Time constrained Influence Maximization on temporal network

Rohit Kumar, Toon Calders

Department of Computer and Decision Engineering
Université Libre de Bruxelles, Belgium
rohit.kumar@ulb.ac.be

1. Temporal Networks?

An temporal network is defined as a sequence of timestamped interactions ϵ over edges of a dynamic graph G = (V,E).

For example:

- Social interaction in social network.
- Email/ Message or call interaction in communication network.
- Data exchange between computer network.
- People contact network.

2. What we want to study?

The main focus of this study is that given a interaction network and a life span of the information or topic :

- Find out the top k influential users.
- Find out the spread of influence given a starting set of nodes or users.
- If the information or influence has reached a particular set of users or nodes find out the possible initiators.

3. Information Channel and Influence reachability set

Information Channel: If there exist a series of time increasing interactions between node u and v such as

$(u,n_1,t_1), (n_1, n_2, t_1), \ldots (n_k, v, t_k)$ where $t_1 < t_2 < .. t_k$

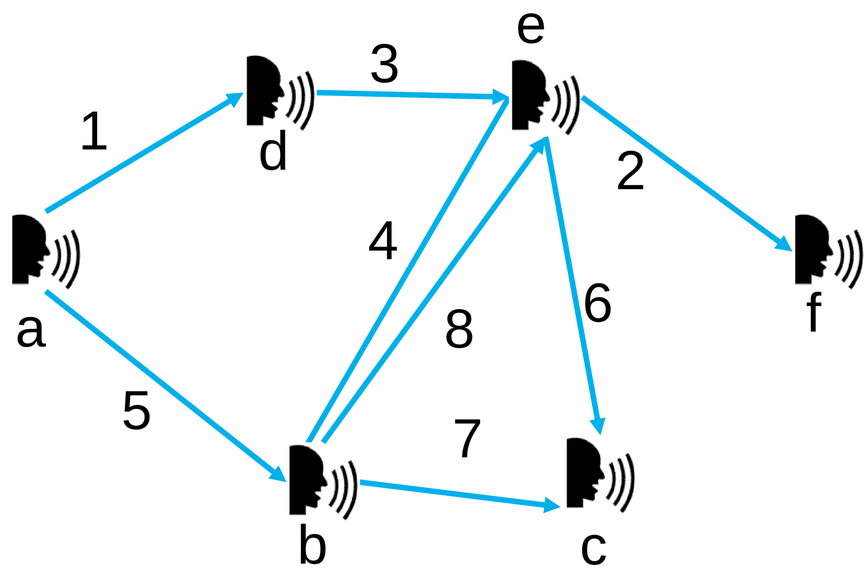
then we say there is an **information channel**, *ic*, between u and v and the length of *ic* = $(t_k - t_1 + 1)$

There could be more then one **information channel** between u and v given as $IC(u, v)$.

Influence reachability Set (IRS) $\rho(u)$ of a node u in a network $G(V,E)$ is defined as the set of all the nodes to which u has a channel i.e., $\rho(u) = \{v \in V \mid IC(u, v) \neq \emptyset\}$.

The influence set for a defined time window length l is given as :

$\rho_l(u) = \{v \in V \mid \exists ic \in IC(u, v) : len(ic) \leq l\}$



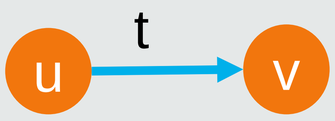
Node (v)	IRS	
	window = 2	window = 3
a	2 (d, b)	**4 (d, b, e, c)**
b	2 (e, c)	2 (e, c)
e	**3 (b, c, f)**	3 (b, c, f)

Influential Node changes with different window length

4. How to Calculate IRS effectively!

$\rho(u) = \rho(u) \cup \rho(v)$

If time window is not considered



$Add\ (S_u, (v, t))$
$For\ All\ (w, t`) \in S_v$
$If\ (t - t`) < l$
$Add\ (S_u, (w, t`))$

If time window is considered

$\rho_l(u) = \{v \mid v \in S_u\}$

- Process edges in decreasing order of time.
- While ***Add*** just make sure only the smallest time stamp is kept.

Time Complexity : O(mn) **!!**

Space Complexity : $O(n^2)$ **!!**

Not very Scalable!!

If we replace the set with a *modified time based* HyperLogLog with *b buckets*

Time Complexity : O(mb log(n))

Space Complexity : O(nb log(n))

Error : ~ 1.04 / √ b

Now this can scale!!

5. Some results on real data

Data Set	#Nodes	# Edges	*processing time(secs)*		
			l =1 %	*l* =5 %	*l* =10 %
DBLP - coauthor	1,314,050	18,986,618	336	1789	2515
arXiv hep-ph	28,093	4,596,803	153	366	421
Enron	87,273	1,148,072	128	186	186
lkml-reply	27,380	1,048,576	168	238	267
Facebook wall posts	46,952	876,993	9	94	160
twitter-higgs	304,691	526,167	2	3	4
Slashdot threads	51,083	140,778	0.3	6	10

6. Application and future work

Application areas :

1) Predicting/maximizing spread of gossip in social network used for campaigns.
2) Predicting spread of contagion through human contact network.
3) Tracing information leakage in private communication network.

Future work :

1) Streaming based solution
2) Making the algorithm parallel and distributed.