Analysis of comlex contagions in random multiplex networks

Osman Yağan and Virgil Gligor

Carnegie Mellon CyLab
CONFIDENCE FOR A NETWORKED WORLD

Cascades in Complex Networks

- The past decade has witnessed a lot of research interest on dynamical processes in real-world complex networks.
- An interesting *phenomenon* that takes place in many such processes is the so-called information cascades.
  Information Cascades: The spreading of an initially localized effect throughout the whole (or, a very large part of the) network.
- **Examples:**
  - ◇ Diffusion of belief, norms, and innovations in social networks
  - ◇ Disease contagion in human and animal populations
  - ◇ Cascading failures in interdependent networks
  - ◇ Global spread of computer viruses or worms on the Web

Our Focus: A class of dynamical processes: **binary decisions with externalities.**

Linear Threshold Model (Watts 2002)

Binary Decisions with Externalities

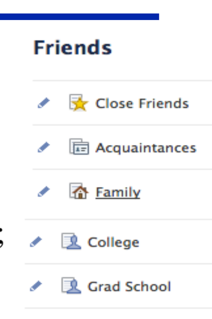
- Each individual must decide between two *actions*, e.g.,
  - ◇ To buy or not to buy a smart phone
  - ◇ To vote for Democrats or Republicans
  - ◇ To join or not to join a dissident movement
- There is an inherent incentive for individuals to **coordinate** their decisions with those of their immediate acquaintances.

Linear Threshold Model (Watts 2002)

- Nodes can be in either one of the two states: *active* or *inactive*.
- Each node is initially given a threshold τ drawn independently from $P_{th}(\tau)$.
- An inactive node with m active neighbors and $k-m$ inactive neighbors will turn activate if its fraction $\frac{m}{k}$ of active neighbors exceeds τ.
- **Global Cascades:** A **linear** fraction of nodes (in the asymptotic limit) eventually becomes active when an arbitrary node is made active.
  - ◇ Watts [PNAS 99, 5766]: **Condition** and **Probability** of global cascades
  - ◇ Gleeson & Cahalane [Pyhs. Rev. E 77, 46117]: **Expected size**

Content-dependent Threshold Model in *Multiplex* Networks

Our Motivation:



- Most existing works consider **simplex** networks that have only a single type of link.
- But, individuals engage in **different** types of relationships; e.g., see the snap-shot from Facebook on the right.
  ⇒ **Multiplex** networks (multiple link types)
- Each link type may play a different role in different cascade processes.
  - ◇ A **video game** would be more likely to be promoted among **high school classmates** rather than among family members; the situation would be exactly the opposite in the case of a new **cleaning product**.
  - ◇ Cascading failures in interdependent networks: **Power links** are vulnerable to **natural hazards**, **computer links** are vulnerable to **viruses**.

Threshold Rule with *perceived* proportion of active neighbors:

- Consider a **multiplex** network where links can be of r different types.
- For a given content (e.g., rumor, product, political view), consider positive scalars $c_1, \ldots, c_r$, such that c_i quantifies the relative bias a type-i link has in spreading this particular content.
- Nodes switch state if their **perceived** proportion of active neighbors exceeds a threshold τ. Namely, an inactive node will become active if

$$\frac{c_1 m_1 + c_2 m_2 + \ldots + c_r m_r}{c_1 k_1 + c_2 k_2 + \ldots + c_r k_r} \geq \tau$$

where m_i (resp. k_i) is the number of **active** neighbors (resp. number of neighbors) that the node is connected via a type-i link.

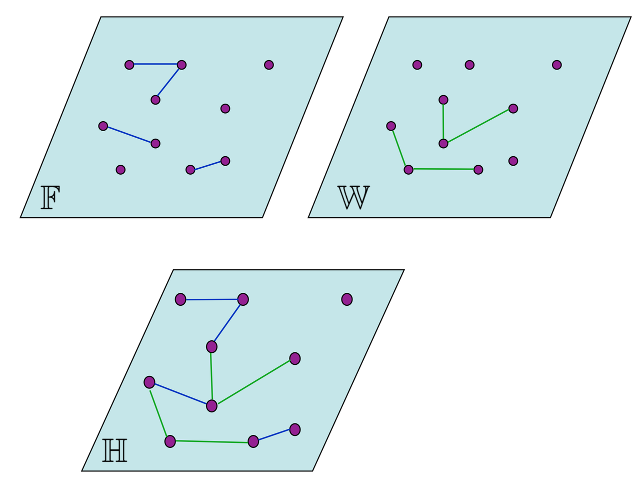
Network Model (An overlay social-physical network)

- Let $r = 2$; i.e., assume that there are only *two* link types.
- $\mathbb{F}$: *Random* network of type-1 links with degree distribution $\{p_k^f\}$.
- $\mathbb{W}$: *Random* network of type-2 links with a degree distributin $\{p_k^w\}$.
- $\mathbb{H}$: The overlay network $\mathbb{F} \cup \mathbb{W}$ with a **colored** degree distribution $\{p_{\boldsymbol{k}}\}$

$$p_{\boldsymbol{k}} = p_{k_1}^f \cdot p_{k_2}^w, \quad \boldsymbol{k} = (k_1, k_2)$$

- With $c := c_1/c_2$ an inactive node will become active with probability

$$\mathbb{P}\left[\frac{cm_1 + m_2}{ck_1 + k_2} \geq \tau\right] := F(\boldsymbol{m}, \boldsymbol{k})$$



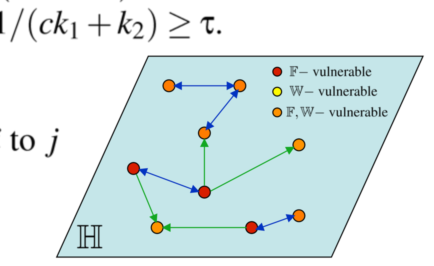
Condition and Probability of Global Cascades

Simplex Networks

- Global cascade condition = Existence of a **giant vulnerable component** (GVC)
  - ◇ A node is deemed **vulnerable** if its state can be changed by a **single** active neighbor; i.e., if $1/k \geq \tau$.
- Probability of global cascades = Fractional size of the **extended** giant vulnerable component
  - ◇ Extended GVC = Nodes that are connected to at least one node in GVC.

Multiplex Networks with Content-dependent Threshold Rule

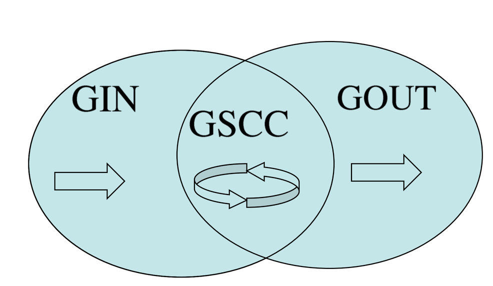
- We need to define *two* notions of vulnerability:
  - ◇ A node is $\mathbb{F}$-**vulnerable** if it becomes active by a single active neighbor in $\mathbb{F}$; i.e., if $c/(ck_1+k_2) \geq \tau$.
  - ◇ A node is $\mathbb{W}$-**vulnerable** if it becomes active by a single active neighbor in $\mathbb{W}$; i.e., if $1/(ck_1+k_2) \geq \tau$.
- If $c \neq 1$, then the subgraph of vulnerable nodes forms a **directed** graph.
  - ◇ A potentially bi-directional $\mathbb{F}$-link between nodes i and j will have the direction from i to j (resp. j to i) only if j (resp. i) is $\mathbb{F}$-vulnerable (similarly for $\mathbb{W}$-links).



$(c > 1)$

Components of a directed network

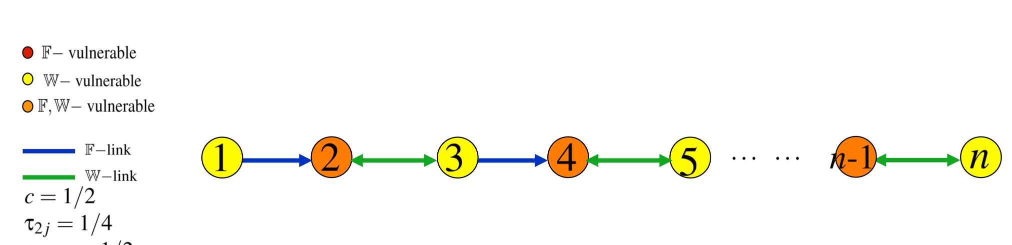
- *Out-component* of a vertex is the set of vertices that are reachable from it.
  - ◇ *In-component* of a vertex is the set of nodes that can reach that vertex.
- Giant out-component (**GOUT**): Set of nodes with **infinite** in-component.
- Giant in-component (**GIN**): Set of nodes that have **infinite** out-component.
- Giant strongly-connected component (**GSCC**): Intersection of GIN and GOUT.



A subtle picture

- **Condition** for global cascades = Existence of **GIN** (Existence of nodes with **infinite out-component**)
- **Probability** of global cascades = Fractional size of the **extended** giant in-component (**EGIN**)
  - ◇ Extended GIN = Nodes in GIN *plus* nodes that, once can activated, can activate a node in GIN
- Giant **vulnerable** component = **GSCC**
  - ◇ A set of vulnerable nodes s.t. activating **any** node (in this set) leads to the activation of **all** nodes in the set.

But, GIN may exist even if GSCC does not!



For $n \to \infty$

- A **positive** fraction of nodes have **infinite** out-component. ⇒ GIN exists!
- The **largest** strongly connected component consists of **two nodes**. ⇒ No GSCC!

Global cascades can take place even without a giant vulnerable component!

◇ **Contradicts all previous models!**

Analytic Results

Branching Process for Exploring Out-Components

- Start by activating an arbitrary node, and then recursively reveal the largest number of vulnerable nodes reached **and** activated by exploring its neighbors.
  - ◇ $g_i(x)$: Generating function for the **finite** number of nodes reached by following a type-i link. $(i = 1,2)$
  - ◇ $\rho_{\boldsymbol{k},i}$: Probability that a node with colored degree $\boldsymbol{k}$ is i-vulnerable. $(i = 1,2)$
  - ◇ $G(x)$: Generating function for the finite number of nodes reached and activated.
- Recursive relations:

$$g_i(x) = x\sum_{\boldsymbol{k}} \frac{k_i p_{\boldsymbol{k}}}{< k_i >} \cdot \rho_{\boldsymbol{k},i} \cdot g_1(x)^{k_1-\delta_{i1}} g_2(x)^{k_2-\delta_{i2}} + x^0 \sum_{\boldsymbol{k}} \frac{k_i p_{\boldsymbol{k}}}{< k_i >}(1-\rho_{\boldsymbol{k},i})$$

$$G(x) = x\sum_{\boldsymbol{k}} p_{\boldsymbol{k}} \cdot g_1(x)^{k_1} g_2(x)^{k_2}$$

- Jacobian matrix:

$$\boldsymbol{J}_p = \begin{bmatrix} \frac{<(k_1^2-k_1)\rho_{\boldsymbol{k},1}>}{<k_1>} & \frac{<k_1k_2\rho_{\boldsymbol{k},1}>}{<k_1>} \\ \frac{<k_1k_2\rho_{\boldsymbol{k},2}>}{<k_2>} & \frac{<(k_2^2-k_2)\rho_{\boldsymbol{k},2}>}{<k_2>} \end{bmatrix}$$

Global Cascade Condition: $\sigma(\boldsymbol{J}_p) > 1$ (spectral radius)

Probability of Global Cascades, P_{trig}: $1 - G(1)$

Expected Cascade Size

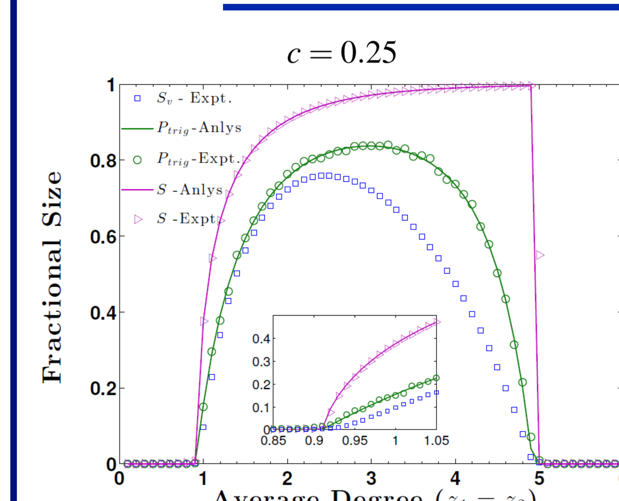
- Construct a tree with a single node at the top level $\ell = \infty$.
  - ◇ $q_{i,\ell}$: Probability that a node at level ℓ, connected to its unique parent by a type-i link, is active given that its parent is not. $(i = 1,2)$
- Recursive relations:

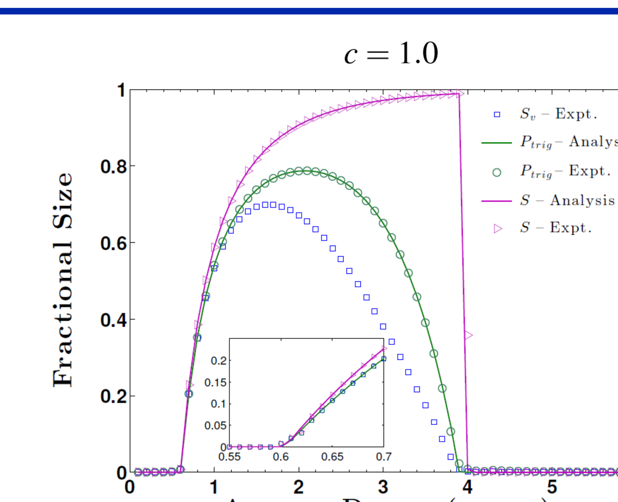
$$q_{i,\ell+1} = \sum_{\boldsymbol{k}} \frac{k_i p_{\boldsymbol{k}}}{< k_i >} \sum_{l=0}^{k_1-\delta i1} \sum_{j=0}^{k_2-\delta_{i2}} F((l,j),\boldsymbol{k}) \binom{k_1-\delta_{i1}}{l} q_{1,\ell}^l \times (1-q_{1,\ell})^{k_1-l-\delta_{i1}} \binom{k_2-\delta_{i2}}{j} q_{2,\ell}^j (1-q_{2,\ell})^{k_2-j-\delta_{i2}}, \quad \ell = 0,1,\ldots$$

$$S = \sum_{\boldsymbol{k}} p_{\boldsymbol{k}} \sum_{l=0}^{k_1} \sum_{j=0}^{k_2} F((l,j),\boldsymbol{k}) \binom{k_1}{l} q_{1,\infty}^l (1-q_{1,\infty})^{k_1-l} \times \binom{k_2}{j} q_{2,\infty}^j (1-q_{2,\infty})^{k_2-j}.$$

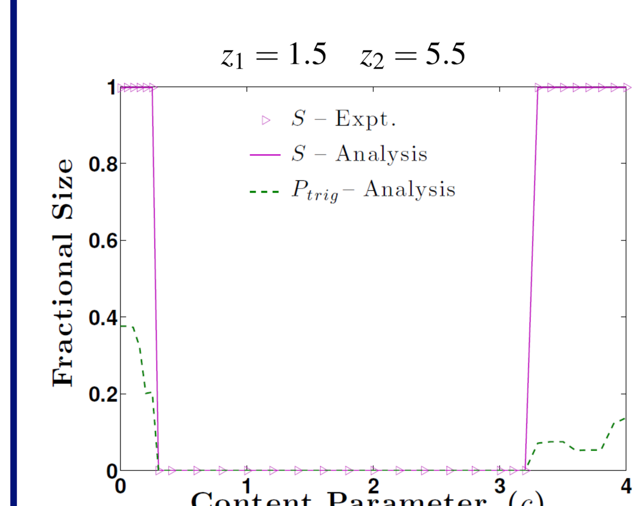
Expected Size of Global Cascades: S

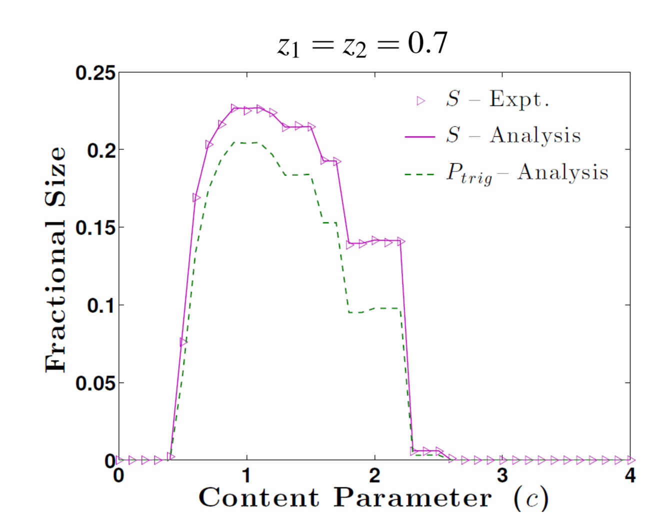
Simulation Results ($\mathbb{F}$, $\mathbb{W}$ ER with average degrees z_1, z_2)

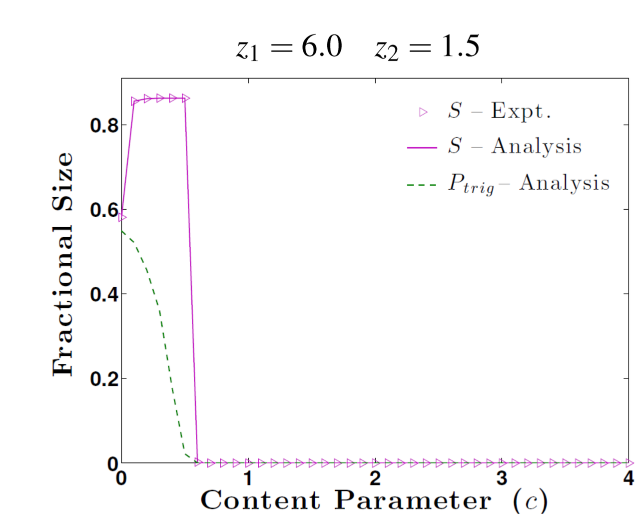


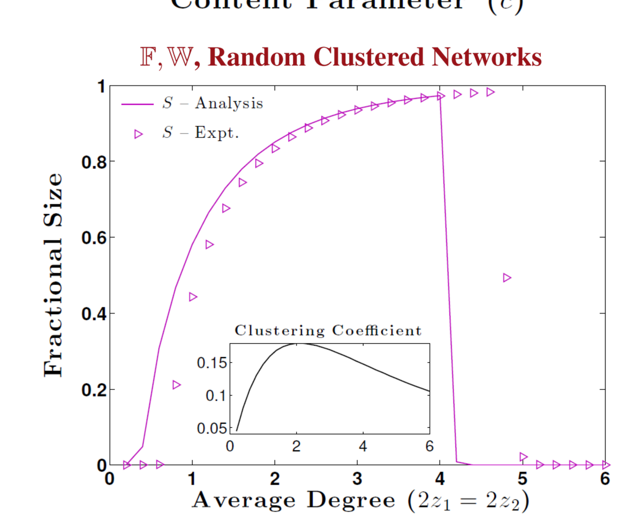


- **Excellent** agreement between analysis and simulations!
- Content parameter c effects the range, probability and size of cascades!
- Parameter ranges that give positive values for S, P_{trig} and S_v (GSCC) are identical for $c \neq 1$.
- Global cascades without a giant vulnerable cluster is **ruled out** in our model, although this possibility exists **in general**.









- Above, all parameters except c are fixed, and the variation of P_{trig} and S are observed.
- Content parameter c can **dramatically** change the dynamics of complex contagions over the **same** network.
- Depending on z_1, z_2, the range of c that favors global cascades changes significantly.
- On the left, we test our theory for random **clustered** networks.
- There is a **disagreement** between the predictions of our analysis **(zero-clustering)** and the actual cascade size from experiments **(positive clustering)**.
- The effect of clustering matches that observed in Watts' model! Low connectivity: $S \downarrow$ if $C \uparrow$; High connectivity: $S \uparrow$ if $C \uparrow$