Meerkat: Community Mining with Dynamic Social Networks

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

***Abstract*—Meerkat is a tool for visualization and community mining of social networks. It is being developed to offer novel algorithms and functionality that other tools do not possess. Meerkat's features include navigation through graphical representations of networks, network querying and filtering, a multitude of graphical layout algorithms, 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 existing in other social network analysis tools.**

***Keywords*-social networks; informational networks; community mining;**

I. Introduction

Conventional data is typically a set of observations, which are considered independent of one another and identically distributed. In reality, data could be highly dependent, with observations relating to one other in a variety of relationships. The analysis, visualizing, and interpreting of such relational data is known as social network analysis (SNA).

Social network analysis allows researchers to study relational phenomena in many domains. Most existing analysis software focuses on statistical modeling and static visualization of networks. With an increasing interest in dynamic networks, these tools come up short in helping researchers to see certain types of features or patterns in their data. We are interested in bridging this gap, by developing an exploratory analysis tool, named Meerkat, based on recent advances in the social network analysis methodology.

Network input format is common to other SNA tools, using a simple pair of vertices to indicate an edge. The network file can optionally include explicit lists of vertices, edge or vertex properties, and timeframe labels for dynamic networks. Meerkat's functionality can be divided into four general categories: (1) interactive network visualization and network metrics; (2) filtering and extraction; (3) community mining; and (4) event analysis. These features allow researchers to discover the importance of entities in their domain networks, and make inferences about algorithmically discovered communities. Being able to track entities and communities over time, observing community evolution, and performing analysis at different hierarchical granularities allows for better leverage in network exploration.

The Meerkat project homepage is currently located at http://aicml.cs.ualberta.ca/?q=node/68, where a demo version is available. A video tutorial is also available for viewing at http://aicml6.cs.ualberta.ca/toolbox/timeframeDemo/meerkatFullDemo.wmv.

II. Interactive Network Visualization and Metrics

The first of these categories, visualization, plays a central role in the exploration style that Meerkat facilitates. Identifying nodes of interest is often made easier by studying a visualization of the network, and studying communities is infeasible without proper visualization. On the level of user interaction, Meerkat allows researchers to zoom, scroll, rotate, skew, or reposition nodes and edges, and use optical lens tools to see dense networks more easily by magnifying regions of interest. Users may also draw shapes and type annotations on graphs, in order to indicate areas of interest in their network. Also, when working with a network, an analyst might want to reduce the complexity; Meerkat can collapse multiple nodes together, either virtually or permanently, thus abstracting a cluster of nodes and aggregating their edge relations to other entities. This node collapse feature is also used in an automated form for hierarchical analysis, which we will discuss later.

Graph layout is a core concern in network visualization. Meerkat utilizes several traditional and novel layout algorithms, and includes Fruchterman-Reingold [5], Circle [8], Spring [8], Self-Organizing [11], and Kamada-Kawai [9] layout algorithms. We are interested in producing layout algorithms with better performance and greater practical utility. Meerkat includes community-based layouts that are driven by community mining, as discussed later.

To complement Meerkat's visualization capabilities, we have placed emphasis on interactivity with the network by implementing a one-touch view of node or edge attributes, and real-time modification of node colour, size, or shape. Researchers can define entity classes, allowing for entity typification, and optionally for entity properties to control node colouration, facilitating quick visual parsing of a network. This gives researchers unprecedented control by allowing them to custom-tailor their own visualization experience.

As is traditional in analysis software, we also provide metrics, or statistical information, about nodes and communities. Entries in the metric table are selectable; clicking on an entity there scrolls the visualization pane to the selected nodes and highlights them for visual clarity. Meerkat's metrics include well-known network measures such as PageRank [13], HITS [10], Normalized-Alpha [6], Betweenness [4], and Centrality [7], which can be used to reveal which nodes are the most important in the global network, or within each community. These metrics can also be visualized in specialized 'dart board' layouts, arranging nodes in descending order of the selected metric relative to the center of the layout.

III. Filtering and Extraction

In addition to visualization, Meerkat facilitates enlightening analysis when studying any social network via 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 creation 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.

When a network has been filtered to include only a subset of entities, or if the user has manually selected entities, these nodes can be extracted to another graph. Meerkat will open a new visualization tab, with only those nodes that passed the filter or were manually selected. This allows an analyst to make small or large changes to a graph, and flip back while comparing any differences in statistics, in node groupings for deterministic layouts, or in the communities discovered by mining algorithms.

IV. Community Mining

Meerkat's community mining abilities are meant to synergize with its other features. Although an analyst might come

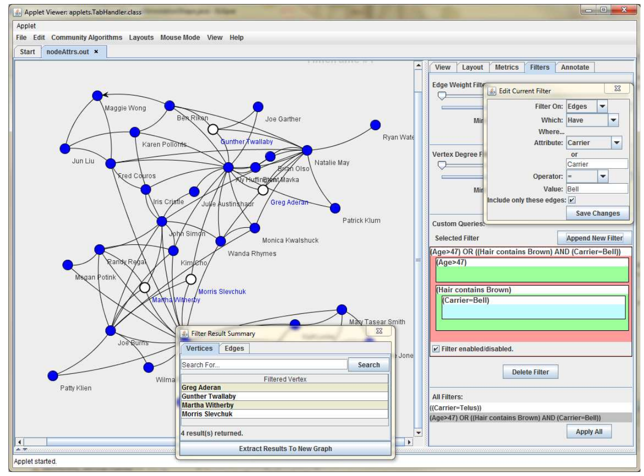

Figure 1. Meerkat's filtering results. A list of filtered nodes is returned in the center window. The window on the right depicts how a new filter is added. The colourful boxes beneath it are a visual representation of the complete filter, with two similar coloured boxes indicating a logical 'OR', and boxes within each other indicating a logical 'AND'. A researcher may interactively add new filters by right-clicking within these boxes.

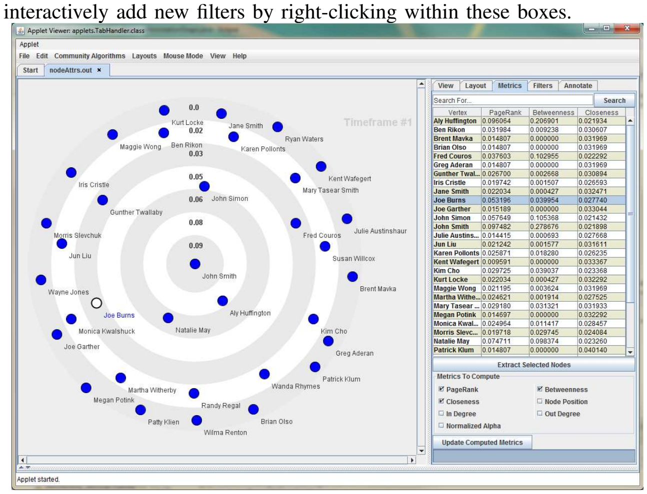

Figure 2. Meerkat's PageRank metric layout, with the highest scoring nodes closer to the center. The 'Metrics' panel can be seen on the right hand side.

to understand their networks through exploration of entities and through patterns visible through a particular layout, having automated community detection can give insight that is otherwise difficult to achieve. Given high quality community groupings based on edge connections, a researcher may be able to alter social policies in government, advise marketing plans in telecommunications, or understand what groups of proteins function together. To facilitate this, Meerkat offers both existing and novel community mining algorithms. This includes several central algorithms, and slight variations of them, including Fast Modularity [12], Max-Min Modularity [3], TopLeaders [15], Clique Percolation [14], Local Mining with or without hubs and overlap [2]. We continue to consider and research other algorithms for inclusion, but

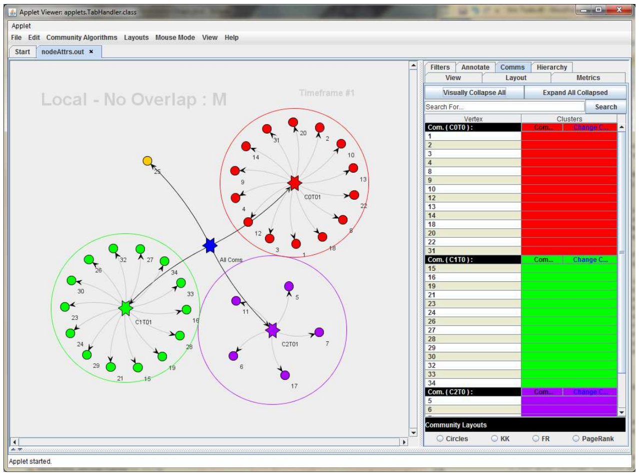

Figure 3. Meerkat's hierarchy layout. The small coloured circles are the nodes, and the edges depict an 'owner' relationship in the hierarchy. Nodes within the larger coloured circles belong to the community indicated by its colour. The star-shaped objects represent a community, with the grey objects being a community made up of other communities and the blue object being the top-level super community. The community panel displayed on the right indicates which nodes belong to which communities.

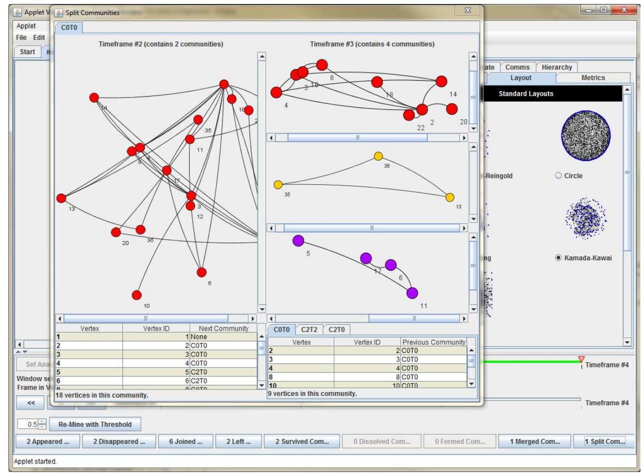

Figure 4. Meerkat's timeframe analysis. The window in the forefront visualizes a split event, where a single community splits into two separate communities in a subsequent timeframe. Here, the largest resulting community retained the colour of the split community. The tables below indicates each node's community after the split. The bottom panel in the background describes which events were detected for this timeframe and contains controls to change timeframes.

seek to include only those that will actually offer some improvement to Meerkat's abilities.

The results of these algorithms are visualized by colouring or labeling the nodes that belong to each community, as well as listing community membership in tabular form. After identifying communities, Meerkat can compute 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. Community layouts are also available following mining, which allows a spatially clustered view of the communities; this is particularly important in larger networks where other layout algorithms might allow too much visual overlap between communities.

Often in very large networks, the granularity provided by standard community mining may be too small, such as in networks containing hundreds or thousands of communities. To help overcome this problem of scale, Meerkat includes 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. Manual rolling up and drilling down of communities may also be undertaken by using the filtering and extraction features discussed in the previous section.

V. Event Analysis in Dynamic Networks

Once a researcher obtains a listing of communities, they may then ask a particularly interesting 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 in many domains are subject to entities dropping in and out of interactions, and thus migrating across communities. 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. Meerkat's event analysis include advanced event definitions and parameterized algorithms [16] which are based on earlier work on event analysis [1]. These enhancements offer a more complete ontology of network events, allowing researchers to have a clear view of their networks evolution.

Community event visualization is also an area of development in Meerkat. It includes basic modes of representing events, by selecting an event type, and showing all of the affected communities before and after the event as applicable, with involved nodes highlighted and listed alongside. It also includes a more advanced view that lets the analyst select a community at some point in time, and see all of the events that occur for it, with visualizations of the community and its changes in one contiguous place. Tracking communities over time like this allows a community-oriented style of interaction that compliments the basic event-oriented visualization.

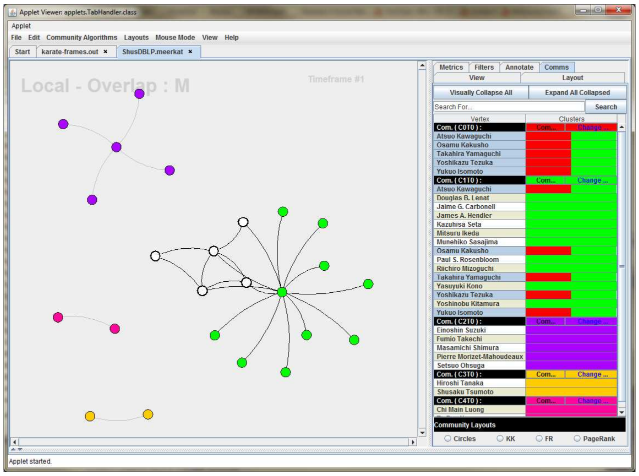

Figure 5. After community mining using the Local Mining [2] algorithm with the overlap option enabled, we see that the red community here is completely overlapped by the green community. We have selected the nodes of the red community, so they are highlighted as white in the left-hand visualization pane. The nodes were layout is Kamada-Kawai [9] algorithm.

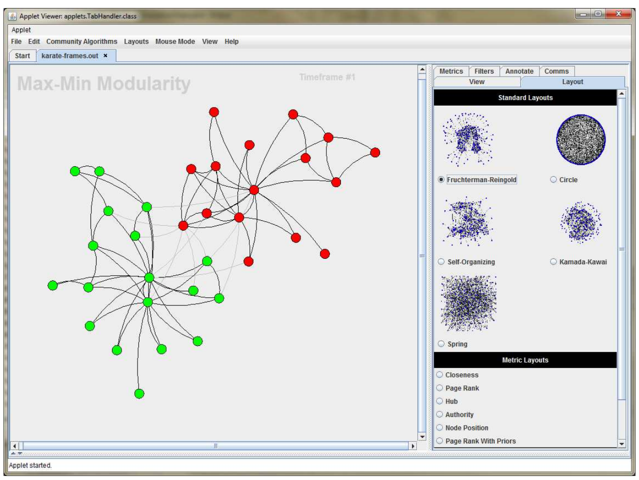

Figure 6. Some karate club members are positioned using the Fruchterman-Reingold [5] algorithm, and are community mined using the Max-Min Modularity [3] algorithm, which results in two communities being identified.

VI. Conclusion

Meerkat facilitates the exploration of social networks, allowing researchers to visualize networks using multiple layout algorithms, compute descriptive node and community metrics, analyze subsets of networks, perform community mining with several algorithms, and reveal network dynamics at the community level. 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 Meerkat's 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

[1] Asur, S., Parthasarathy, S., and Ucar D., "An Event-Based Framework for Characterizing the Evolutionary Behaviour of Interaction Graphs," *ACM SIGKDD*, 2007.

[2] Chen, J., Zaïane, O., and Goebel, R., "Detecting Communities in Large Networks by Iterative Local Expansion," *International Conference on Computational Aspects of Social Networks (CASoN)*, Fontainebleau, France, June 24-27, 2009.

[3] Chen, J., Zaïane, O., and Goebel, R., "Detecting Communities in Social Networks using Max-Min Modularity," *SIAM International Conference on Data Mining (SDM'09)*, Sparks, Nevada, USA, April 30- May 2, 2009.

[4] Freeman, L. C., "A set of measures of centrality based on betweenness," *Sociometry*, 40:35-41, 1977.

[5] Fruchterman, T. M., and Reingold, E. M., "Graph Drawing by Force Directed Placement," *Software Practise and Experience*, 21(11), 1991.

[6] Ghosh, R., and Lerhman, K., "Predicting Influential Users in Online Social Networks," *Proceedings of KDD Workshop on Social Network Analysis (SNA-KDD)*, Washington DC, USA, 2010.

[7] Hage, P., and Harary, F., "Eccentricity and centrality in networks," *Social Networks*, 17:57-63, 1995.

[8] JUNG. [Online]. Available: http://jung.sourceforge.net/

[9] Kamada T., and Kawai, S., "An Algorithm for Drawing General Undirected Graphs," *Information Processing Letters*, 31:7-15, 1988.

[10] Kleinberg, J. M., "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, 46(5):604632, 1999.

[11] Meyer, B., "Self-organizing graphs: A neural network perspective of graph layout," *Graph Drawing (GD'98)*, 242-262, 1998.

[12] Newman, M. E. J., "Finding community structure in networks using the eigenvectors of matrices," *Physical Review E*, 74, 2006.

[13] Page, L., Brin, S., Motwani, R., and Winograd, T., "The pagerank citation ranking: Bringing order to the web," Technical Report, Computer System Laboratory, Stanford University, 1998.

[14] Palla, G., Dernyi, I., Farkas, I., and Vicsek, T., "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, 435, 2005.

[15] Rabbany Khorasgani, R., Chen, J., and Zaïane, O., "Top leaders Community Detection Approach in Information Networks," *4th SNA-KDD Workshop on Social Network Mining and Analysis*, Washington, DC, July 25 2010.

[16] Takaffoli, M., Sangi, F., Fagnan, J., and Zaïane, O., "A Framework for Analyzing Dynamic Social Networks," *Applications of Social network Analysis (ASNA)*, Zurich, Switzerland, September 2010.