Meerkat: Community Mining in Social Networks

Jiyang Chen, Justin Fagnan, Randy Goebel, Reihaneh Rabbany, Farzad Sangi, Mansoureh Takaffoli, Eric Verbeek, Osmar Zaiane
Department of Computing Science
University of Alberta, Edmonton, Alberta, Canada T6G 2E8
{jiyang, jfagnan, goebel, rabbanyk, fsangi, takaffol, everbeek, zaiane}@cs.ualberta.ca

Abstract—Meerkat is a tool for the visualization and community mining in social networks. It is being developed to fill a gap in algorithms present in other social network tools. It offers easy navigation through graphical representations of a network, network querying, community mining using recently developed algorithms, and dynamic network event analysis using recently published algorithms. These features will allow more insightful exploratory analysis and more robust inferences about communities and the significance of entity relationships. Meerkat is under active development, and future features will include additional options for community mining and visualization, focusing on algorithms and user interface designs not extant in the existing social network analysis tools.

Index Terms—social networks, informational networks, community mining

A. Information

Contact:
Dr. Osmar R. Zaiane
Telephone: 1-780-492-2860
Fax: 1-780-492-1071

Category:
Mining in Emerging Domains: Mining and link analysis in networked settings: web, social and computer networks, and online communities.

Equipment We Will Bring:
Laptop

Equipment We Need:
Table, Poster Board, Power Socket, Internet Connection

Present State:
Meerkat is currently functional, but unreleased and under development. All features described have been implemented to a completely functional level.

I. INTRODUCTION

Social network analysis allows researchers to pull apart relational phenomena in many domains. Currently, most existing analysis software focuses on statistical modeling and static visualization of networks. With an increasing interest in dynamic networks, these tools draw up short in helping researchers to see certain types of features in their data. We are interested in bridging this gap, by offering exploratory analysis functionality based on recent advances in the social network analysis methodology. This functionality can be divided into four general categories: dynamic network visualization and network statistics; filtering and extraction; community mining; and event analysis. These features allow researchers to discover important entities, and infer the communities that they belong to, while tracking them over time and through hierarchical exploration.

II. NETWORK VISUALIZATION AND STATISTICS

The first of these categories, visualization, plays a subtle yet important role when analyzing social networks in the search of interesting patterns or features. Identifying outliers and potential nodes of interest is often made easier by studying a visualization of the network. To facilitate better analysis we have made visualization a key-component of Meerkat and in doing so have improved upon existing tools by making our visualization fully malleable, allowing the researcher to zoom, scroll, rotate, skew, or reposition nodes and edges. Additionally we have enabled both traditional and novel layout algorithms.

To complement the visualization we have placed emphasis on interactivity with the network by implementing a one-touch view of node or edge attributes, scroll to selected node, and real-time modification of a nodes color, size, or shape. This gives researchers unprecedented control by allowing them to custom-tailor their own visualization experience.

As is traditional in analysis software, we also provide an array of statistical information to compliment the visualizations. This includes many well-known network measures such as PageRank, HITS, and Centrality, which can be used to reveal which nodes are the most important in the global network, or within each community. These statistics can also be visualized as a dart board to facilitate analysis for the visually inclined.

III. FILTERING AND EXTRACTING

In addition to visualization we also offer what we believe should be a fundamental operation when studying any social network: interactive what-if analysis. What happens to the community if a node or edge is removed? What communities are formed if we only consider a sub-set of this network? For the first time in social network analysis questions like these can be answered interactively by two separate, yet related, concepts.
Before we introduce those concepts we should first notice that many social networks contain multiple dimensions of data outside of the standard node and edge relationship. For example, an email social network may contain information on when the email was sent, the subject of the email, or the full name of the recipient. This information can be captured in Meerkat by defining attributes that can be specific to a particular node or edge, or a group of nodes or edges.

Once defined, these attributes can be employed by the first of our interactive what-if concepts: filtering. A researcher may define a filter on any user-defined attributes or on a selection of pre-defined attributes, such as node degree and edge weight. The filter is created through an intuitive user interface that allows for a series of nested conjunctive and disjunctive Boolean operators. Furthermore, multiple filters may be defined and then enabled or disabled to create either a complex or simple query, without needing to redefine everything twice.

IV. COMMUNITY MINING

Many of the previously discussed features have been designed to synerize well with the core functionality of Meerkat: social network community mining. As a motivation for this functionality we believe that by understanding the communities in their network, a researcher may be able to alter policies in government, advise marketing plans in telecommunications, or understand which group proteins may be responsible for malignant side effects. To facilitate this, Meerkat offers both existing and novel community mining algorithms that discover communities based on the relationships between nodes. This includes algorithms that operate globally or locally and with or without: overlap, hub nodes, and edge weights. As far as we know it is the only tool to offer such a vast selection of community mining algorithms. The results of these algorithms are visualized by coloring (or numbering) the nodes that belong to each community and via a listing of each community’s members.

Identifying communities is often only the first step in analysis. To aid in further analysis we provide common statistics such as density, diameter, and cohesion for each community. Along with these we also provide visualizations depicting the most statistically important nodes in each community and how many inter-community relations exist.

Often in very large networks the granularity provided by standard community mining may be too small; such as networks containing hundreds or thousands of communities. To combat this we have drawn on our experiences in hierarchical clustering to produce a generalized hierarchical community mining algorithm. By discovering the communities at each level of the hierarchy we are able to provide the ability for the researcher to view various levels of granularity, including an aggregate view where the entire hierarchy is displayed at once in an intuitive yet informative layout. Additionally, if the smallest granularity is too large, the researcher may extract a single community and find its sub-communities.

V. EVENT ANALYSIS

Once a researcher obtains a listing of communities, they are often tasked with answering an important question: How are the communities changing over time? This idea of analysis over time, or dynamic network analysis, is the final pillar on which Meerkat is built. Social networks are only infrequently invariable. To promote and facilitate community mining across time, Meerkat offers event analysis functionality. Researchers can see how the communities discovered in their network change over time, via gaining or losing members, forming afresh or dissolving entirely, splitting into multiple schism
communities, merging together into a combined community, or surviving mostly intact across time. Researchers need the ability to clearly see the evolution of communities over time, or they will be missing out on an essential relational aspect of their domain networks. Based upon earlier work on event analysis [1], additional event definitions and computational parametrizations [2] have been implemented for Meerkat. These enhancements offer a more complete ontology of network events, allowing researchers to have a clear view of their networks evolution.

VI. CONCLUSION

Meerkat facilitates the exploration of social networks, allowing researchers to visualize networks using multiple layout algorithms, compute descriptive node statistics, analyze subsets of networks, perform community mining with several algorithms, and reveal network dynamics at the community level. Many of these tasks are designed to be as unobtrusive as possible to the researcher through prolific use of threading. Most importantly, Meerkat uses recently developed layout, mining and event detection algorithms currently unavailable elsewhere, and will continue to include cutting edge algorithms as they are developed. These new developments will be immediately available to clients through Meerkats web based user-interface, meaning there will be no need to run an update installer in order to receive new features when they are available.

REFERENCES
