A Framework for Analyzing Dynamic Social Networks

Mansoureh Takaffoli, Farzad Sangi, Justin Fagnan, Osmar R. Zaïane

Department of Computing Science

University of Alberta, Edmonton, Canada

{takaffol, fsangi, fagnan, zaiane}@ualberta.ca

Abstract

Social network analysis has emerged as a set of methods for the analysis of social structures and uncovering the patterning of interactions among the entities. In the past, social network analysis was mainly a static investigation by considering independent graphs at different snapshots or one aggregated graph over the time period. However, for the dynamic social networks that change over time, the static analysis misses the opportunity to capture evolutionary patterns. A community is one of these patterns, and it is affected by changes in the underlying population in the dynamic social networks. In the literature there has been a considerable amount of work done to detect communities in social networks. However, the communities are independently detected at each snapshot regardless of the structural relationship between consecutive snapshots. In this paper we present a framework for modeling and detecting community evolution in social networks. This framework allows tracking of events related to communities as well as events related to individual nodes. These events can be considered as building blocks for pattern detection in networks with evolving communities. We illustrate the capabilities and potential of our framework by applying it to a real dataset consisting of emails from the Enron Corporation. The evolution of the Enron communities is detected with the events defined in our framework.

1 INTRODUCTION

Information networks, also known as social networks, are interconnected records typically represented by a graph where nodes are data points and edges represent relationships. These networks are not independent and identically distributed (i.i.d.) and thus necessitate a different analysis approach than those used on i.i.d. data. A social network represents an interaction network of individuals that are often connected to each other by a relationship. Although at first glance these relationships may seem simple, upon further inspection they can form interesting structures. For instance, a relevant structure that materializes in social networks is the notion of a community. There is no clear definition for a community, but a reasonable explanation proposes that the nodes in a community are more connected to each other than to other nodes of the network. In particular, the relationships between nodes in a community have something in common that distinguishes them from other nodes.

Communities in a real social network are affected by changes in the underlying population. Tracking and understanding these changes plays an effective role in areas such as sociology, anthropology, bioinformatics, sociolinguistics, geography, information science, politics, marketing, etc. As an example, the evolution of informal groups within a large organization can provide insight into the organization's global decision-making behaviour. Another example includes tracking the early stages of an epidemic disease in a specific subpopulation. While some work has been done on community mining in static social networks, very little has been done on dynamic social networks. However, tracking the evolution of a community, or the relationships of an individual, over time is important for many applications.

We propose an event-based framework to categorize and track how communities evolve in social networks. Our framework takes the detected communities at consecutive snapshots as an input and provides a mapping of how each community evolved at each snapshot. It also allows one to follow the events pertaining to individual nodes across each snapshot.

The rest of the paper is organized as follows. First in Section 2, we discuss related work. In Section 3, we describe our event-based framework in detail. Experimental studies on the Enron email datasets are provided in Section 4. Finally in Section 5, we discuss the conclusion and future work.

2 RELATED WORK

In the literature there has been a considerable amount of work done to detect communities in social networks [1]–[4]. A common issue in the previous work is that the analysis of social networks was mainly a static investigation of the aggregated graph of the network across multiple snapshots. Hence, in the noticeable effect of time was neglected. However, a large number of social networks are continuously changing over time, thus they require a dynamic analysis.

Recently there has been some work on analyzing communities and their evolutions in dynamic social networks. Leskovec et al. [5] studied the patterns of growth for graphs based on various topological properties, such as the degree of distribution and small-world properties of large networks. They also proposed a graph generation model, called the Forest Fire model, to produce graphs exhibiting the discovered patterns. Backstrom et al. [6] proposed using structural features of communities and individuals and then applying decision-trees to approximate the probability of an individual joining a community. They also tried to identify communities that are more likely to grow over time and predicted the movements between communities based on the same features. Tantipathananandh et al. [7] presented frameworks and algorithms to determine the evolution of communities in social networks. Although they assumed all groups are disjoint and explicitly defined, they tried to identify the notion of a community over all snapshots based on the changes in those groups. They focused mostly on tracking the membership of an individual across all snapshots. Asur et al. [8] analyzed the behavior of interaction graphs by defining critical events and computing them in an efficient manner. They also introduced novel behavioral measures such as stability, sociability, influence and popularity for nodes and an incremental way to calculate them over time. Falkowski et al. [9] analyzed the evolution of communities and studied their stability and fluctuation by defining similarity between them. Moreover, in order to identify persistent communities, they applied standard statistical measures.

3 EVENT-BASED FRAMEWORK TO DETECT EVOLUTIONS IN SOCIAL NETWORKS

Detecting the evolution of communities by monitoring when they form, dissolve, and reform can provide great insight into a dynamic social network. Asur et al. [8] proposed an event-based framework to capture and identify events on communities and individuals. Based on these events, the behavioural patterns of communities over time can be characterized. Although they formulate the critical events for communities, and propose behavioural measures for individuals, the presented events are too restricted to cover all of the changes that a community may experience.

In this paper we present a framework for modeling and detecting community evolution in social networks. The framework allows tracking of events related to communities as well as events related to individual nodes. In order to define events that cover all possible transitions of a community, a new term called the community flag is defined. Base on this concept, we propose event definitions that cover all possible transitions of a community.

Naturally, individuals in a community have mutual common interests and interact with each other around those interests. For example, members gather physically, or virtually, to share an idea or to discuss about a topic. This is exactly what identifies members from non-members. Although this is more sensible for human communities, artificial communities have the same patterns in their structure. Thus one can assume an independent identity for a community based on the interests that members share with each other. We call this identity the c*ommunity flag*, which shows characterization of the community and its members. A community flag is unique and cannot be divided or cloned.

The life cycle of a community is defined as follows. A community forms in a snapshot: *Flag has been raised*. It may be stable from a snapshot to another: *Flag is still there*. It could attract new members or lose some members: Flag is waving. It may incorporate another community: *Dominant flag takes control*.

It may divide into two or more smaller communities, with each new part having its own independence: *The most significant part carries the flag with itself*. Finally it can break apart into pieces while no piece preserves the identity of the community: *Flag has been vanished*. The identity of a community is defined by a significant portion of that community. However, this portion could be different in various contexts. Thus our new event definitions are parametric based on this portion, denoted by k.

In order to use our proposed framework, the social network should first be converted into a time series graph, where the static graph at each time captures the information at that specific moment. Then, based on a community mining algorithm, the communities in each snapshot are obtained independently. Finally the transition of the communities between two consecutive snapshots will be obtained by the critical events defined in the framework.

In the following, $G = (V, E)$ denotes a dynamic social network where V and E are the total individuals and total interactions respectively. A snapshot $S_i = (V_i , E_i)$ of G represents a graph only with the set of individuals and interactions at a particular time interval i. Each snapshot S_i contains k_i communities $C_i = \{C^1_i , C^2_i , \ldots, C^{ki}_i\}$ where the community C^j_i is also a graph denoted by (V^j_i , E^j_i). For each two consecutive snapshots a total of 11 events are defined with seven events involving communities and four other involving individuals in the network.

3.1 EVENTS INVOLVING COMMUNITIES

In order to categorize the changes of communities that evolve over time, we consider seven events including form, dissolve, continue, split, merge, shrink, and reform. These events are based on the relationship between communities and are parameterized based on the portion k.

A community splits if it fractures into more than one community and one of these communities carry the flag of the former community. In the case where it fractures into more that one community but none of these communities carry the flag, a dissolve event is occurred. A community continues if there exists a community in the future that contains all the nodes of the former community. A community may shrink or reform when it loses a portion of its members but this portion is not significant enough to be detected as a split. In the case where new individuals join to the community, the community is marked as reformed, while it shrinks when no one has joined to it. Two or more communities are marked as merge if a major portion of at least one of these communities involve in the merge. Furthermore at any snapshot there may be newly formed community that does not carry the flag of any community at previous time. For two consecutive snapshots S_i and S_{i+1} where C_i and C_{i+1} denoting the set of their communities respectively, the formal definitions of the seven events involving communities are as follows:

***k*-form:** A new cluster C^k_{i+1} is marked as formed if at least k% of its nodes have not been a member of the same community at the previous time. Thus C^k_{i+1} is formed if

$$\nexists\ C_i^j\ such\ that\ \frac{|V_{i+1}^k \cap V_i^j|}{\max(|V_i^j|, |V_{i+1}^k|)} \geq k\%$$

***k*-dissolve:** A community C^k_i is marked as dissolved if at least k% of its nodes will not be a member of the same community in the next snapshot. Thus, the conditions for the dissolved is

$$\nexists\ C_{i+1}^j\ such\ that\ \frac{|V_i^k \cap V_{i+1}^j|}{\max(|V_i^k|, |V_{i+1}^j|)} \geq k\%$$

***k*-continue:** A community C^k_i is marked as continued if there exists a community C^j_{i+1} that contains all the nodes of C^k_i and at least k% of its nodes are belonging to C^k_i. In other words, the two conditions for continue are as follows:

$\exists\ C_{i+1}^j\ such\ that$

1) $V_i^k \subseteq V_{i+1}^j$

2) $\frac{|V_i^k|}{|V_{i+1}^j|} \geq k\%$

***n*-*k*-merge:** A set of communities $\{C^1_i, C^2_i \ldots, C^n_i\}$ are marked as merged if there exists a community C^j_{i+1} in the next snapshot that for any community C^k_i, the following conditions are held:

$\exists\ C_i^m\ such\ that$

$$\frac{|(V_i^k \cup V_i^m) \cap V_{i+1}^j|}{Max(|V_i^k \cup V_i^m|, |V_{i+1}^j|)} \geq k\%$$

$\frac{|V_i^k \cap V_{i+1}^j|}{|V_i^k|} \geq k\%$ The flag of C^k_i has been moved into C^j_{i+1}

$\frac{|V_i^m \cap V_{i+1}^j|}{|V_i^m|} \geq k\%$ The flag of C^m_i has been moved into C^j_{i+1}

Also at least one flag in C^j_{i+1} has to be dominant in order to distinguish this case and the case that a new community has been formed from small pieces of some other communities. Thus the following condition should be held for $\{C^1_i, C^2_i \ldots, C^n_i\}$:

$$\exists\, C_i^m \textit{ such that } \frac{|V_i^m \cap V_{i+1}^j|}{|V_{i+1}^j|} \geq k\%$$

***n*-*k*-split:** A community C^j_i is marked as split if there is a set of communities $\{C^1_{i+1}, C^2_{i+1} \ldots, C^n_{i+1}\}$ in the next snapshot that for any community C^k_{i+1} the following conditions are held:

$\exists\, C_{i+1}^m \textit{ such that}$

$$\frac{|(V_{i+1}^k \cup V_{i+1}^m) \cap V_i^j|}{Max(|V_{i+1}^k \cup V_{i+1}^m|, |V_i^j|)} \geq k\%$$

$\frac{|V_{i+1}^k \cap V_i^j|}{|V_{i+1}^k|} \geq k\%$ There is a potential of raising the flag of C^j_i in C^k_{i+1}

$\frac{|V_{i+1}^m \cap V_i^j|}{|V_{i+1}^m|} \geq k\%$ There is a potential of raising the flag of C^j_i in C^m_{i+1}

Also the flag of C^j_i has to be carried into one of $\{C^1_{i+1}, C^2_{i+1} \ldots, C^n_{i+1}\}$ and it has to be dominant there:

$$\exists\, C_{i+1}^m \textit{ such that } \frac{|V_{i+1}^m \cap V_i^j|}{|V_i^j|} \geq k\%$$

If the above condition is not held, the community C^j_i undergoes the dissolve event.

A community may shrink or reform if it loses a portion of its members but this portion is not significant enough to be detected as a split. In the case where new individuals join to the community, the community is marked as reformed. On the other hand, it shrinks when no one has joined to it.

***k*-shrink:** . A community C^k_i is marked as k-shrink if there exists a community C^j_{i+1} that its set of nodes is a subset of the nodes in community C^k_i and also contains at least k% of the nodes from C^k_i . Thus the community is marked as k-shrink if

$\exists\, C^j_{i+1}$ *such that*

1) $V^j_{i+1} \subseteq V^k_i$
2) $\frac{|V^j_{i+1} \cap V^k_i|}{|V^k_i|} \geq k\%$

***k*-reform:** A community C^k_i is marked as k-reform if there exists a community C^j_{i+1} that at least contains k% of the nodes from C^k_i but its set of nodes is not a subset of the nodes in community C^k_i :

$\exists\, C^j_{i+1}$ *such that*

1) $V^j_{i+1} \not\subset V^k_i$
2) $\frac{|V^j_{i+1} \cap V^k_i|}{|V^k_i|} \geq k\%$

3.2 EVENTS INVOLVING INDIVIDUALS

In order to analyze the behaviour of individuals in communities, four events involving individuals are defined. The taxonomy we use here is the same as Asur et al. [8]. However, unlike [8] we define the join and leave events parameterized based on the portion *k*. For two consecutive snapshots S_i and S_{i+1}, the events involving individuals are defined as follows:

Appear: A node *v* is marked as appeared when it is in the current snapshot but it was not in the previous snapshot *i.e.* $v \notin V_i$ *and* $v \in V_{i+1}$.

Disappear: A node *v* is marked as disappeared when it existed in the previous snapshot but it does not exist in the current snapshot *i.e.* $v \in V_i$ *and* $v \notin V_{i+1}$.

***k*-join:** A node *v* joined to community C^j_{i+1} if it exists in this community at snapshot *i+1* but was not in C^k_i in the previous snapshot where C^j_{i+1} carries the flag of C^k_i. Thus, the conditions for the join event are as follows:

$\exists\ C_i^k\ such\ that$

1) $\frac{\left|V_i^k \cap V_{i+1}^j\right|}{\left|V_i^k\right|} \geq k\%$

1) $v \notin V_i^k$

2) $v \in V_{i+1}^j$

***k*-leave:** A node v left community C^k_i if it existed in this community at snapshot i but it does not exist in C^j_{i+1} in the next snapshot where C^j_{i+1} is sufficiently similar to C^k_i. In other words, the conditions for the leave event are as follows:

$\exists\ C_{i+1}^j\ such\ that$

2) $\frac{\left|V_i^k \cap V_{i+1}^j\right|}{\left|V_i^k\right|} > k\%$

3) $v \in V_i^k$

4) $v \notin V_{i+1}^j$

4 EXPERIMENTS

We have tested our framework on Enron email dataset in order to show the feasibility of the proposed events. To visually track the evolution of communities, we have integrated our code into Meerkat [10]. This tool enables us to preview the graph of each timeframe and have the communities at each timeframe marked with different colours. In fact these colours are the notion of Community Flag and they come from the results of our event-detection formulas.

The Enron email dataset contains the emails between employees of Enron Corporation. The entire dataset includes a period of 15 years and the corresponding graph for the data has over 80,000 nodes and several hundred thousand edges, where nodes are individuals and edges are emails between them.

Without loss of generality, we chose the last year (2001) to reduce the graph size, and only considered people who had sent at least one email per day to filter out non-informative nodes. The resulting graph has almost 250 nodes and 1500 edges. We set the snapshots to be 1 month each and found the communities on each month by a local community mining algorithm with no overlap between communities [4], provided in Meerkat.

In order to evaluate our framework we have also implemented the event-based framework by Asur et al. [8] which is the only framework that has an event based approach similar to our framework. Figure 1 shows the general view of communities in each snapshot. The area of each community in the figure is proportional to the number of its members. By assigning colours to the different flags found by our framework, one can easily make a map between communities through snapshots (the communities without any color are the ones that only exist for one snapshot). For example, appearance of a colour in a snapshot means a community has been formed and similarly disappearance of a colour shows the end of life for that community. Also tracking the community transitions such as reformation, shrinkage, merger, and split are almost possible by looking at Figure 1. However, since there is no notion of a community identity in Asur framework, determining a map from communities in one snapshot to another is impossible. Also there is no way to keep track of a specific community and its transitions over time when using Asur framework. Thus, in order to compare the results found by the two frameworks, the number of events found by Asur and our framework are provided in Table 1 and Table 2 respectively. Using Asur framework, most of the communities are not marked by any event. On the other hand, our framework detects exactly one of the continue, reform, shrink, split, or dissolve events. Thus, the number of communities at each snapshot is the same as the total number of continue, reform, shrink, split, and dissolve events. From Table 2, we can observe that for the Enron dataset, the reform and dissolve events far outnumber the other events. The high number of reform event indicates that most communities do not change greatly between two consecutive snapshots. However, the relatively high number of dissolve event denotes that most of the Enron communities have short life cycles. So we can conclude that in the Enron dataset most of the communities have a short life cycle and do not change drastically.

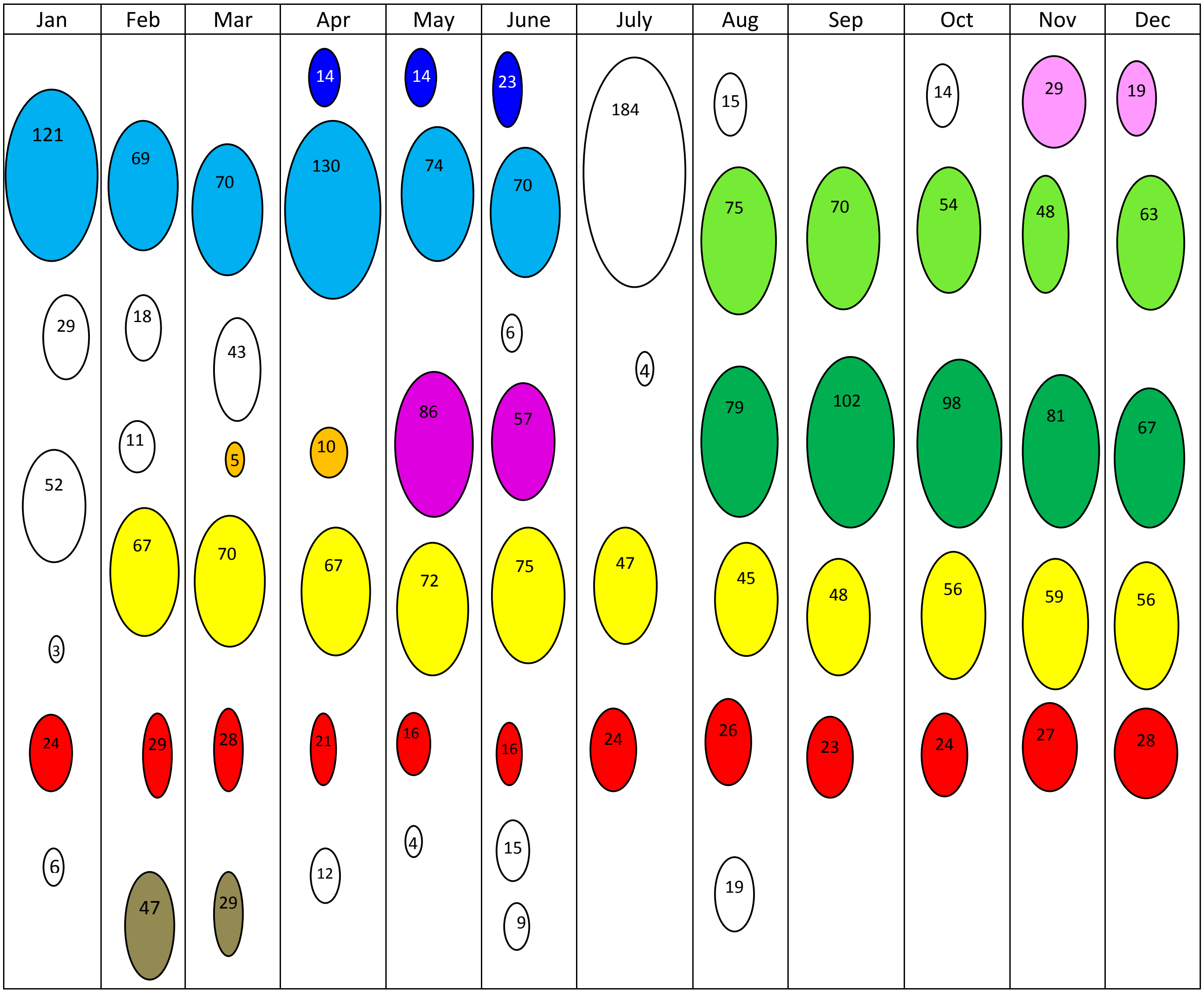

Figure 1. General view of communities in each snapshot

Table 1. Number of events occurred for the Enron using Asur Framework

Month	Communities	Continue	Split	Dissolve	Merge	Appear	Disappear	Form
January	6	0	2	0	1	---	2	---
February	6	0	0	0	0	8	7	0
March	6	0	0	0	1	11	1	0
April	6	0	1	0	0	10	11	0
May	6	0	3	0	0	13	6	0
June	8	0	0	0	2	11	16	0
July	4	0	1	0	0	4	11	0
August	6	0	0	0	1	11	22	0
September	4	0	1	0	0	6	3	0
October	5	0	0	0	0	6	6	0
November	5	0	0	0	0	4	12	0
December	5	---	---	0	---	1	---	0

5 CONCLUSION AND FUTURE WORK

Although dynamic network analysis is required by a wide range of applications, this field of study suffers from lack of comprehensive work. In this paper, we presented an event-base framework to analyze different types of dynamic social networks. Defining the concept of a Community Flag allows us to capture all of the possible events among communities. This includes tracing the formation, continuation and dissolution of communities. Moreover, it detects events involving individuals in the network and tracks their behaviour. Applying our framework on the Enron email dataset, we visualized the Life-Cycle of all communities and the events that occurred in Enron Corporation’s final year. Our results on the

Enron dataset indicate that most of the detected communities in Enron have short life cycle while having stable members during their life.

Most existing community mining algorithms find separated set of communities, where every individual is a member of exactly one community. However, in social networks individuals may belong to different communities which results in highly overlapping and nested communities. One possible future research direction is to analyze the evolutions of overlapping communities based on the proposed events in a dynamic social network. Furthermore in our work, we only consider the events between two consecutive snapshots. However, it is possible to detect events for any number of contiguous timeframes. Considering more than two snapshots at a time would enable us to detect communities that are inactive in a time frame which may reactivate again later on.

Table 2. Number of events occurred for the Enron using our Framework

Month	Communities	Continue	Reform	Shrink	Split	Dissolve	Merge	Appear	Disappear	Form
January	6	0	1	0	1	4	0	---	2	---
February	6	0	4	0	0	2	0	8	7	2
March	6	1	3	0	0	2	1	11	1	1
April	6	0	3	0	1	2	0	10	11	2
May	6	0	3	0	2	1	0	13	6	1
June	8	0	2	0	0	6	1	11	16	2
July	4	0	2	0	0	2	0	4	11	1
August	6	0	3	1	0	2	1	11	22	4
September	4	1	2	0	1	0	0	6	3	0
October	5	1	3	0	0	1	0	6	6	1
November	5	0	5	0	0	0	0	4	12	1
December	5	---	---	---	---	0	---	1	---	0

REFERENCES

[1] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Physical Review, E 69(026113), 2004.

[2] Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In SIAM International Conference on Data Mining, 2005.

[3] Kai Yu, Shipeng Yu, and Volker Tresp. Soft clustering on graphs. In Advances in Neural Information Processing Systems, 2005.

[4] Jiyang Chen, Osmar R. ZaÏane, and Randy Goebel. Local community identification in social networks. In International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Athens, Greece, 2009.

[5] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In KDD ’05: Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining, pages 177–187, 2005.

[6] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: membership, growth, and evolution. In KDD ’06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 44–54, 2006.

[7] Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. In KDD ’07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 717–726, 2007.

[8] Sitaram Asur, Srinivasan Parthasarathy, and Duygu Ucar. An event-based framework for characterizing the evolutionary behavior of interaction graphs. In KDD ’07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 913–921, 2007.

[9] Tanja Falkowski, Anja Barth, and Myra Spiliopoulou. Studying community dynamics with an incremental graph mining algorithm. In Proceedings of the 14th Americas Conference on Information Systems (AMCIS), 2008.

[10] Alberta Ingenuity Centre for Machine Learning (AICML) at the University of Alberta. Meerkat. http://www.machinelearningcentre.ca/meerkat.cfm, February 2010.