The past decade has witnessed the emergence of participatory Web and social media, bringing together people in many creative ways.
Millions of users are playing, tagging, working, and socializing online, demonstrating new forms of collaboration, communication, and intelligence that were hardly imaginable just a short time ago.
Social Media refers to interaction among people in which they create, share and exchange information and ideas in virtual communities and networks.
Social Media also helps reshape business models, sway opinions and emotions, and opens up numerous possibilities to study human interaction and collective behavior in an unparalled scale.

In the study of complex networks, a network is said to have community structure if the nodes can be easily grouped into sets of nodes (even overlapping) such that each set of nodes is densely connected internally.
Community structure are quite common in real networks.
Social Networks often include community groups based on common location, interests, occupation etc.
Metabolic Networks have communities based on functional groupings.
Citation Networks form communities by research topic.
Being able to identify these sub-structures within a network can provide insight into how network function and topology affect each other.

In this thesis, we design an end-to-end framework for identifying communities from raw, noisy social media data.
The framework is composed of two important phases.
First, we introduce a new method of converting the raw, noisy social media data into a weighted entity-entity co-occurrence based consistency network.
This includes a simple iterative noise removal procedure for cleaning the entity consistency network by removing noisy entity pairs.
Secondly, we propose an approach for identifying coherent communities from the weighted entity network, by introducing novel notions of community-ness and community, based on eigenvector centrality.

We use this framework to solve three different problems from two distinct domains.
The first problem involves detecting communities from raw social media data and showing the application of the communities discovered in a recommendation engine setting.
We use the framework for converting the raw data into a clean network and propose a highly parallelizable seed based greedy algorithm to detect as many communities as possible from the weighted entity consistency network.
Our framework for community detection is unsupervised, domain agnostic, noise robust, computationally efficient and can be used in different Web Mining applications like Recommendation Systems, Topic Detection, User Profiling etc.
We also design an recommendation system to evaluate our framework with existing state-of-art frameworks~\cite{LDA,Farkas,bigclam} on a

variety of large real-world social media data - Flickr, IMDB, Wikipedia, Bibsonomy, Medline.
Our results outperform other frameworks by a huge margin.

The second problem is, given a set of communities of discovered by traditional community detection methods~\cite{Palla,Lancichinetti09}, we need to identify loose communities among them and partition them into compact ones.
Here, we use the second phase of our framework to identify such loose communities using our notion of community-ness and propose an algorithm for partitioning such loose communities into compact ones.
We illustrate the results of our algorithm over Amazon Product and Flickr Tag data and compare its superiority over the traditional community detection methods in a recommendation engine setting.

The third problem is about showing the application of such framework in an Image Annotation scenario in presence of noisy labels.
The problem of image annotation is defined to be, given an unknown image, we need to predict labels which best describes the semantics of the image.
This problem is best solved in a supervised nearest neighbor setting, and we show how our framework can be used to address this problem, when the labels associated with training images can be noisy and redundant.