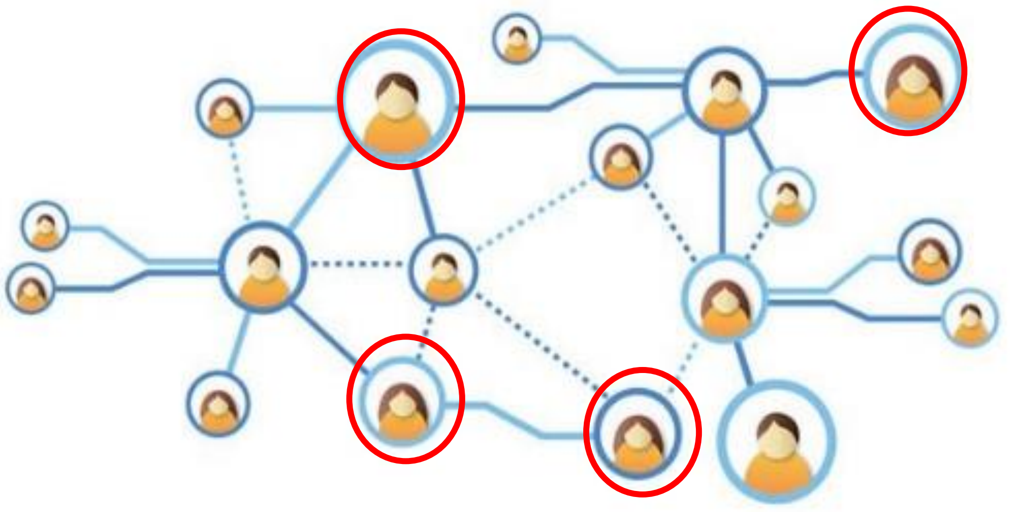
COMP810 Data Warehousing and Big Data

Complex Network with Big Data - Mining Large-scale and Dynamic Social Networks with Evolutionary Computation

Dr Weihua Li

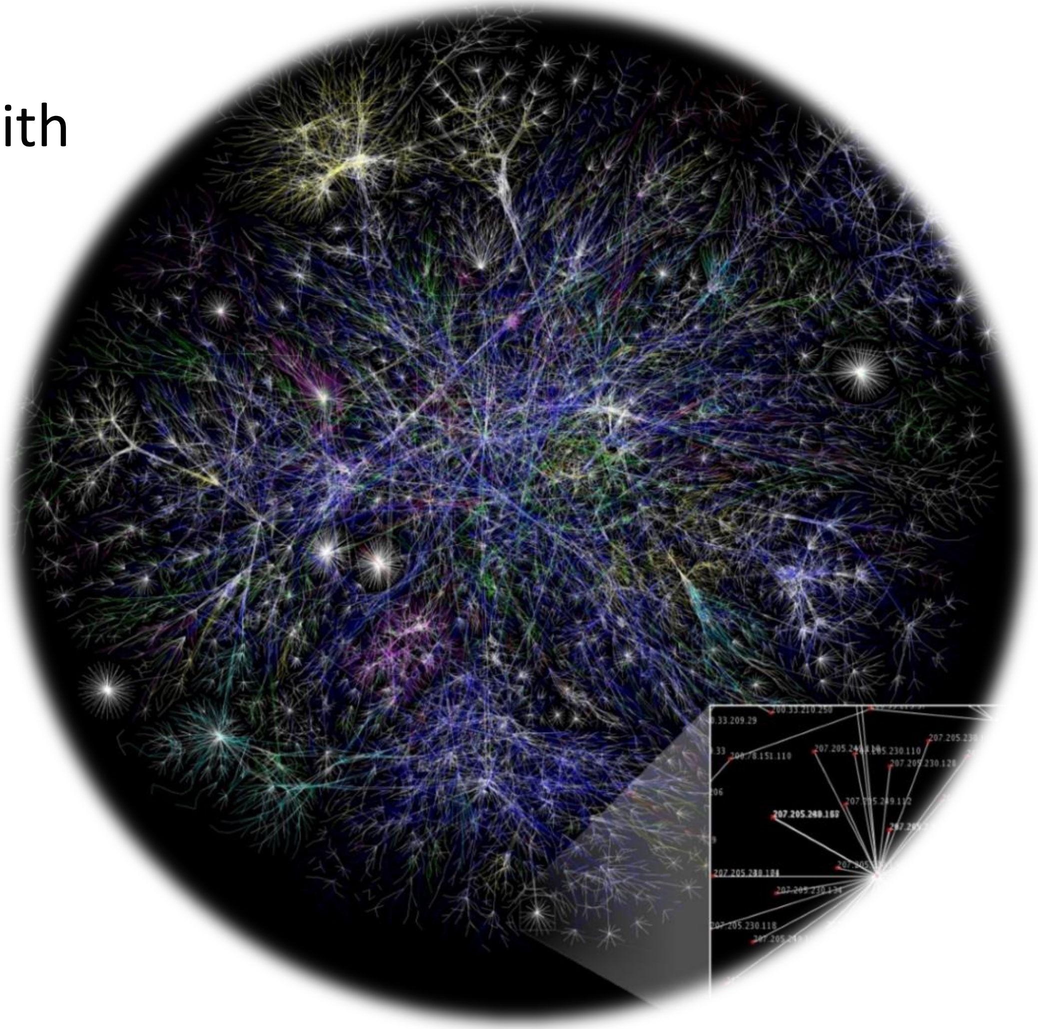
Outline

- Complex Network
- Social Influence Diffusion
- Classic Influence Diffusion Models
- Influence Maximization Problem
- Key Challenges
- Stigmergy-based Influencers Miner
- Experiments and Results



Complex Network

- A complex network is a graph (network) with non-trivial topological features
- Node: vertex, entity
- Link: edge, relationship
- Key Characteristics:
  - Large scale
  - Dynamic
  - Heterogeneity
- Social network is a typical example of complex network.



Social Influence Diffusion

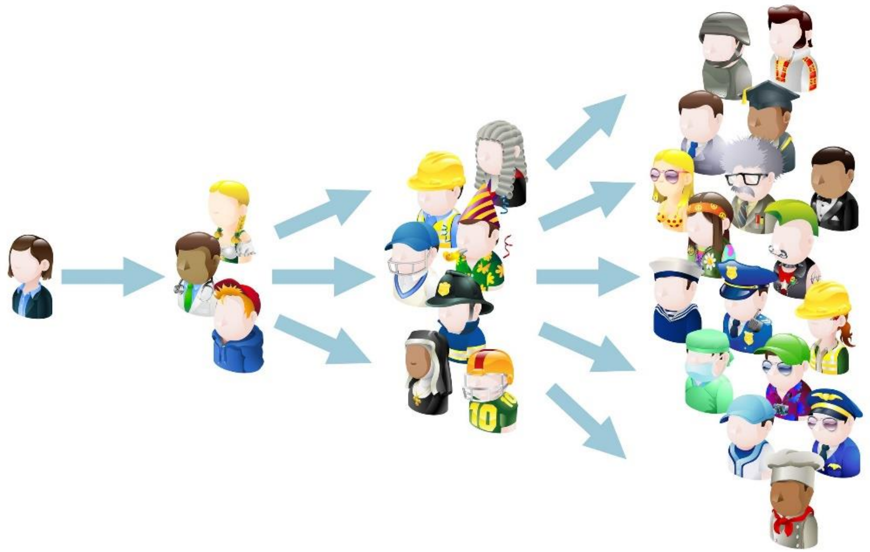
- Social influence
  - Emotions, options or behaviours are affected by others
  - Conformity, socialization, peer pressure, obedience, etc.
- Viral marketing
  - Drive network evolution direction
  - Word-of-mouth effect
- Influence propagation
  - Independent Cascade (IC) Model
  - Linear Threshold (LT) Model
  - Agent-based Influence Diffusion Model

Viral Marketing

Identify influential customers

Convince them to adopt the product – Offer discount/free samples

These customers endorse the product among their friends

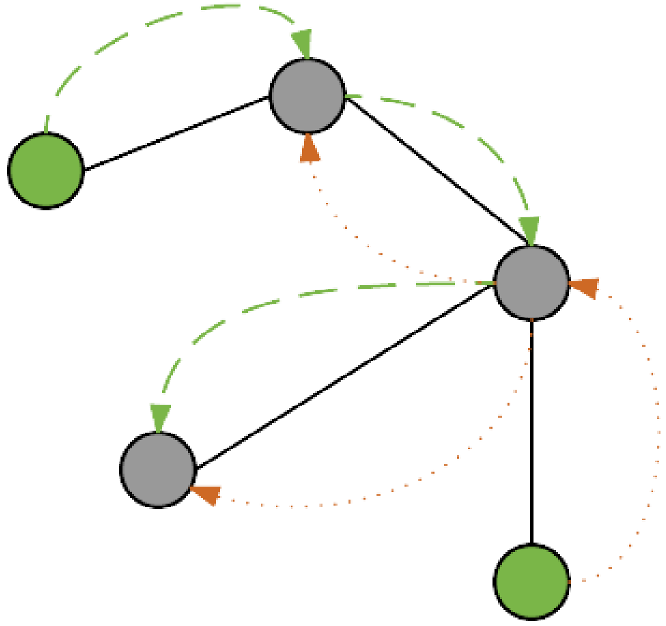


Classic Influence Diffusion Models and Influence Maximization

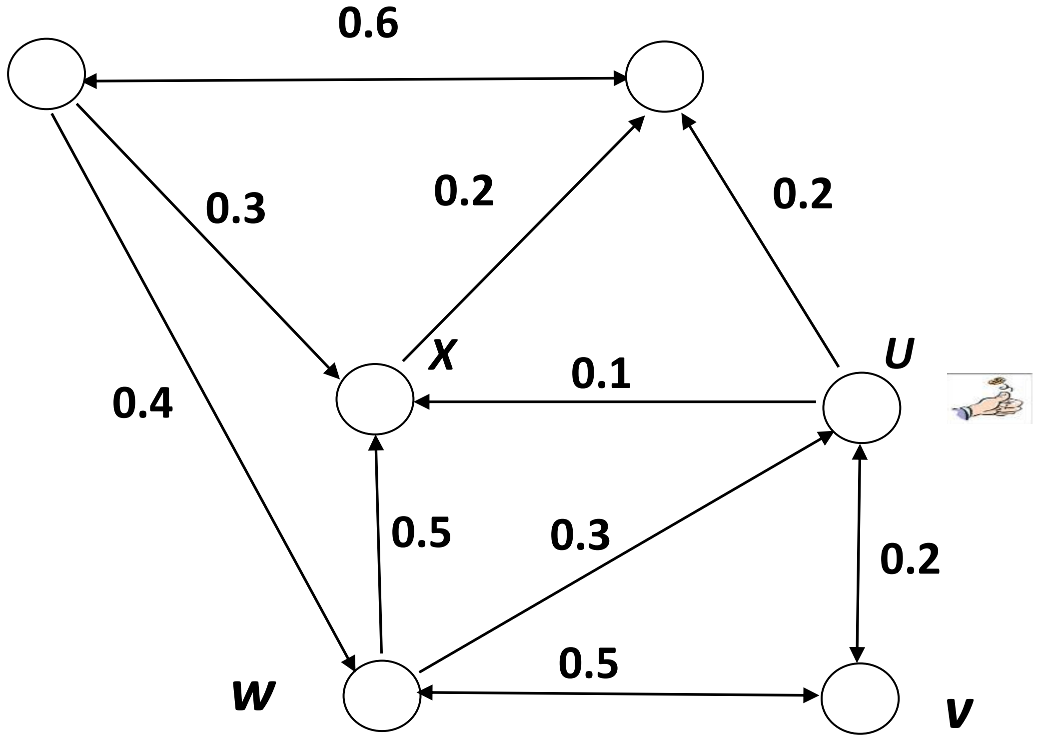
Independent Cascade (IC) Model

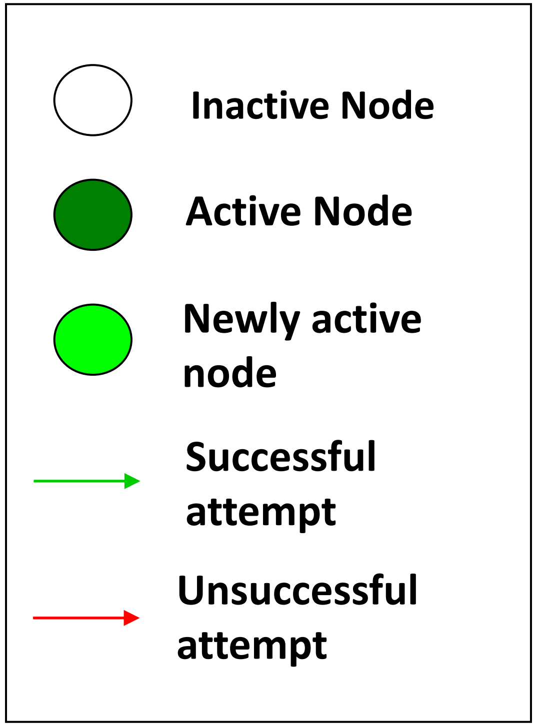
- When node v becomes active, it has a single chance of activating each currently inactive neighbours w.
- The activation attempt succeeds with probability p_{vw} .

- Key features
  - Propagation
  - Attenuation



IC Propagation Example





Stop!

By Kempe, KDD 2003

Linear Threshold (LT) Model

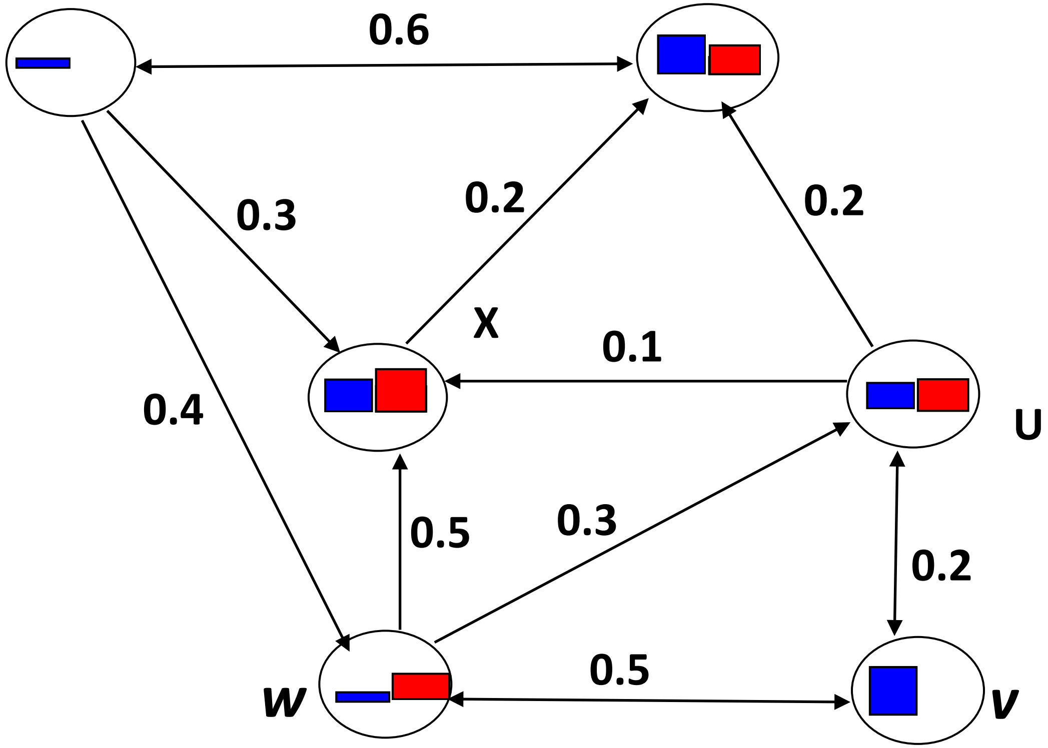
- Linear Threshold (LT) Model.
  - A node v has random threshold $\theta v \in [0, 1]$
  - A node v is influenced by each neighbour $w, w \in \Gamma(v)$, according to a weight $b_{v,w}$

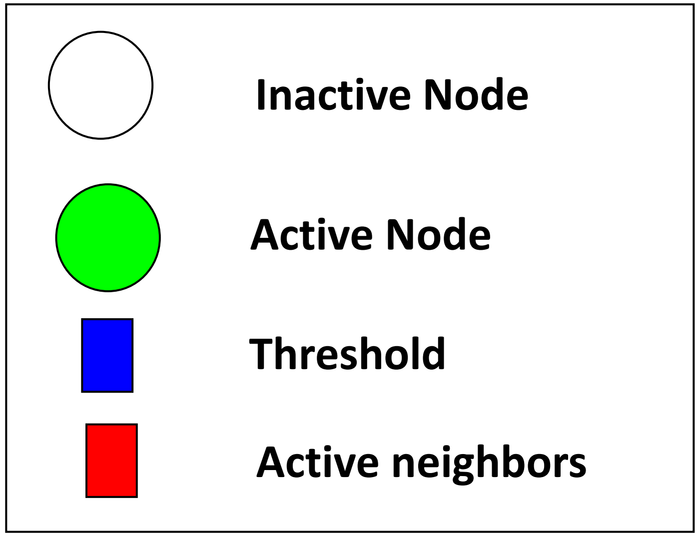
$$\sum_{w \in \Gamma(v)} b_{v,w} \leq 1$$

  - A node v becomes active when at least (weighted) θ_v fraction of its neighbours are active

$$\sum_{w \in \Gamma(v)} b_{v,w} \geq \theta_v$$

LT Propagation Example

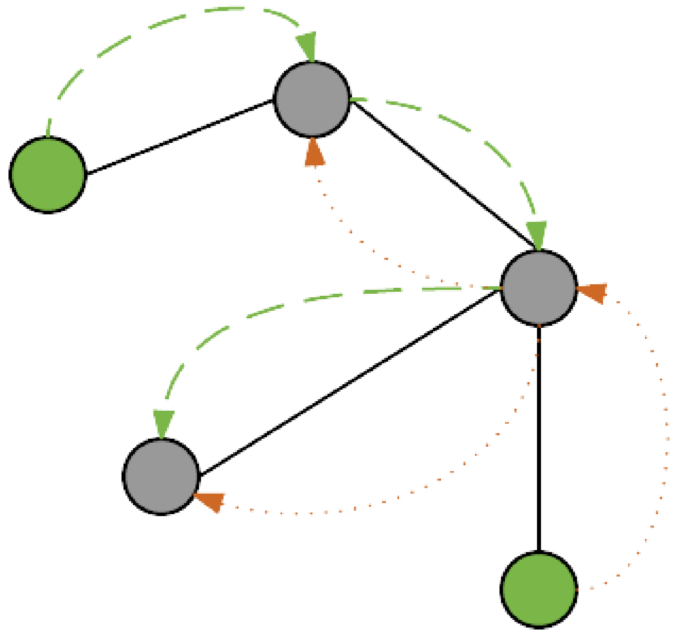




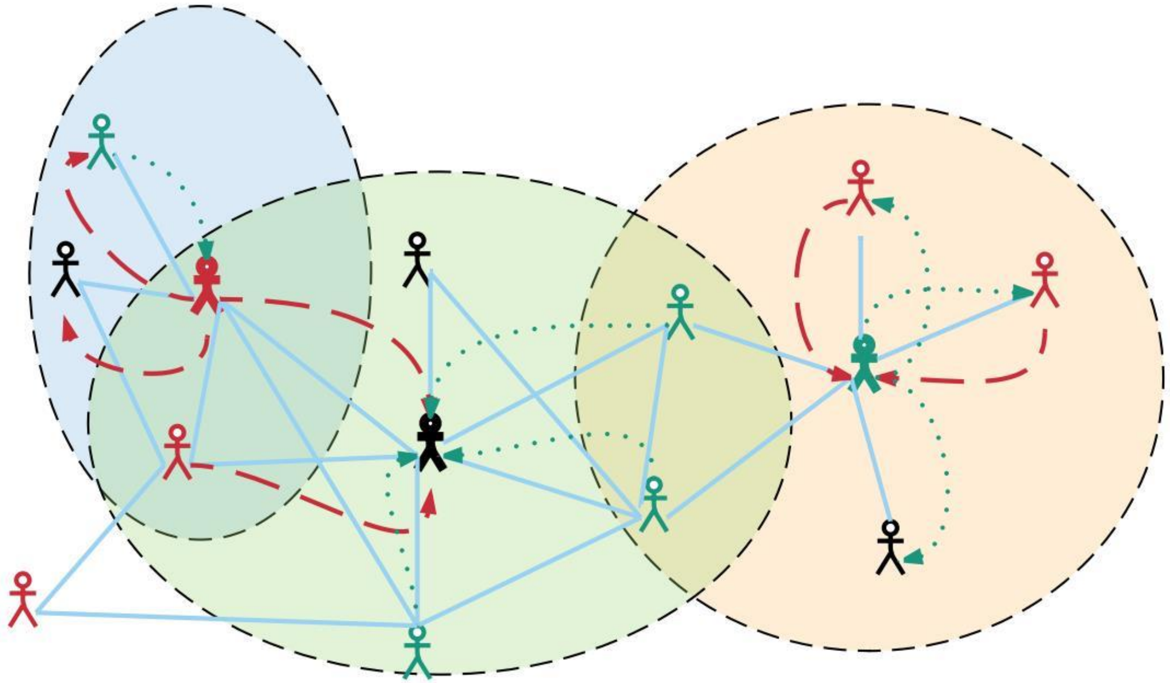
Stop!

By Kempe, KDD 2003

Agent-based Influence Diffusion

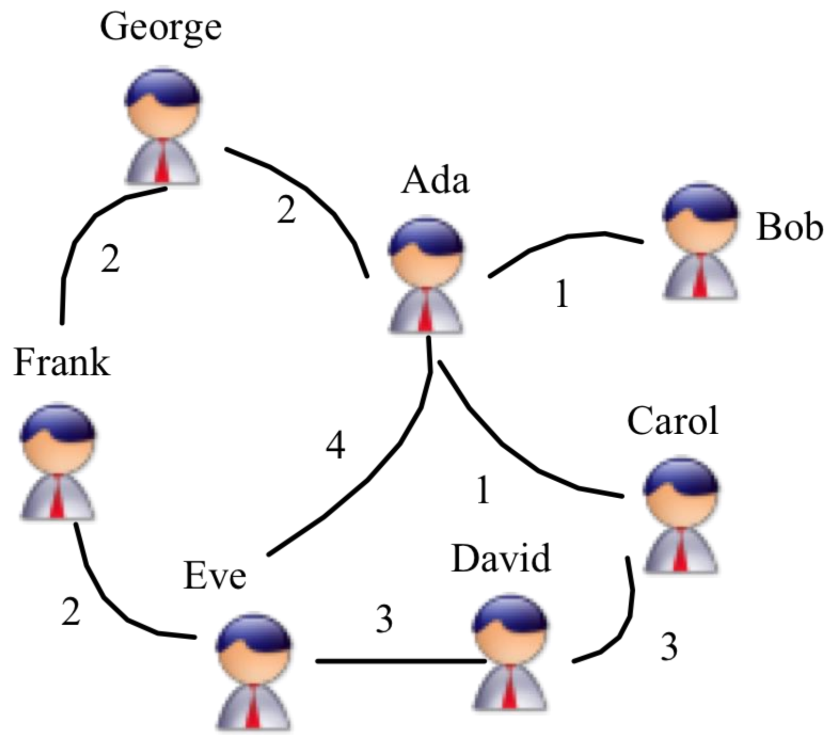


The classic influence diffusion models
Independent Cascade (IC) Model / Linear Threshold (LT) Model



Agent-based Influence Diffusion Model (Li proposed in IEEE International Conference on Agents, ICA 2016)

Influence Maximization Problem

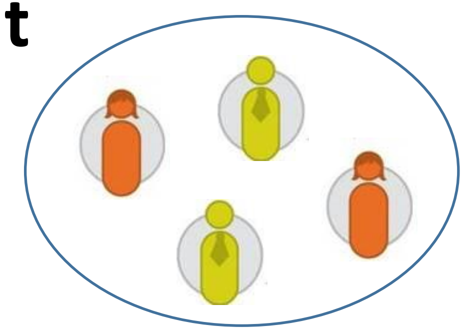


Social graph with influence probabilities of edges

Problem

- Select *k* individuals such that by activating them, the expected spread of influence is maximized.
- The selection process is called seed selection
- The identified set of users is named as seed set

Output



Seed set of size k

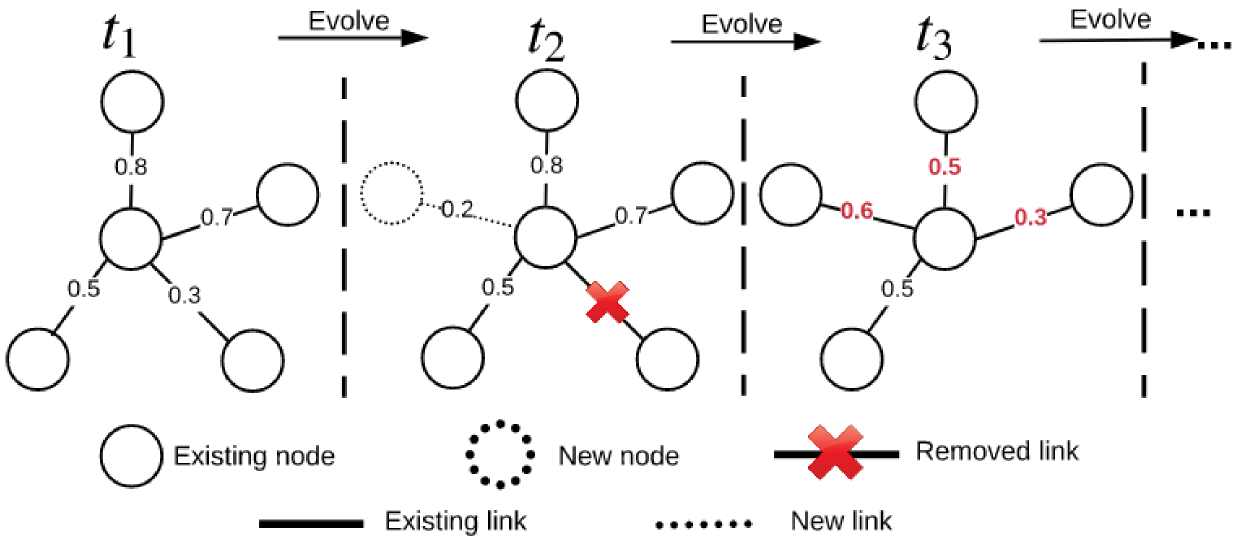
Classic Seed Selection Algorithms

- **Objective**
  - Achieve the maximum influence by selecting a set of limited users as seed set.
  - The overall influence through the network impact as many users as possible.
  - The seed set is regarded as an assembled “**team**” and influence propagation becomes a “**team work**”, rather than individual performance.

- **Traditional Seed Selection Approaches**
  - Random Selection: Select the users randomly.
  - Ranking-based Selection: Rank the users based on the node degree and select top k users.
  - Greedy Selection: Obtain the maximum influence marginal gain in selecting each seed.

Key Challenges

- Large-scale social networks
  - Huge number of nodes
  - Huge number of links
- Dynamics
  - Users join and quit
  - Links form and vanish
- Recalibration
  - Network evolves faster than some of the traditional seeding algorithms
  - Re-launch seeding algorithms takes time



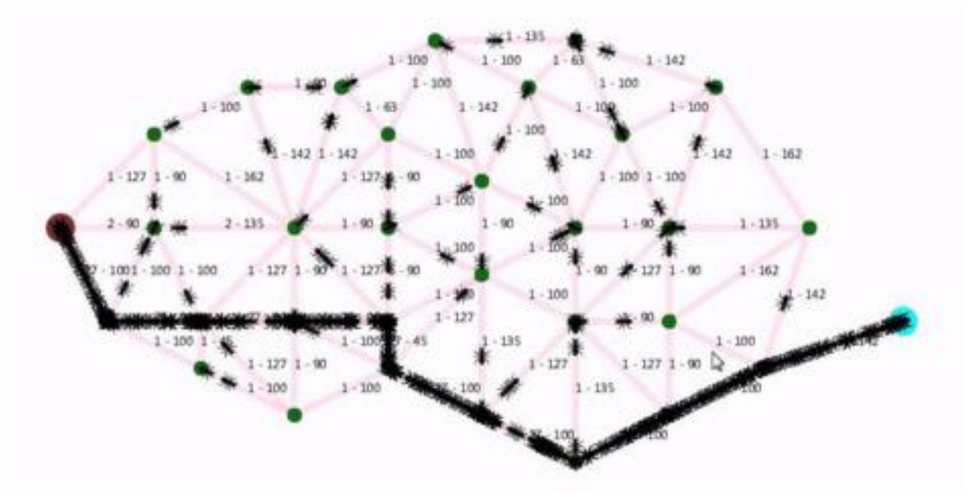
Proposed Approach

- Decentralised agent-based modelling:
  - Share the computational cost
  - Multiple agents work in parallel
- Evolutionary computing:
  - Continuously optimise the solution and handle the dynamics of large-scale social network.

Stigmergy-based Influencers Miner

Preliminary – Ant and Stigmergy Algorithms

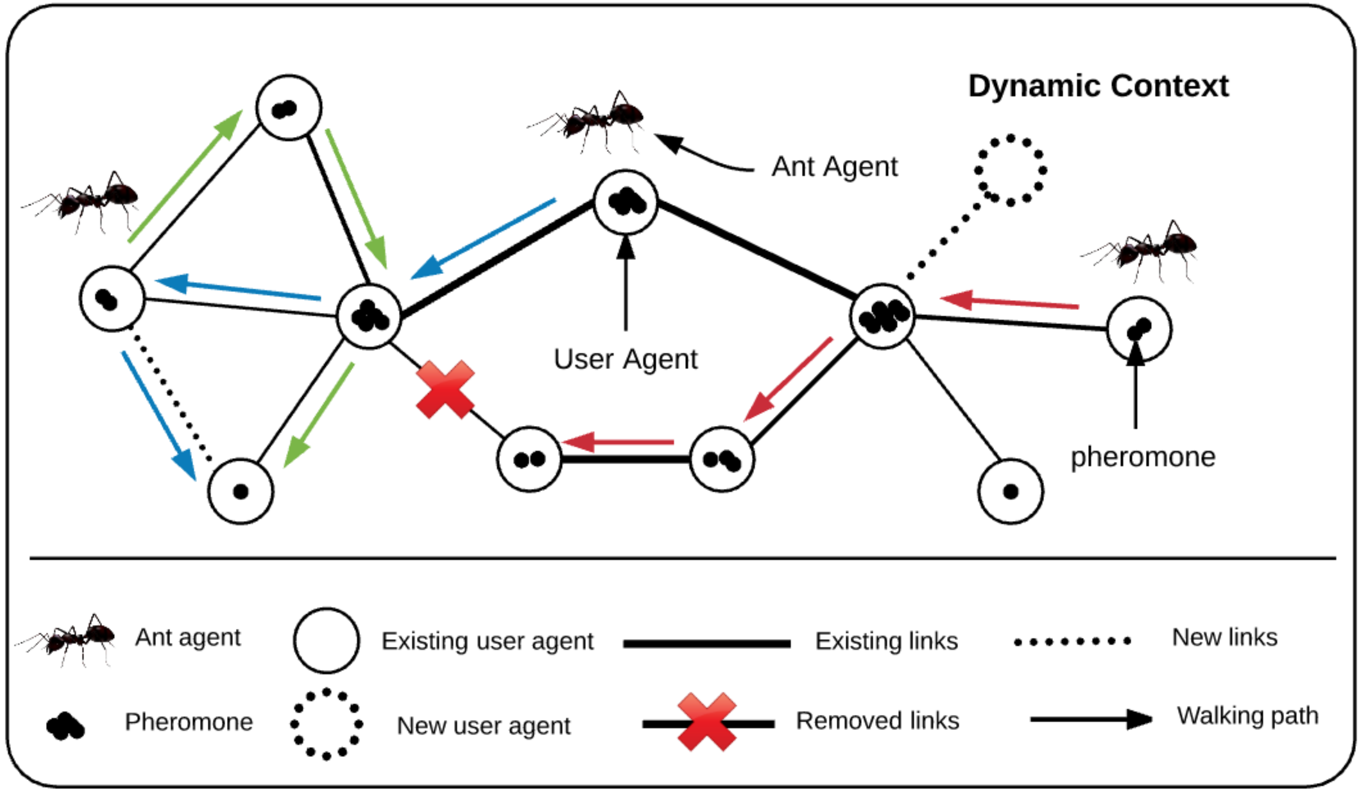
- Multi-Agent System (MAS)
- Indirect communications - stigmergic interactions
- The individuals of the species conduct communications and pass signals to others via a chemical substance, i.e., pheromone



SIMiner – Stigmergy-based Influencers Miner

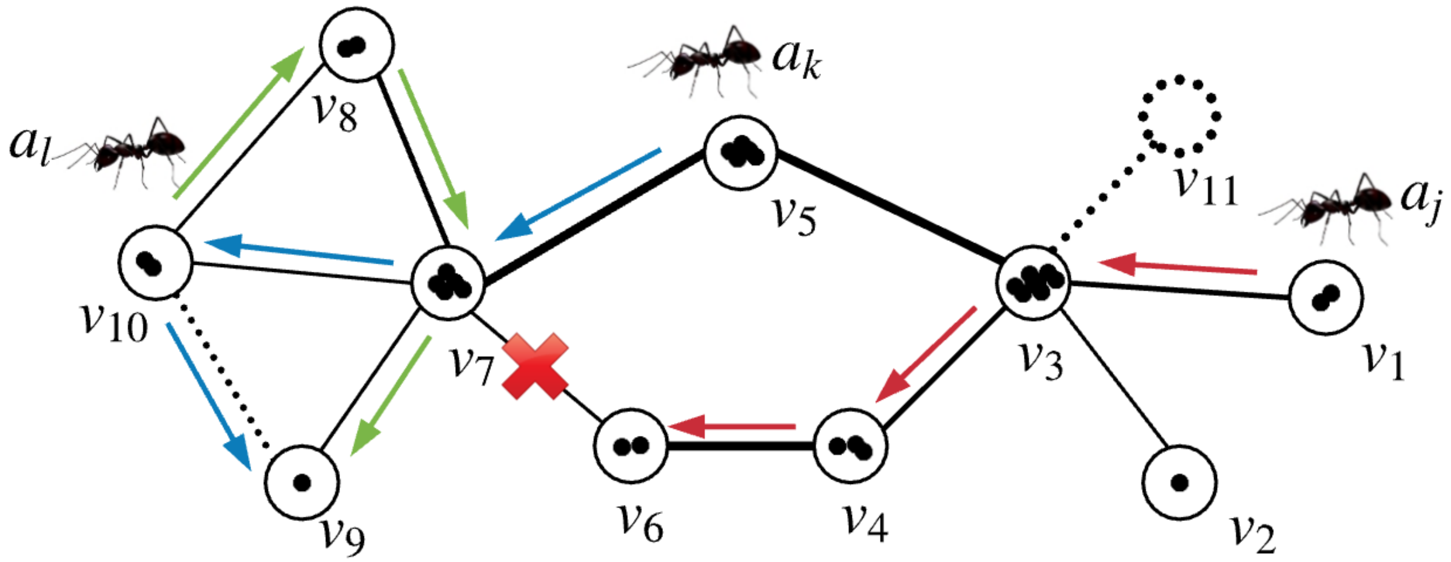
- Social network is modelled as the environment of ant agents.
- Influence diffusion process is modelled as ants' crawling behaviours.

- The objective of an ant is to explore the potential influential nodes from the network.
- Ants leave the pheromone on the nodes as messages, which can be referred by their peers.

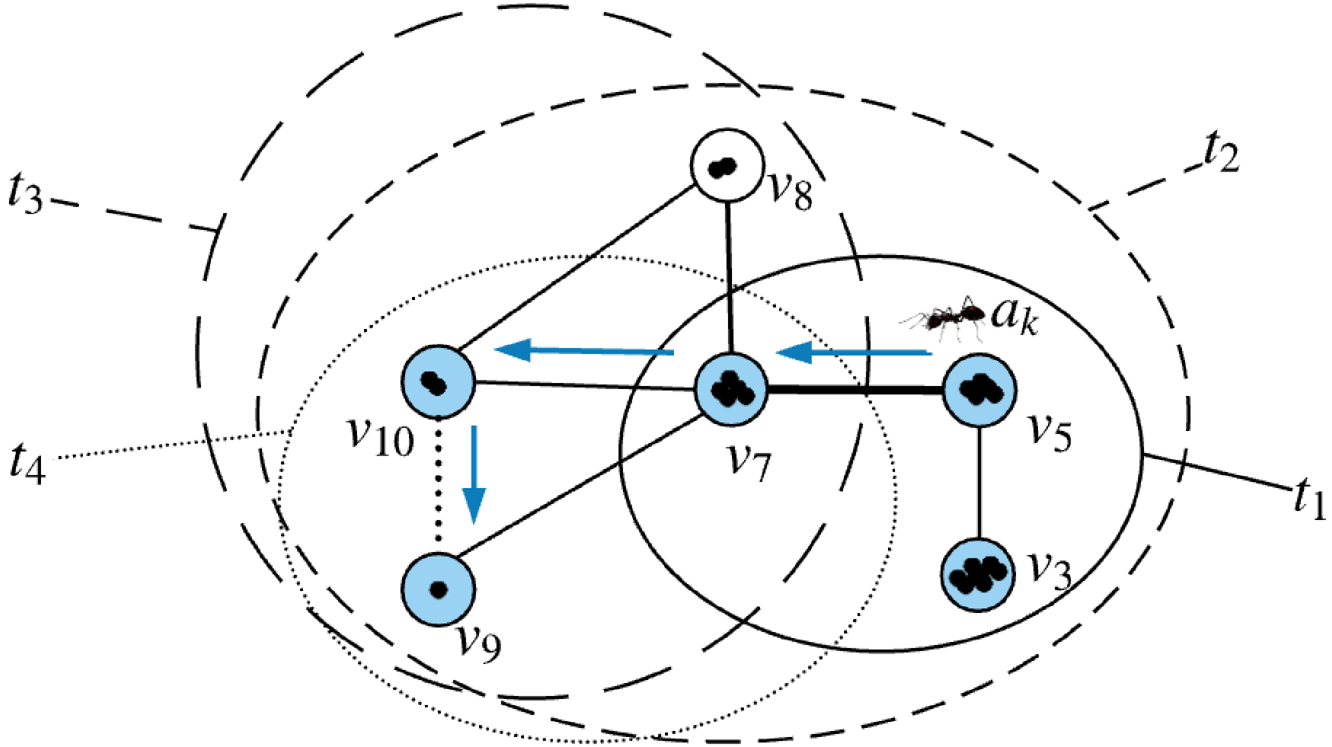


An Annotated Dynamic Social Network

- Numerous ant agents crawl simultaneously and update the shared environment by allocating pheromones on the nodes
- The pheromone amount distributed by an ant to a node is proportional to the contribution of the node in the tour.



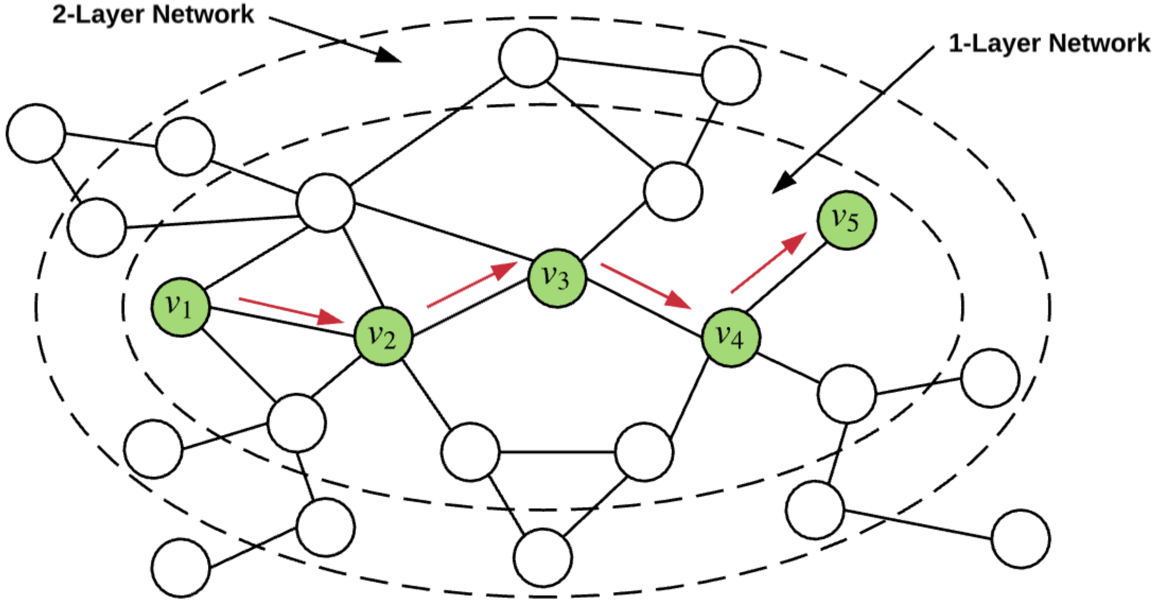
SIMiner – Agent's Local View

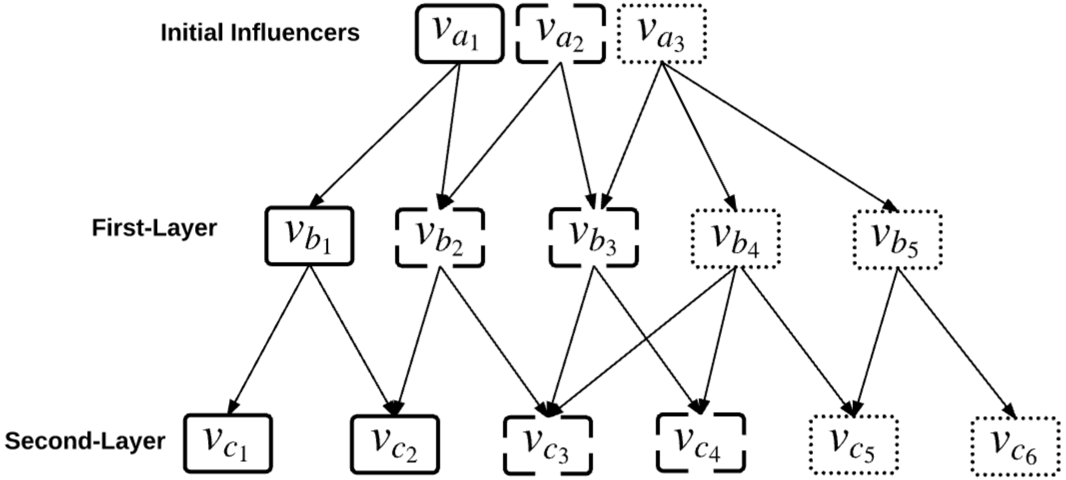


- Each ant agent has a local view, which keeps changing over time.
- The local view only cover the node and its corresponding neighbours.
- Ants keeps walking until reach the "end" of the network.

SIMiner – Sub-networks Generation

- Each ant agent captures a N-layer sub-graph after completing a tour.
- Simulate the number of influence cascade levels
- We use the following diffusion model to estimate the contribution of each node





SIMiner – Operations

- **Tour Formation**
  - ***Path Selection.*** Probability of an ant clawing from node i to node j is affected by both the edge weight and pheromone amount

$$q_{ij} = \begin{cases} \dfrac{(e_{ij}.w)^{\alpha} \cdot (v_j.p)^{\beta}}{\sum_{v_x \in \Gamma(v_i) \backslash T} (e_{ix}.w)^{\alpha} \cdot (v_x.p)^{\beta}}, & e_{ij} \in v_i.E \\ 0, & e_{ij} \notin v_i.E, \end{cases}$$

  - ***Skip.*** Many central nodes are clustered together and it is not necessarily to target all of them. Skip the node is common neighbour ratio is greater than a threshold

$$\eta_{ij} = \frac{|(v_i) \cap (v_j)|}{|(v_j)|}.$$

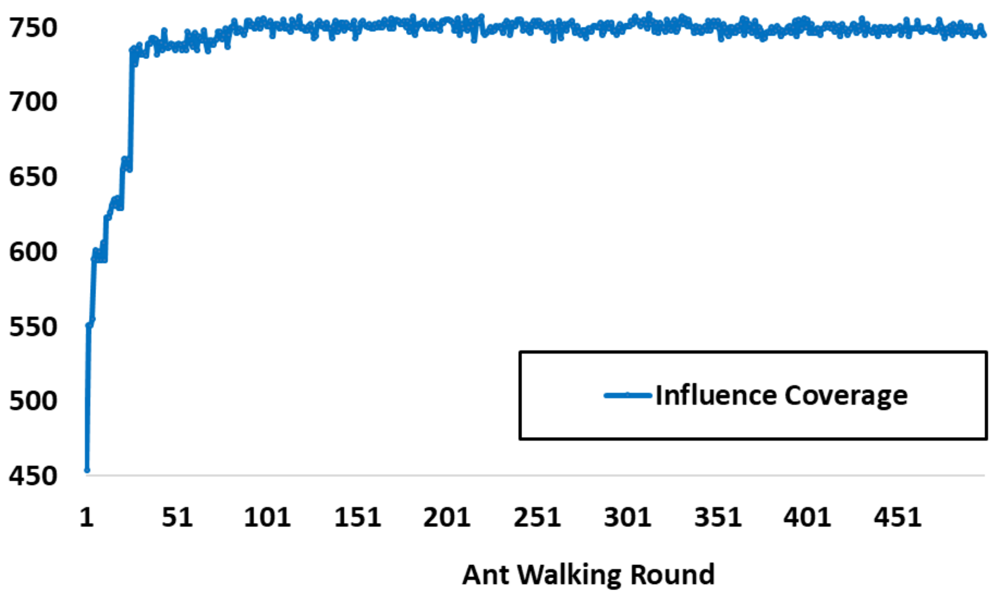
SIMiner – Operations (cont.)

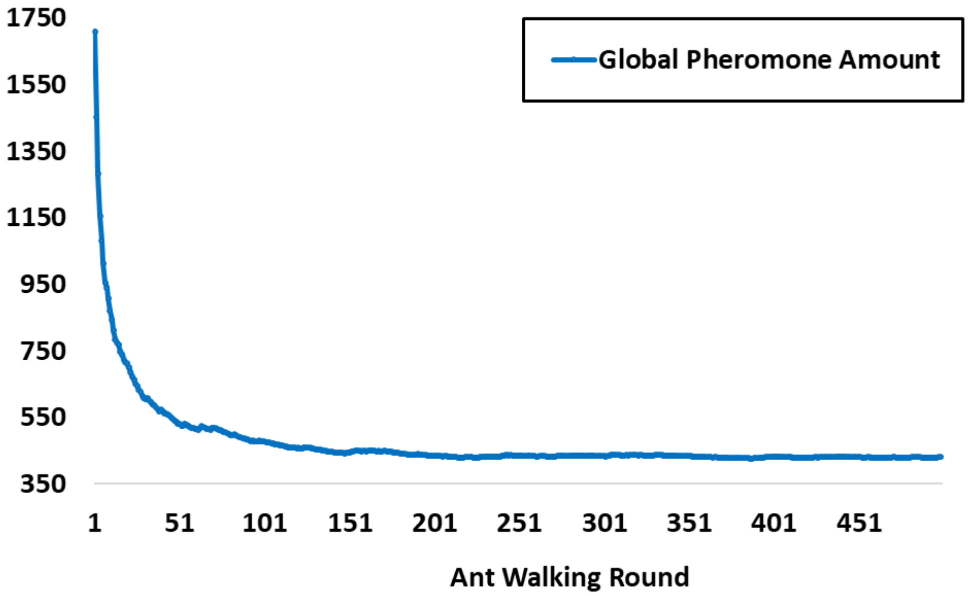
- **Pheromone Evaporation**
  - Amount of allocated pheromone decreases over time
  - Delay the faster convergence
  - Avoid to converge to a locally optimal solution

$$v_j.p(n+1) = v_j.p(n) \cdot \rho_n$$

$$\rho_n = 1 - \sqrt{\frac{n}{n+a}}.$$

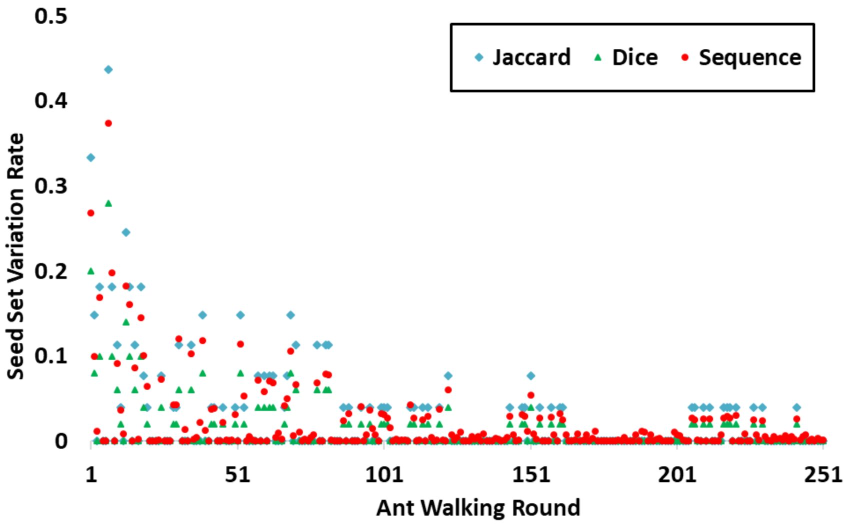
Experiment 1 - Convergence Analysis



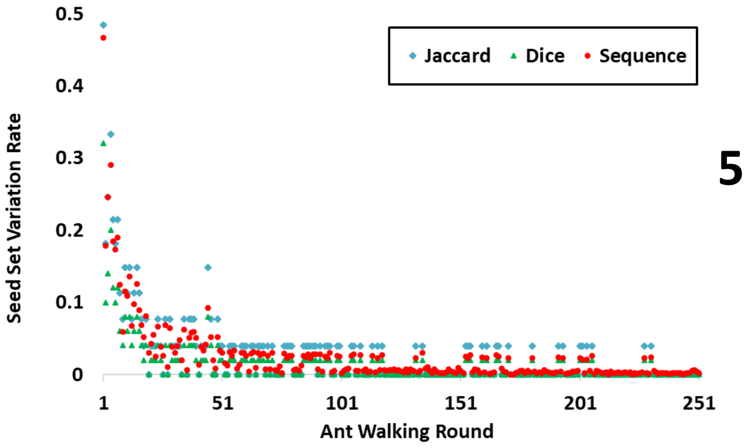


Experiment 1 - Convergence Analysis

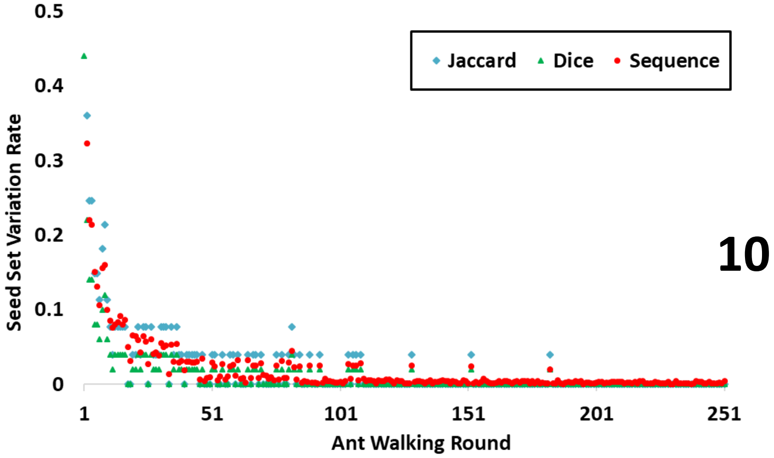
- Seed set (solution) variation



Single ant

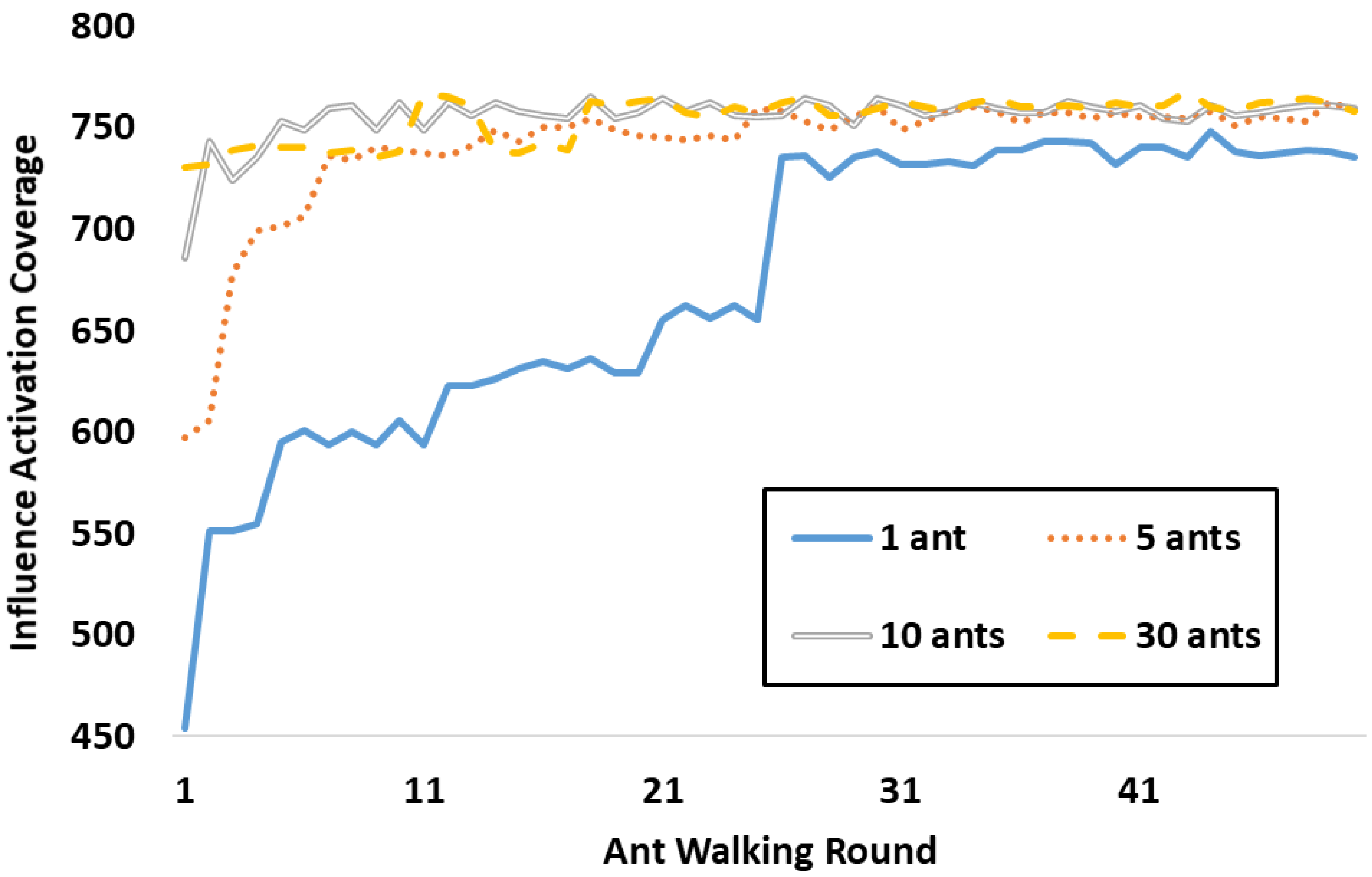


5 ants

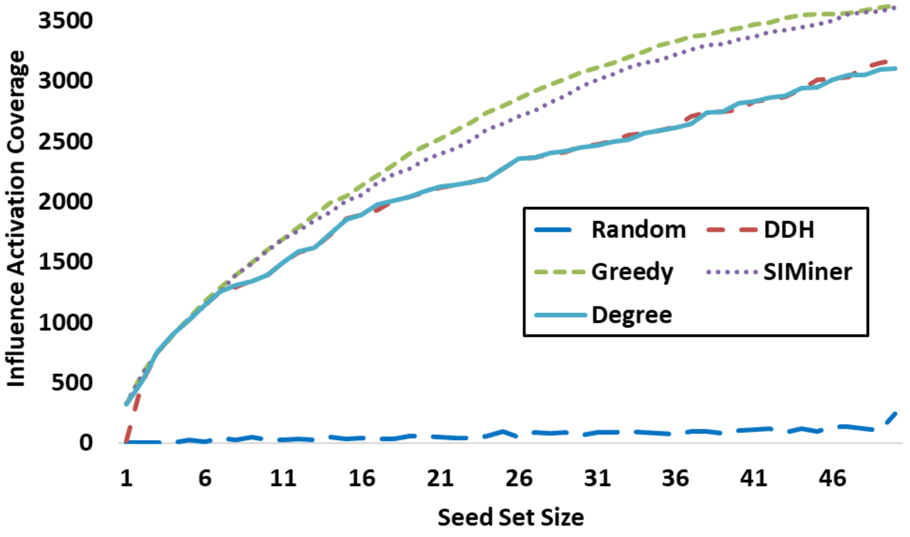


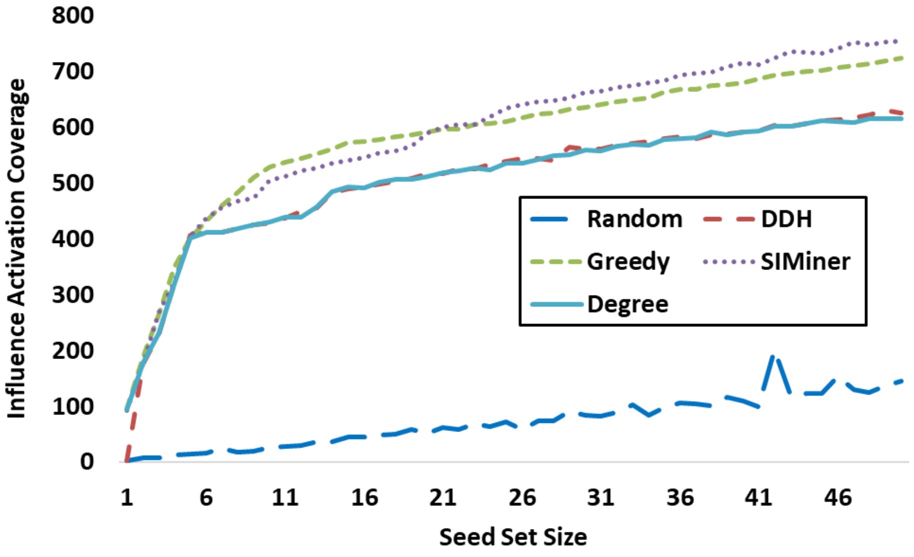
10 ants

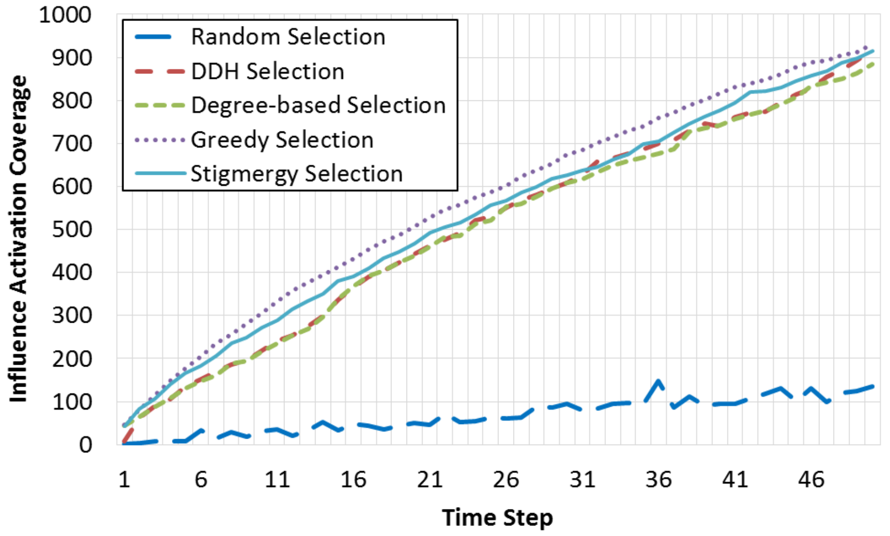
Experiment 2 – Time to Converge

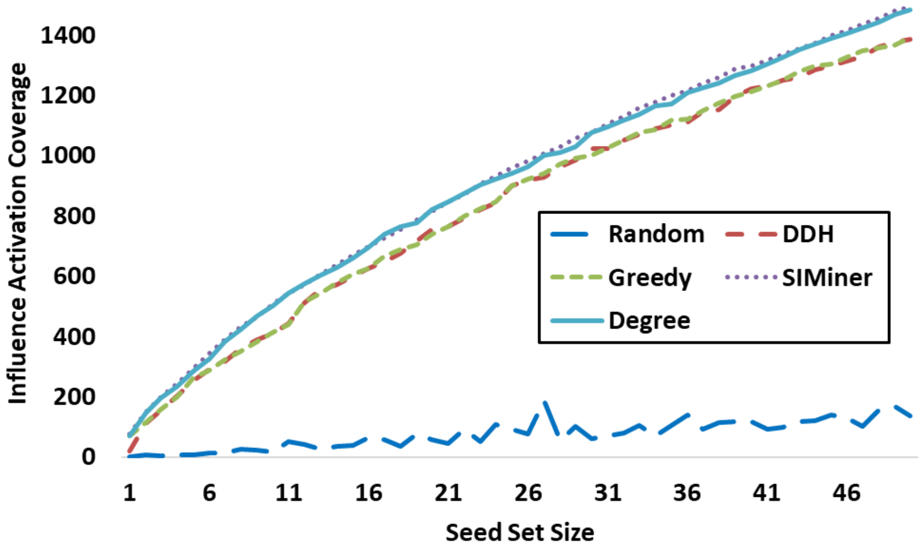


Experiment 3 – SIMiner in Static Networks

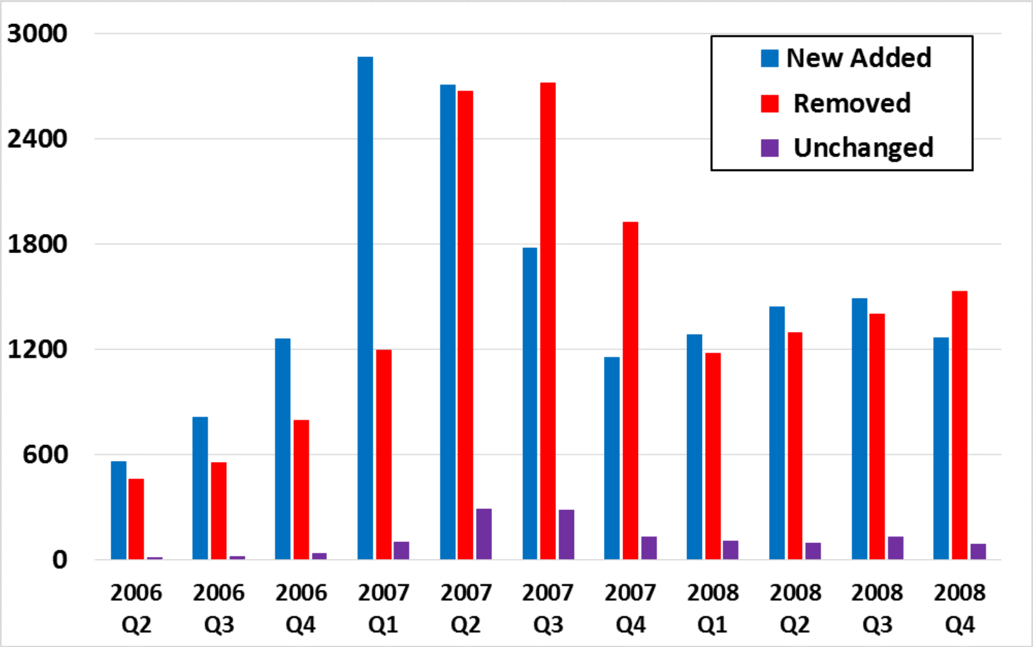


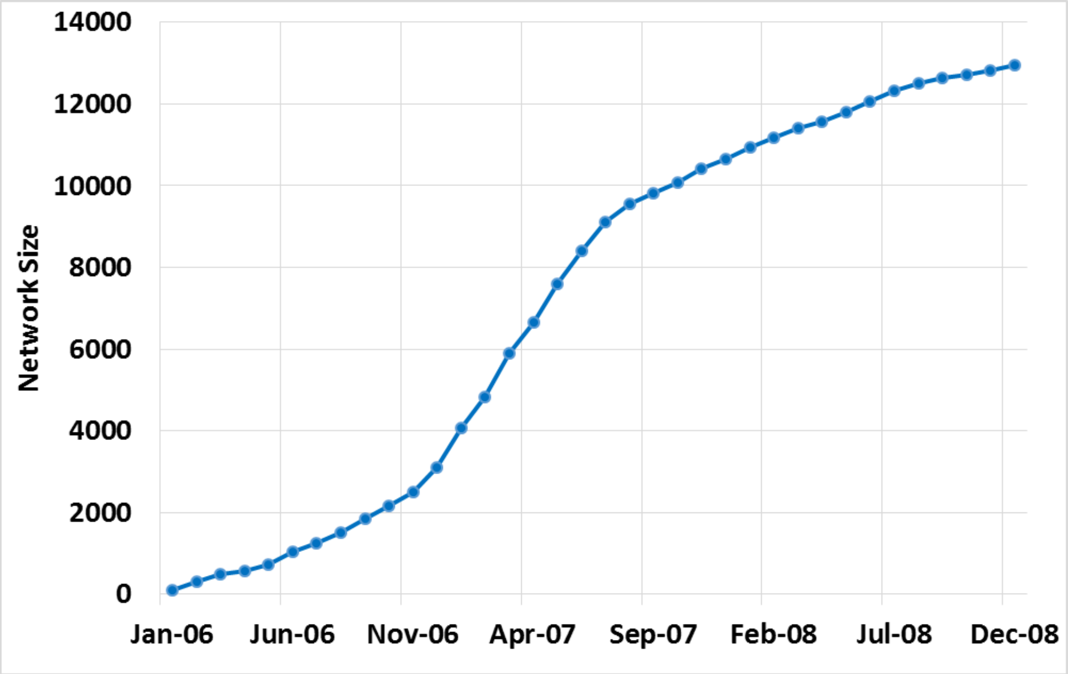




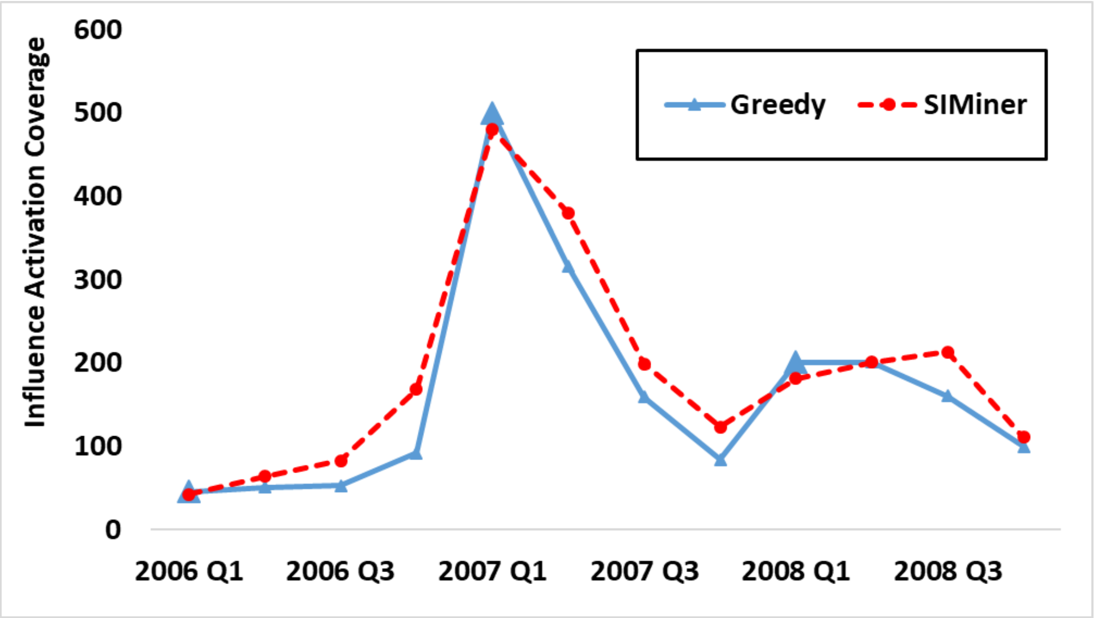


Experiment 4 – Datasets

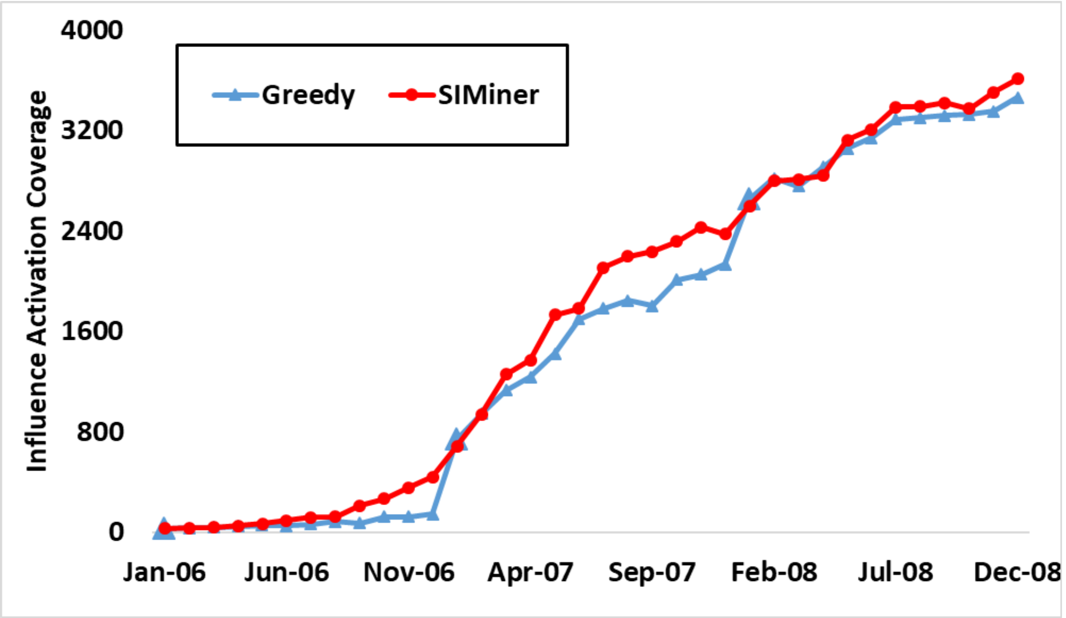




Experiment 4 – SIMiner in Dynamic Networks



Quarterly changes



Monthly changes

Conclusion

- Continuously adapt to the dynamic environment (social networks)
- Multiple ants work together
- Distribute the computational cost and handle dynamics
- Identification of influencers based on the pheromone amount
- Better performance
- More efficient – handle Big Data

Reference

- Li, W., Bai, Q., & Zhang, M. (2019). SIMiner: a stigmergy-based model for mining influential nodes in dynamic social networks. *IEEE Transactions on Big Data*, *5*(2), 223-237.
- Li, W., Bai, Q., & Zhang, M. (2019). A multi-agent system for modelling preference-based complex influence diffusion in social networks. *The Computer Journal*, *62*(3), 430-447.
- Li, W. (2018). *Comprehensive modelling of influence diffusion in complex social networks, an agent-based perspective*. (Auckland University of Technology, New Zealand). Retrieved from http://156.62.60.45/handle/10292/11904