Discussion on “Influence/ Influencer”

Presented by Junfeng He
Outline

One paper as an example: “Everyone’s an influencer: Quantifying influence on Twitter”

- What: Problem definition
- Why: Motivation
- How: Techniques and results

General discussion
Outline

One paper as an example: “Everyone’s an influencer: Quantifying influence on Twitter”
  – What: Problem definition
  – Why: Motivation
  – How: Techniques and results

General discussion
What kind of social network?

What social network?
- Twitter

What kinds of propagation?
- Copying post (Repost) on Twitter
- More specifically, Repost the same url
What is the definition of influence?

- If we can define “A influenced B”, then
- Influence of a seed post: given a seed node with a url, how many nodes are influenced in the follower graph
- Influence of an individual: the average Influence of seed post by him/her
- However, what is the definition of “A influenced B”? 
What is the definition of influence?

The definition of “A influenced B”
- If B follows A
- A posted the url earlier than B
- A is the only B’s friend to post the url

If B has multiple friends who posted the url
- Influencers to be the first, or the latest, split equally
- Almost the same for the above three in experiments
What is the data set?

- 1.6M users
- 74 million diffusion events during 9/2209/11-2009

Table 1: Statistics of the Twitter follower graph and seed activity

<table>
<thead>
<tr>
<th></th>
<th># Followers</th>
<th># Friends</th>
<th># Seeds Posted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>85.00</td>
<td>82.00</td>
<td>11.00</td>
</tr>
<tr>
<td>Mean</td>
<td>557.10</td>
<td>294.10</td>
<td>46.33</td>
</tr>
<tr>
<td>Max.</td>
<td>3,984,000.00</td>
<td>759,700.00</td>
<td>54,890</td>
</tr>
</tbody>
</table>
What is the task?

- Predict the one’s influence
- How to use influence for marketing
Outline

One paper as an example: “Everyone’s an influencer: Quantifying influence on Twitter”
- What: Problem definition
- Why: Motivation
- How: Techniques and results

General discussion
Why Twitter?

- The network over word-of-mouth is usually difficult to observe/record
- Observations is usually biased, i.e., only the successful diffusions will be recorded
- Twitter provides a natural laboratory for study of diffusion, overcoming the above two problems
- Since twitter post short, easy to collect and study large scale network
Why choose the propagation and the definition of influence?

- It may not exactly the same as the real-world influence, such as those on purchasing behavior, political viewpoints
- But
  - No need for labeling
  - No ambiguity
  - Not difficult to compute
Why this work is unique from the previous ones

Definition of influence

- Previous work:
  - Pagerank
  - The number of (in-coming edges) followers

- The proposed work
  - Defined on the diffusion trees/cascades

Task

- Previous work:
  - find the most influential one
  - Or classify one is influential or not

- The proposed work
  - Predicting influence for all users
Outline

One paper as an example: “Everyone’s an influencer: Quantifying influence on Twitter”
- What: Problem definition
- Why: Motivation
- How: Techniques and results

General discussion
How to predict influence

Features

1. Seed user attributes
   (a) # followers
   (b) # friends
   (c) # tweets,
   (d) date of joining

2. Past influence of seed users
   (a) average, minimum, and maximum total influence
   (b) average, minimum, and maximum local influence,
How to predict influence

Model: regression tree

Figure 5: Regression tree fit for one of the five cross-validation folds. Leaf nodes give the predicted influence for the corresponding partition, where the left (right) child is followed if the node condition is satisfied (violated).
Results

Only two features are selected in the tree

Figure 6: Influence as a function of past local influence and number of followers for (a) all users and (b) users with the top 25 actual influence. Each circle represents a single seed user, where the size of the circle represents that user’s actual average influence.
Results

Looks very good, but…
Results

On average within each leaf node, quite good: \( R^2 = 0.98 \)

Without average in each leaf node, quite good: \( R^2 = 0.31 \)

In other words, Influencers are those
  – Who have been influential in the past
  – Who have many followers

But this is true only on average
Results

- Individual-level predictions of influence is unreliable
- Marketers should not rely on exceptional influencers, but target many potential influencers to rely on average performance
Targeting Strategies

Cost function

\[ e_i = c_a + f_i c_f \]

\[ c_f = \$0.01 \]

\[ c_a = \alpha c_f \]
Results
Critical thinking

What else, why not, how about …
Outline

One paper as an example: “Everyone’s an influencer: Quantifying influence on Twitter”
- What: Problem definition
- Why: Motivation
- How: Techniques and results

General discussion
What

What social network, propagation?

- Blogger network + blog link
- Game role (in Second life) + content diffusion (e.g., dancing gestures, clothes)
- Face book user+ software installation diffusion
- Any other examples?
  - e.g., flickr, youtube netflix, publication networks
What is the definition of influence?

- page rank or its adaptation
- Number of edges (incoming inlinks)
- Anything else…
What is the task?

- Identifying the most influential nodes
- Categorization/Analysis of the influencers
- ...
Why and how