Y}$ meets $\mathcal{X}_d$	$\lambda_N(X_d(k))(S(k) - Y(k))$	$(0, 1, 0, 1, 0)$ w.p. Γ $(0, 0, 0, 0, 1)$ w.p. $(1 - \Gamma)\sigma$

$$\dot{b} = \beta d$$

$$\dot{d} = -\beta d + \lambda \Gamma d s$$

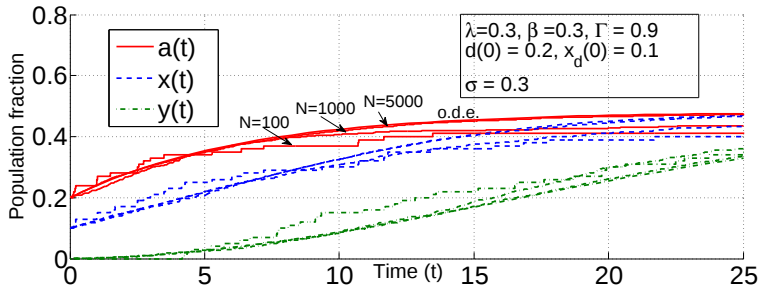
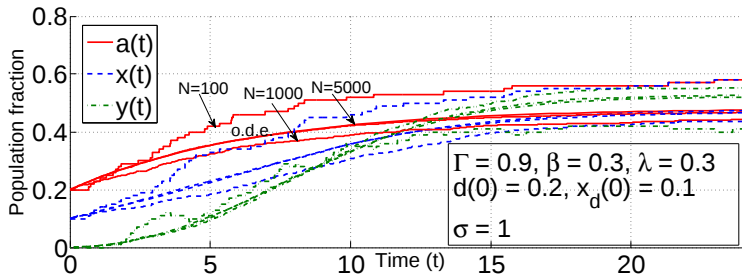
$$\dot{x}_b = \beta x_d + \lambda(b - x_b)(x + y)$$

$$\dot{x}_d = \lambda(d - x_d)(x + y) + \lambda \Gamma d y + \Gamma \lambda x_d(s - y) - \beta x_d$$

$$\dot{y} = -\Gamma \lambda d y + \lambda \sigma(s - y)(x_b + y + (1 - \Gamma)x_d)$$

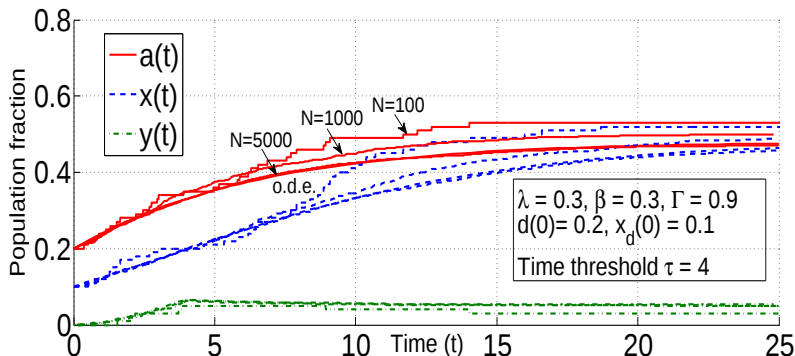
where $s(t) = 1 - b(t) - d(t)$

Convergence of the CTMC to O.D.E. Limit



SIR-SI Model: Dynamic Control of Copying

- Dynamic control $\sigma : Z(k) \rightarrow [0, 1]$
- CTMDP for each N : Obtaining optimal control is difficult
- Replace probabilistic control σ in the ODE by $\sigma(t)$ (controlled ODE)
- Optimal (deterministic, open-loop) control for the controlled ODE
- Can be shown to be asymptotically optimal for the finite size problem



Optimal Control

Target time: $T_\sigma = \inf\{t : x_\sigma(t) \geq \alpha a(\infty)\}$

- $a(\infty) = b(\infty) + d(\infty)$: terminal fraction of destinations
- $x_\sigma(t)$: fraction of destinations that have the coupon at time t

Cost function: $C_\sigma = \psi y_\sigma(T_\sigma) + T_\sigma = \psi y_\sigma(T_\sigma) + \int_0^{T_\sigma} 1 dt$

- $y_\sigma(T_\sigma)$: fraction of relays that have the coupon at time T_σ

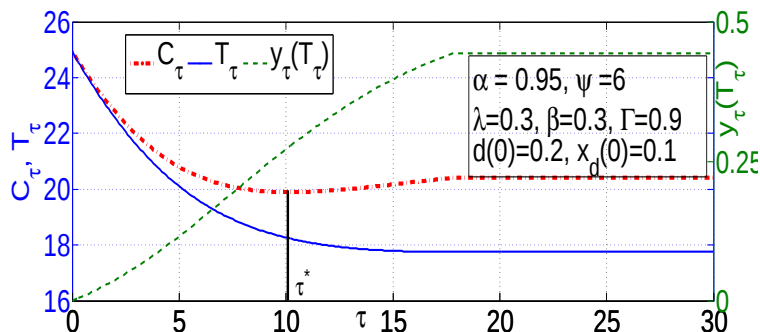
Theorem

For the above o.d.e. system, for the cost function displayed earlier, there exists an optimal control of the form,

$$\sigma_\tau(t) = \begin{cases} 1, & 0 < t < \tau \\ 0, & t \geq \tau \end{cases}$$

- Distributed implementation: Time stamping of the coupon

Optimal Control for the Running Example



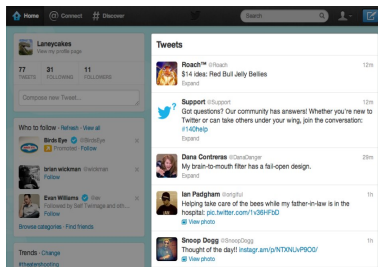
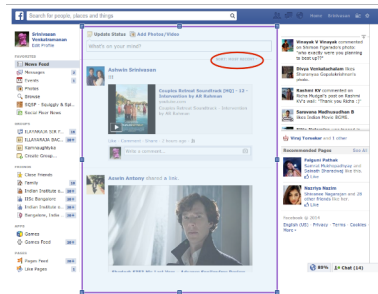
- $\alpha = 0.95, C_\sigma = 6y_\sigma(T_\sigma) + T_\sigma$
- The optimal control is to copy until $\tau = 10.1$ and then stop copying

Outline

- 1 Introduction
- 2 Influence Maximization: Linear Threshold Model
- 3 Fluid Limit Characterization: Linear Threshold Model
- 4 Coevolution of Influence and Content in Mobile P2P Networks
- 5 Competition over Timelines

Timelines on Social Networks

- Facebook, Twitter use a **Timeline** based social feed
 - Reverse chronological** - latest entries pushing out older entries
 - Similar to an email inbox
- Google+, Pinterest employ **parallel timelines**
- Recently, OSNs also **sort the entries** according to user preference
 - Priority Inbox, Facebook's EdgeRank, etc.



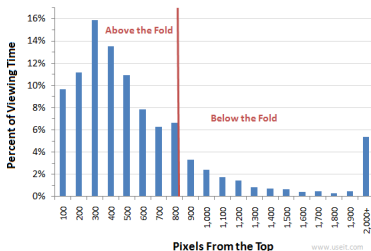
Social Advertising

- **Traditional online ads:** sponsored search, featured links, banner ads
 - Consumers are becoming more immune; Ads cannot be shared
- **Social ads:** Customized suggestions based on personal/social history
 - Brands have their own pages/accounts on the social network
 - Consumers can share or retweet the promotional content



Limited User Attention

- 80 % of the users' viewing time is spent on the contents *above the fold*
- Timeline: *User attention is limited to the top few items*

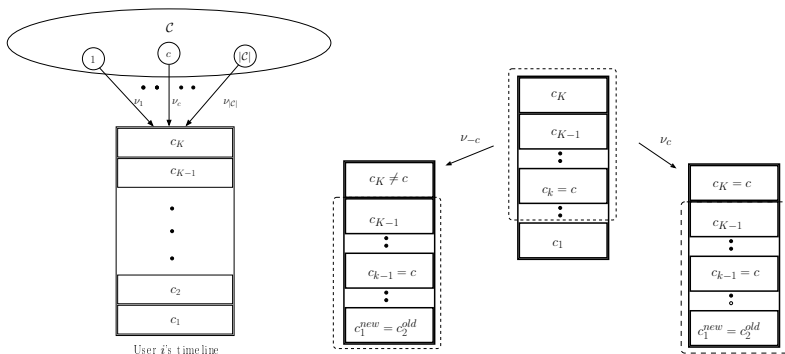


- Economy of user attention [Wu-Huberman 2007], Clash of the contagions [Myers-Leskovec 2012], Timeline model [Weng et al. 2012]
- Competition between content creators for timeline space yet to be analytically modeled

Source: <http://www.nngroup.com/articles/scrolling-and-attention/>

User's Timeline

- Reverse chronological timeline of size K
- Items of content c generated at points of a Poisson process of rate ν_c
- $\nu := \sum_c \nu_c$, $\nu_{-c} := \sum_{c' \neq c} \nu_{c'}$



Stationary distribution: $\pi_c = \prod_{k=1}^K \frac{\nu_{c_k}}{\nu}$

Expected ic -Busy Period

ic-busy period: The duration for which *at least one item* of c 's content is present in user i 's timeline *after* first entering the head of the timeline

$$\alpha_c := \frac{\nu_c}{\nu}$$

Theorem

The expected ic-busy period is given by

$$E[T_{ic}(K)] = \frac{1}{\nu_c} \left(\frac{1 - \alpha_c^{-K}}{1 - \alpha_c^{-1}} \right)$$

Proof sketch: Recursive equation for $E[T_{ic}(k)]$, the duration for which content c stays on user i 's timeline, starting at position k .

$$E[T_{ic}(0)] = 0,$$

$$E[T_{ic}(k+1)] = \frac{1}{\nu} + \frac{\nu_c}{\nu} E[T_{ic}(K)] + \frac{\nu - \nu_c}{\nu} E[T_{ic}(k)]$$

Probability of Finding Content c on a User's Timeline

$$p_{ic} = \frac{E[T_{ic}(K)]}{E[T_{ic}(K)] + 1/\nu_c} = 1 - \alpha_c^K$$

Game formulation:

- **Players:** Content creators $c \in \mathcal{C}$, $|\mathcal{C}| = N$
- **Strategies:** $\nu_c \in [\phi, \infty)$, $\forall c \in \mathcal{C}$
- **Utility of player c :** $p_{ic} - \gamma_c(\nu_c - \phi)$
 - Models payment to the social network for additional promotion

$$\begin{aligned} \max \quad & 1 - \alpha_c^K - \gamma_c(\nu_c - \phi) \\ \text{s.t.} \quad & \nu_c \geq \phi \ (\geq 0) \end{aligned}$$

Best response rate for content creator c obtained by solving for ν_c in

$$\nu_c = \nu \left[1 - \left(\frac{\gamma_c \nu}{K} \right)^{\frac{1}{K}} \right]$$

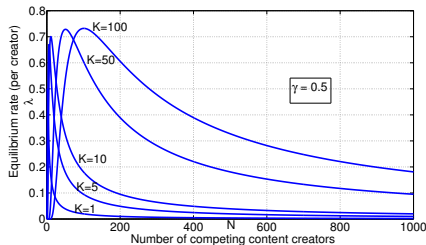
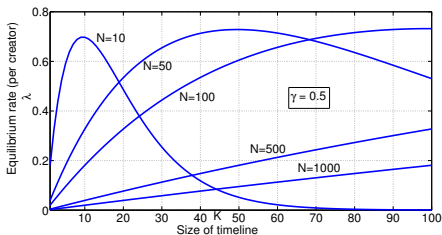
if this solution $> \phi$, else best response is ϕ

Equilibrium rate

Theorem

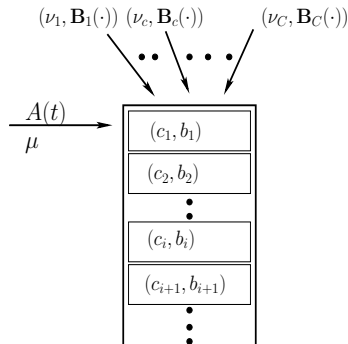
In a symmetric game ($\gamma_c = \gamma, \forall c \in \mathcal{C}$), restricting to equal rate profiles, at equilibrium each player sends either $\nu_c = \frac{K}{\gamma} \frac{(N-1)^K}{N^{(K+1)}}$ or ϕ , whichever is larger.

- For large N this is maximised when $\frac{K}{N} \approx 1$
- When $K \ll N$, too much competition $\Rightarrow$ use small ν
- When $K \gg N$, little competition $\Rightarrow$ use small ν



Infinite Timeline Model: Content Generation and Influence

- **Timeline size not restricted to K**
- Each content creator is characterized by
 - **content generation rate ν_c**
 - **influence weight distribution $B_c(\cdot)$**
- User visits timeline at points of a **Poisson process** of rate μ
- On each visit, number of posts seen is **geometrically distributed**
 - stops at position j , $j \geq 1$ with prob. $\alpha^{j-1}(1 - \alpha)$



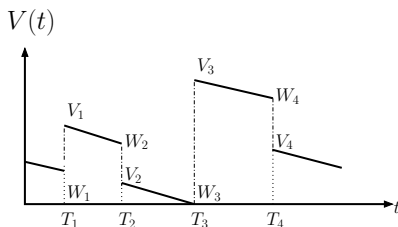
- Expected influence of content c is $\nu_c = \sum_{j=1}^{\infty} \mathbb{I}_{\{c_j=c\}} b_j \alpha^{j-1}$
- **Influence of the timeline at t : $V(t)$**

Evolution of Expected Influence, $V(t)$

- Focus on single content case
- Jumps in $V(t)$ occur only at the content arrival instants $T_k, k \geq 1$

$$V(T_k+) = b + \alpha V(T_k)$$

- Between these instants we assume that the value of $V(t)$ decreases at a constant rate (1 after appropriate scaling of time)



- Average influence on the user over the visits to the timeline

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \int_0^t V(u) dA(u)$$

Existence of Stationary Distribution, $F(\cdot)$

$V(t)$ is a piecewise deterministic Markov process. If $V(t)$ is asymptotically stationary and ergodic, the earlier limit can be obtained *a.s.* as the expectation w.r.t the stationary distribution

Theorem

Let $0 \leq \alpha < 1$ and $B(\cdot)$ be supported on $[0, b_{\max}]$. Then, for the process $V(t)$, there exists a probability distribution $F(\cdot)$ on $[0, \frac{b_{\max}}{1-\alpha})$, such that,

$\forall v \in [0, \frac{b_{\max}}{1-\alpha})$

① Almost surely,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t I_{\{V(u) \leq v\}} du = F(v)$$

② Almost surely,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t V(u) du = \int_0^\infty (1 - F(u)) du$$

Level Crossing Analysis for $F(\cdot)$

We established existence of $F(\cdot)$: point mass p_0 ; density $f(x)$, $x \in [0, \frac{b_{\max}}{1-\alpha})$
Using level-crossing analysis[Brill-Posner 1977], we obtain

$$f(x) + \nu \int_x^{\frac{x}{\alpha}} B(x - \alpha y) f(y) dy = \nu p_0 B(x) + \nu \int_0^x B(x - \alpha y) f(y) dy$$

- **LHS: Downcrossing rate of level x**
 - unit rate of decay with time
 - arrival of new content with influence that does not compensate for the α discount
- **RHS: Upcrossing rate of level x**
 - arrival of new content which sees the influence process at level 0 (with prob p_0), or
 - at level y , so that the incoming influence large enough to cause an upcrossing of x
- Taking Laplace transforms across the integral equation yields

$$\tilde{f}(s)(s - \nu) + \nu \tilde{b}(s) \tilde{f}(\alpha s) = \nu p_0 (1 - \tilde{b}(s))$$

Moments of the Stationary Distribution

$$\begin{aligned}p_0 &= 1 - \nu EB + \nu(1 - \alpha)EV \\ &= 1 - \nu(1 - \alpha) \left(\frac{EB}{1 - \alpha} - EV \right)\end{aligned}$$

where EV is the expectation of $F(\cdot)$. Similarly, we can get

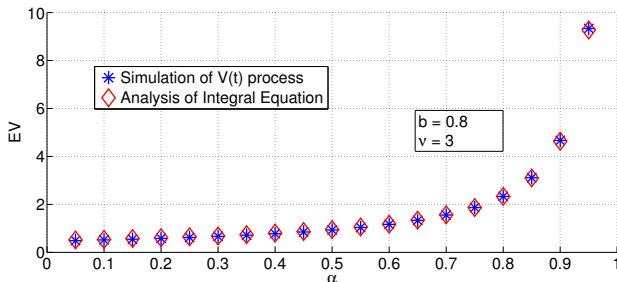
$$EV = \frac{\nu EB^2 - \nu(1 - \alpha^2)EV^2}{2(1 - \alpha\nu EB)} \quad (7)$$

When $\alpha = 1$, the process resembles the work-in-system of an $M/G/1$ queue

- p_0 equals the probability of finding the queue empty
- EV equals the expected work-in-system
- $V(t)$ can be thought of as work-in-system of an $M/G/1$ queue with *immediate discount at arrivals*

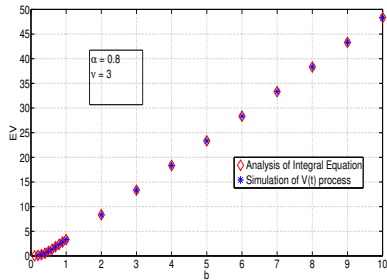
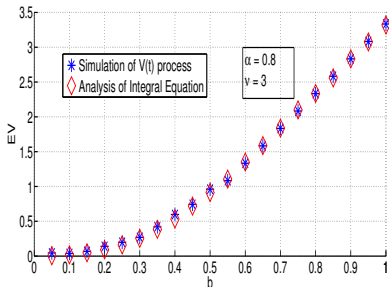
Expected Influence EV - Effect of α

- Single publisher; content generation rate ν
- Each item brings a fixed influence b
- Recall that α is the parameter of the geometrically distributed number of timeline entries seen by the user on each visit



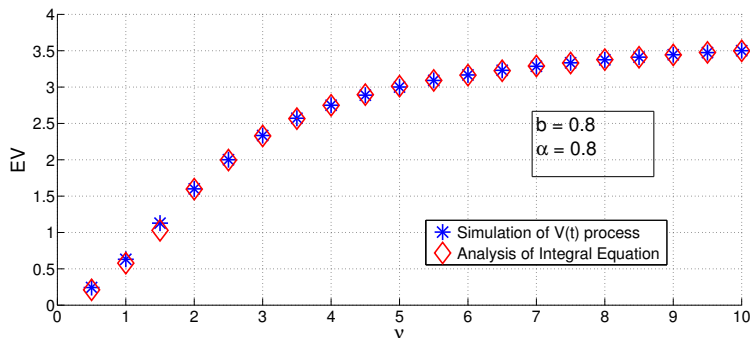
- For α close to zero, only the top entry's influence matters
- As α increases, the number of entries seen by the user increases, thus increasing the expected influence of the content

Expected Influence EV - Effect of b



- b , the influence weight is indicative of **quality**
- Increasing marginal returns for small ranges of b
 - Increasing b offsets the effect of decrease of influence with time
- As b increases, $p_0 \rightarrow 0$
- As p_0 approaches 0, the effect of b on EV becomes linear (with slope $\frac{1}{1-\alpha}$)

Expected Influence EV - Effect of ν



- ν , the content generation rate is indicative of **quantity**
- Diminishing marginal returns for increasing values of ν
- EV approaches $\frac{EB}{1-\alpha}$ as $\nu \rightarrow \infty$

Expected Influence: Multiple Content Case

- $\mathcal{C}$, the set of content creators with rates ν_c
- Define $\nu := \sum_{c \in \mathcal{C}} \nu_c$ and $\nu_{-c} = \nu - \nu_c$
- $V_c(t)$: the average influence process for content c
- Whenever a content $c' \neq c$ arrives, the net influence of content c is scaled down by α
 - From the perspective of content c , this is equivalent to an arrival with 0 influence
 - Reduce the analysis to the single content case, by introducing point mass at 0 in the arrival influence distribution
- If $b_c(x)$ is the probability density function of arrival influence of content c , then the modified distribution would be

$$\frac{\nu_{-c}}{\nu} \delta(x) + \frac{\nu_c}{\nu} b_c(x)$$

where $\delta(x)$ indicates point mass at 0

Level Crossing Analysis: Multiple Content Case

Integral equation for $f_c(\cdot)$

$$\begin{aligned} f_c(x) + \nu_{-c} \int_x^{\frac{x}{\alpha}} f_c(y) dy + \nu_c \int_x^{\frac{x}{\alpha}} B_c(x - \alpha y) f(y) dy \\ = \nu_c p_{c,0} B_c^c(x) + \nu_c \int_0^x B_c^c(x - \alpha y) f_c(y) dy \end{aligned}$$

On taking Laplace transform:

$$\tilde{f}_c(s)(s - \nu) + (\nu_c \tilde{b}_c(s) + \nu_{-c}) \tilde{f}(\alpha s) = \nu_c p_{c,0} (1 - \tilde{b}_c(s))$$

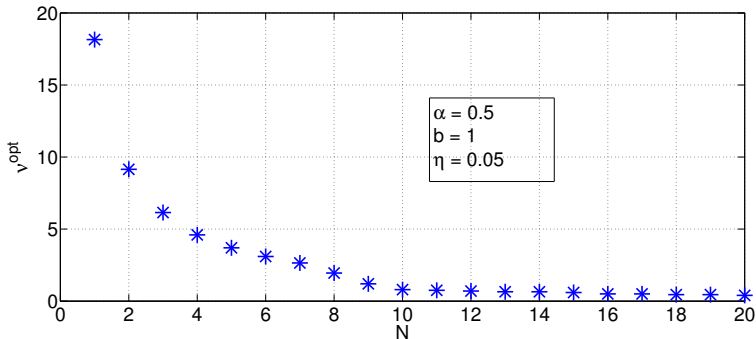
Expressions for $p_{c,0}$ and EV_c

$$p_{c,0} = 1 - \nu_c EB_c + \nu(1 - \alpha)EV_c$$

$$EV_c = \frac{\nu_c EB_c^2 - \nu(1 - \alpha^2)EV_c^2}{2(1 - \alpha\nu_c EB_c)}$$

N -Player Symmetric Game: Optimal rate vs. N

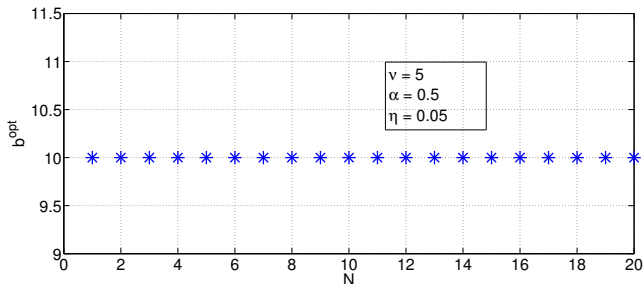
$$U_i = EV_i - \eta_i \nu_i b_i$$



- Equilibrium rate decreases monotonically with number of players N
- Contrast with finite timeline case (with a different objective function), where the behavior was non-monotonic with N

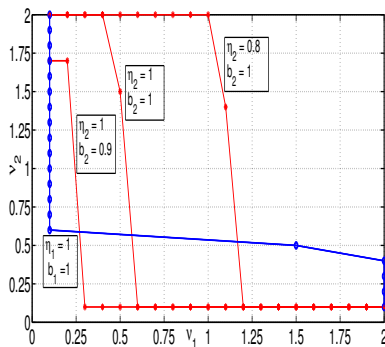
N -player symmetric game: Optimal weight

$$U_i = EV_i - \eta_i \nu_i b_i$$



- In this example, $b \in [0.1, 10]$
- All players operate at $b_{\max}$, the maximum possible influence weight
- The number of players N has no effect on b^{opt}

Two-Player Rate Competition: Examples of Equilibria



$$U_1 = EV_1 - \eta_1 b_1 \nu_1$$

$$U_2 = EV_2 - \eta_2 b_2 \nu_2$$

- Best response curves of player 1 (blue) and player 2 (red) for various values of (η_2, b_2) and $(\eta_1 = 1, b_1 = 1)$
- In this example, $\{\nu_{\min}, \nu_{\max}\} = \{0.1, 2\}$

- If $\eta_1 = \eta_2$ and $b_1 = b_2$, the equilibrium is symmetric $\nu_1^* = \nu_2^*$
- Decrease in 2's cost rate, η_2 , allows use of larger ν_2
 - Causes player 1 to use higher content generation rate ν_1 at equilibrium
- Decrease in 2's influence, b_2 , causes player 2 to use lower ν_2
 - Permits player 1 to use lower content generation rate ν_1 at equilibrium

Influence maximization

- Analytical characterization of LT model
- Optimal seeding set for simple networks
- Interpretation via APP in Markov chains
- G1-sieving algorithm

Fluid limit characterization

- Fluid limit o.d.e. for HILT model
- Equivalence to SIR epidemic model
- Effect of threshold distribution
- Fluid limit with heterogeneous communities

Coevolution of Interest and Content

- Application framework for coupon delivery
- SIR-SI model for the joint spread
- Kamke-dominance condition for o.d.e.s
- Delay-cost optimal time-threshold policy

Content Competition on Social Networks

- Competition over multiple social networks
 - Threshold structure of best response
- Competition over timelines
 - Finite timeline model
 - Infinite timeline model

Publications

Journals

- Srinivasan Venkatramanan and Anurag Kumar, *"Spread of Influence and Content in Mobile Opportunistic Networks,"* to appear in IEEE Transactions on Mobile Computing.
- Srinivasan Venkatramanan and Anurag Kumar, *"Fluid Limit Characterization of the Linear Threshold Model for Influence Spread on Social Networks,"* in preparation, 2014

Conference/Workshop Publications

- Srinivasan Venkatramanan and Anurag Kumar, *"Competition for Content Spread over Multiple Social Networks,"* Workshop on Social Networks in Science and Engineering (SCINSE'14), co-held with 6th Intl. Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, 2014. (Recipient of the Best presentation award, sponsored by Adobe Research)
- Eitan Altman, Parmod Kumar, Srinivasan Venkatramanan and Anurag Kumar, *"Competition over Timeline in Social Networks,"* Workshop on Social Network Analysis and Algorithms (SNAA), co-held with IEEE/ACM ASONAM, Niagara Falls, Canada, 2013.
- Srinivasan Venkatramanan and Anurag Kumar, *"Co-evolution of Content Popularity and Delivery in Mobile P2P Networks,"* IEEE Infocom 2012 mini-Conference, Orlando, FL, 2012.
- Srinivasan Venkatramanan and Anurag Kumar, *"Information Dissemination in Socially Aware Networks under the Linear Threshold model,"* National Conference on Communication(NCC), IISc. Bangalore, 2011.

Technical Reports

- Srinivasan Venkatramanan and Anurag Kumar, *"Co-evolution of Content Popularity and Delivery in Mobile P2P Networks,"* arXiv:1107.5851
- Srinivasan Venkatramanan and Anurag Kumar, *"New Insights from an Analysis of Social Influence Networks under the Linear Threshold model,"* arXiv:1002.1335

*Thank
You*