Monitoring Information Diffusion in Micro-Blogs

Introduction

Introduction

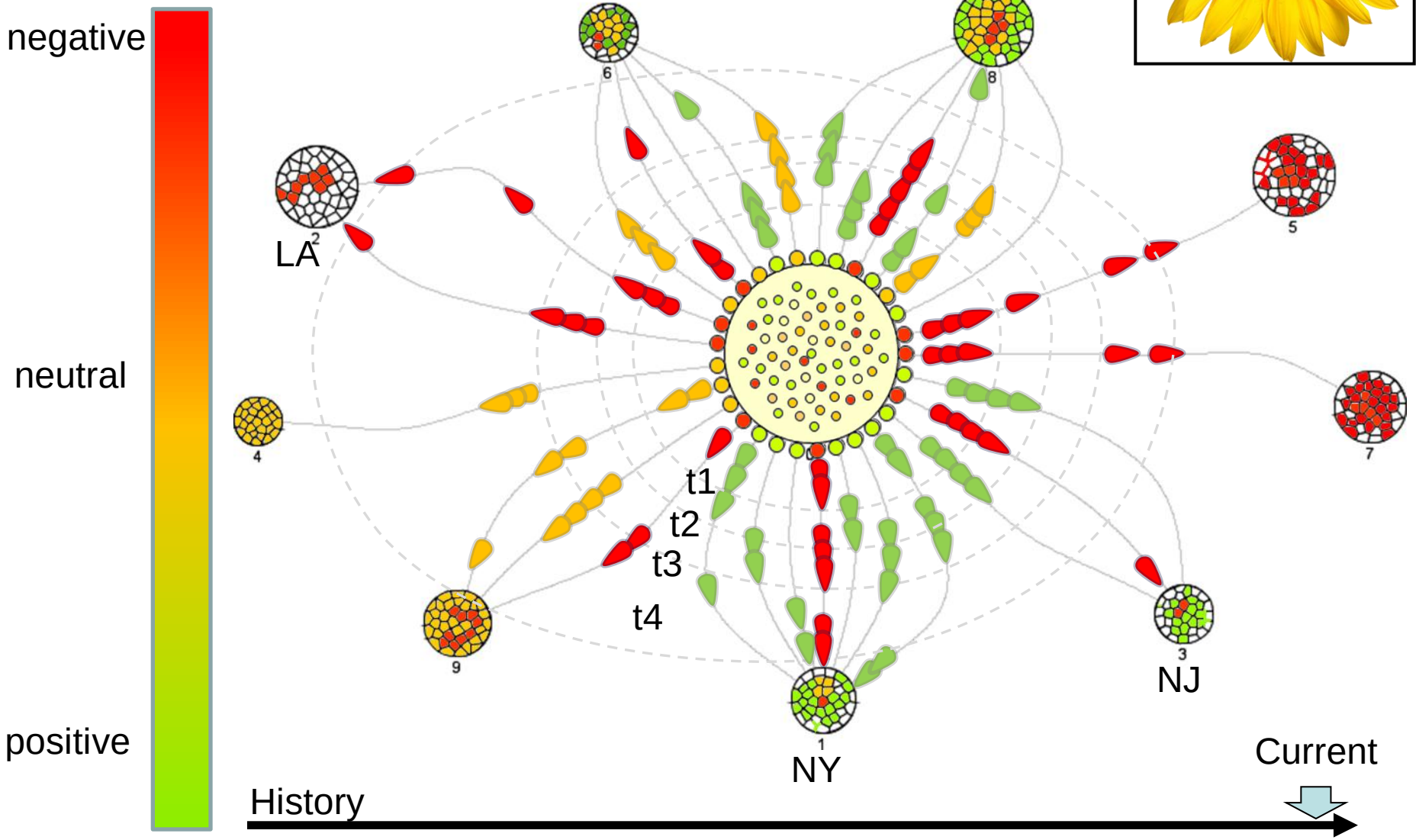
- Background and Motivation
  - Twitter, Google+, Sina Weibo already become part of our life. It provides a platform for breaking out news and other timely information
  - Efficient filtering and monitoring micro-blogs will help on detecting the “public evaluations” about certain people, movies, products, brands, markets, etc. Thus will help on decision making and problem solving
  - Filtering and monitoring trending topics in micro blogs is always a challenge problem since the data is streaming in nature, huge in amount and dynamic in real time

Introduction

- The specific problem you want to address
  - Dynamic monitoring of the information diffusion as well as the opinion diffusion
  - Efficient sampling of huge amount micro-blog data
  - Visual comparing on diffusion patterns of related topics
    - Ex: “Nick” vs “Adidas”, “Obama” vs “XXX” check the topic hotness and public opinions over them

- Is it a new problem or old problem?
  - Both filtering and monitoring are old problems however the idea of diffusion pattern comparison is novel.
  - Most of the existing methods are designed based on tag clouds or original tweet items which provide limited information and lack of relational context

Our Methods


negative
neutral
positive
6
8
LA
2
5
4
7
t1
t2
t3
t4
9
3
NJ
1
NY
Current
History

Important Queries

- Detailed Visual Patterns:
  - What kind of people likes/dislikes the specified topic?
  - How fast/many the information is diffused among different user groups ?

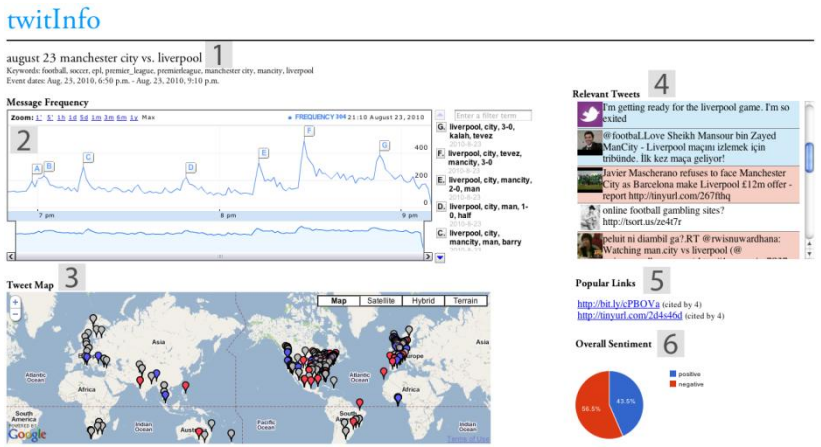
- Comparison
  - What's the diffusion difference of multiple related topics? E.g. "iPhone" vs "Sunsung Galaxy" ?
    - Speed / Amount / Opinion / Interested User Group

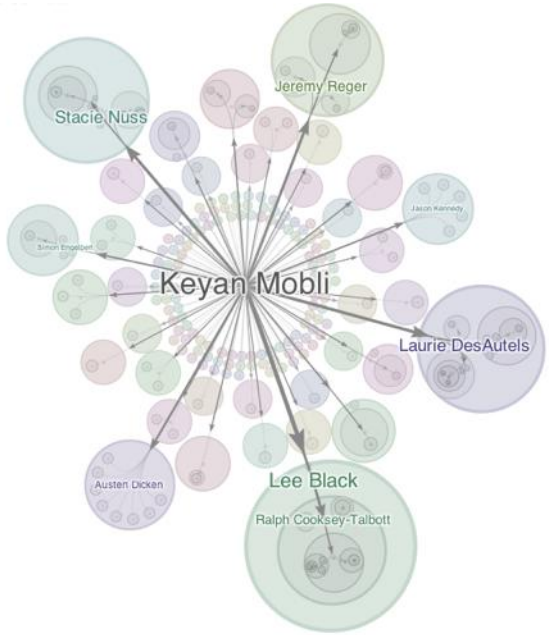
Contributions

- The major contributions of this paper are
  - A configurable bipartite graph design based on magnitude field that minimizes information overlaps as well as encodes both information and opinion diffusion patterns for real time micro-blogs monitoring on users' interested topics
  - A sunflower liked graph layout and topic icon design that automatically distinguishes active data items from data spams that help on huge data filtering and diffusion pattern comparison
  - A curved multi-thread timeline visualization combined with tweet glyphs that indicates the diffusion details like direction, speed, amount as well as opinions that help on pattern detection and comparison

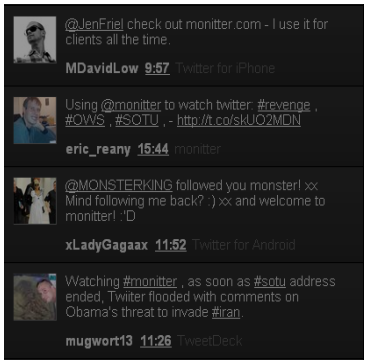
Related Work

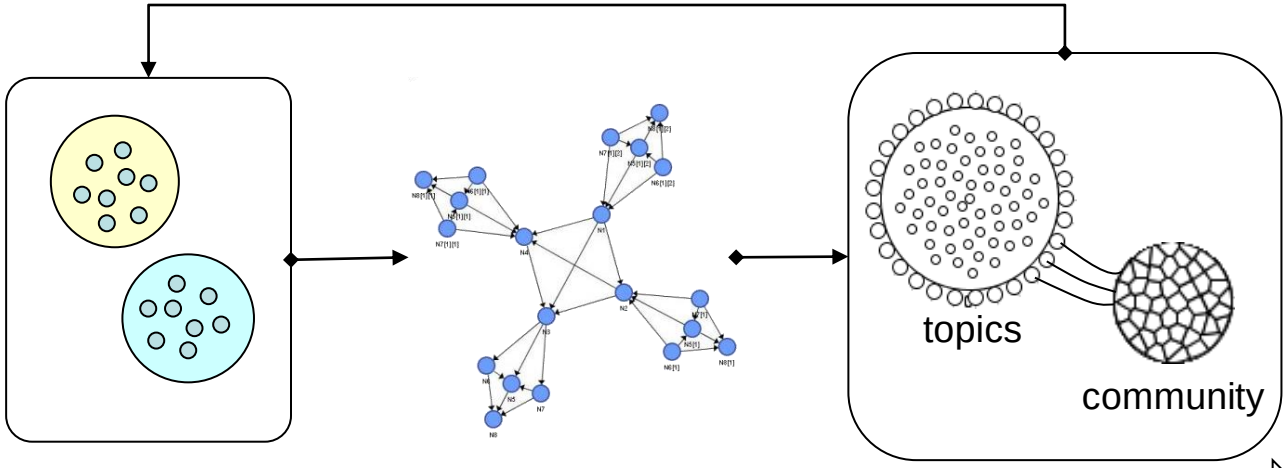
- Related work to your paper
  - TwitInfo: Aggregating and Visualizing Microblogs for Event Exploration, CHI 2011
  - Google+ Ripples
  - Others:
    - http://www.cse.ust.hk/~nancao/download/Twitter-RW.doc





System Pipeline





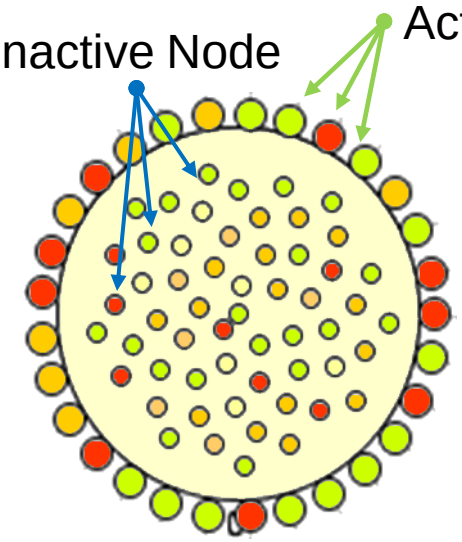
configure to monitor

bi-partite clustering

layout

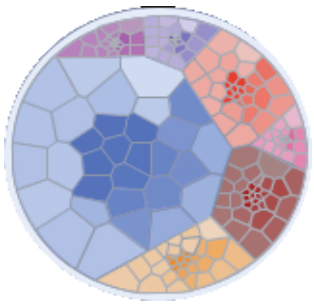
visualization

Design Details: Clusters

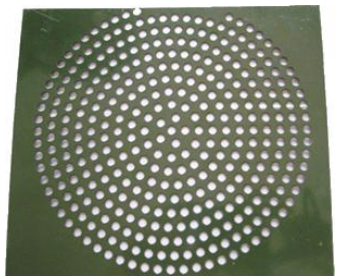


Topic Sunflower

- inactive nodes: when the user posts a new tweet. The tweet is represented as an node in the center
- active nodes: when a tweet is captured by another user, is will move to the boundary and become an active node
- colors: opinions

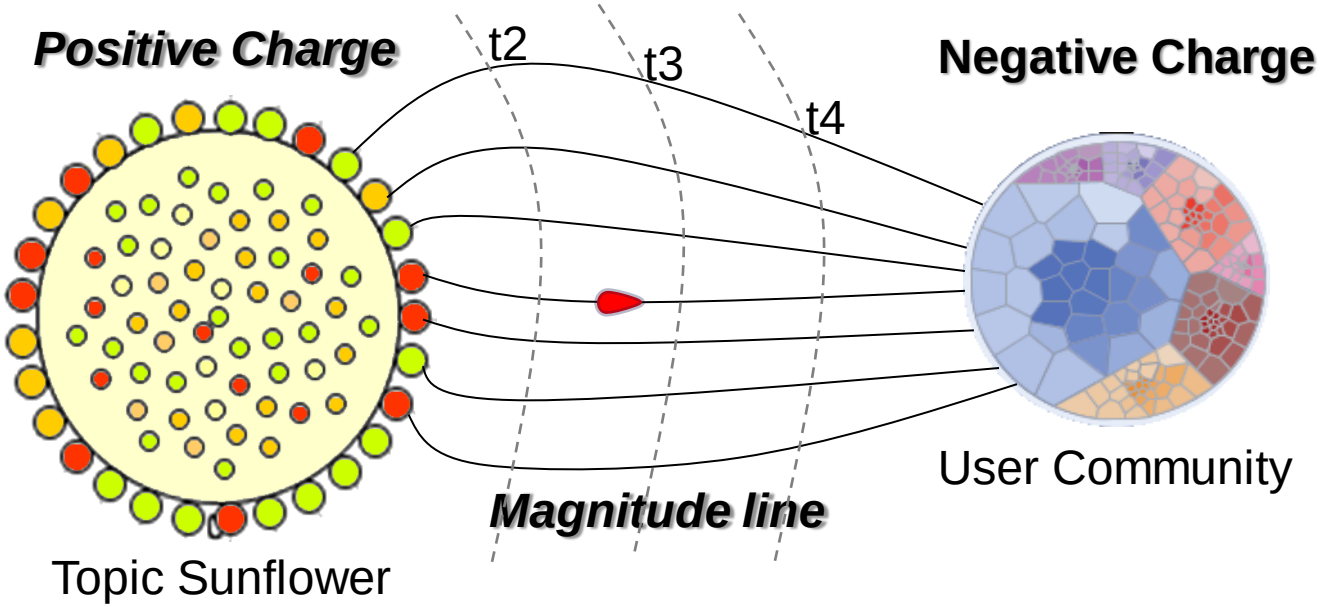


User Community



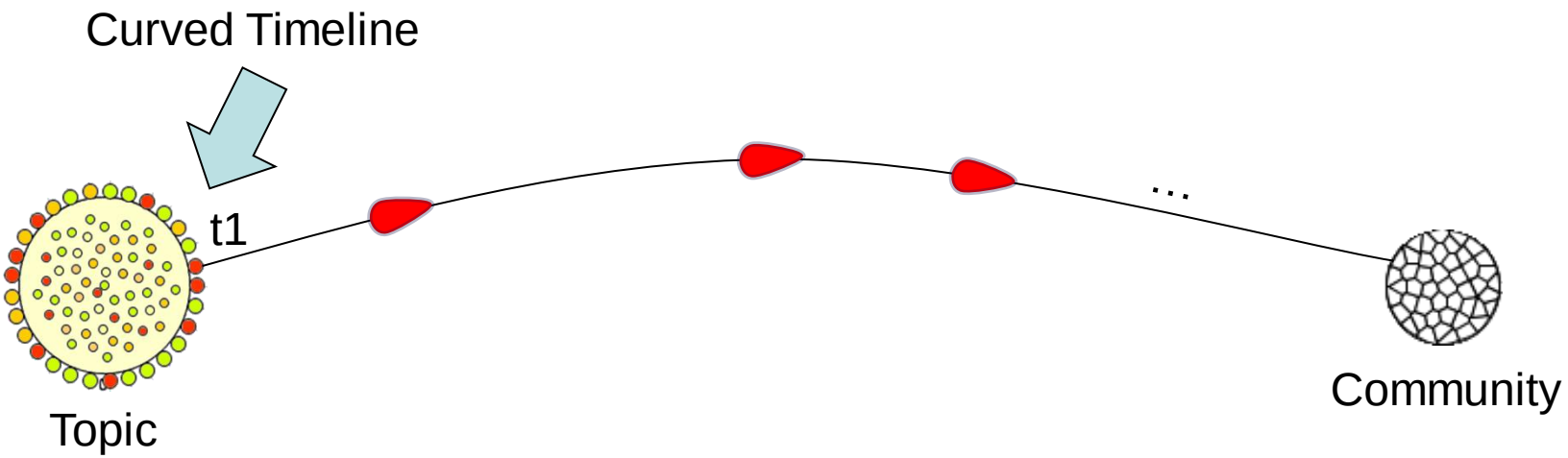
- Use icons to represent user community
- Each cell is an user that reweet a certain tweets in the related topics
- Users are grouped by their locations

Design Details: Magnitude Field



- Magnitude Field
  - Motion intuitiveness : from
  - Avoid overlaps
- Contours are used to mark different time

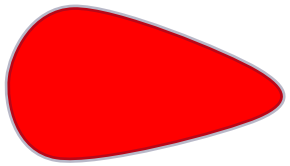
Design Details: Retweet Links

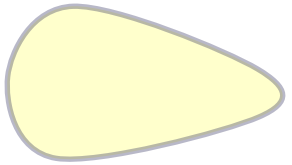


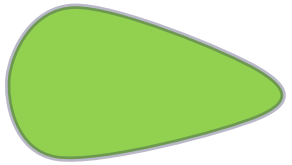
- Each link represents the behavior of reweeting
- It is visualized as a curved timeline along a magnitude line in the field

Design Details: Tweet Glyph

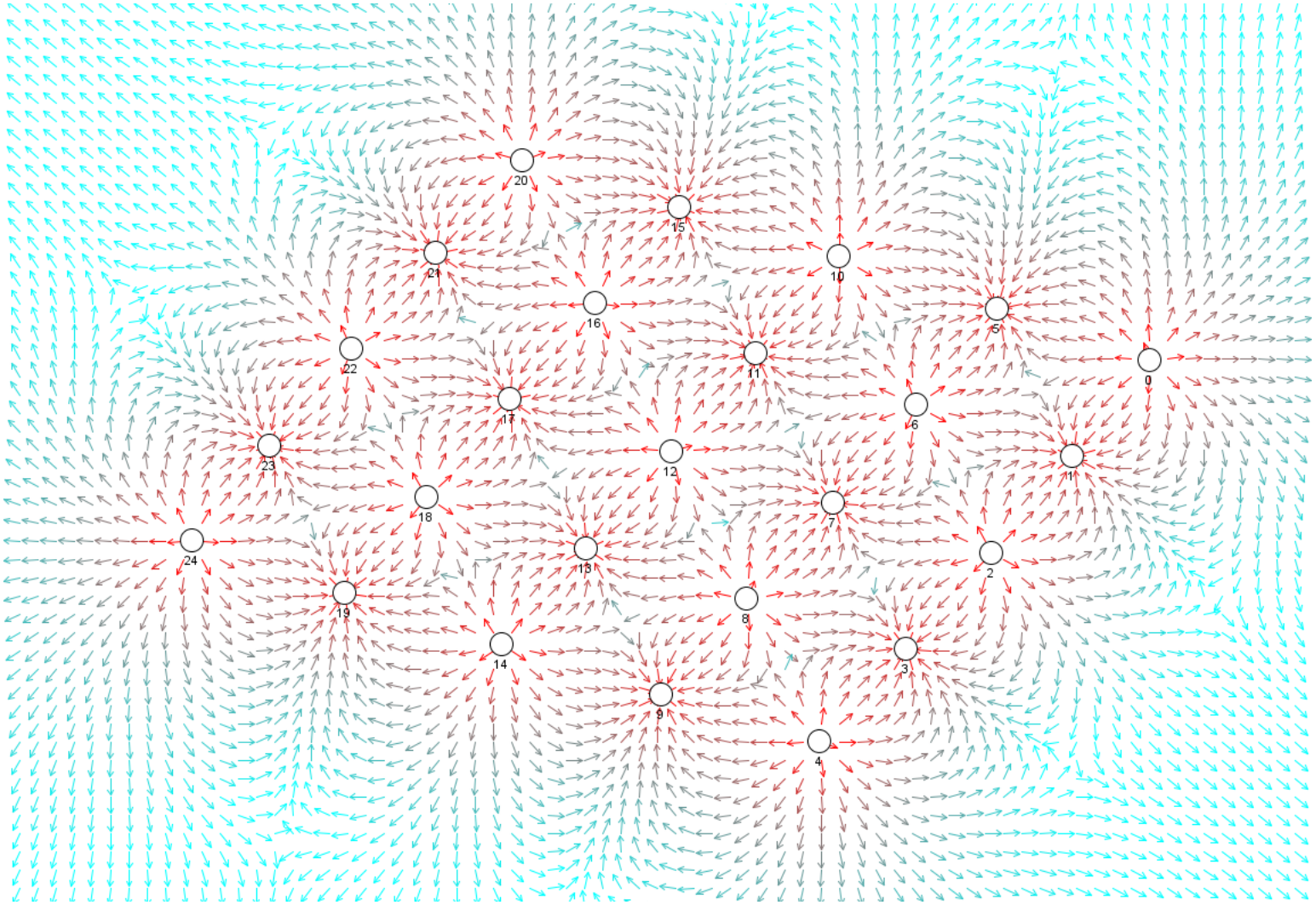
- Tweet Glyph
  - Size
    - Consistent with same size to avoid overlaps crossing different thread
  - Color
    - Hue: opinion
    - Saturation: value of the opinion evaluation
  - Direction
    - From topic to users to indicate the retweet behavior



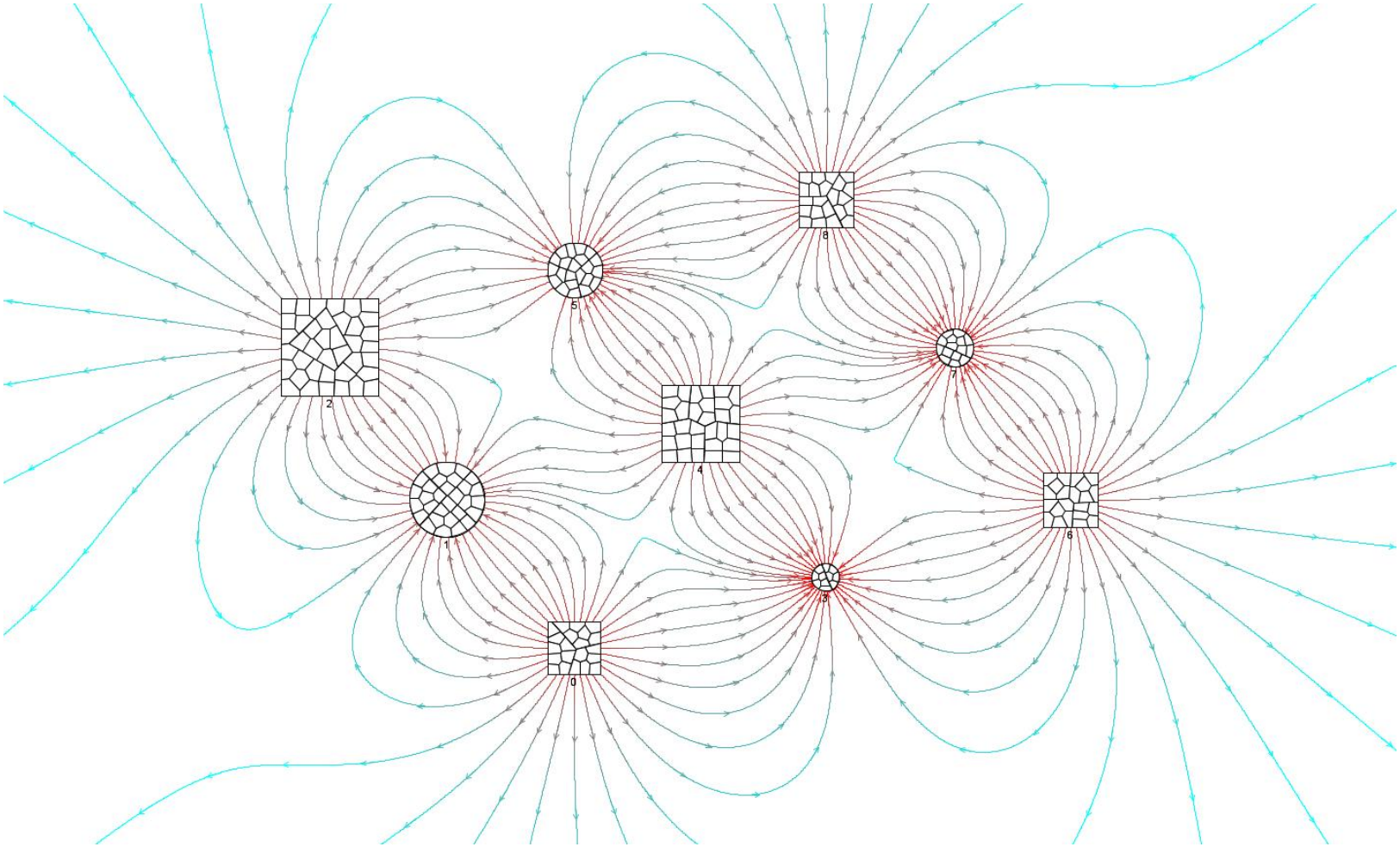




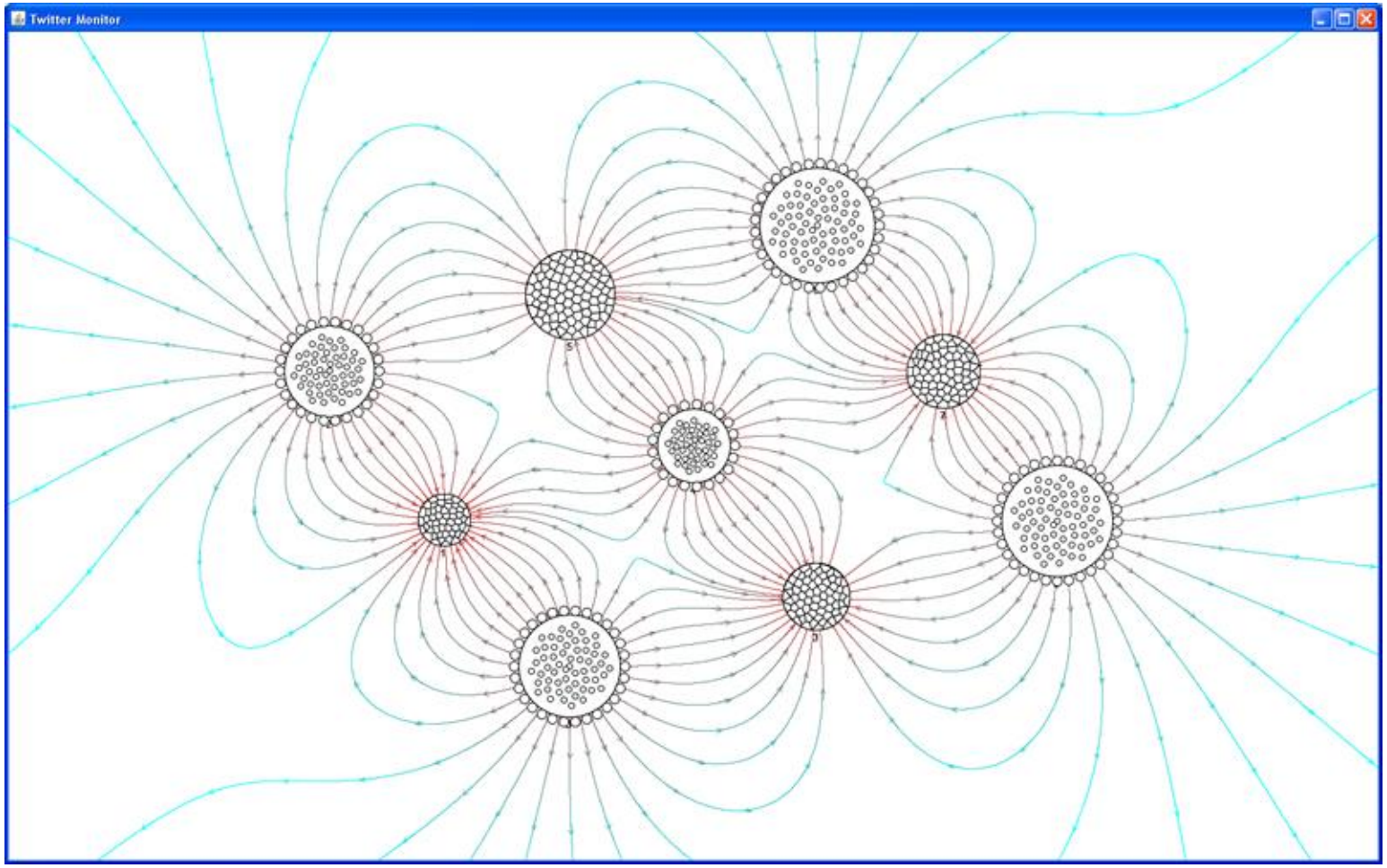
Initial Results



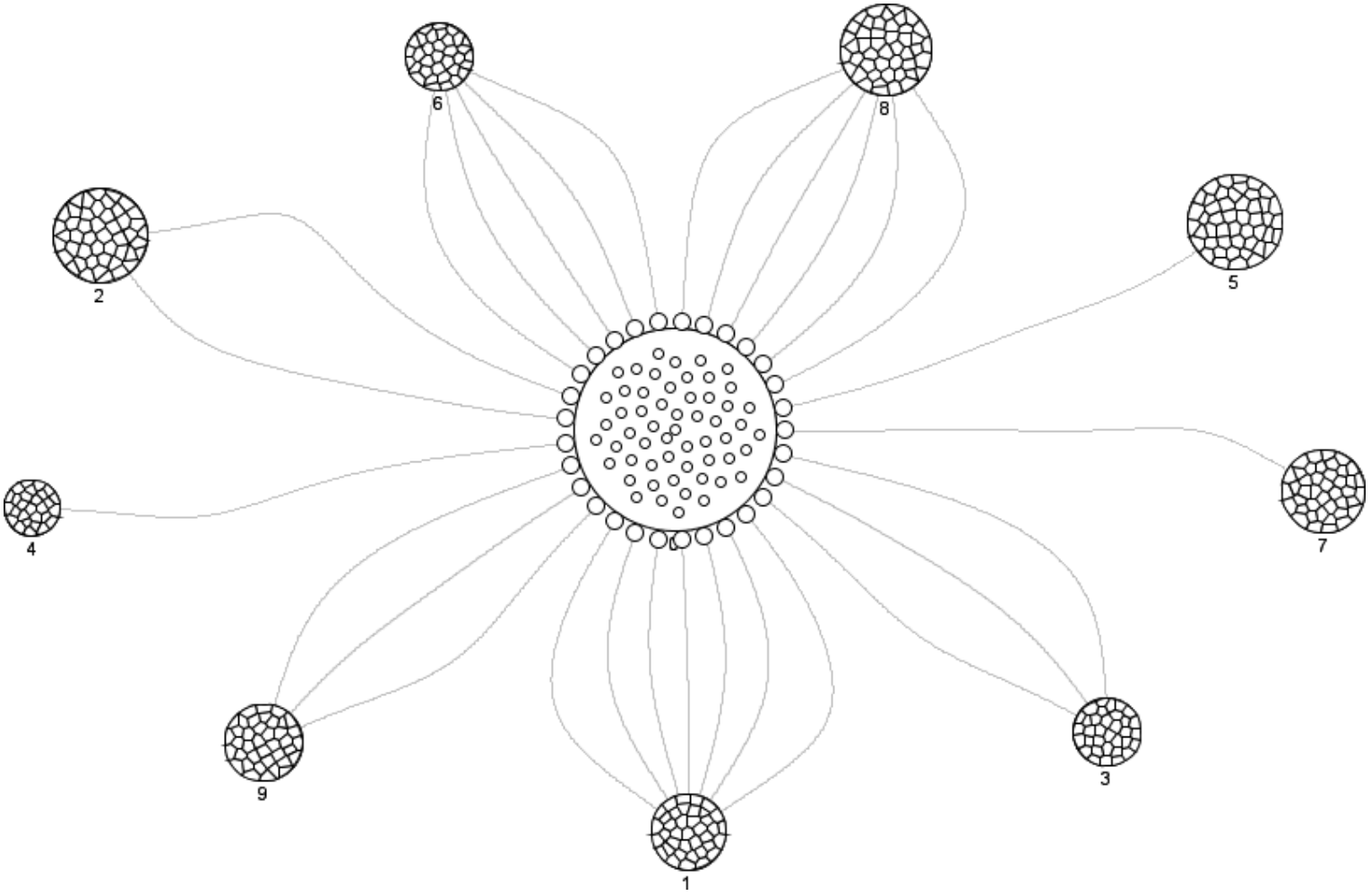
Initial Results



Initial Results



Initial Results



Discussions

- Advantages
  - Compress the dynamic information and opinion diffusion patterns within one consistent design which can easily generate comparative diffusion icons

- Limitations
  - Fail to visualize the information diffusion over different people
  - Lost the diffusion details such as “who reweet whom”

Expected Results

- Dataset(s) to use : Twitter
- Platform, language, and software to use: Java / JavaScript
- How many figures in your experiments?
  - Several case studies on the comparison of different diffusion patterns
  - Fig.1 : Given a topic, discuss it diffusion pattern among different user groups
  - Fig.2 : Compare two related topics (Eg. iPhone vs Nokia vs Sunsung Galaxy) , one is get more and more attention but others lose more and more attentions
  - Fig.3 : Not decided yet

Expected Results

- Evaluation
  - Cast Study: see previous slides
  - User study
    - TASK 1: given two topics, monitor them and their related user communities
    - TASK 2: given two topics, monitor them and identify which one is hotter
    - TASK 3: given two topics, monitor them and identify which one get more positive retweets
  - Interview with domain experts
    - Try to find some scientists who work on social media analysis

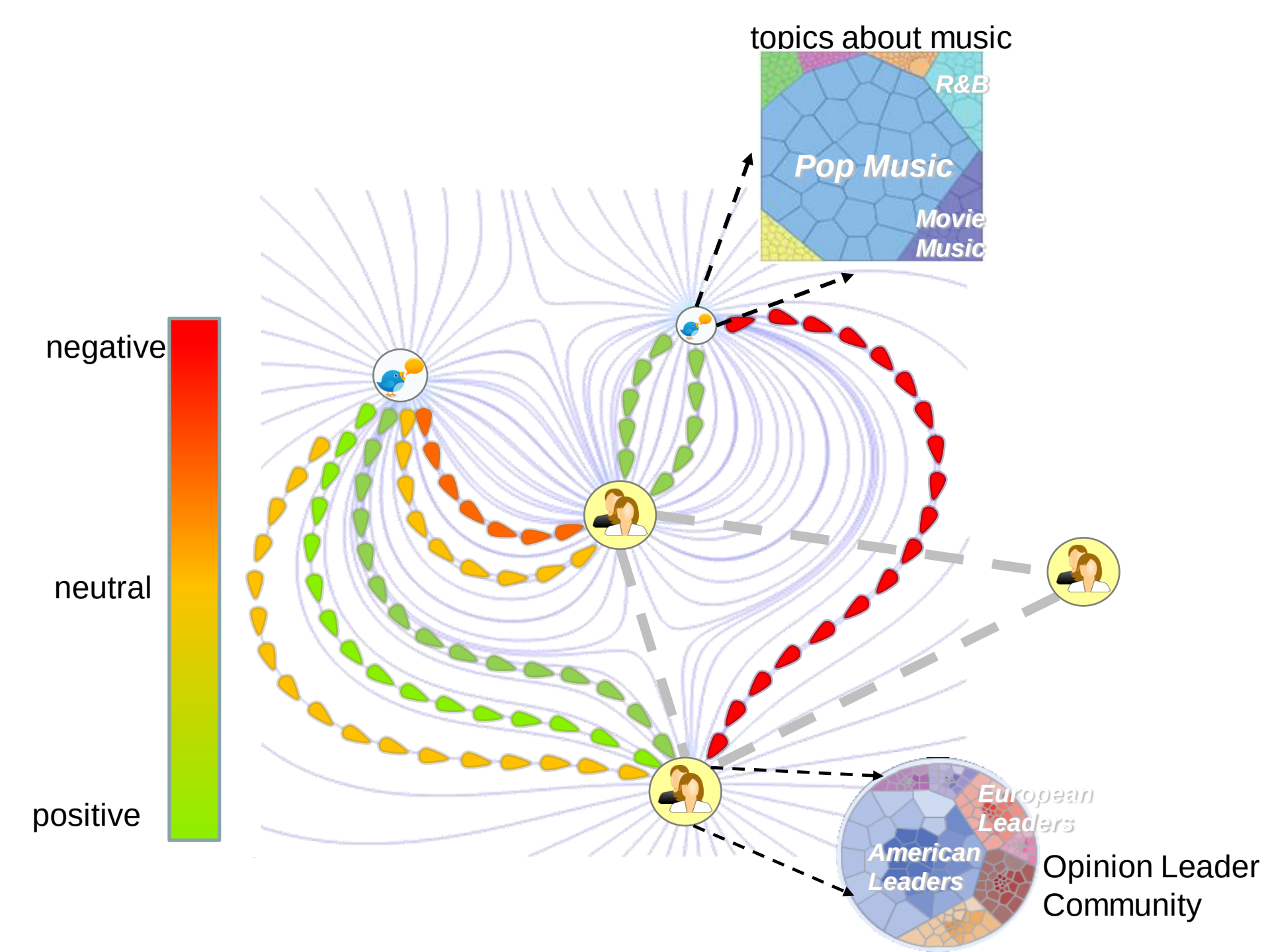
Remaining Tasks

- Unsolved major problems
  - Twitter data online processing and integration
  - Some layout problem
    - Directed magnitude line layout
    - Curved timeline
  - Evaluation
  - Paper writing

Plans

- Workload
  - Coding, case study should be finished at the end of February
  - Finish user study and writing in March

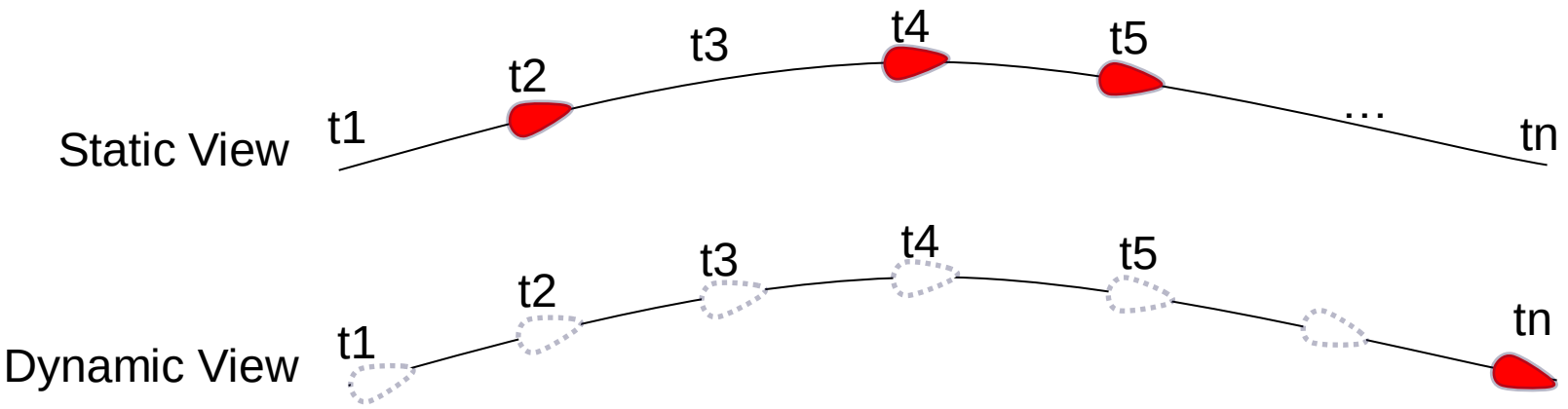
Monitoring Information Diffusion in Micro-Blogs


topics about music
R&B
Pop Music
Movie Music
negative
neutral
positive
European Leaders
American Leaders
Opinion Leader Community

Features

- A bipartite-graph design that provides an overview of both topics and related user communities as well as their social relations in view of information diffusion
- An efficient visual data sampling mechanism based on users' behaviors such as post, retweet and reply
- A user configurable monitoring that reveals the diffusion patterns of multiple streaming trending topics

Design Details



- Each link is a time line which supports both static view and dynamic view
  - Static view for analysis: the tweets are highlighted at the time point where they are retweeted
  - Dynamic view for online monitoring: the tweets are moved in an animation once it is reweeted