

MOTIVATION

Big Data

- Popularity of Social Networks has generated high volume, high variety and high velocity data
- Substantial hype about how Big Data can help facilitate business decisions

Viral Marketing

- Viral Marketing exploits Social Networks to spread marketing information leading to reduced marketing costs
- Using Viral Marketing, a campaign can target only a small subset of a network and lead to huge exposure
- Seed selection is the process of selecting the best subset of nodes from the network that will begin the spread of information

RESEARCH QUESTION

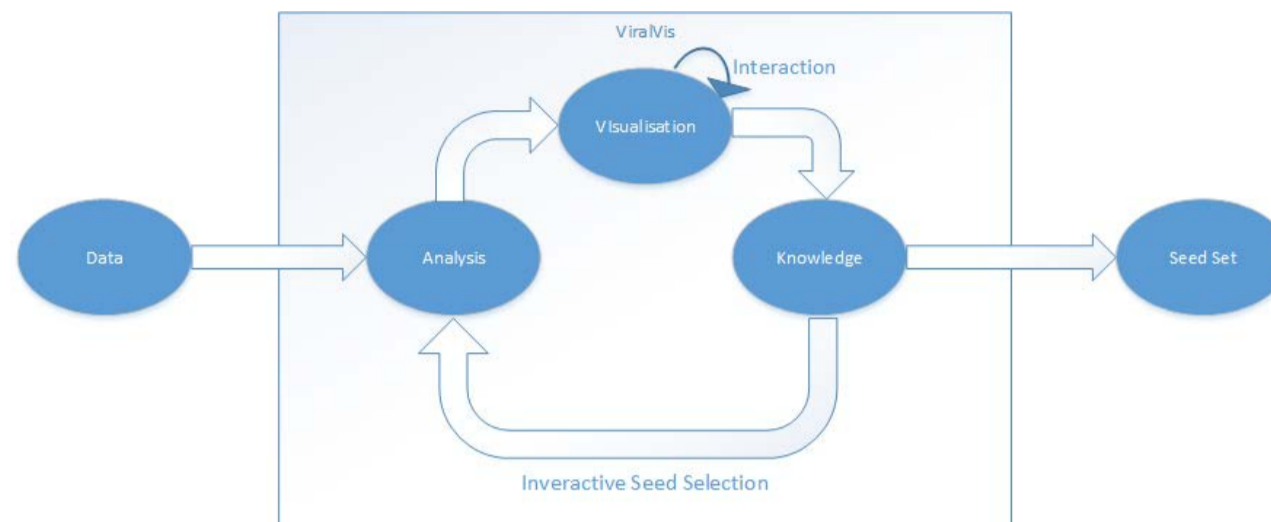
- The current state of research offers automated analysis techniques that generalize marketing campaigns
- Marketing Campaigns are highly complex activities and therefore require tailored methods of seed selection
- Can the selection of seed sets for Viral Marketing be improved by allowing more control to the analyst?

RESEARCH CONTRIBUTION

- Designed a Visual Analytic Framework that combines automated analysis and human cognition to improve seed set selection
- Developed a prototype for the Visual Analytic framework
- Evaluated the framework using case studies on two real datasets

VIRALVIS FRAMEWORK

- This framework aims to achieve tight integration between Visualisation, Analysis and Interaction



- The goal is to allow a process of discovery that helps analyst distribute seeds throughout a network

ANALYSIS

Diffusion Models

- Two Models are utilised in this framework to simulate propagation throughout the network
 - Independent Cascade Model
 - Linear Threshold model

Community Detection

- Community Structure has a significant impact on diffusion throughout networks
- Communities are natural inhibitors to diffusion and weak links or bridges between communities are key components to diffusion
- Multilevel Community Detection is used for scalability with a linear worst case runtime

Structural Characteristics

- Size and density of communities impact upon diffusion, hence seeding strategies need to be modified for communities with vastly different characteristics

Centrality Analysis

- Betweenness Centrality
- Degree Centrality

VISUALISATION

- Multi-view Visualization method
- Community View, Detail View, Simulation

INTERACTIVE SEED SELECTION

- Analysts can interactively distribute seeds throughout the communities
- Rule based Method of selection allows different distribution methods for communities

EVALUATION

- The framework is evaluated on synthetic and real datasets
 - Flickr Data
 - Facebook Data
- Two case studies are applied to the datasets

Flickr Dataset

- Flickr is a photo sharing social network
- The dataset consists of 35 000 nodes
- Undirected Social Network and comment ship

Facebook Dataset

- Facebook is an online social network
- The dataset is mined through a breadth first search on the network
- The Facebook dataset consists of 100 000 nodes
- Links represent mutual friendships between nodes
- Undirected Social Network

Case Studies

- Two case studies with different marketing objectives
 - White goods start up implementing a new viral marketing campaign to reduce costs
 - Technology company looking to achieve maximum penetration in a time constrained environment

Facebook Dataset

Size	Degree	BT	Density	Selection	NumNodes	Distribution
4 - 2582	0 - 38	7 - 220.91351116204058	1 - 4	degree	1000	Temp
3538 - 8297	6 - 19	38 - 220.91351116204058	1 - 4	degree	2500	Temp