Diffusion Maximization in Evolving Social Networks

Nathalie Gayraud, Evaggelia Pitoura, Panayiotis Tsaparas

Department of Computer Science and Engineering

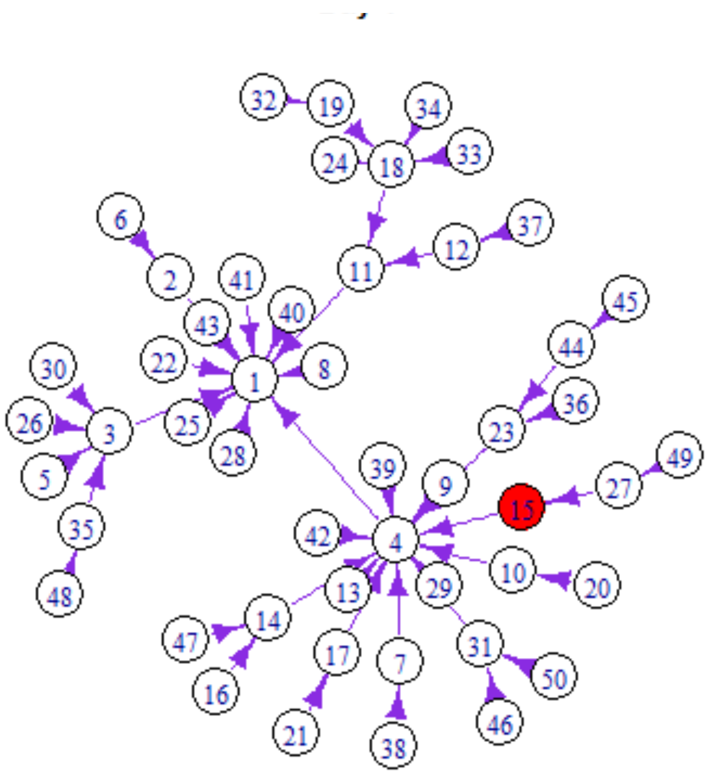
University of Ioannina

Greece

PEOPLE

Diffusion in Networks

- The process by which an item propagates through the edges of a network.
  - Rumors
  - Brand awareness
  - Product adoption
  - Hashtags
  - etc…
- Nodes are active or inactive
  - The number of active nodes is the spread of the diffusion.

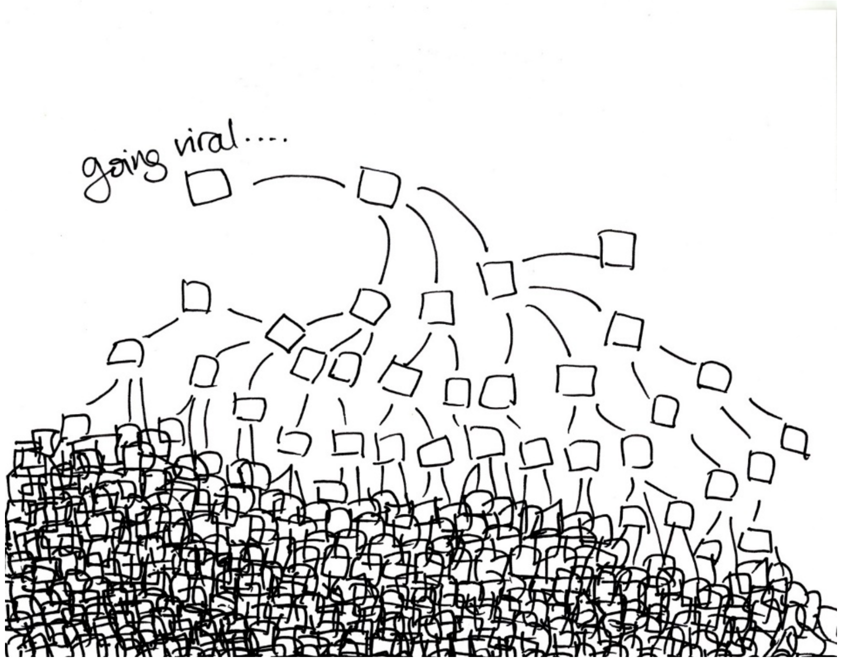


Diffusion Maximization Problem

[Kempe, Kleinberg, Tardos – KDD 2003]

Given a graph G, a diffusion model D, and a budget k, select a subset of k initiator nodes $I = (v_1, \dots, v_K)$ to activate such that the spread of the diffusion $\sigma_D(I)$ is maximized.

- Applications in viral marketing



Diffusion Models

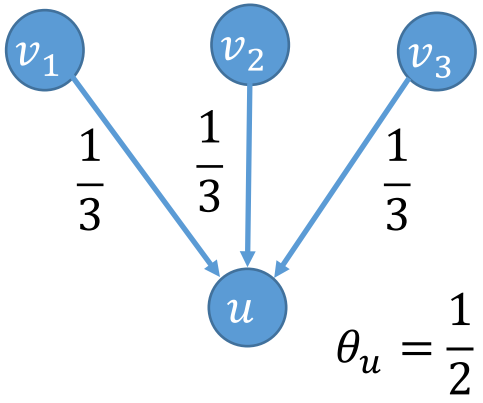
- Independent Cascade model
  - Diffusion proceeds in steps.
  - Nodes activated at step $t-1$ are infectious at step t.
  - An infectious node u at time t will try to activate an inactive neighbor v with probability p_{uv}.
  - After time-step t node u is no longer infectious.

Diffusion Models

- Linear Threshold model
  - Each node u has a tolerance threshold θ_u uniformly distributed in [0,1]
  - The edges incoming to node u have weights b_{vu} that sum to at most 1

$$\sum_{v \in N(u)} b_{uv} \leq 1$$

  - Diffusion proceeds in steps
  - At each step we sum the incoming weight to node u from active nodes and if the weight exceeds the threshold the node gets activated.



Diffusion Maximization Problem

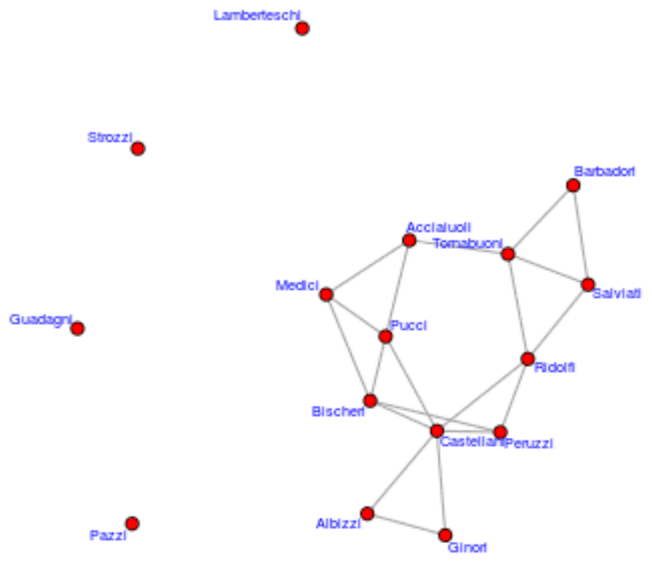
- Diffusion maximization is NP-hard

- The spread function $\sigma_D(I)$ is monotone and submodular for both the Independent Cascade and the Linear Threshold models.
  - Monotone: $\sigma(A) \leq \sigma(B)$ for all $A \subseteq B$
  - Submodular: $\sigma(A \cup \{x\}) - \sigma(A) \geq \sigma(B \cup \{x\}) - \sigma(B)$ for all $A \subseteq B$

- The Greedy algorithm gives a $\left(1 - \frac{1}{e}\right)$-approximate solution.

Static vs Dynamically Evolving Graphs

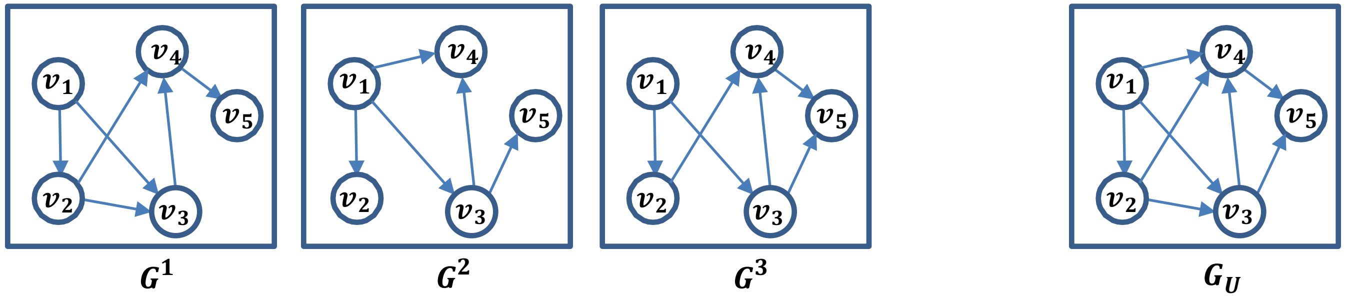
- Most prior work assumes that the graph G is static
  - Connections remain unchanged as the diffusion process unfolds.
- This assumption is not realistic
  - Many graphs evolve dynamically over time
    - Collaboration networks
    - Social networks
    - Mobility networks
    - Biological networks



- How does graph evolution affect diffusion and diffusion maximization?

Evolving (Dynamic) graphs

- An evolving graph is a sequence $\mathcal{G} = (G^1, \ldots, G^n)$ of graphs defined over the same set of nodes V, i.e., $G^t = (V, E^t)$.
  - Graph G^t defines a step in the graph evolution.



- Union graph: The graph $G_U = (V, E_U)$, where $E_U = \bigcup_{t=1}^{n} E^t$

- Most prior work uses the union graph as a simplified model for the evolving graph

Time

- How do we synchronize diffusion and evolution time tracks?
  - Philosophical and practical question
- We assume that evolution and diffusion time track run in lock step
  - One diffusion step happens in a single graph snapshot.
  - We can still capture faster or slower diffusion rates.

Persistent vs Transient items

- In a static graph the activation of a node has a clear effect on its neighbors
  - When the neighbor set changes this is not clear
  - We need to determine the temporal nature of the diffusion

- Transient Diffusion:
  - An active node affects its neighbors only in a single snapshot

- Persistent Diffusion:
  - An active node has an effect over multiple snapshots.

Evolving diffusion models

- Evolving Independent Cascade model (EIC):
- Transient EIC (tEIC):
  - A node u that becomes active at time t-1 is infectious at time t, but it can only activate its neighbors in the graph snapshot G^t
- Persistent EIC (tEIC):
  - A node u that becomes active at time t tries to activate node v in the first graph snapshot $G^{t'}$ in which they become neighbors, for $t' > t$. Once it meets v it will never try again.

Evolving Diffusion Models

- Evolving Linear Threshold model (ELT)
  - We assume again that each node has a threshold chosen uniformly at random
  - The weights on the edges are defined on the union graph G_U as before.
- Transient ELT (tELT)
  - Node u becomes active at time t, if the incoming weight from active nodes in the graph G^t exceeds the threshold θ_u.
- Persistent ELT (pELT)
  - Node u becomes active at time t, if the accumulated incoming weight from active nodes up to time t exceeds the threshold θ_u.

Diffusion Maximization in Evolving Graphs

- When the graph changes over time the question is not only which nodes to activate, but also, when to activate them

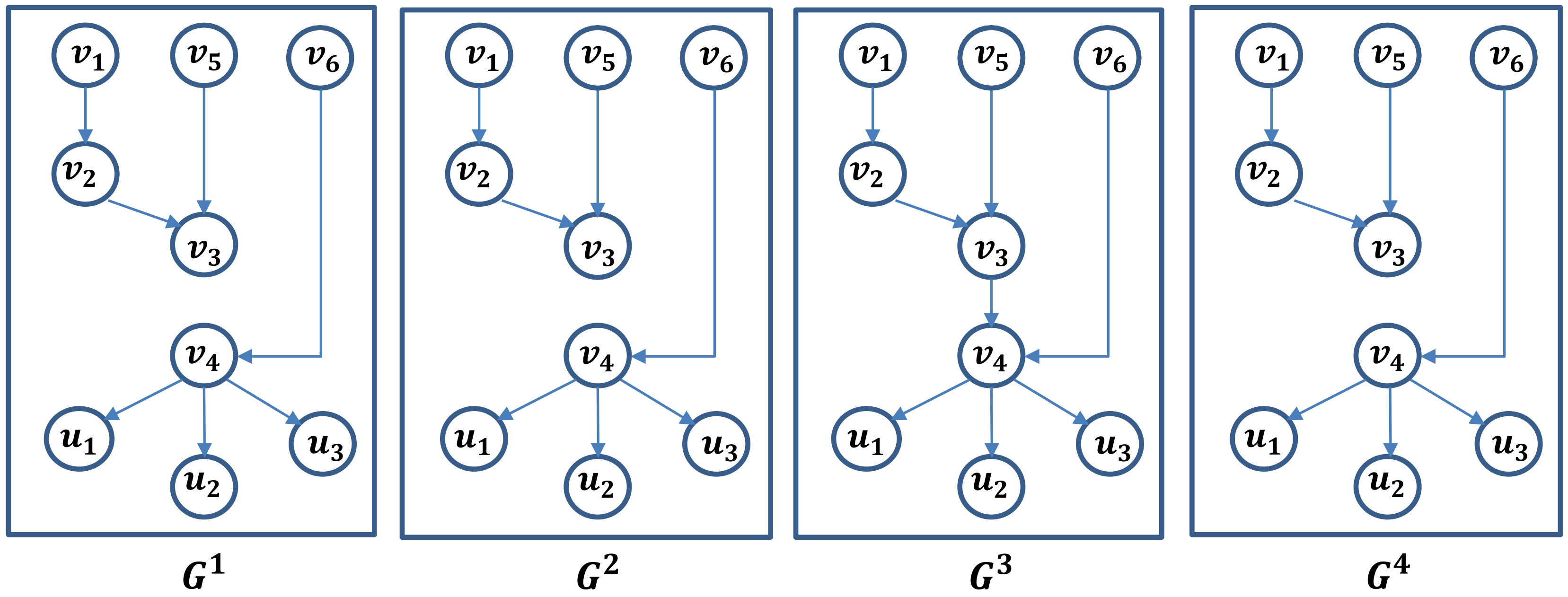
- Given a graph sequence $\mathcal{G}$, a diffusion model D, and a budget k, select a subset of k initiator nodes $I = (v_1^{t_1}, \dots, v_k^{t_k})$ and the time of their activation so as to maximize the spread $\sigma_D(I)$.

- Timing sensitive
  - For some models the best spread is achieved by activating nodes at time t=0. In this case, we say that the model is timing-insensitive. Otherwise, we say that the model is timing-sensitive

Model properties

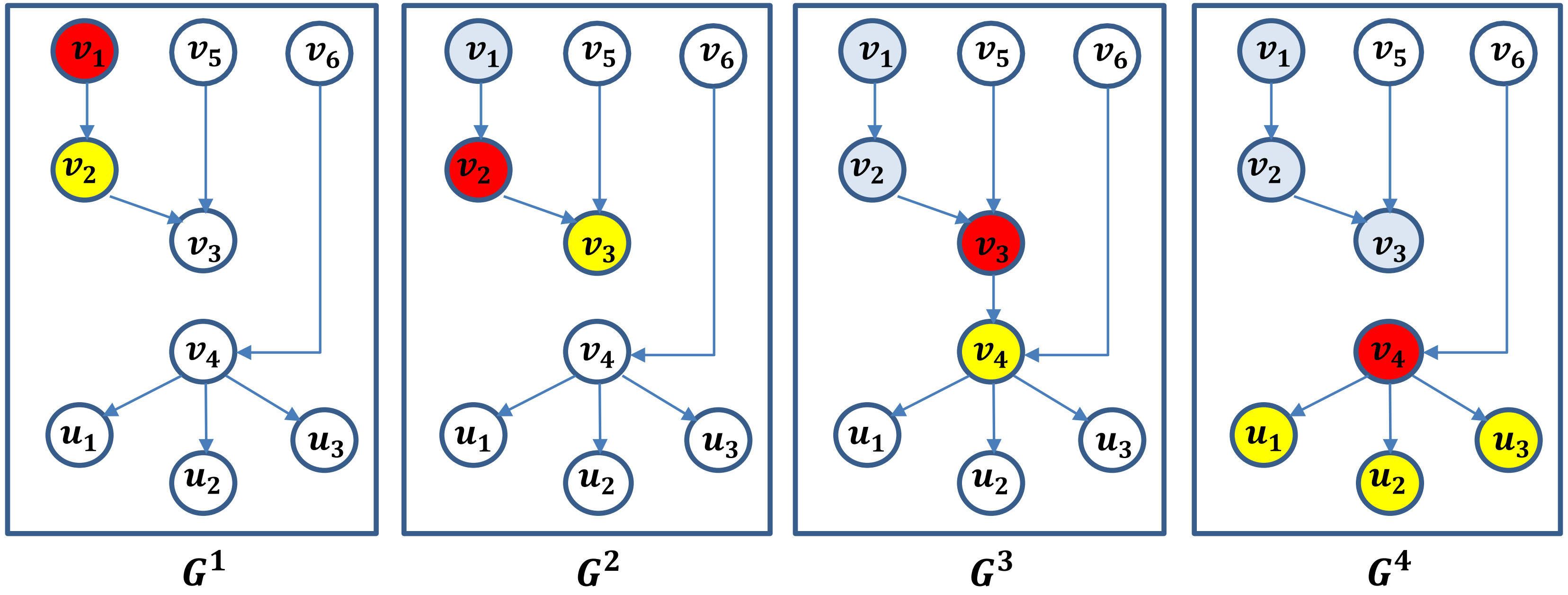
Evolving Independent Cascade	Transient EIC	Not Monotone Not Submodular Timing Sensitive
	Persistent EIC with variable probabilities	Not Monotone Not Submodular Timing Sensitive
	Persistent EIC with equal probabilities	Monotone Submodular Timing Insensitive
Evolving Linear Threshold	Transient ELT	Monotone Not Submodular Timing Insensitive
	Persistent ELT	Monotone Submodular Timing Insensitive.

Monotonicity of Transient EIC



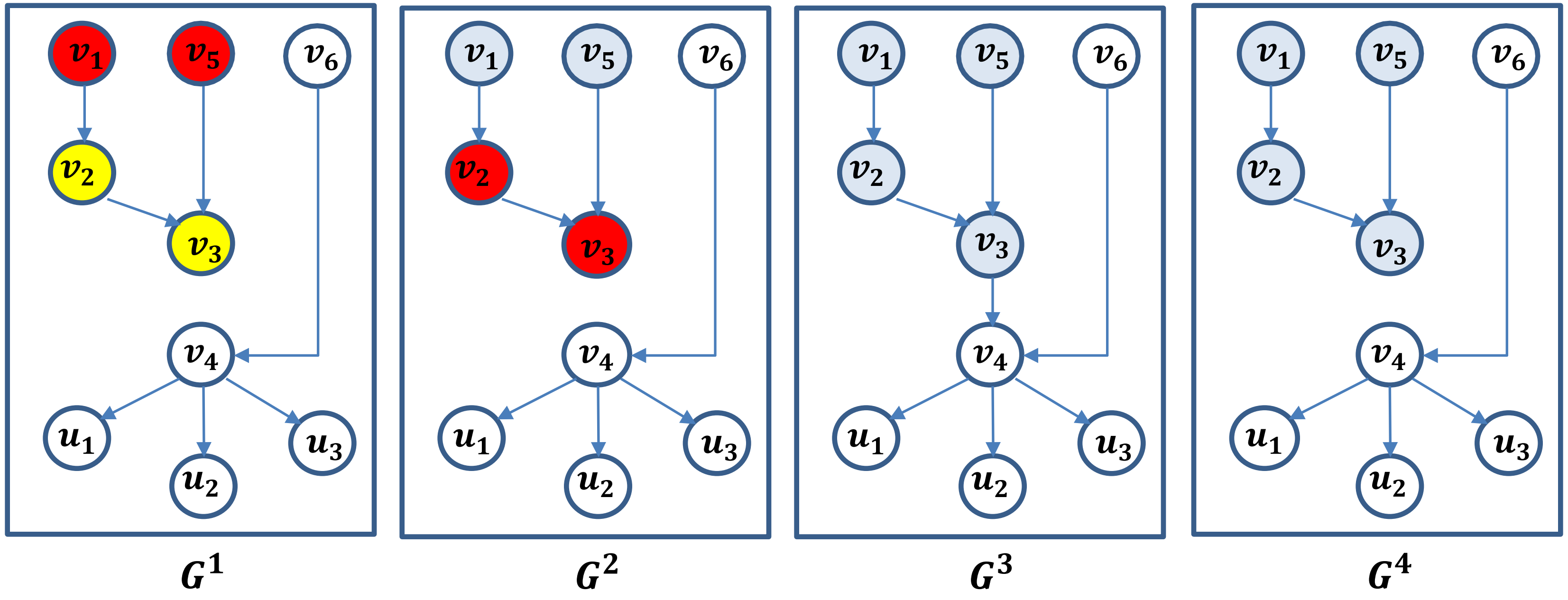
Monotonicity of Transient EIC

- Activating node v_1 at time t = 0 has spread 7



Monotonicity of Transient EIC

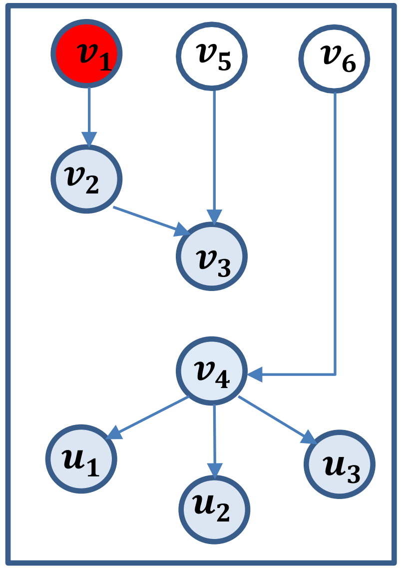
- Activating node v_1 and v_5 at time t = 0 has spread 4



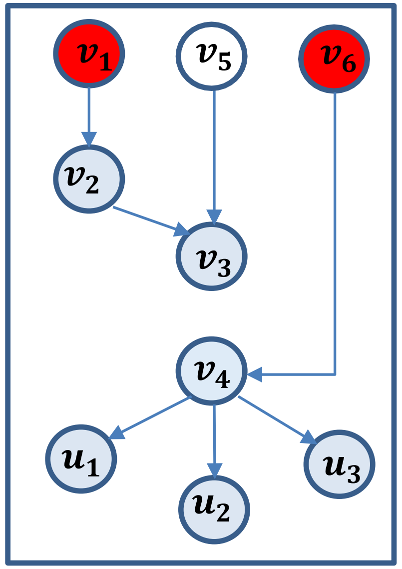
- The activation of v_5 causes node v_3 to be activated too early and block the spread originated from v_1.
    - Node v_3 must be activated at exactly time t= 2.

Submodularity of Transient EIC

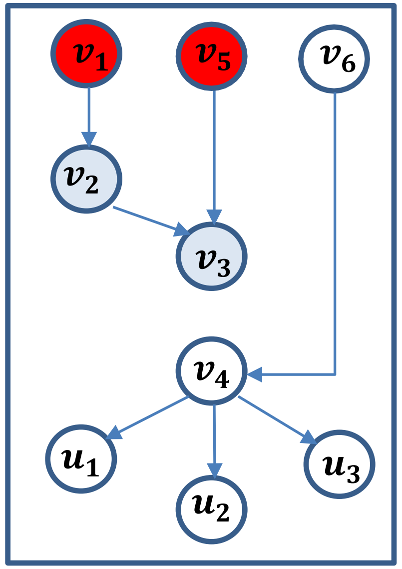
- Adding v_6 to the initiator set $I = \{v_1\}$ has no effect
- Adding v_6 to the initiator set $I = \{v_1, v_5\}$ increases the spread by four.



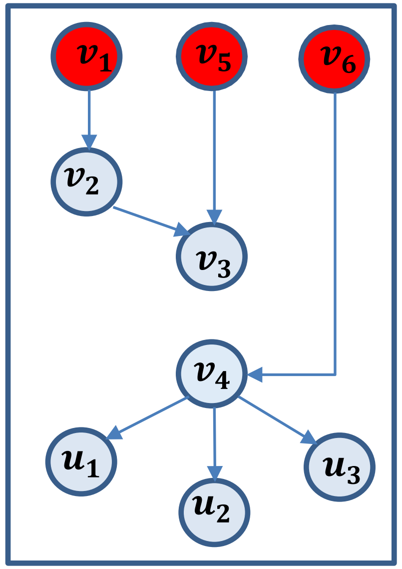
$\boldsymbol{G^4}$



$\boldsymbol{G^4}$



$\boldsymbol{G^4}$

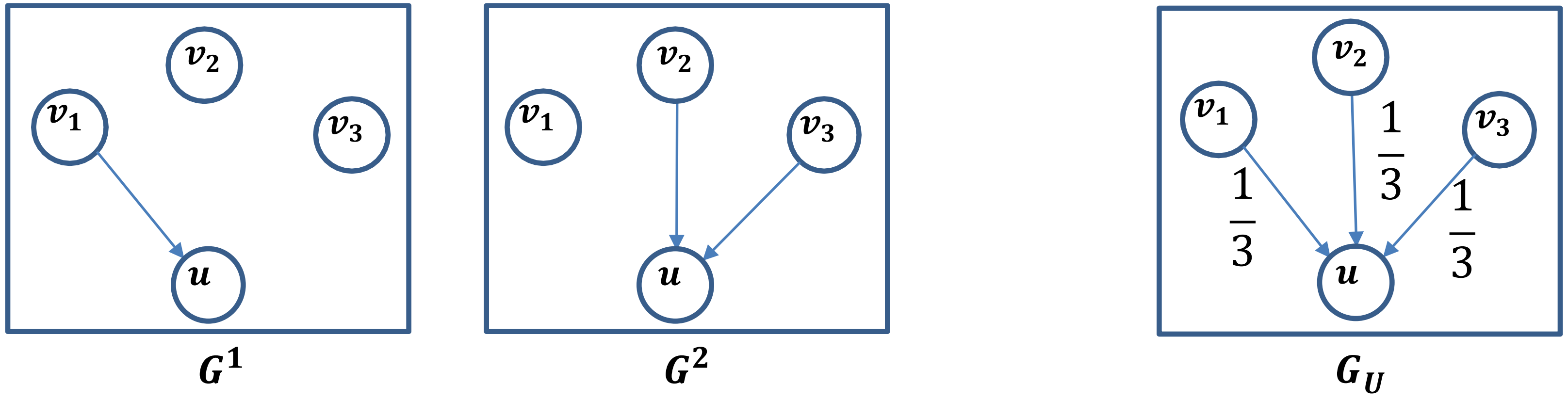


$\boldsymbol{G^4}$

$$\sigma(\{v_1, v_6\}) - \sigma(\{v_1\}) < \sigma(\{v_1, v_5, v_6\}) - \sigma(\{v_1, v_5\})$$

Submodularity of Transient ELT

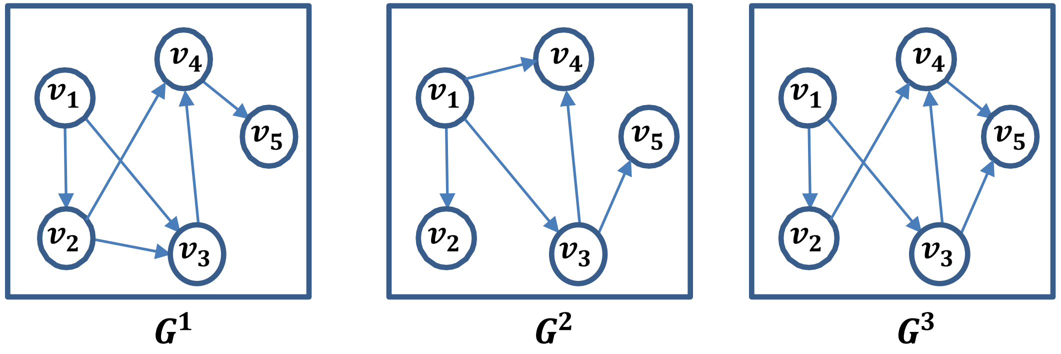
- Transient ELT is montone but not submodular

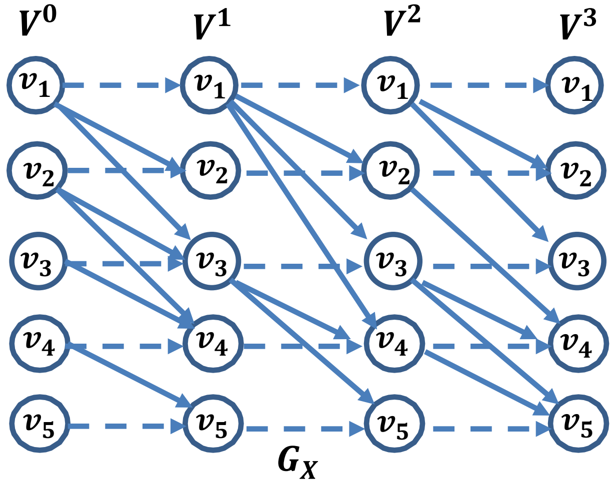


- Expected Spread: the probability that u gets infected
  - Adding node v_3 to the initiator set $\{v_1\}$ has no effect to the expected spread.
  - Adding node v_3 to the initiator set $\{v_1, v_2\}$ increases the spread from $\frac{1}{3}$ to $\frac{2}{3}$
  - $\sigma(\{v_1, v_3\}) - \sigma(\{v_1\}) < \sigma(\{v_1, v_2, v_3\}) - \sigma(\{v_1, v_2\})$

The Persistent models

- The Persistent EIC with fixed probabilities, and the Persistent ELT models are monotone and submodular.
- The proof follows by showing equivalence with reachability in the expanded graph G_X
  - The idea is similar to that in KKT03



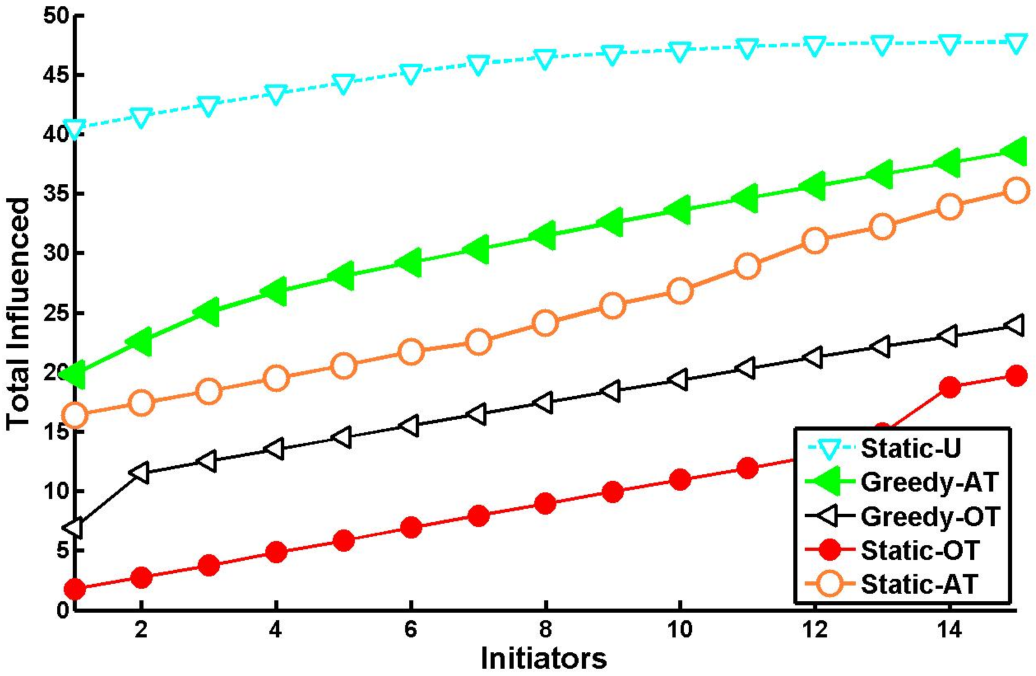


Experiments

- Goal of the experiments
  - How does graph evolution affect the estimation of diffusion spread and the initiator set?
  - How does timing of activation affect the spread of diffusion for timing-sensitive models?

Results on Transient EIC

MIT reality dataset:
Daily contacts of students for the period of one week



Static-U: The spread of the Greedy algorithm on the static union graph G_U

Static-OT: The spread for the nodes selected by Static-U when used on the graph sequence at time t=0.

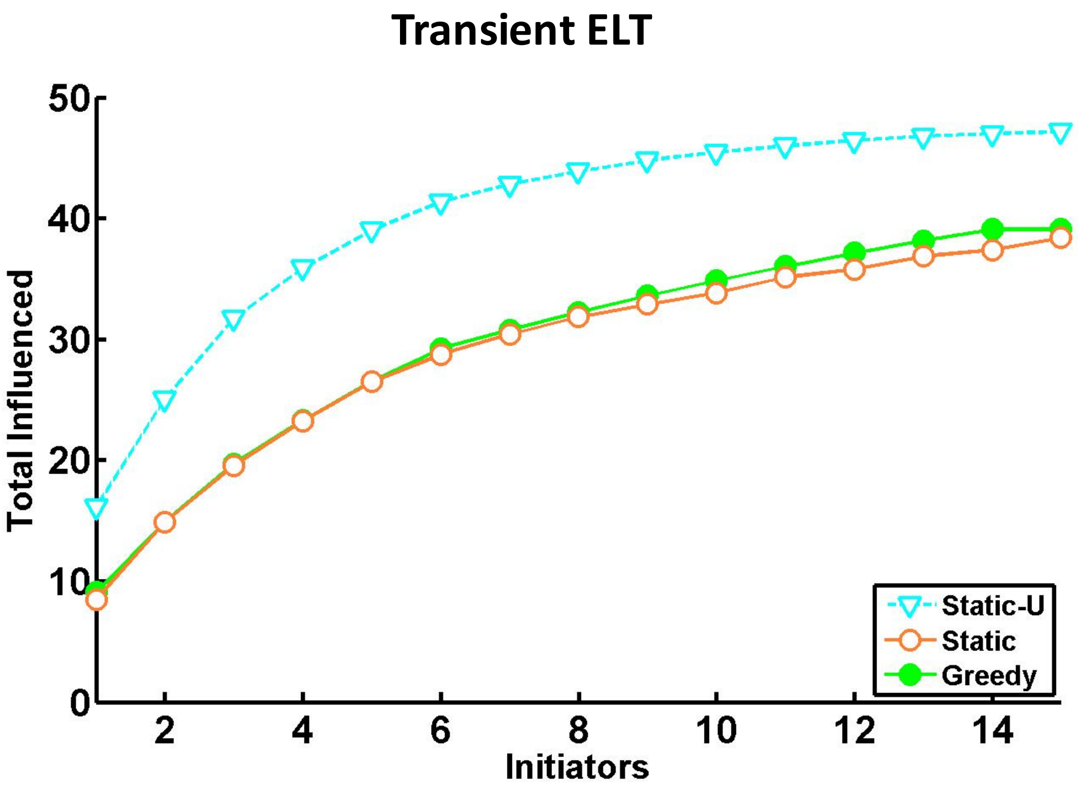
Greedy-AT: The Greedy algorithm that selects nodes at any time instance.

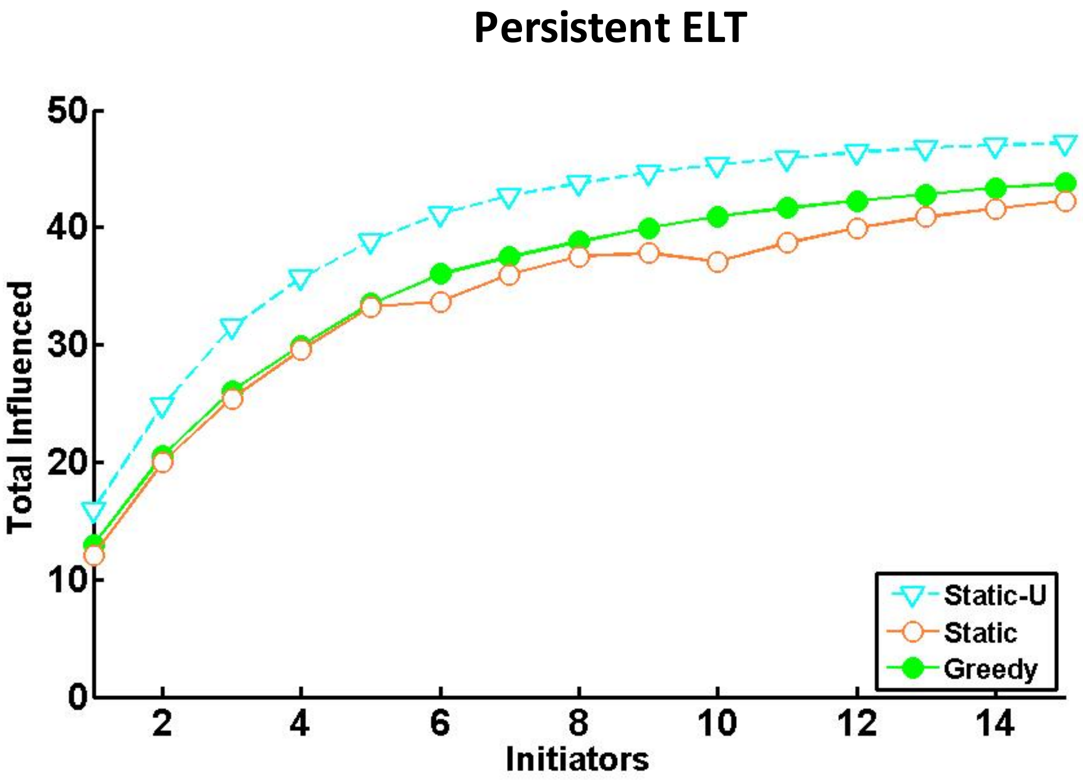
Greedy-OT: The Greedy algorithm that selects nodes always at time t=0.

Static-AT: The spread for the nodes selected by Static-U when used on the graph sequence at the best graph instance.

- Estimating diffusion spread on the static union graph G_U severely overestimates the spread on the evolving graph sequence
- The initiators selected for the static union graph G_U perform poorly on the evolving graph sequence
- Timing of activation makes a big difference for the Transient EIC model

Results on Transient ELT





Conclusion

- We studied the problem of Diffusion Maximization in Dynamically Evolving Networks
  - Surprisingly the spread function is not always monotone and submodular.
  - We demonstrated the importance of timing of activation in practice.

- Future work:
  - Can we apply known techniques to speed up the computation of the spread?
  - Can we prove something more interesting for special cases of dynamic graphs?
  - Study deeper the interplay of diffusion and evolution time