A catchy title

Anonymous Author(s)

ABSTRACT

A convincing abstract

ACM Reference Format:

Anonymous Author(s). . A catchy title. In *Proceedings of (The Web Conference'19).* ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

- Motivations
- Background
- Contributions

2 BACKGROUND AND RELATED WORK

3 MODEL SPECIFICATION

Topic-aware influence propagation models share three main hypothesis:

i Users have different interests
ii Items have different characteristics
iii Similar items are likely to interest same users

These assumptions entirely describe a propagation cascade through a social influence which is strictly dependent on the topics and interests. To account preferences we enrich the notion of node defining it with the following triplet:

Definition 1. *(Node)* $< i,\ \mathbf{f}_i,\ \mathbf{t}_i >$

where i is the identification index while $\mathbf{f}_i$ and $\mathbf{t}_i$ are respectively the influence vector and interests vector of node i in a K dimensional topics space.

Definition 2. *(Influence vector)* $\mathbf{f}_i = [f_i^1, f_i^2, .., f_i^K]$

Definition 3. *(Interests vector)* $\mathbf{t}_i = [t_i^1, t_i^2, .., t_i^K]$

Each component represents respectively the degree of influence that the node can exerts on a particular topic and the degree of interest of the node in that topic; for these reasons these entries are non-negative defined.

The graph is defined by the set of nodes V and the set of edges $E \subseteq V \times V$: $G = (V, E)$. The topic-aware perspective implies the generalization of its adjacency matrix entry:

Definition 4. *(Link weight vector)* $\boldsymbol{p}_{uv} = [p_{vu}^1, p_{vu}^2, .., p_{vu}^K]$

where vu is the direct link from u to v, $z = 1, .., K$ is the topic, and $p_{vu}^z = t_u^z + f_v^z \in [0, +\infty)$. In this way the directed link from node v to u on a particular topic z has an importance proportional to the interest of node u on that topic, incremented by the amount of influence that node v can exert on u.

Moreover the necessity of initial conditions in the diffusion process is solved by introducing a special node in the network, called *god node*; it is connected with all other nodes and denoted with a negative index *-1*.

Definition 5. *(God node)* $v_{-1} \mid E^+(v_{-1}) \equiv V \wedge E^-(v_{-1}) \equiv \emptyset$

where: $E^+(u) := \{v \in V | v \neq u \wedge e_{uv} \in E\}$ is the set of nodes followed by u, and symmetrically: $E^-(u) := \{v \in V | v \neq u \wedge e_{vu} \in E\}$.

The corpus I, i.e. the set of items with cardinality I involved in the cascades, is described by probability distributions and scalars. As a main concept LDA assumes that each topic is described by a superposition of words while a document can be statistical interpreted by a composition of topics. This particular model defines these distributions as Dirichlet distributions in order to incorporate in the model the assumption that few topics can well describe a text. The document (or *item*) is entirely defined by its distribution over the topic space and a scalar value that defines its propensity to propagate, which we denote as *virality* parameter.

Following the standard notation of the generative LDA model we have:

- α := prior topic distribution
- β := prior word distribution
- γ := document's distribution over topics; each entry is the probability of topic z occurring in document i: $p(k|i)$
- φ := topic's distribution over vocabulary words; each entry is the probability of word w occurring in topic z: $p(w|z)$

As a final step towards cascades modeling we extended the Topic-aware Linear Threshold Model. The activation process is driven by the exceeding of a deterministic threshold. Firstly, we considered the LDA output as item's topic distribution. Secondly, we defined the social pressure on a particular node proportional to her level of interests and the degree of influence of her neighboors. Finally, we bounded the activation parameter introducing a sigmoid function[1]. The activation parameter for node u at time t over item i is:

$$W_i^t(u) = \Theta\left[\sum_{z=1}^{K}\left(\gamma_i^z \sum_{v \in F_i(u,t)} p_{vu}^z\right)\right] \tag{1}$$

where: γ_i^z is the item's distribution over the topics; $p_{v,u}^z$ is the strength of influence exerted by v on u on topic z; $F_i(u, t)$ is the set of users active on i at time t and that have a link with u. If $W_i^t(u)$ is greater or equal to θ_i threshold, it becomes active item i. A couple of considerations: [a] γ_i^z is the z-th coefficient of item-i's linear combination of topics; [b] p_{vu}^z corresponds to the link weight of the directed graph from u to v for topic z (*def 4*), remember weights are K-dim vectors (one weight for each topic). Moreover we introduced a set of activation thresholds that is equal for each node: one item has the same threshold for each individual. This quantity is linked to what we call *virality* parameter; in particular the item's threshold is the inverse of its virality. In this way we are introducing the independence of the diffusion strength of a document, which represents the real structural characteristics of the texts.

The following box summarizes the presented variables:

[1] $\Theta(x) = 1/(1 + e^{-x})$

The Web Conference'19, May 13 - 17, San Francisco, CA, USA
.

$\mathbf{f}_u$, $\mathbf{t}_u$:influence and interests vectors of node u
α, β: prior topic, word distribution
φ: topic's distribution over words
γ: item's distribution over topics
p_v: strength of influence exerted by v on an other node;
W^t: node's influence weight at time t for an item
θ_i: item i threshold (inverse of its virality)
$W_i^t(u)$: activation parameter for node u at time t on item i

- WoMG from thesis

4 GENERATING INTERESTS

Intro: what is the challenge here? What are we going to propose to solve it?

4.1 A propagation model

4.2 Node2vec

Simple changes:

- Prior
- Positivity

4.3 Matrix Factorization

5 IMPLEMENTATION

6 EXPERIMENTAL ASSESSMENT

RQ1: Which is the best method to represent nodes in our model? RQ2: Is this model tunable for different assumptions? RQ3: Is this model realistic?

6.1 Interest generation comparison

What is *best*?

- Scalability
- Tunable homophily

6.2 Parameters analysis

Here we fix one or two interests generation methods and we explore the range of properties for the synthetic propagation the model can generate (RQ2).

6.3 Real data

Experiments whwere we compare with real data set (RQ3).

7 CONCLUSIONS AND FUTURE WORK

REFERENCES