

# Visualize the Dynamics of Information Diffusion in Digg Social Network

## Introduction

- Digg is a major social news aggregation website where users can submit links of news stories and the friends of the news initiators can vote and comment on submitted stories. The most popular stories are promoted to the front page therefore can be accessed and voted by all Digg users.
- The spreading of information, known as Information Diffusion, has become more effective with social media sites like Digg and Twitter.
- Understanding the spreading of information and its influence in online social networks has become increasingly important as they grow in size and economic power.

## Goal

- The goal of this project is to create a method of processing data from social news site Digg and generating a visualization displaying the dynamics of information diffusion seen within the website. Visualization of diffusion process plays an important role to help identify the diffusion pattern and furthermore to design mathematical models to describe the pattern.

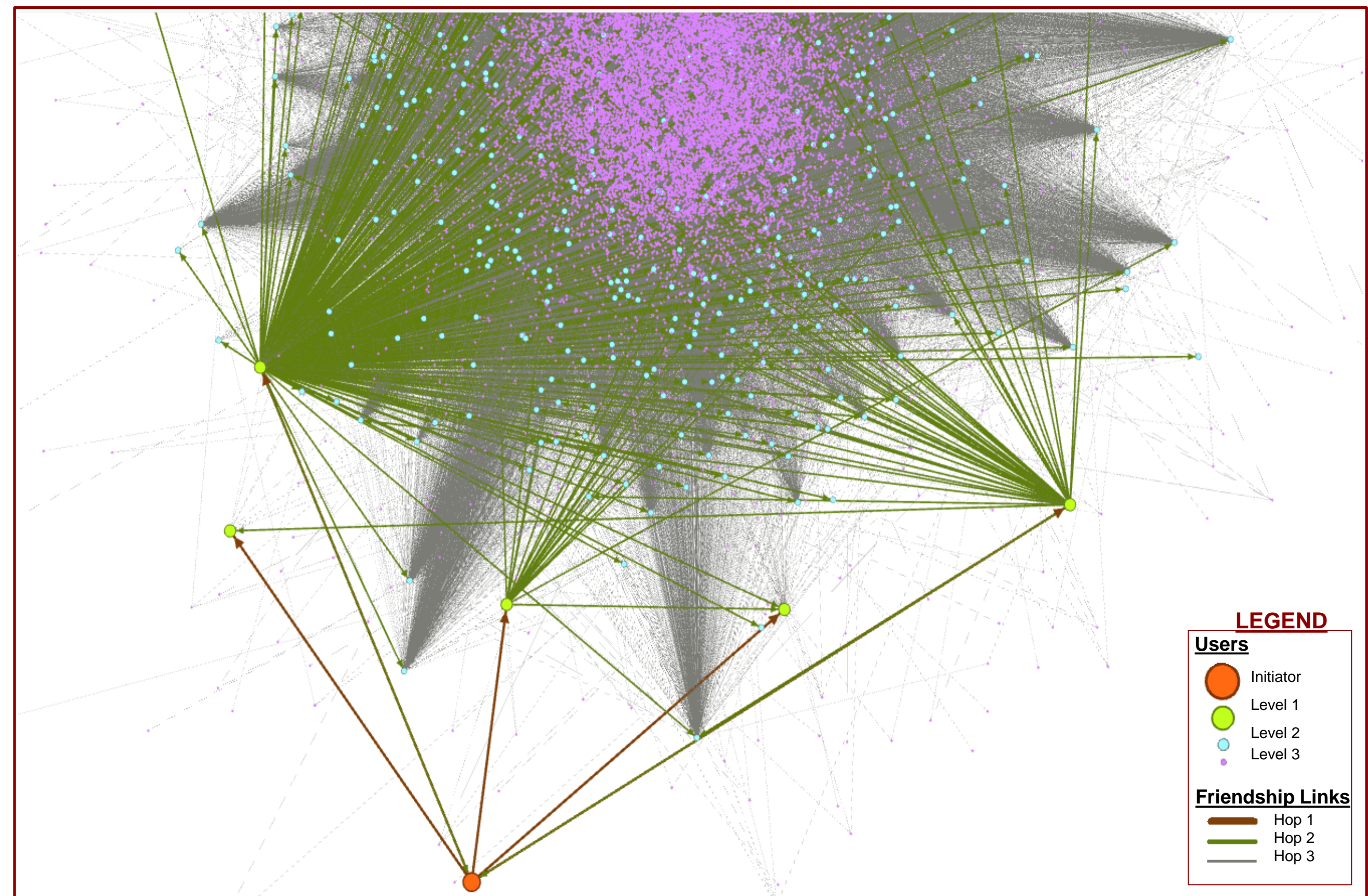
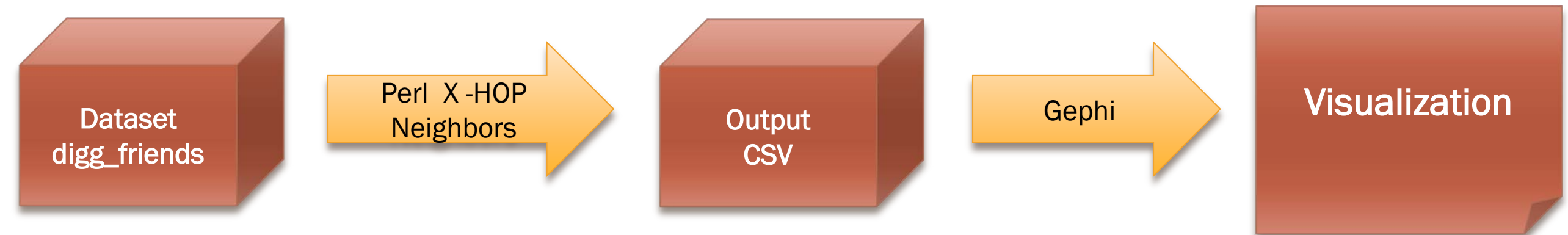
## Digg Dataset

- Dataset being processed contains stories promoted to Digg's front page over a period of a month in 2009. Each story collected contains a list of all the Digg users who voted this story up to collection date, the timestamp of each vote, and the friendship links of each voter. The Dataset is made up of two files:
  - **digg\_votes** contains 3,018,197 votes on 3553 popular stories made by 139,409 distinct users.
  - **digg\_friends** contains 1,731,658 friendship links of 71,367 distinct users.

## Approach

- The challenge of this project is to visualize the **dynamics** of information diffusion over friendship links in a **large scale** social network. We first process the two Digg files to generate **snapshots of information diffusion status and network topology** at different time, then load the snapshots to **Gephi** to animate the diffusion process.
- To illustrate network topology, we implement an algorithm for finding x-hop neighbors of a user. For example, 2-Hop Neighbors of a starting user id include user's friends (1st hop) and his friend's friends (2nd hop). We adopt a modified breadth first search graph traversal algorithm for this purpose. We then mark users at different hops from the initiator with different colors.
- Gephi
  - Open source software for graph and network analysis. Uses 3D render engine to display graphs in real-time and to speed up exploration of data. Gephi provides features to layout, analyze, specialize, filter, manipulate and export all types of graph data. Gephi is the ideal platform for dynamic network analysis, dynamic structures, such as social networks can be filtered on a timeline component.
  - Data is represented in Gephi as Nodes and Edges and can be imported in many formats from CSV to Graph Data Types such as GEXF, GraphML, etc.
  - Gephi can be used in this project to help visualize the Digg topology by importing the processed Digg data into Gephi and using its suite of features to layout and analyze the Digg topology.

## Process



## Configuration of Environment

- Windows 7 Professional SP1 64bit
- Intel® Core™ i7-920 Processor(8M Cache, 2.66 GHz, 4.80 GT/s Intel® QPI)
- Seagate 500GB 7200 RPM 16MB Cache SATA 6.0Gb/s
- 18GB RAM
- ATI Radeon HD5800 Series Graphics Card

## Conclusion

- Overall initial research proved to be successful with the help of Gephi.
- Successfully visualized Digg social network topology.

## Future Work

- Begin to incorporate the vote timestamp to create animated visualizations based on timeline of votes.
  - Given Story ID, visualize the flow of voters over time.
  - Visualize the flow of votes through friendship network.