Community Dynamics: Event and Role Analysis in Social Network Analysis

Justin Fagnan, Reihaneh Rabbany, Mansoureh Takaffoli,
Eric Verbeek, and Osmar R. Zaïane

Department of Computing Science,
University of Alberta,
Edmonton, Alberta, Canada
`{fagnan,rabbanyk,takaffol,everbeek,zaiane}@ualberta.ca`

Abstract. Social networks are analyzed and mined to find communities, or groupings of interrelated entities. Community mining provides this higher level of structure and offers greater understanding, but networks change over time. Their constituent communities change, and the elements of those communities change over time as well. By performing event analysis, the evolutions of communities are abstracted in order to see structure in the dynamic change over time. This higher level of analysis has a counterpart that deals with the fine grain changes in community members with relation to their communities or the global network. We discuss here an approach to analyzing community evolution events and entity role changes to uncover critical information in dynamic networks.

1 Introduction

Many complex information networks and social networks can be modelled by graphs of interconnected nodes to represent the interaction of individuals or entities with one another. For instance, the graph of co-authorship relationships between researchers, the interaction between posters on an on-line forum, the graph of web pages inter-connected through hyperlinks, and Protein-Protein Interaction (PPI) networks are examples of complex networks. In these networks, understanding the underlying structure and determining the structural properties of the network facilitates the global understanding of the system and benefits applications such as targeted marketing and advertising, influential individuals identification, information diffusion modelling, and much more. One way to gain information about the network is the identification of communities, which are sets of densely connected individuals that are loosely connected to others [18]. There has been a considerable amount of work done to detect communities in static graphs, such as modularity methods [18,8,13], stochastic methods [2,9], and heterogeneous clustering methods [24,3]. For a comparative surveys, see, for example [15,19].

Social networks usually model systems that are evolving over time as the entities change their activities and interactions (authors publish papers with new

co-authors, old pages are deleted while new ones are added to the web, etc.). Furthermore, the communities in these dynamic networks usually have fluctuating members and these communities grow and shrink over time. However, studying these dynamic networks as static graphs discards the temporal information associated with the interaction. In order to explicitly address the dynamic nature of the interactions, the dynamic social network can be modelled by a series of static snapshots. In these models, each snapshot corresponds to a discrete time interval, constituted of the interactions during that specific interval. In some scenarios the size of such a time interval is determined, while for many cases the interval size is arbitrary. However, the size of a time interval has a great impact on the observation found by the dynamic network analysis. Recently, Caceres et al. [6] propose an algorithm to determine the appropriate time interval by finding a balance between minimizing the noise and loss of temporal information. After modeming the dynamic social network with the appropriate time interval, the temporal evolution of the network can be studied. Leskovec et al. [16] study the patterns of growth for large social networks based on the properties of large networks, such as the degree of distribution. The problem of mining patterns of link formations and link predictions in a time evolving graph is proposed in [21]. All the aforementioned studies considered the macroscopic properties on the graph level and overlooked the mesoscopic properties on the level of communities.

Tantipathananandh and Berger [27] formulate the detection of dynamic communities as a graph colouring problem and prove that their algorithm is a small constant factor approximation. Falkowski et al. [10] discover the evolution of communities by applying clustering on a graph formed by all detected communities at different time points. A number of researchers are working on identifying critical events that characterize the evolution of communities in dynamic social networks. Palla et al. [20] identify events by applying Clique Percolation Method (CPM) community mining on a graph formed by the communities discovered at two consecutive snapshots. Then, based on the results of the community mining algorithm, events pertaining to the communities are specified. Asur et al. [4] define critical events between detected communities at two consecutive snapshots which are implement in the form of bit operations. However, these events do not cover all of the transitions that may occur for a particular community. Greene et al. [12] describe a weighted bipartite matching to map communities and then characterized each community by a series of events.

In all of the above work, the communities in each snapshot are mined independently without considering the temporal information and their relationship to communities at previous snapshots. This independent community mining approach is suitable for the social networks with highly dynamic community structures. Another approach is to use incremental community mining, where the community mining at a particular time snapshot is influenced by the communities detected in previous timeframes. Thus, the incremental community mining approach finds a sequence of communities with temporal similarity and hence, is suitable for networks with community structures that are more stable over

time. Incremental community mining approaches, which consider both current and historic information in the mining process, are proposed in [17,28,5,25].

In our previous work, we provide an event-based framework, MODEC (Modelling and Detecting the Evolutions of Communities), to capture the events and transitions of communities and individuals over the entire observation time [26]. However, in the MODEC framework, the reason why a community or an individual experiences a specific event is not addressed. The changes in the role of individuals in a community can have a high influence on the development of the community and can act as triggers to evoke community changes. For example, if the leader of a community leaves, it might cause the remaining community members to become less active or disperse to other communities.

In this paper, we illustrate event analysis and change of individual's roles. Our main contribution is to propose a framework to reveal the relation between the structure of communities and the behavior of individuals in a dynamic scenario. In the following section, we present how the evolutions of communities are abstracted in order to see structure in the dynamic change over time. Then we describe different roles that an individual can play in the whole network and in their communities and also how these roles change after events. Finally through the visualizations in our last section, we demonstrate that analyzing community evolution events and entity role events uncovers critical information in dynamic networks. We implemented the proposed visualizations in our tool Meerkat, a social network analysis system that encompasses our MODEC framework [7].

2 Community Dynamics Modelling

In order to analyze dynamic social networks and study the evolution of their communities and individuals, we propose a two-stage framework, called MODEC, that analyzes the dynamic evolution of communities [26]. Our framework assumes that the communities are independently extracted in each snapshot by an arbitrary community mining algorithm. In the first stage of the framework we employ a one-to-one matching algorithm to match the communities extracted in different snapshots. A meta community, which is a series of similar communities detected by the matching algorithm in different timeframes, is then constructed. In the second stage, we identify a series of significant events and transitions which are used to explain how the communities and individuals of a meta community evolve over time. In this section, we review the MODEC framework and the event definitions proposed to track communities or individuals over time.

We model the dynamic social network as a sequence of graphs $\{G_1, G_2, ..., G_n\}$, where $G_i = (V_i, E_i)$ denotes a graph containing the set of individuals and their interactions at a particular snapshot i. The set $C_i = \{C_i^1, C_i^2, ..., C_i^{n_i}\}$ denoted the n_i communities detected at the ith snapshot, where community $C_i^p \in C_i$ is also a graph represented by (V_i^p, E_i^p). Here, we distinguish between a community and a meta community. A community contains individuals that are densely connected to each other at a particular snapshot, whereas a meta community is a series of similar communities from different snapshots which represents the evolution of its constituent communities over time.
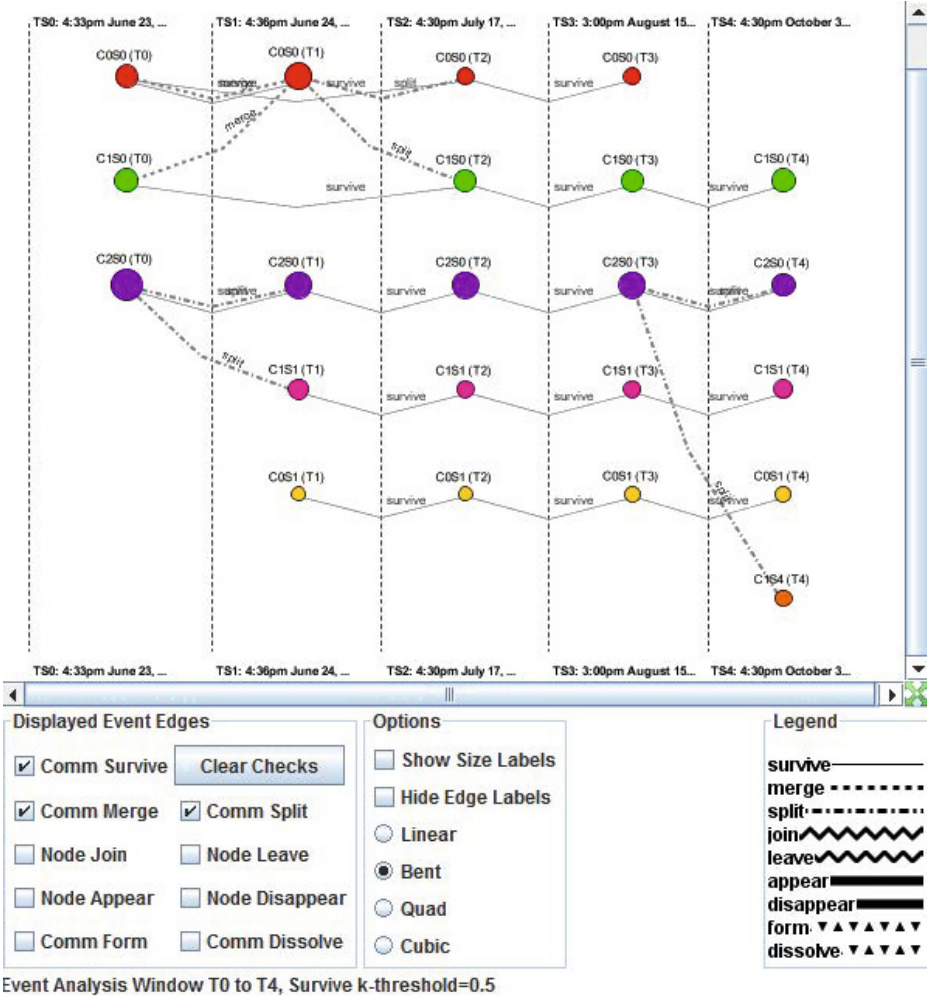
Fig. 1. Visualization of captured MODEC defined events along the time dimension (Visualization by Meerkat http://meerkat.aicml.ca/)

In order to capture the changes that are likely to occur for a community, we consider five events (form, dissolve, survive, split, and merge) and define four transitions (size, compactness, leader, and persistence transition). A community may *split* at a later snapshot if it fractures into multiple communities. It can *survive* if there exists a similar community in a future snapshot. In the case where there is no similar community at a later snapshot, the community *dissolves*. A set of communities may also *merge* together at a later snapshot. Finally, at any snapshot there may be newly *formed* communities which are defined as communities that have no similar community in any previous snapshots. The meta community is then a sequence of *survival* communities ordered by time, from

the timeframe where it first appears to the timeframe where it is last observed. Furthermore, a community may undergo different transitions at the same time. The size transition occurs when the number of nodes of a community increases (i.e. expand) or decreases (i.e. shrink) over time. Moreover, a community compacts or diffuses at a later snapshot if its normalized number of edges increases or decreases respectively. For the case when the number of nodes and edges of a community remains the same, the community persists. Finally, when the most central member of a community shifts from one node to the other, the community experiences leader shift. Figure 1 presents an example of such events and changes in communities while the complete definition of our proposed events and transitions can be found in [26].

The key concept for the detection of the events, meta community, and also transitions is the notion of similarity between communities from different snapshots. The similarity between different communities can be determined using the similarity measures such as Jaccard, correlation-based, and more. However, in this paper, we consider two communities discovered at different snapshots as similar if the percentage of their mutual members exceed a given threshold $k \in [0, 1]$. After selecting the similarity measure, the set of communities extracted by a community mining algorithm at a given snapshot have to be matched to the communities at previous snapshots based on their similarity. This matching is non-trivial, because a community may be similar to several communities at the same time. We use greedy matching, and a weight bipartite matching [26] to match communities at different snapshots.

In order to analyze the behaviour of individuals in communities, we define four events involving individuals (appear, disappear, join, leave) [26]. A node appears at a snapshot when it exists in that snapshot but was not present in the previous snapshots. It may disappear from one snapshot if it exists in that snapshot but will not occur in the next snapshots. A node joins to a community if it exists in that community but did not belong to a community with the same meta community in the previous snapshots. Finally, it leaves a community if it exists in that community but will not belong to a community with the same meta community in the next snapshots.

To capture the behavioural characteristics of the individuals, we define two metrics [26]: the stability metric calculates the tendency of an individual to interact with the same nodes over the observation time; the influence metric determines how one individual influences others to join or leave a community. In the next section we present a more comprehensive list of possible roles that an individual can play both within its community and in the whole network and how these roles are related to the temporal events.

3   Role Analysis and Event Triggers

Walton [29] states that the changing nature of an individual community and its leadership are of central importance to the explanation of community action. Here, we explore different roles associated with individuals in a social networks,

and in the later section we will provide examples of how the changes in these roles affect the events detected for communities.

In our discussion of roles within dynamic community mining, we explore some domain agnostic, generic network roles. These roles should be interpretable in any network, although the names may not apply all of their connotations. The presence of these roles in a given analysis is dependent upon the specific dataset, the community mining algorithm used, the event detection framework applied, and any thresholds applied for the roles.

We define these generic roles across two possible role scopes. The role scopes are either global to the network, or limited to the community the individual is in. Some of the defined roles have both global and community bound versions, while some roles cannot have a community bound version. There is also some notion of connectivity incorporated in the definition of roles; which could be the betweenness metric, node degree, or another appropriate metric. The choice of metric will have an effect on the results of the analysis, but the role definitions do not rely on the selected metric.
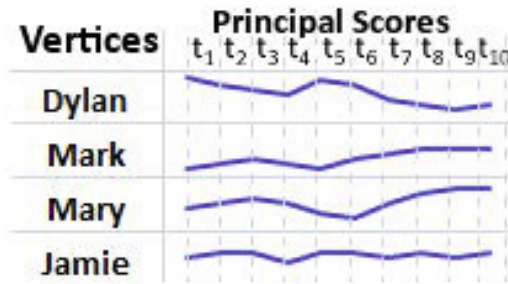
Fig. 2. An token example showing a change of role of individuals vis-à-vis principal role. While Dylan was authoritative in t1, he lost his principal role by t10. Mary on the other hand became authoritative in t10 when she had a low principal score in t5.

We will start with the more well-studied role, the **"outlier role"**. The outlier role reflects individuals who are not a part of any community. Gao et al. [11] and Aggarwal et al. [1] introduced the concept of community outliers and then proposed algorithms to find these community outliers. However, not all mining algorithms allow individuals to be excluded from all communities, so its presence is mining algorithm dependent. This role does not have an associated threshold, nor is there a community role equivalent.

The **"principal role"** (or authority role) indicates an individual who is very central due to high connectivity. *Community leaders* are evaluated within the community, the induced subnetwork, they belong to, whereas *global principals* have high centrality within the entire network. Note that communities with flat intra-community degree distribution for their member nodes will be without a principal, which is acceptable. In the case the distribution is nearly flat, with a

few extremes with only slightly higher degree, we see a need to scale the threshold relative to the community density to prevent assignment of a principal by blind faith. When looking at the principal role in a social domain, we might expect some followers to move with the principal if a principal changes communities (see Figure 2).The concept of principle role in general and more specifically community leader is also used to detect communities in the social network [22,23].

The **"peripheral role"** indicates individuals with the least connectivity. These individuals have the lowest levels of centrality within their community, or centrality within the network for the global version. To identify these individuals that are not well integrated into a community, we apply a threshold to select the individuals with the least connectivity, relative to the network or community density, in the same way as for the principal role. If the threshold is not relative to edge density, sparsely and evenly connected communities could consist entirely of members who are both peripheral and principal. One could argue that this would be acceptable, but our perspective is that such communities are best described as having no member in either of the principal or peripheral roles.

The **"mediator role"** indicates an individual which has high centrality but does not belong solely to one specific community. This can include individuals who happen to belong to multiple communities when the mining algorithm supports overlap and hub detection, or those that are excluded from community membership but are still highly connected, but insufficiently to be a community member.

The community based equivalent of the mediator role is the **"extrovert role"** whom must belong to some community. Unlike the rest of the roles defined so far, the extrovert is detected by comparing their inter-community connections to their intra-community connections. The threshold can be set somewhere slightly above 1.0, since community members tend to have many of their connections within their community. Having slightly more reaching out indicates an extrovert. Depending on the community mining algorithm used, extroverts could have significantly more inter-community edges. If this is the case, the threshold can be made much higher than 1.0, particularly when the mining algorithm detects many communities where conservative algorithms detect very few.

Note that both the mediator and extrovert roles differ from hubs as defined by the HITS algorithm [14], because a hub is defined as connecting multiple authorities, whereas these two roles do not have to connect identified authorities but any set of individuals from more than one community. Likewise, the principal role is not defined as resulting from connections from multiple hubs.

It is desirable to scale the roles' thresholds by the community density as they are applied to each community. This is defined as the number of edges present divided by the number of possible edges density. When this is not done, very sparse or very dense communities may have a preponderance or lack of members filling the roles, or members filling the roles when our intuitive sense of the roles would expect none.

Events usually indicate structural change in a network, excepting the survive event, which can occur with no accompanying changes such joining and leaving
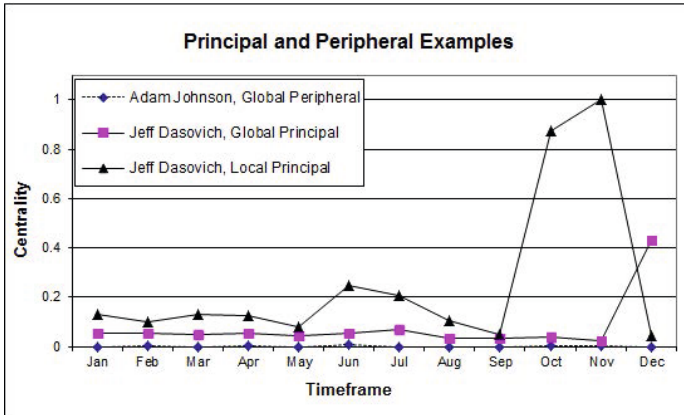
Fig. 3. An example showing how one can monitor changes in roles over time

members (or it can occur despite accompanying changes). For the remaining community events (merge, split, form, dissolve), significant structural changes are almost always indicated, and individual roles are dependent upon network and community structures. When individual events occur (join, leave, appear, disappear), the likelihood of an impact upon individual roles is less likely, or at least dependent upon there being a large number of individual events, or individuals with more central roles coming into competition.

Some clear examples of the effects of events upon role changes may be described. A community authority (or having a principal role) may leave one community, and afterwards they cannot be an authority for the community they just left. If they join another community, they might or might not have a principal role in the new community.They may also have been a global authority, but the changes that led to them leaving their original community might also have caused them to lose connectivity to the greater network, and they may no longer be globally authoritative. The same sort of considerations apply to merge and split events, where the community context has changed an old principal now has competition from others, or has lost connections vital to their authoritative role.

For any given role change, we can identify the community events involved. Our current method of role change attribution consists of associating each significant role change with all events involving that individual directly, as well as events involving the communities that the individual was involved with before and after the role change. This will not capture domain dependent attributions, that is, attributions that require domain knowledge. For example, a change in authority may be associated with a merge of the authority's community with two other communities, resulting in a loss of authority role for that individual. There might also have been splits and joins elsewhere in the network, not involving this individual or these three networks. If the domain contained causal connections between domain specific events and the change in role, and if those domain specific events had no clean relation to the generic network events, the
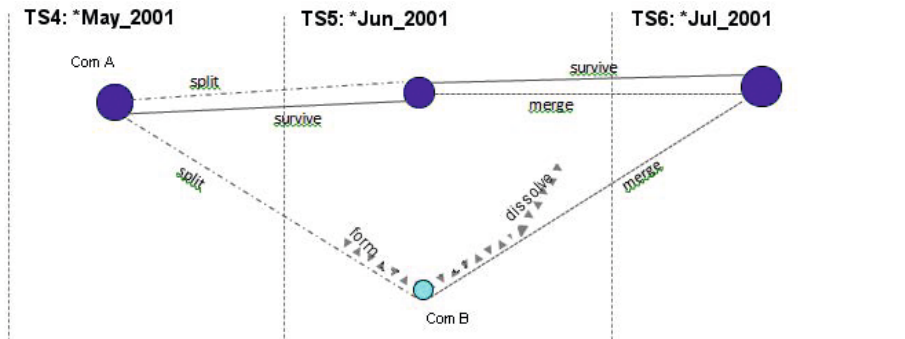
Fig. 4. An example of event visualization

causal connection would go undetected in the network analysis. This is a domain dependent risk, and it is not clear that a generic analysis can ever deal with all such possibilities.

Other than these community-event-individual associations, it might be possible to give deeper attributions of events upon role changes. This would require a much more in depth analysis framework, and would involve questions of ambiguous causality at the generic network level. If tentative causal links are identified, they could then be associated with real domain events that are not represented in SNA models. For example, if an individual changes from being a community authority to being an authority in another community following a leave/join event, there might be a discrete real-world event corresponding with this change, such as the individual being promoted from being director of one department to being a director of another department. It is our position that to understand network dynamics at the domain level, it is necessary to have the generic, quantitative backing from dynamic SNA, including both community event analysis as well as individual role analysis. The generic network methods are intended to offer evidence and act as a modelling lens for domain specific hypotheses and descriptions. These two sources of knowledge may be compared and contrasted with the combination of theoretical analysis and statistical modelling found in experimental scientific disciplines. Without network analytic evidence to support domain hypotheses or measure domain events, full understanding cannot occur.

4   Visualization and Analysis

In this section, we illustrate the practical application of the proposed role analysis on the Enron email dataset. The Enron email dataset contains the emails exchanged between employees of the Enron Corporation. The entire dataset includes a period of 15 years and the corresponding graph for the entire data has over 80,000 nodes and several hundred thousand edges, where nodes are employees and edges are emails between them. We study the year 2001 and consider a

(a) January   (b) February   (c) March
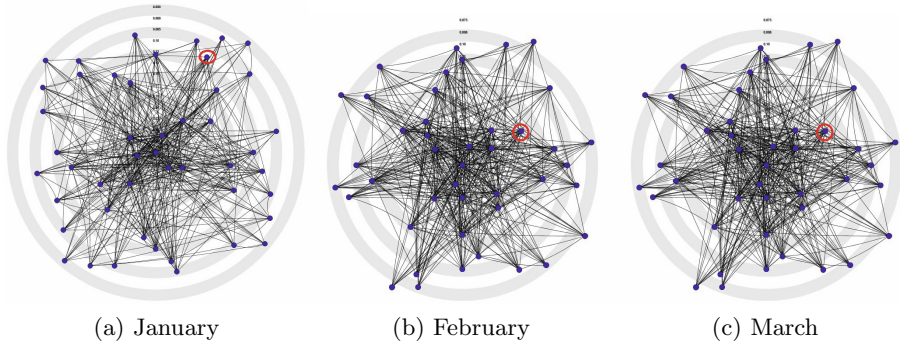
Fig. 5. Changing the role of an individual, Paul Kaufman, from January till March, changing from a peripheral role to principal role (Visualization by Meerkat http://meerkat.aicml.ca/).

total of 285 nodes and 23559 edges, with each month being one snapshot. For each of the 12 snapshots, one graph is constructed with the extracted employees as the nodes and email exchanged between them as the edges. Due to computational efficiency, we apply the weighted local community mining algorithm [8] to produce sets of disjoint communities for each snapshot.

Figure 3 illustrated an example of the changes in roles over time. Here we focus on two individuals. One of them, Adam, is globally peripheral for all timeframes. The other, Jeff, is relatively central both globally and locally to his communities, but displays changes in his principal role value over time. In May, Jeff is a member of the blue community. However, in June, he is a member of the cyan community. Due to his move, the blue community itself splits to the blue and green communities. Then in July when Jeff comes back to the blue community, the communities that had split, merge back into one. So it seems that Jeff acts as a glue between the two, which shows an example of the involvement of a role in the community evolutions.

Community events can be visualized as a second order network. Vertices correspond to communities, with names and colours identifying those that correspond across multiple timeframes. Events themselves are portrayed as edges that connect the communities. We see in Figure 4 a subset of communities from the Enron data. This diagram depicts community events occurring across timeframes May, June, and July only for communities, blue, cyan, and green. The split that occurs here, where blue community splits and cyan community is formed, corresponds with the move of Jeff from community blue to green, and other members following him there. The subsequent merge also relates to Jeff's move.

These event diagrams facilitate the abstraction of communities over time into communities that evolve via community events. Furthermore, it allows users to explore and pick out interesting and critical changes occurring in their data over time. The depicted diagram is a filtered subset for purposes of illustration.

A more refined level of detail could be obtained by a series of concentric circles that visually show the role metric scores for all vertices in a selected community,

as depicted in Figure5. Those vertices closer to the centre of the circles have high metric scores and those on the outskirts have low metric scores. Here, we only provide the circles for blue community in the first three months. Using these concentric circles we can visually observe that for instance an individual name Paul Kaufman (marked by red circles), at the beginning is among peripheral nodes, but as times goes he develop his relationship and become more central within his community.

5   Conclusion

Social networks are dynamic. The change affects the network as a whole, the individual communities, as well as the particular entities in the network. We depicted in this paper the MODEC defined events and transitions that pertain to the evolution of communities and the events associated to individuals moving between these communities. We also described the detectable roles individuals may take globally or within their respective community (principal, peripheral, outlier, mediator, etc.). We explained the relationship between the evolution of communities, the movement of individuals between these communities and changes in the role of those individuals. We further illustrated that changes in the role of individuals in a community have a direct relationship with the development of the community. The role change can act as triggers to evoke community changes. Role modification can affect the dynamics of communities and the events in the communities can alter the role of individuals. Through our visualizations, we demonstrated that analyzing community evolution events and entity role events gives us valuable insights on the dynamics of networks.

References

1. Aggarwal, C.C., Zhao, Y., Yu, P.S.: Outlier detection in graph streams. In: Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, ICDE 2011 (2011)
2. Airoldi, E.M., Blei, D.M., Fienberg, S.E., Xing, E.P.: Mixed membership stochastic blockmodels. Journal of Machine Learning Research 9, 1981–2014 (2008)
3. Alqadah, F., Bhatnagar, R.: A game theoretic framework for heterogenous information network clustering. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2011)
4. Asur, S., Parthasarathy, S., Ucar, D.: An event-based framework for characterizing the evolutionary behavior of interaction graphs. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2007)
5. Aynaud, T., Guillaume, J.-L.: Multi-step community detection and hierarchical time segmentation in evolving networks. In: Proceedings of the Fifth International Workshop on Social Network Mining and Analysis (2011)
6. Caceres, R.S., Berger-Wolf, T., Grossman, R.: Temporal scale of processes in dynamic networks. In: Proceedings of the IEEE ICDM 2011 Workshop on Data Mining in Networks, DaMNet (2011)

7. Chen, J., Fagnan, J., Goebel, R., Rabbany, R., Sangi, F., Takaffoli, M., Verbeek, E., Zaiane, O.R.: Meerkat: Community mining with dynamic social networks. In: IEEE International Conference on Data Mining (ICDM), Sydney, Australia (December 2010)

8. Chen, L., Roy, A.: Event detection from flickr data through wavelet-based spatial analysis. In: Proceedings of the ACM CIKM International Conference on Information and Knowledge Management (2009)

9. Choi, D.S., Wolfe, P.J., Airoldi, E.M.: Stochastic blockmodels with growing number of classes. CoRR, abs/1011.4644 (2010)

10. Falkowski, T., Bartelheimer, J.: Applying social network analysis methods to explore community dynamics. In: Serdult, U., Taube, V. (eds.) Applications of Social Network Analysis 2005, pp. 189–212. Wissenschaftlicher, Berlin (2008)

11. Gao, J., Liang, F., Fan, W., Wang, C., Sun, Y., Han, J.: On community outliers and their efficient detection in information networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2010 (2010)

12. Greene, D., Doyle, D., Cunningham, P.: Tracking the evolution of communities in dynamic social networks. In: Proceeding of the International Conference on Advances in Social Networks Analysis and Mining (2010)

13. Kenley, E.C., Cho, Y.-R.: Entropy-based graph clustering: Application to biological and social networks. In: Proceedings of the 11th IEEE International Conference on Data Mining (2011)

14. Kleinberg, J.M.: Hubs, authorities, and communities. ACM Comput. Surv (December 31, 1999)

15. Lancichinetti, A., Fortunato, S.: Community detection algorithms: a comparative analysis (September 2010)

16. Leskovec, J., Kleinberg, J., Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (2005)

17. Lin, Y.-R., Chi, Y., Zhu, S., Sundaram, H., Tseng, B.L.: Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In: Proceeding of the 17th International Conference on World Wide Web (2008)

18. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Physical Review E 74(3), 036104 (2006)

19. Orman, G.K., Labatut, V., Cherifi, H.: Qualitative comparison of community detection algorithms. In: Cherifi, H., Zain, J.M., El-Qawasmeh, E. (eds.) DICTAP 2011 Part II. CCIS, vol. 167, pp. 265–279. Springer, Heidelberg (2011)

20. Palla, G., Barabasi, A.-L., Vicsek, T.: Quantifying social group evolution. Nature 446(7136), 664–667 (2007)

21. Prakash, A.K.J., Comar, M., Tan, P.-N.: Linkboost: A novel cost-sensitive boosting framework for community-level network link prediction. In: Proceedings of the 11th IEEE International Conference on Data Mining (2011)

22. Rabbany, R., Chen, J., Zaïane, O.R.: Top leaders community detection approach in information networks. In: The Fifth ACM workshop on Social Network Mining and Analysis, SNA-KDD (2010)

23. Shah, D., Zaman, T.: Community detection in networks: The leader-follower algorithm. In: Workshop on Networks Across Disciplines: Theory and Applications (2010)

24. Sun, Y., Yu, Y., Han, J.: Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2009)
25. Takaffoli, M., Rabbany, R., Zaïane, O.R.: Incremental local community identification in dynamic social networks. In: International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2013 (2013)
26. Takaffoli, M., Sangi, F., Fagnan, J., Zaïane, O.R.: Tracking changes in dynamic information networks. In: The International Conference on Computational Aspects of Social Networks (2011)
27. Tantipathananandh, C., Berger-Wolf, T.: Constant-factor approximation algorithms for identifying dynamic communities. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2009)
28. Tantipathananandh, C., Berger-Wolf, T.Y.: Finding communities in dynamic social networks. In: Proceedings of the 11th IEEE International Conference on Data Mining (2011)
29. Walton, J.: Differential patterns of community power structure: An explanation based on interdependence. The Sociological Quarterly 9, 3–18 (1968)