# Intertwined Viral Marketing in Social Networks

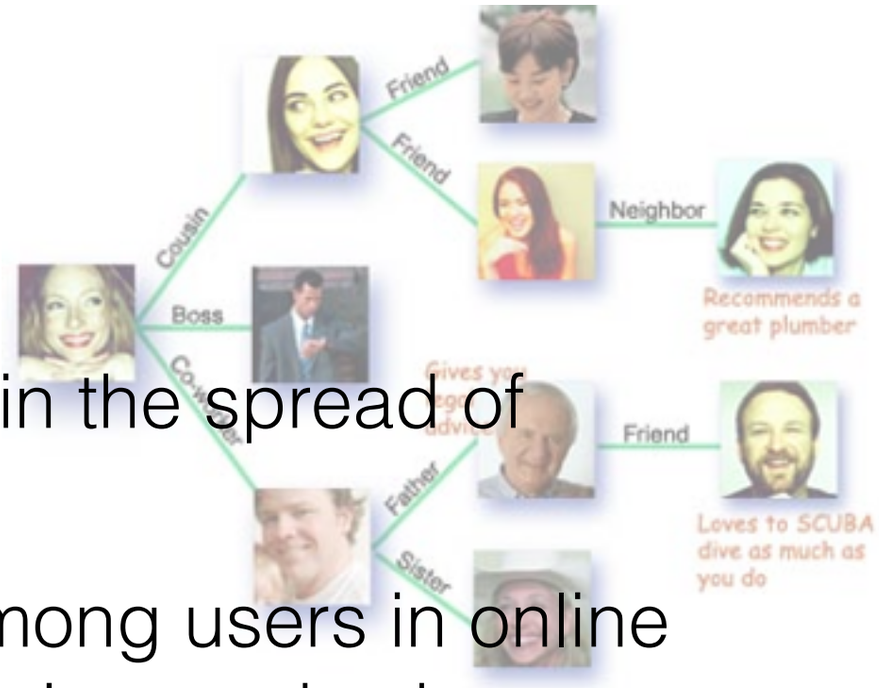
**Jiawei Zhang**[1], Senzhang Wang[2], Qianyi Zhan[3], Philip S. Yu[1]

[1]University of Illinois at Chicago, Chicago, IL, USA
[2]Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China
[3]Nanjing University, Nanjing, Jiangsu, China

# Outline

- Background Knowledge Introduction
- Intertwined Information Diffusion Model
- Problem Formulation of TIM
  - C-TIM vs J-TIM
- Proposed Methods: TIER
  - C-TIER vs. J-TIER
- Experimental Results
- Summary

# Viral Marketing Problem

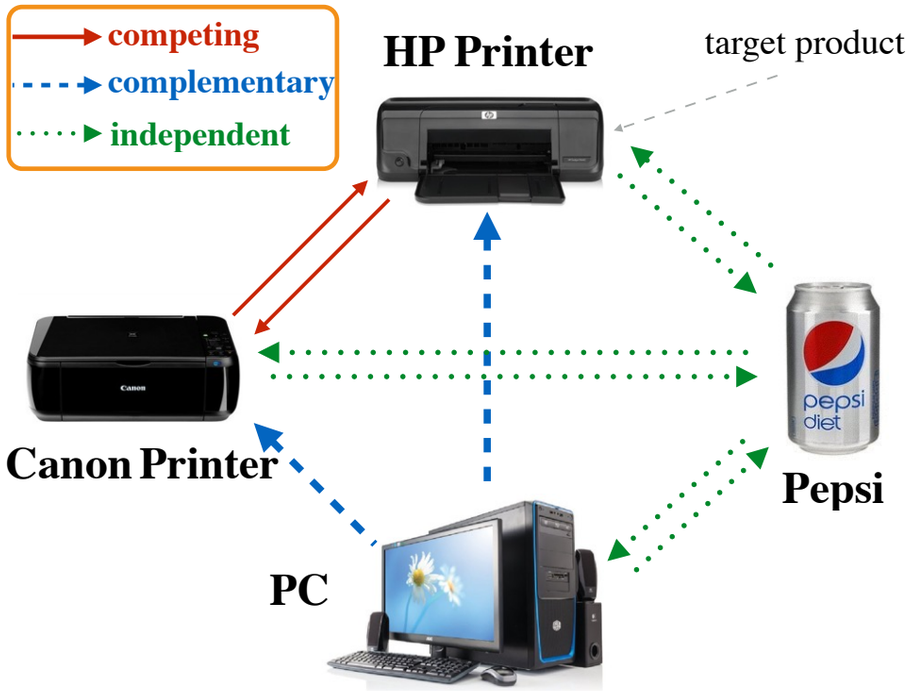



- Social networks play a fundamental role in the spread of information among the users.
- To model how information propagates among users in online social networks, several information diffusion modes have been proposed:
  - IC model, LT model, SIR model, etc.

- Viral marketing problem
  - *Given*: Advertising budget, and information diffusion model in online social network
  - *Objective*: Achieve the maximum influence in the social network
  - *Problem*: Which set of users should be targeted at initially?
  - *Application*: commercial promotion, election campaign

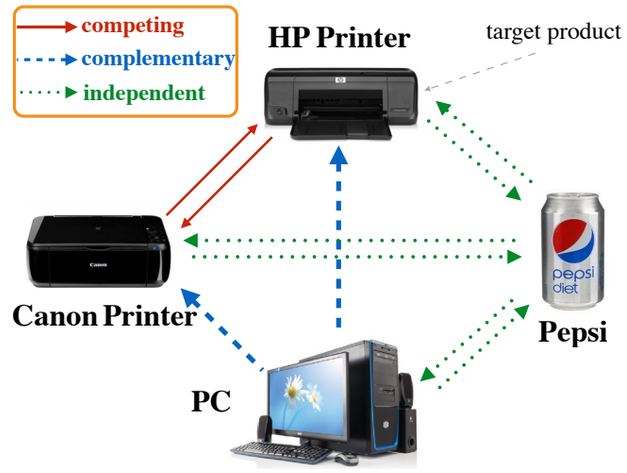
# Intertwined Viral Marketing Problem

- *Observation*: Multiple products are being promoted in the social network at the same time.



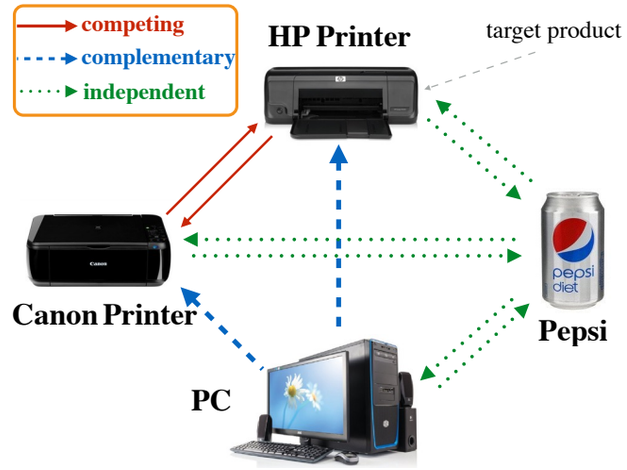

- Inter**T**wined **I**nfluence **M**aximization Problem (**TIM**)
  - *Given*: target product, advertising budget, *information diffusion model in the network*, product relationships;
  - *Objective*: Achieve the maximum influence for the target product;
  - *Problem*: Identify the optimal initial seed user for the target product.

# Intertwined Product Relationships




- The product relationships are intertwined:
  - *Competing*: Canon Printer <—> HP Printer;
    - Individuals who have Canon printer will be less likely to buy HP printer, and vice versa.
    - The purchase of Canon printer will **decrease** users' chance to buy HP printer.
  - *Complementary*: PC —> Canon Printer, PC —> HP Printer;
    - Individuals who have PC are more likely to buy a Canon printer or HP printer.
    - The purchase of PC will **increase** users' chance to buy printers.
  - *Independent*: PC <—> Diet Pepsi, Printer <—> Diet Pepsi.
    - The likelihood for individuals to buy PC/Printers has nothing to do with the purchase on Diet Pepsi.
    - The purchase of PC/Printers **doesn't affect** users' chance to buy Diet Pepsi.

# Intertwined Product Relationships




- Intertwined Product Relation Definition

**Definition** (Independent, Competing and Complementary Products): Let $P(s_i^j = 1)$ (or $P(s_i^j)$ for simplicity) denote the probability that $u_i$ is activated by product $p^j$ and $P(s_i^j|s_i^k)$ be the conditional probability given that $u_i$ has been activated by $p^k$ already. For products $p^j, p^k \in \mathcal{P}$, the promotion of $p^k$ is defined to be (1) *independent* to that of $p^j$ if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) = P(s_i^j)$, (2) *competing* to that of $p^j$ if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) < P(s_i^j)$, and (3) *complementary* to that of $p^j$ if $\forall u_i \in \mathcal{V}$, $P(s_i^j|s_i^k) > P(s_i^j)$.

**Definition** (Threshold Updating Coefficient): Term $\phi_i^{l\to j} = \frac{P(s_i^j)}{P(s_i^j|s_i^l)}$ is formally defined as the "*threshold updating coefficient*" of product $p^l$ to product $p^j$ for user $u_i$, where

$$\phi_i^{l\to j} \begin{cases} < 1, & \text{if } p^l \text{ is } \textit{complementary} \text{ to } p^j, \\ = 1, & \text{if } p^l \text{ is } \textit{independent} \text{ to } p^j, \\ > 1, & \text{if } p^l \text{ is } \textit{competing} \text{ to } p^j. \end{cases}$$

# Intertwined Information Diffusion Model

- Intertwined Information Diffusion Model (TLT)
  - Given network structure $G = (\mathcal{V}, \mathcal{E})$, product set $\mathcal{P}$, users activation thresholds $\{\theta_i^j\}_{u_i \in \mathcal{V}, p^j \in \mathcal{P}}$, user influence weight $\{w_{i,k}^j\}_{(u_i,u_k)\in\mathcal{E}, p^j \in \mathcal{P}}$ .
  - At step 1, information propagates from the seed user sets $\{\mathcal{S}^j\}_{p^j \in \mathcal{P}}$
  - At step t (t>1), all active users at step t-1 remain active, and inactive user $u_i$ will be activated by their neighbors($\Gamma_{out}(u_i)$) to buy product if

$$\sum_{u_l \in \Gamma_{out}(u_i)} w_{l,i}^j \geq \theta_i^j$$

  - For user $u_i$, who has been activated by products $p^{\tau_1}, p^{\tau_2}, \cdots, p^{\tau_m} \in \mathcal{P} \setminus \{p^j\}$ in a sequence, $u_i$'s threshold toward product $p^j$ will be

$$(\theta_i^j)^{\tau_1} = \theta_i^j \frac{P(s_i^j)}{P(s_i^j|s_i^{\tau_1})}, (\theta_i^j)^{\tau_2} = (\theta_i^j)^{\tau_1} \frac{P(s_i^j|s_i^{\tau_1})}{P(s_i^j|s_i^{\tau_1}, s_i^{\tau_2})}, \cdots, (\theta_i^j)^{\tau_m} = (\theta_i^j)^{\tau_{m-1}} \frac{P(s_i^j|s_i^{\tau_1}, \cdots, s_i^{\tau_{m-1}})}{P(s_i^j|s_i^{\tau_1}, \cdots, s_i^{\tau_{m-1}}, s_i^{\tau_m})},$$

  - In this paper, to simplify the calculation, we assume only the most recent activation has an effect on updating current thresholds:

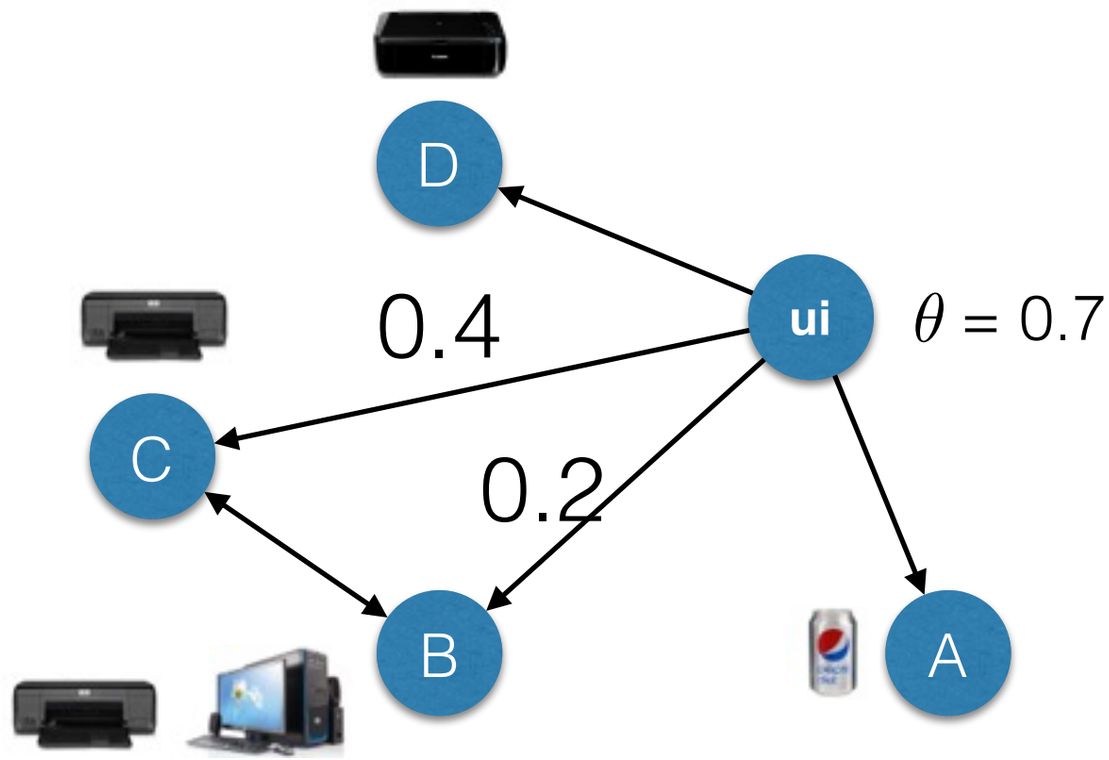
$$\frac{P(s_i^j|s_i^{\tau_1}, \cdots, s_i^{\tau_{m-1}})}{P(s_i^j|s_i^{\tau_1}, \cdots, s_i^{\tau_{m-1}}, s_i^{\tau_m})} \approx \frac{P(s_i^j)}{P(s_i^j|s_i^{\tau_m})} = \phi_i^{\tau_m \to j}.$$

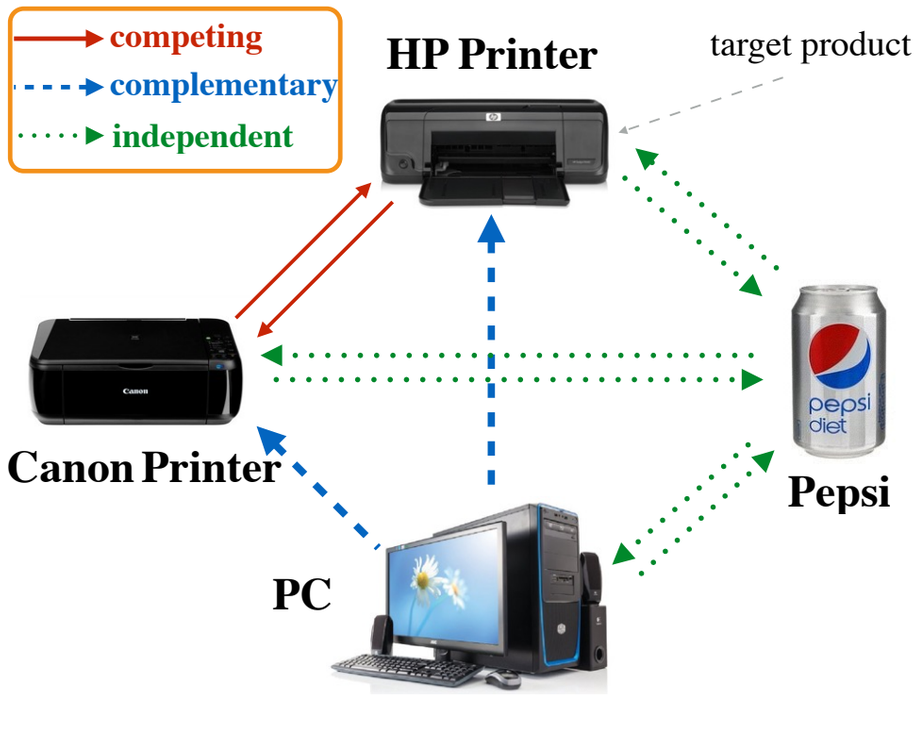
  - Therefore, we have

$$(\theta_i^j)^{\tau_m} \approx \theta_i^j \cdot \phi_i^{\tau_1 \to j} \cdot \phi_i^{\tau_2 \to j} \cdots \phi_i^{\tau_m \to j}.$$

  - The diffusion process stops if no further activation is possible.

# Intertwined Information Diffusion Model Example




- Initially, ui's threshold to HP can be $\theta$ = 0.7;
- ui cannot be activated to buy HP, because

$$w_{B,u_i}^{hp} + w_{C,u_i}^{hp} < \theta_j^{hp}$$

- Assume ui is activated by A to buy Pepsi, the new threshold will be $\theta \cdot \phi_i^{pepsi \to hp}$ = 0.7;
- If ui is activated by B to buy PC, the new threshold will be $\theta \cdot \phi_i^{pepsi \to hp} \cdot \phi_i^{pc \to hp}$ = 0.35
- Therefore, user ui will be activated by B, C to buy HP printer, since the influence is greater than the updated threshold

$$\phi_i^{pc \to canon} = 0.5$$

$$\phi_i^{pc \to hp} = 0.6$$

$$\phi_i^{canon \to hp} = 1.6$$

$$\phi_i^{hp \to canon} = 2.5$$

$$\phi_i^{\cdot \to pepsi} = \phi_i^{pepsi \to \cdot} = 1.0$$

# Intertwined Viral Marketing Problem

- Two variants of the TIM problem:
  - *Conditional TIM problem: C-TIM*
    - The other products are promoted ahead of the target product.
    - Information about other products have been propagated to users in the network already.
    - E.g., Apple to announce *iPhone 7* long after the release of iPad Pro, Samsung Galaxy S7, etc.
  - Joint TIM problem: J-TIM
    - The other products are being promoted simultaneously with the target product in the network.
    - Information about all the products have not be spread to users in the network yet.
    - E.g., Apple and Samsung will release the *new iPhone* and new Galaxy phone to compete for the market share.

# Conditional TIM Problem

- After the spread of information about the other products, we can update the users' thresholds towards the target product.
- Based on the updated network, we can carry the promotion of the target product.
- Conditional Intertwined Influence Function Definition

**Definition** (Conditional Intertwined Influence Function): Let $\mathcal{S}^{-j} = (\mathcal{S}^1, \cdots, \mathcal{S}^{j-1}, \mathcal{S}^{j+1}, \cdots, \mathcal{S}^n)$ be the known seed user sets selected for all products in $\mathcal{P} \setminus \{p^j\}$, the *influence function* of the target product $p^j$ given the known *seed user sets* $\mathcal{S}^{-j}$ is defined as the *conditional intertwined influence function*: $I(\mathcal{S}^j|\mathcal{S}^{-j})$.

- C-TIM Problem Definition

**C-TIM Problem**: The C-TIM problem aims at selecting the optimal *marketing strategy* $\bar{\mathcal{S}}^j$ to maximize the *conditional intertwined influence function* of $p^j$ in the network, i.e.,

$$\bar{\mathcal{S}}^j = \arg_{\mathcal{S}^j} \max I(\mathcal{S}^j|\mathcal{S}^{-j}).$$

# Conditional TIM Problem Analysis and Solution

- C-TIM Problem Analysis

  **Theorem**: The C-TIM problem is *NP-hard* based on the TLT diffusion model.

- Conditional Intertwined Influence Function Property

  **Theorem**: For the TLT diffusion model, the *conditional influence function* is *monotone* and *submodular*.

- Solution: Conditional interTwined Influence EstimatoR (C-TIER)
- step-wise greedy method, which selects users who will introduce the maximum influence increase in each step

**Algorithm 1** The C-TIER Algorithm

**Input:** input social network $G = (\mathcal{V}, \mathcal{P}, \mathcal{E})$
target product: $p^j$
known seed user sets of $\mathcal{P} - \{p^j\}$: $\mathcal{S}^{-j}$
conditional influence function of $p^j$: $I(\mathcal{S}^j|\mathcal{S}^{-j})$
seed user set size of $p^j$: $k^j$
**Output:** selected seed user set $\mathcal{S}^j$ of size $k^j$

1: initialize seed user set $\mathcal{S}^j = \emptyset$
2: propagate influence of products $\mathcal{P} - \{p^j\}$ with $\mathcal{S}^{-j}$ and update users' thresholds with intertwined threshold updating strategy
3: **while** $\mathcal{V} \setminus \mathcal{S}^j \neq \emptyset \wedge |\mathcal{S}^j| \neq k^j$ **do**
4: pick a user $u \in \mathcal{V} - \mathcal{S}^j$ according to equation $\arg\max_{u \in \mathcal{V}} I(\mathcal{S}^j \cup \{u\}|\mathcal{S}^{-j}) - I(\mathcal{S}^j|\mathcal{S}^{-j})$
5: $\mathcal{S}^j = \mathcal{S}^j \cup \{u\}$
6: **end while**
7: return $\mathcal{S}^j$.

# Experimental Results of C-TIM Problem

- Experimental Datasets
  - Facebook Network
  - Wikipedia Network
  - arXiv Collaboration Network
  - Epinions Network

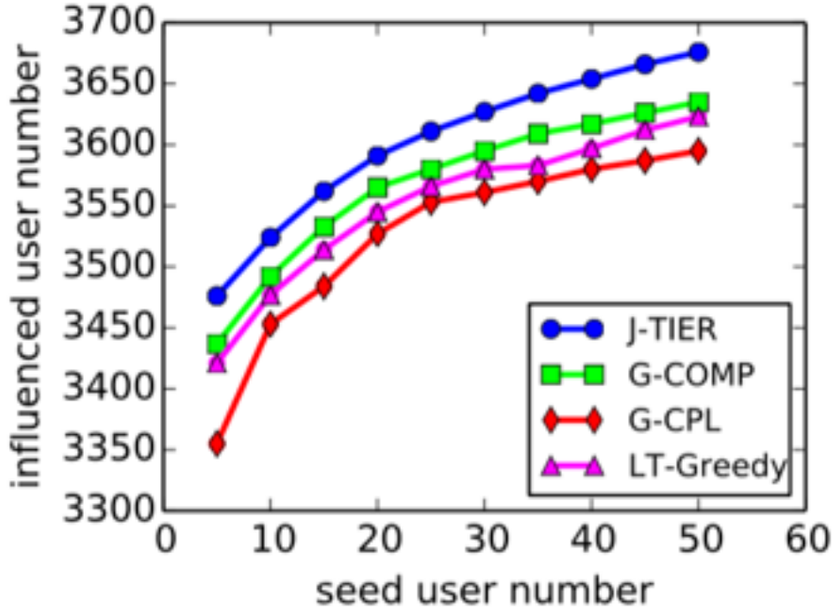
TABLE I
PROPERTIES OF THE DIFFERENT NETWORKS

| network | # nodes | # links | link type |
|---|---|---|---|
| Facebook | 4,039 | 88,234 | undirected |
| Wikipedia | 7,115 | 103,689 | directed |
| arXiv | 5,242 | 14,496 | undirected |
| Epinions | 7,725 | 82,861 | directed |

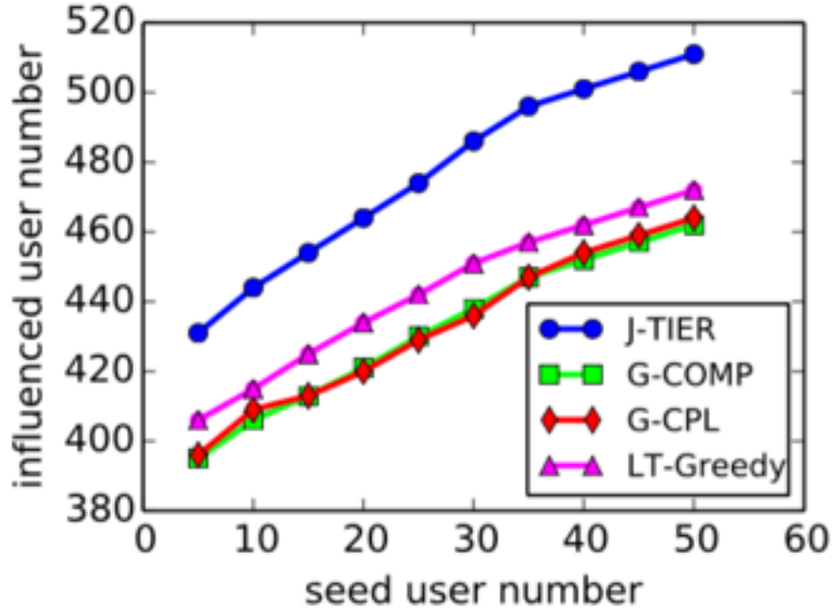
- Comparison Methods
  - C-TIER: Step-wise greedy seed user selection method based on TLT diffusion model
  - LT-Greedy: Step-wise greedy seed user selection method based on traditional LT diffusion model without considering product relationships
  - LT-PageRank: Select nodes with the top K PageRank scores
  - LT-Degree: Select nodes with the top K degree scores
  - LT-Random: Randomly select K nodes

# Experimental Results of C-TIM Problem

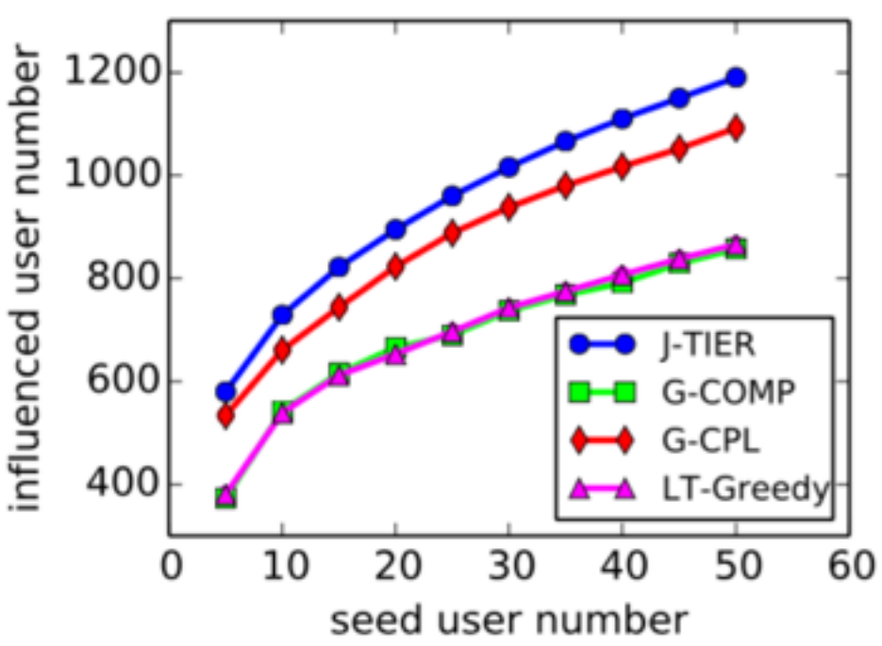
- Experimental Results



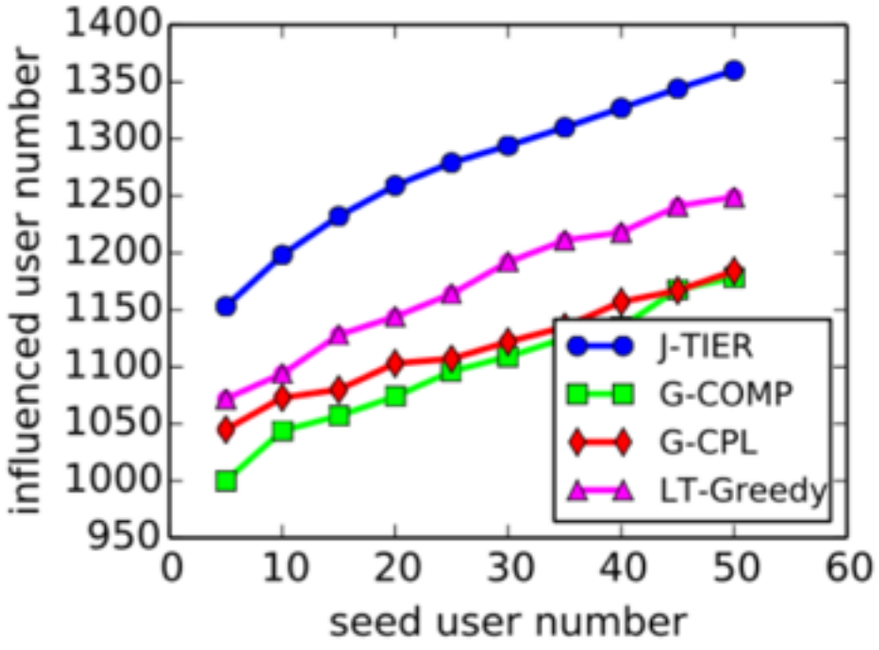

(a) Facebook




(b) Wikipedia




(c) arXiv




(d) Epinions

# Joint TIM Problem

- Products with Intertwined relationships are being promoted in online social networks at the same time.
- The seed users selected by other products are unknown, and the information about other products has not been propagated yet.
- Joint Intertwined Influence Function Definition

**Definition** (Joint Intertwined Influence Function): When the seed user sets of products $\mathcal{P} \setminus \{p^j\}$ are unknown, i.e., $\mathcal{S}^{-j}$ is not given, the *influence function* of product $p^j$ together with other products in $\mathcal{P} \setminus \{p^j\}$ is defined as the *joint intertwined influence function*: $I(\mathcal{S}^j; \mathcal{S}^{-j})$.

- C-TIM Problem Definition

**J-TIM Problem**: J-TIM problem aims at choosing the optimal *marketing strategy* $\bar{\mathcal{S}}^j$ to maximize the *joint intertwined influence function* of $p^j$ in the network, i.e.,

$$\bar{\mathcal{S}}^j = \arg_{\mathcal{S}^j} \max I(\mathcal{S}^j; \mathcal{S}^{-j}),$$

where set $\mathcal{S}^{-j}$ can take any possible value.

# Joint TIM Problem Analysis

- J-TIM Problem Analysis

  **Theorem**: The J-TIM problem is *NP-hard* based on the TLT diffusion model.

- Joint Intertwined Influence Function Property

  **Theorem**: Based on the TLT diffusion model, the *joint influence function* is *monotone* and *submodular* if all the other products are *independent* to the target product $p^j$.

  **Theorem**: Based on the TLT diffusion model, the *joint influence function* is not *monotone* nor *submodular* if there exist products which are either *competing* or *complementary* to the target product $p^j$.

- No theoretic performance guarantee exists for the step-wise greedy seed user selection algorithm in the J-TIM problem if there exists one products either competing or complementary to the target product.

# J-TIM Problem Solution: J-TIER

- Joint interTwined Influence EstimatoR (J-TIER)
  - In J-TIER, all the products are assumed to be “selfish” and aims at maximizing their influence gain, which leads to a “game” among products.
  - Formally, the seed users to be selected by all the products can be represented as set

$$\{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_j, \cdots, \mathcal{S}_{|\mathcal{P}|}\}$$

  - J-TIER lets the products to select seed users alternatively in random order step by step. Let $(\mathcal{S})^{\tau-1}$ be the seed users selected by all the products after round $\tau - 1$
  - If product $p^j$ is randomly picked to select seed users in round $\tau$, the selected seed user will be

$$\arg \max_{u \in \mathcal{V} - (\mathcal{S}^j)^{\tau-1}} [I\left((\mathcal{S}^j)^{\tau-1} \cup \{u\}; (\mathcal{S}^{-j})^{\tau-1}\right) - I\left((\mathcal{S}^j)^{\tau-1}; (\mathcal{S}^{-j})^{\tau-1}\right)].$$

  - If product $p^i$ is randomly picked to select seed user after $p^j$, the selected seed user will be

$$\hat{u}^i = \arg \max_{u \in \mathcal{V} - (\mathcal{S}^i)^{\tau-1}} [I\left((\mathcal{S}^i)^{\tau-1} \cup \{u\}; \bar{\mathcal{S}}^{-i}\right) - I\left((\mathcal{S}^i)^{\tau-1}; \bar{\mathcal{S}}^{-i}\right)].$$

  - Such a process stops until all the products finish the seed user selection process.

# J-TIM Problem Solution: J-TIER

- Joint interTwined Influence EstimatoR (J-TIER)
  - In J-TIER, all the products are assumed to be “selfish” and aims at maximizing their influence gain, which leads to a “game” among products.
  - Formally, the seed users to be selected by all the products can be represented as set

$$\{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_j, \cdots, \mathcal{S}_{|\mathcal{P}|}\}$$

  - J-TIER lets the products to select seed users alternatively in random order step by step. Let $(\mathcal{S})^{\tau-1}$ be the seed users selected by all the products after round $\tau - 1$
  - If product $p^j$ is randomly picked to select seed users in round $\tau$, the selected seed user will be

$$\arg \max_{u \in \mathcal{V} - (\mathcal{S}^j)^{\tau-1}} [I\left((\mathcal{S}^j)^{\tau-1} \cup \{u\}; (\mathcal{S}^{-j})^{\tau-1}\right) - I\left((\mathcal{S}^j)^{\tau-1}; (\mathcal{S}^{-j})^{\tau-1}\right)].$$

  - If product $p^i$ is randomly picked to select seed user after $p^j$, the selected seed user will be

$$\hat{u}^i = \arg \max_{u \in \mathcal{V} - (\mathcal{S}^i)^{\tau-1}} [I\left((\mathcal{S}^i)^{\tau-1} \cup \{u\}; \bar{\mathcal{S}}^{-i}\right) - I\left((\mathcal{S}^i)^{\tau-1}; \bar{\mathcal{S}}^{-i}\right)].$$

  - Such a process stops until all the products finish the seed user selection process.

# Experimental Results of J-TIM Problem

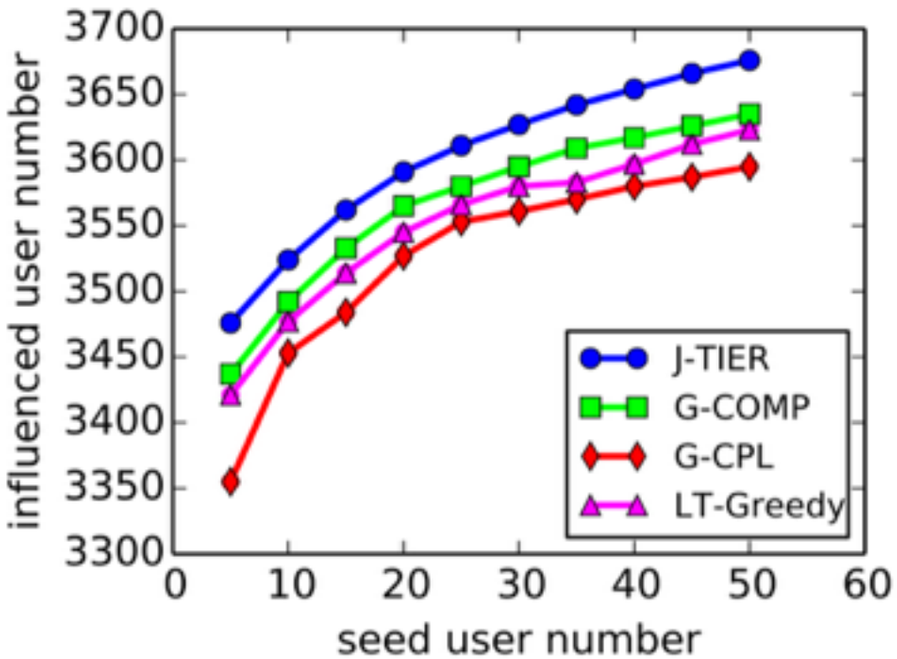
- Experimental Datasets
  - Facebook Network
  - Wikipedia Network
  - arXiv Collaboration Network
  - Epinions Network

TABLE I
PROPERTIES OF THE DIFFERENT NETWORKS

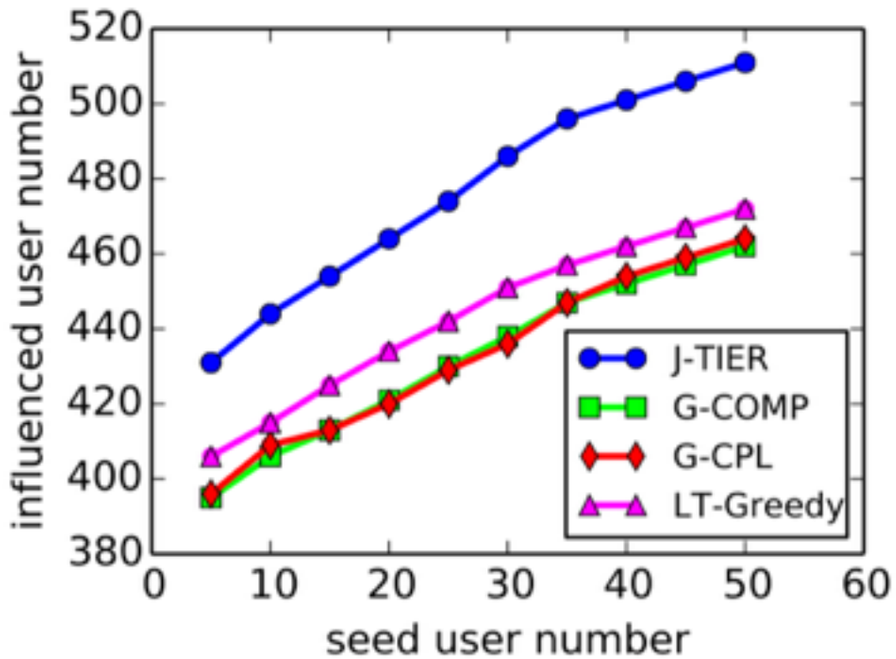
| network | # nodes | # links | link type |
|---|---|---|---|
| Facebook | 4,039 | 88,234 | undirected |
| Wikipedia | 7,115 | 103,689 | directed |
| arXiv | 5,242 | 14,496 | undirected |
| Epinions | 7,725 | 82,861 | directed |

- Comparison Methods
  - J-TIER: Iterative seed user selection method based on TLT diffusion model, which considers all products in the game.
  - G-COMP: Seed user selection considering the competing products only in the game.
  - G-CPL: Seed user selection considering the complementary products only in the game.
  - G-INDEP: Seed user selection considering the independent products only in the game.

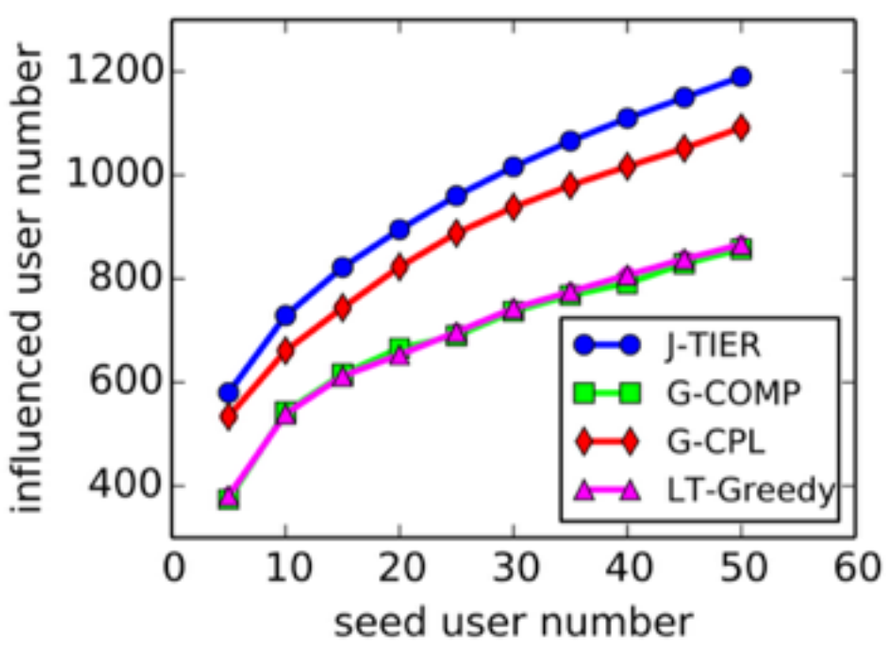
# Experimental Results of J-TIM Problem



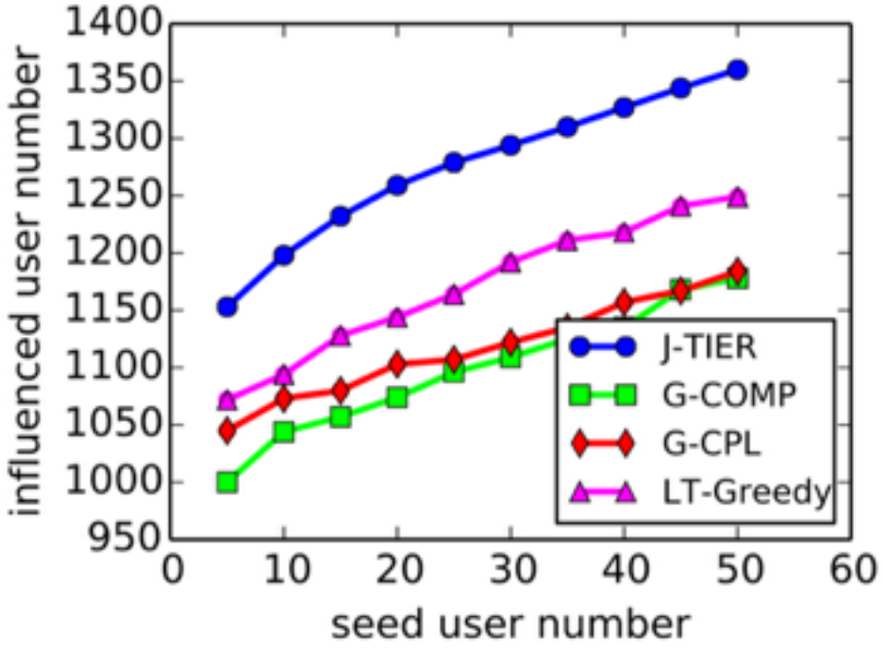

(a) Facebook




(b) Wikipedia




(c) arXiv




(d) Epinions

# Summary

- **Problem Studied**
  - *Intertwined viral marketing problem* in social networks with multiple products being promoted at the same time
- **Proposed Method**
  - *TLT Diffusion Model*: depicts the information diffusion process in online social networks considering the intertwined relationships among the products
  - *C-TIER for C-TIM problem*: step-wise greedy seed user selection, achieve 1-1/e approximation of the optimal result
  - *J-TIER for J-TIM problem*: game based alternative seed user selection, considers the ***competing***, ***complementary*** and ***independent*** products simultaneously

# Intertwined Viral Marketing in Social Networks

# Q&A

Jiawei Zhang[1], Senzhang Wang[2], Qianyi Zhan[3], Philip S. Yu[4]

jzhan9@uic.edu, szwang@nuaa.edu.cn, zhanqianyi@gmail.com, psyu@uic.edu

UIC
UNIVERSITY OF ILLINOIS AT CHICAGO

NANJING UNIVERSITY 1902

南京航空航天大学 1952 NUAA