Temporal Analysis of Individual Influence in Twitter

Motivation

* Understand an individual activity's influence on its neighbors and the social network
  * How many retweets/favorites interaction can I get?
  * How long will my tweet be active in the network?
  * How do i fare in reach compared to other information sources?
  * Will my friends like what I post?
  * What to post to make me popular?
* Analyze the temporal pattern of retweets (favorites)

Hypothesis

* The temporal variation of number of retweets follows certain patterns
* Certain factors have effect on the pattern of retweet
  * Time of day
  * Initial reaction
  * Content
  * Personal weight in the network
  * immediate network…
* Time series of reaction can be predicted

Pattern Classification

* Classify temporal patterns of number of retweets
* How to measure similarity of time series
  * Total number of retweets
  * Shape of the time series
* Clustering algorithms

Pattern Prediction

* Predict what pattern the time series will follow based on the observation in the beginning
  * Post time
  * Number of retweets/favorites
  * Users who retweeted
  * Number of followers of the user who retweet
  * Hashtag ..?
  * ...

Data

* Twitter
* ~20 known active users from different fields
  * active users are the ones who post consistently, have a healthy following and interact actively.
* 2 weeks of monitor. status check every 30 minutes ?
* @ 12 tweets per profile. ~ 240 tweets per day.
* Aim to collect 2000 tweet charts.
* Select only tweets with activity.
* Train on subset and test on the rest.

Data

* post time (hour)
* number of rewteets in the first hour
* number of favorites in the first hour
* number of followers
* active time (if the number of retweets remains the same for an hour)
* total number of retweets (the number of retweets during active time)
* …

Related Work

* "Modeling information diffusion in implicit networks". *ICDM' 10*
  * Not require the knowledge of the social network
  * Model the number of newly infected nodes as a function of nodes got infected in the past
* "Patterns of temporal variation in online media". WSDM'11.
  * Define the similarity metric based on scaling and shifting.
  * Develop a clustering algorithm to finds cluster centroids