Eurographics / IEEE Symposium on Visualization 2011 (EuroVis 2011)
H. Hauser, H. Pfister, and J. J. van Wijk
(Guest Editors)

Volume 30 (2011), Number 3

Visualizing the Evolution of Community Structures in Dynamic Social Networks

Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh, and Tanya Berger-Wolf

Department of Computer Science, University of Illinois at Chicago, USA

Abstract

Social network analysis is the study of patterns of interaction between social entities. The field is attracting increasing attention from diverse disciplines including sociology, epidemiology, and behavioral ecology. An important sociological phenomenon that draws the attention of analysts is the emergence of communities, which tend to form, evolve, and dissolve gradually over a period of time. Understanding this evolution is crucial to sociologists and domain scientists, and often leads to a better appreciation of the social system under study. Therefore, it is imperative that social network visualization tools support this task. While graph-based representations are well suited for investigating structural properties of networks at a single point in time, they appear to be significantly less useful when used to analyze gradual structural changes over a period of time. In this paper, we present an interactive visualization methodology for dynamic social networks. Our technique focuses on revealing the community structure implied by the evolving interaction patterns between individuals. We apply our visualization to analyze the community structure in the US House of Representatives. We also report on a user study conducted with the participation of behavioral ecologists working with social network datasets that depict interactions between wild animals. Findings from the user study confirm that the visualization was helpful in providing answers to sociological questions as well as eliciting new observations on the social organization of the population under study.

Categories and Subject Descriptors (according to ACM CCS): H.5.m [Information Interfaces and Presentation]: Miscellaneous—

1. Introduction

Social networks are abstract representations of social interactions between individuals. They can be used to capture a wide range of interactions and relationships between social entities. As such, social networks have attracted attention from sociologists [WF94, Fre04], epidemiologists [Het00, And99], ecologists [KLJ09], and political scientists [ZFT*08], among others.

Graphs are considered the de facto tool to computationally model and visually depict social networks. Nodes represent individuals and edges represent interactions between these individuals. The majority of social network models aggregate interactions over the entire observation period without regard to their chronological order, representing the network using one static graph. But thanks to technological developments that have allowed sampling of populations at temporally high resolutions, a new paradigm has been emerging in the social network analysis community in the recent years. Scientists are transitioning from static to dynamic models that incorporate time, and emphasize the importance of the temporal ordering of interactions [Car03]. These new models offer an arguably better approximation to social systems, which more often than not are characterized as being dynamic. They have also enabled scientists to address questions relating to dynamic social processes, such as spread of diseases and rumors.

Visualization has always played a central role in social network analysis [Fre00]. Yet in spite of this, visualization techniques have not kept pace with the recent developments in the field. Most visualization tools used by social network analysts are based predominantly on automatic graph layout algorithms. Although there is a large body of literature on the layout of static graphs, dynamic graphs have received far less attention. Moreover, there appears to be a consensus that dynamic graphs face significant challenges to overcome before becoming a standard tool for visualizing dynamic social networks [BdM06].

Sociologists have long recognized an important phenomenon in social systems, namely, the emergence of communities characterized by groups of individuals who interact frequently [Fre93, NBW06]. The identification of communities often reveals profound knowledge about the social system. Therefore, an important goal of social network visualization is to facilitate the discovery of communities and patterns that govern their evolution. While many tools have been developed to visualize communities in static social networks, the problem remains largely unresolved in dynamic networks.

In this paper we introduce an interactive visualization methodology aimed at dynamic social networks. In contrast to earlier work, our technique can be used to obtain a stable and coherent view of the evolving social structure while enabling the user to inspect the dynamic network at the finest temporal scale. We preprocess the network using a community identification framework to detect persistent communities. Individuals are represented with threads that join together forming tight bundles which in turn represent communities. This depiction enables analysts to uncover individual-community association patterns, as well as observe the evolving community structure as it emerges from the collective interactions of individuals. To illustrate our technique, we use it to visualize voting records of legislators in the US House of Representatives and analyze the resulting community structure. We also describe a user study we conducted with the participation of behavioral ecologists. Results from the user study reveal that the new visualization was helpful in providing answers to sociological questions as well as eliciting new observations. In summary, the contributions of this paper are:

- An interactive visualization methodology for dynamic social networks with a focus on revealing the community structure implied by the observed interactions, and how that structure evolves over time.
- A real-world case study in which the visualization was used and evaluated by a team of behavioral ecologists who regularly work with dynamic social network models of animal behavior.

We believe that a fresh look at the problem of dynamic networks visualization will not only benefit sociologists, but will also be of interest to the wider community of domain scientists who increasingly rely on dynamic social network models. The rest of the paper is divided as follows. We survey the literature in section 2. We introduce our visualization in section 3 along with a discussion of task requirements. The implementation details including the layout algorithm are discussed in section 4. In section 5 we use our visualization to analyze the community structure in the US House of Representatives. We present the user study in section 6. We discuss current limitations and suggest future research directions in section 7, and conclude the paper in section 8.

2. Related work

The majority of literature on social network visualization is based on static graph drawing. A good survey of graph layout algorithms can be found in [BETT99]. Fewer techniques have been proposed to visualize dynamic networks that explicitly contain a notion of change over time. Perhaps the most straightforward method is to produce an animation to depict the evolving structure by drawing the network at different timesteps using standard graph layout algorithms [MMBd05]. Minimizing movement of nodes throughout the animation has been shown to contribute to clarity for some tasks, a principle that has been referred to as maintaining the ‘mental map’ [FE02, PHG07]. Extensions to straightforward animation have been proposed. For example, Kumar et al. combined dynamic graph layout with clustering and edge filtering to reveal more prominent structural changes [KG06]. Although there is some evidence to suggest that animation could help the viewer perceive structural change in the network [MB06], these results were limited to small datasets with few tens of nodes. It remains to be seen whether animated graphs are effective for visualizing large dynamic social networks with hundreds or even thousands of individuals. However, evidence points to limitations of animation when used to illustrate trends in the data [RFF*08]. Moreover, there is usually a conflict between maintaining the mental map and other agreed upon aesthetic principles such as minimizing edge crossing [SP08], potentially resulting in poor local layouts throughout the animation.

An alternative to animation is unrolling the dynamic network in space [UC03]. In these techniques, the network is drawn on a translucent plane at each timestep using a graph layout algorithm. The planes are stacked on top of each other, simulating passage of time. This method however is limited to visualizing few timesteps, as simply stacking up slices indefinitely makes it difficult to see ones that are further back in time. Moreover, this representation leads to substantial redundancy with each individual (or cluster) encoded independently at every snapshot, placing sharp constraints on the scalability of this method. A common extension to stacked graphs is the ‘tube’ or ‘worm’ metaphor which is used to link nodes or clusters across time slices to make them salient [GHW09, ADM*04]. These extensions however typically employ a force-directed layout or some form of a simple 2D euclidean-space clustering to group individuals in every time slice independently. This usually results in suboptimal clustering of nodes, and potentially an incoherent layout across time slices. Moreover, the 3D layout typically used in stacked graphs suffer form artifacts such as occlusion and perspective distortion.

Some techniques have been proposed to address the limitations of stacked graphs. Shen et al. describe a hybrid visualization that combines graphs with 'behavior rings' [SM08]. This technique can be used to study activity patterns within communities. However, one cannot use this technique to observe structural changes that led to the formation of these communities in the first place. Falkowski et al. detect similar groups across different timesteps, drawing a link between instances of the same community [FBS06]. This technique enables the viewer to notice major structural changes, but makes it difficult to observe slow, long-term evolution of communities. Ogawa et al. detect and visualize clusters at each timestep, representing them using Sankey diagrams and drawing edges only when individuals leave their cluster to join a new one [OMB*07]. Our visual representation is semantically similar with its focus on revealing communities. A potential problem with Ogawa et al's approach is their time-insensitive clustering of every snapshot independently, which could introduce abrupt changes to the structure and obscure smooth transitions. We overcome this problem using a community identification framework, which produces a temporally fine-grained but coherent view of the structure. Variations of the Sankey diagram have also been used to visualize multi-dimensional data. The Clustergram shows data points moving between clusters under various clustering conditions, with the X axis representing the number of clusters and the Y axis representing the cluster centroid [Sch02]. The Clustergram however can not be used to show time-varying data. Our visualization adapts this layout to dynamic social networks by employing an adaptive clustering algorithm instead of varying the number of clusters. This frees up the X axis, which we use to depict time.

Our technique is also visually similar to parallel coordinates diagrams [Ins09], some variations of which have been proposed to visualize time varying data with the horizontal axis depicting time. Examples include History Flow which visualizes the evolution of Wikipedia documents to illustrate trends in collaborative editing [VWD04]. Another notably similar visualization is TimeNets, which visualizes genealogical data emphasizing the temporal ordering of marriage, divorce, and consanguine relationships [WKCH10]. Individuals are represented using lines, which converge together to depict marriage and diverge to depict divorce. Our design utilizes the same metaphor and extends it several individuals who converge to depict a formation of a community.

In contrast to earlier dynamic social network visualization tools which attempt to adapt existing clustering and graph drawing algorithms to dynamic networks, our technique employs a community identification framework specifically designed for dynamic social networks to derive a high-level description of the social structure. This framework offers a sociological interpretation of the raw interactions, which tend to be noisy and incoherent, providing a better approximation to the social reality. The visualization exploits those interpretations to offer a coherent, yet temporally fine-grained depiction of the evolving social structure.

3. Design

In this section we first formulate the design goals and describe the decisions we took when designing the visualization. We also briefly describe the community identification process that precedes the visualization. We then introduce our visual metaphor and describe the interactive features of the visualization. We motivate the discussion by outlining the challenges analyst face when working with dynamic social networks and justify our design decisions by addressing the requirements of sociologists and other domain scientists working with social networks.

3.1. Design goals

One phenomenon social scientists are interested in is the emergence of communities in social networks, which manifest as groups of individuals interacting closely and frequently with each other [Fre93,NBW06]. Communities tend to evolve gradually as opposed to forming and disbanding spontaneously [BHKL06]. Therefore, it is important for social network visualizations tools to facilitate the analysis of community structures, and how these structures evolve over time. A related task is identifying patterns that govern community evolution in the social system under study. Thus, our main goal is enabling the user to quickly perceive the various communities present in the network at different times. The design should also facilitate perception of patterns with respect to community evolution.

The passage of time by itself however does not necessarily provide a compelling explanatory narrative. A secondary goal is a design that allows the integration of additional domain data that could help explain how and why the observed social structure evolved in the way it did. We represent the temporal dimension in space, which makes it easier to encode both static and time dependent individual or group attributes. This also makes it possible to represent important events on the timeline. For example, the visualization of the US House of Representatives community structure in section 5 depicts the political party of legislators using color. The user can also see a description of the resolutions being voted on by hovering the mouse pointer over the timeline. In section 6, we describe how we added a second view depicting the movement of groups of animals in the environment alongside the social structure. This coupling turned out to be helpful to behavioral ecologists who were able to integrate their understanding of the habitat and come up with ecologically plausible narratives that explain the observed communities in Grevy's zebras.

3.2. Challenges with dynamic social networks

Dynamic social networks present a number of challenges. A significant number of interactions captured by the network could be purely incidental. For example, a network that derives ties from physical proximity of subjects might

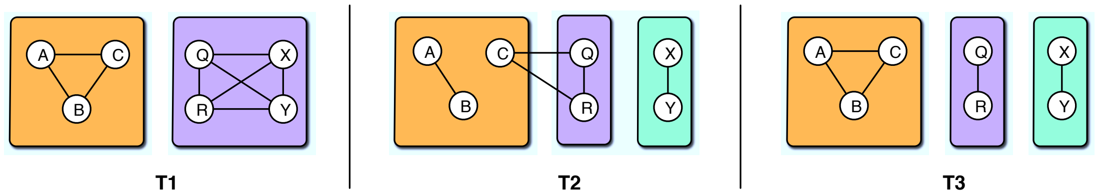

Figure 1: *Simplified example of community identification. In timestep T1 there are two communities: orange and purple. In timestep T2, C remains affiliated with the orange community despite its transient interaction with Q and R. On the other hand, X and Y split from the purple community to form the green community*

contain a large number of unintended interactions between individuals simply because they happen to get close to each other with no intention to interact. While it might be interesting to analyze these incidental interactions in some circumstances, social scientists are often interested in persistent interactions that recur frequently between closely-knit social groups [Fre93]. Identifying and visualizing these groups is not a straightforward task. In a static network, the analyst resorts to applying one of many available graph clustering algorithms. Alternatively, she/he can visualize the graph using a force-directed layout to reveal clusters by automatically moving highly linked nodes close to each other. In dynamic networks on the other hand, applying a clustering algorithm to snapshots of the network at different timesteps results in independent clusters that are hard to link across timesteps. Since most communities tend to evolve gradually over time as opposed to forming and disbanding spontaneously [BHKL06], clustering snapshots of the network independently imposes artificial discretization, making it more difficult to analyze the evolution of communities spanning multiple timesteps.

3.3. Community identification

To filter out transient interactions and deduce a higher-level interpretation of the data, we preprocess the input network using the community identification framework proposed in [TBW09]. The framework infers persistent social groups in the network and establishes a stable labeling for instances of these groups across all timesteps. The dynamic groups are referred to as ‘communities’. A community is a fluid grouping of individuals that persists and evolves over time, allowing new members to join and existing members to leave. An individual can be affiliated with different communities over the observation period, however the individual can only be affiliated with a single community at a particular point in time. Figure 1 illustrates this through an example. We give an overview of the community identification framework below. Complete description can be found in [TBW09].

Given a set of N individuals $\{I_1, I_2, \ldots, I_N\}$ and a sequence of M timesteps, the input to the algorithm comes in the form of an interaction sequence $< H_1, H_2, \ldots, H_M >$ over M timesteps. Each element $H_t = \{g_{t,1}, \ldots, g_{t,k(t)}\}$ is the collection of disjoint groups of individuals that were observed at timestep t. At every timestep, an individual can be in only one of the groups $g_{t,i}$, although not all individuals are required to be present at a given timestep. This makes the framework robust when the sampling fails to locate all individuals. The algorithm detects communities that are latent in the interaction sequence and generate a *community interpretation*. The community interpretation is essentially a table with rows listing timesteps $(T_1 \ldots T_M)$ and columns listing individuals $(I_1 \ldots I_N)$. A cell (t,i) is a single integer value indicating the community ID with which the individual i is affiliated at timestep t.

Calculating a community interpretation is formulated in terms of a combinatorial optimization problem. Costs are assigned to the actions of individuals over time. Three types of costs are defined: switching, visiting, and absence costs, which are treated as input parameters to the algorithm. An individuals who switches his/her community affiliation incurs a switching cost at every switch. Individuals at times do interact with groups outside their communities, in which case they incur a visit cost. Lastly, an absence cost is incurred when an individuals is absent from his/her community. The optimal community interpretation is the one the assigns individuals to communities at a minimum cost observing the condition that an individual can be in one community only at a given time. The three parameters (switching, visiting, and absence costs) allow the algorithm to be used to model a wide range of social systems with varying degrees of dynamism. For example, in some social settings such as informal parties, it is acceptable for strangers who are not members of the hosting community to attend, although attending a party in most cases does not necessarily imply that the individual has has been accepted as a member of that community. In this social setting, the visiting cost could be set to a low value where as the switching cost (that is the cost of joining a new community) could be set to a relatively high value. In other settings such as legislatures, voting against the party’s stated position can have several consequences for a party member. Therefore we expect equally high visiting and switching costs in political social networks of legislators.

One limitation of the current community identification algorithm is the assumption that individuals interact in disjoint groups at every point in time, though these groups can and do change over time. While this limits the type of networks that can be visualized using our tools, the constraint of disjoint interaction groups holds in a wide range of commonly studied social activities such as academic co-authorship relationships.

3.4. Visual metaphor

The layout of our visualization is similar to a timeline chart depicting a time series. We use the X axis to represent time, while the Y axis is used to position individuals in their appropriate communities. This allows us to use the vertical space in the chart to group individuals rather than drawing edges between them. The inspiration for our visual metaphor was a Randall Munroe chart depicting the proximity of movie characters throughout the narrative [Mun]. Social scientists often use the term 'social fiber' when referring to the cohesiveness and persistence of relationship between individuals in a particular social setting. Expanding on this, we can think of individuals as thin threads that are intricately woven together to form the social fiber. As ties between individuals strengthen, their threads join together to form thicker bundles, which in turn represent communities. We use a single thread to visually plot the community affiliation of an individual over time. When a group of individuals form a community, their threads come together forming a bundle, which we use to denote that community. When a community starts disbanding, its threads start to separate join other communities. This phenomenon is evident in the visualization in the form of gradual thinning of communities over time. Figure 2 visualizes the dataset presented earlier in Figure 1.

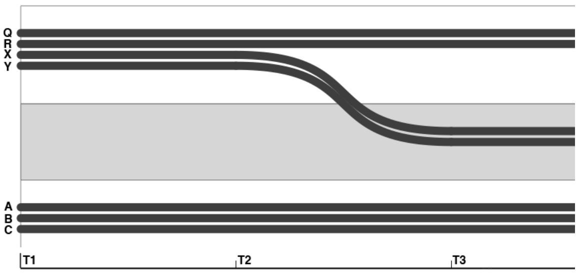

Figure 2: *Visualization of the example presented in Figure 1. There are two communities at timesteps T1 and T2. At T3 two individuals, X and Y, leave their community to form a new community at T3.*

3.5. Interaction

The default view of the visualization shows an overview of community structure throughout the entire observation period. Figure 3 shows the community structure of the 2010 US House of Representatives between March 3 and 18, 2010. The visualization can be manipulated using conventional scaling and translation. Zooming in allows one to see the composition of communities in detail. It also enables the analyst to focus on a shorter period of time. Figure 4 shows a detailed segment of the 2010 House of Representatives. The diagram can also be scaled none-uniformly in the time dimension. This is equivalent to adjusting the aspect ratio in a time-series chart to make trends more obvious.

Another way to inspect the network is by looking at the community affiliation patterns of individuals. We allow the user to select an individual by brushing its corresponding thread. This causes other threads to be drawn translucently, emphasizing the selected individual. Although other threads fade out, community bundles remain visible, allowing one to see the affiliation pattern of the focal individual in relation to the community structure without loosing context. Figure 5 illustrates this feature. One can identify the various communities an individual chose to affiliate with by tracing its thread. A straight, horizontal thread indicates that the individual remains affiliated with its current community. When an individual leaves its community and joins another one, its thread starts to separate from its community to join the new one. Selection of multiple individuals is possible by holding down the shift key and clicking more individuals, or by drawing a selection box. This feature can be used to compare affiliation patterns of multiple individuals.

4. Implementation

After preprocessing the network with the community identification framework (Section 3.3), the layout algorithm calculates a vertical position for each community, as well as vertical offsets for the threads of individuals within communities. The ordering of communities is based on an influence factor, which we define as the cumulative number of individuals affiliated with the community over all timesteps (individuals may appear more than once in this count). The communities are then assigned to rows. A single row can contain multiple communities as long as they do not overlap in time. We loop through all communities in decreasing order of influence and assign each community to the first available row, taking care to not overlap with existing communities. New rows are added as needed to accommodate all communities. This results in large and active communities being placed closer to the top of the visualization. The layout algorithm also orders the individual threads inside each community. To allow the viewer to make correlation between community structure and other domain variables, we have chosen to order the threads based on individual attributes. For example, in the visualization of congressional voting record, individuals were grouped according to their party (Republicans, Democrats, then independents), and within each party individuals were ordered alphabetically. On the other hand, in the visualization of animal communities used in the user

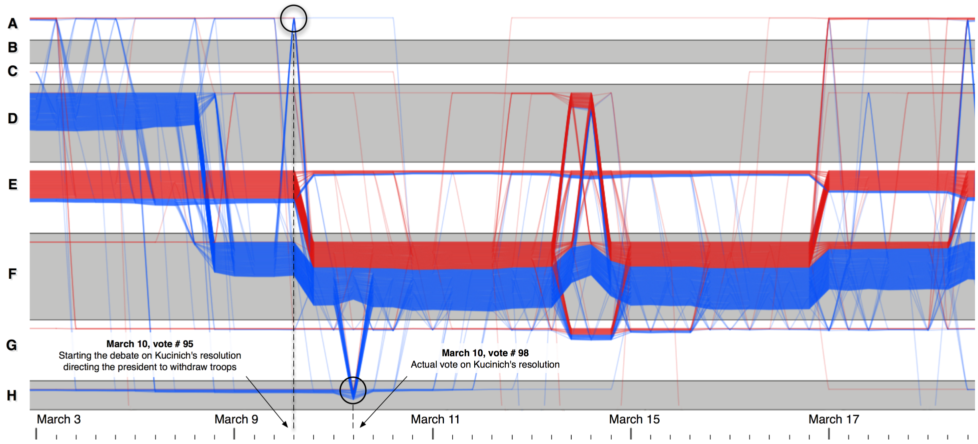

Figure 3: *Portion of US House of Representatives roll call votes covering the period of March 3 through 18, 2010. Red threads represent Republicans while the blue threads represent Democratic representatives. Hovering over the timeline with the mouse brings up a description of the proposed resolution at the highlighted timestep.*

study, we ordered individual animals based on the code assigned by the ecologists during data collection. This allowed the ecologists to find an individual quickly and relate their previous knowledge of that individual's behavior. While the current layout algorithm does not optimize the view for minimum thread crossing, earlier experiments have shown that optimizing the layout for minimum thread transition often resulted in poor local layouts where some areas suffer form excessive crossings. Instead of implementing optimizations, we included a feature that allows the viewer to interactively manipulate the vertical position of communities. This accomplished by holding down the control key on the keyboard while dragging and dropping the desired row, causing it to be and moved vertically to the new position. This feature turned up to be useful while inspecting the interactions happening during a short period of time. In these cases, the viewer can interactively arrange the communities to obtain a good local layout that minimizes thread crossings.

During our experiments, we varied the size of the individual threads based on the size of the dataset. We used thin, translucent threads to depict legislators in the congressional voting records dataset in order to depict all of the 434 members of the house. In the visualization of animal communities, we increased the thickness of the threads to make them salient, as the ecologists were interested in the behavior of individual animals and relating it to the overall community structure. The spacing between the lines was also increased to make the individual threads more clear.

5. Example: analyzing legislature voting record

In this section we show how our visualization can be used to analyze a dynamic social network that consists of 'hidden' communities embedded in larger groups. In a typical legislative body such as the US House of Representatives, legislators cast their votes (supporting, opposing, or not voting) on proposed resolutions. This activity can be regarded as a form of social interaction and can be represented with a dynamic social network. Legislators who cast the same vote on a particular issue can be thought to be interacting with each other. Their collective interaction results in the emergence of various communities representing different political or opinion groups in the legislature. Identifying these groups from the voting record is a daunting task. This is because, taken together, the voting record shows only three groups (Yes, No, and not voting). The various opinion groups are embedded in this larger context. Moreover, observing how these opinion groups evolve yields insights into the dynamics of the legislature.

The input dataset was obtained from the clerk office of the House of Representatives [Cle]. The dataset contains 434 individuals who cast 500 roll call votes on the house floor from January 13 to July 30, 2010 . For each vote, representatives were grouped into 3 groups: Yes and No groups, and those who voted present or were absent during the vote. In each of the three groups, representatives were considered to be interacting with each other forming 3 disjoint cliques. Each vote was considered to be happening in a distinct timestep. Representatives were color coded according to their political party. Red threads represent Republicans while blue threads represent Democratic representatives. Figure 3 shows an overview of the resulting community structure between March 3 and 18, 2010.

A look at the 2010 House of Representatives in Figure 3 reveals two major communities corresponding to members of Democratic (blue) and Republican (red) party. A small group of conservative Democrats who vote more closely with Republicans can be seen in community E. Community H on the other hand contains progressive Democrats who have a distinct voting record from more establishment-leaning Democrats (communities D and F). At times, Democrats and Republicans converge into a single community, reflecting agreement between the two major parties. At other times, the community structure reveals more diverse opinion groups within the Democratic party. An example of this happens in the period between March 9 and 11, leading up to a resolution proposed by Democratic representative *Kucinich* of Ohio who is known for his liberal views. The proposed resolution directs the president to withdraw American troops from Afghanistan by the end of 2010. Although the resolution was seen as largely symbolic since it was almost guaranteed to fail, the events leading up to it were seen as a measure of the house of representatives' view on president Obama's Afghanistan troop surge announced in December, 2009 [Hul10]. The vote on whether to consider Kucinich's resolution for debate comes on March 10 in roll-call #95. The majority of Democrats voted in favor of starting the debate, and can be seen in community F. All Republicans alongside some conservative Democrats were opposed to even debating the resolution and voted No, forming an opposing opinion group that can be seen in community E. However, one can also see a relatively small splinter of Democrats who branched off from community E to form a short lived community in row A (top circle in Figure 3). This community depicts a second opposing opinion group comprised of moderate and centrist Democrats who normally vote with the party's establishment but thought the Kucinich resolution was too radical to even qualify for debate, and thus voted against considering it. The opinion group in community A was very short-lived, and formed only in opposition to the controversial resolution. Thus it was distinct from the more persistent opposing opinion group in community E. Both parties converge to a single community F on votes #96 and #97 for a ceremonial recognition and an act to prevent fraudulent census materials, respectively. The actual vote on whether to approve the Kucinich resolution comes in roll-call #98. At this point, Republicans joined the majority of Democrats in community F to vote against the resolution. However, this time a group liberal Democrats diverged from community F to join progressives in community H (bottom circle), including representative Kucinich (Figure 5) and voted in favor of the troops withdrawal resolution.

Figure 4 on the other hand shows a detailed look at the period between March 11 and 15. The first vote was a ceremonial gesture recognizing Greece independence anniversary. At this point, the two parties form one bundle. One March 12, the combined community split again into 4 opinion groups only to converge again on March 15 for another ceremonial vote.

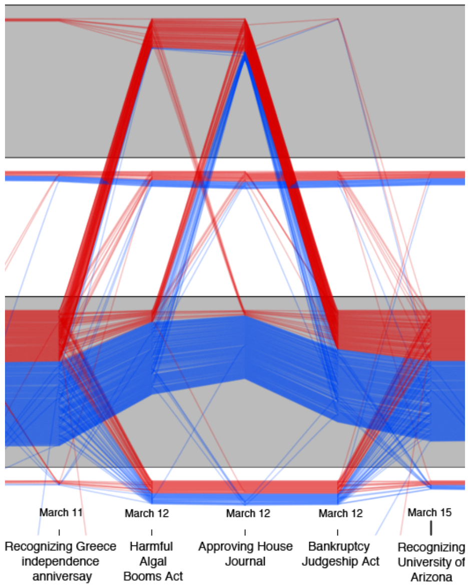

Figure 4: *Detailed view of the community structure of the House of Representatives between March 11 and 15, 2010.*

6. Case study: visualizing animal community structure

To evaluate the usefulness of our visualization to domain scientists, we conducted a user study with the participation of a team of behavioral ecologists studying the social behavior of two endangered species: Grevy's zebra *(Equus grevyi)* in Laikipia, Kenya and Onagers *(Equus hemionus)* in the Kutch desert, Gujarat, India. The two species exhibit distinctively different social behavior due to ecological and phenotypic variation. The datasets consisted of geographically tagged sightings of animals. Census loops were driven ap-

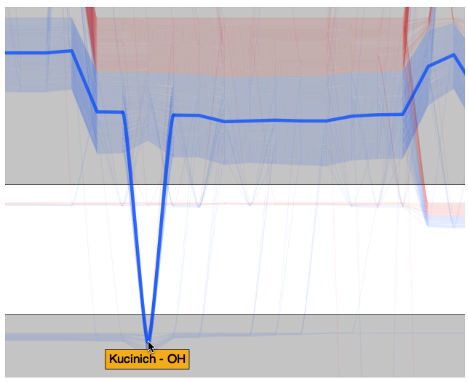

Figure 5: *Selection of a single thread enables one to see the corresponding individual's affiliation over time within the context of the community structure.*

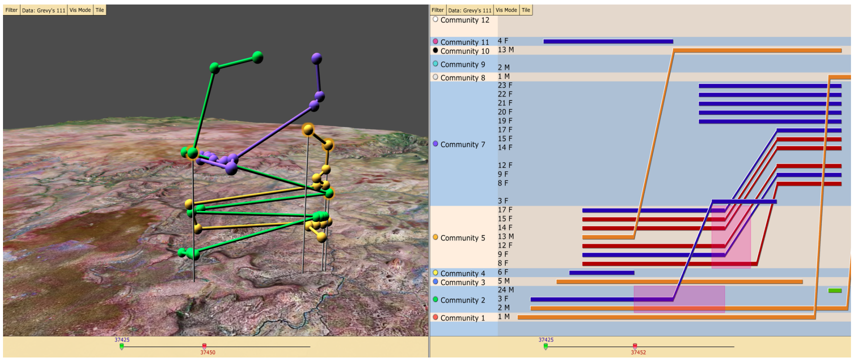

Figure 6: *The visualization environment used in the user study showing the Grevy's zebra dataset. The left view is the space-time cube showing movement of zebra communities in space-time, while the view on the right shows the community structure timeline. Color pins in the community structure timeline (to the left of community labels) associate a unique color with every community for cross-reference with the space-time cube. Individuals were labeled with an ID number followed by their sex code (F/M). The size of individual lines have been increased to make them more clear.*

proximately 5 times a week to sample the population. 35 Grevy's zebra individuals were observed over a period of 3 months in 2002, and 41 Onagers were observed over 5 months in 2003. Individuals were uniquely identified from their striping patterns and other unique features, and their location and time of sighting was recorded. A dynamic social network was constructed from these observations. Animals who were in close proximity were considered to be interacting with each other at the time of sighting.

The visualization environment consisted of two views (Figure 6). The view on the right shows the community structure. Individuals were color coded according to their reproductive state (Bachelors, Stallions, Lactating and Non-lactating females colored with green, orange, red, and navy blue, respectively). To integrate additional domain data, a second view employed a space-time cube [Kar03], showing the movement of communities in space and time over an aerial photograph of the landscape. A cross highlighting feature was added to allow points/regions selected in one view to be automatically highlighted in the other. We hypothesize that this integration of movement trajectories for communities helps ecologists in asking and finding answers to questions about to the environmental and ecological factors that influence the social structure.

Four researchers studying the social behavior of Grevy's zebras and Onagers used the environment to explore their datasets. The participants were not asked to perform any specific task. Rather, they were encouraged to freely interact with the visualization to explore the datasets and discuss observations among themselves. The study lasted for approximately two hours, and was video/audio taped. Participants were accustomed to looking at a variant of stacked-graphs illustrated in [TBW09] to depict the social behavior of animals in their datasets, and thus were familiar with the notion of a dynamic community. The new visualization provided them with an alternative perspective that encoded the community affiliation of individuals along with their reproductive status. Although this encoding of domain attribute could be achieved with stacked graphs, the community structure timeline allows the viewer to easily correlate structural changes in the network with these attributes. For example, during the user study, the ecologists were able to quickly find individuals who shift their community affiliation by visually searching for crooked lines. They were then able to correlate patterns of affiliation shifts with the reproductive status of individuals. One participant had the following comments:

> "We are looking at a different projection that shows the individual by [reproductive] state moving in and out of the community." "This says what the males, lactating, and non-lactating females are doing. It is a very powerful analysis to see when the switch happens."

One problem with stacked graphs is their inherently unstable layout causing nodes to shift their position over time even in the face of small structural changes, and making it difficult to follow the group affiliation of individuals across time. The community structure timeline provides a more stable layout, with individuals shifting their vertical position only when their community affiliation changes. There is evidence that participants found the community structure timeline easy to interpret, suggesting that the new visualization is more intuitive to domain scientists than stacked graphs:

> "Once we know this is a community, to see the individuals aligned very consistently like this in almost what looks like a British subway map with simple angles is very useful." "This is a very clean depiction of community membership".

The coupling of the space-time cube and the community structure timeline enabled new observations about the data that would have been difficult to perceive otherwise. For instance, two Grevy's zebra communities attracted a participant's attention after looking at their movement trajectories in the space-time cube. These two communities seemed to oscillate periodically between two sites in the landscape, which the ecologists identified as a grazing and a drinking site. Although the two communities were sharing the two sites and moving between them periodically, there movement was one day apart. Furthermore, looking at the community timeline confirms that these communities remained separate for some time before merging together to form a third community.

> "Even though these communities lineup in space very close to each others, they are off by a day. This is very robust at telling us they are avoiding each others."

The user study provided evidence that combining the community timeline with additional domain data (in this case, group movement) helped scientists in generating and testing hypothesis that attempt to explain how and why the observed social structure emerged and evolved in the way it did. One participant had the following comments on the visualization:

> "[This visualization] finally put time and space together. This allows us to understand the physical decision making that lead to the shaping of communities. The dynamic community analysis gave us a better picture for understanding zebra dynamics. The space will give us even a better picture of that temporality."

We conclude from the user study that the community structure timeline was effective at conveying a dynamic picture of the changing social structure to the participants. There is evidence that the visual encoding was easily perceived, suggesting that it is more effective than stacked graphs. Moreover, the integration of additional domain data (namely, the space-time cube) was helpful in illuminating additional observations, and in shedding a light on external factors not necessarily captured by the social interaction model, but nevertheless have some influence social behavior.

7. Limitations and future work

The user study suggests that the visualization could be useful for social network analysts and other domain scientists, however there are some important limitations that need to be addressed. One potential issue is scalability. While the visualization appears to scale well with the number of individuals, scalability seems to be influenced by the relative stability of community membership. Our experience shows that the visualization can cope well with about 20-30 communities at a given time with moderate membership changes. However in social systems where a large portion of individuals switch their community affiliation often, the visualization will suffer from excessive thread crossing, resulting in a cluttered view. One potential solution in these limiting the number of switching threads (when individuals move between communities) to a subset of individuals. This subset could be selected based on a degree-of-interest criteria. Alternatively, the user could interactively specify this subset. For datasets with a large number of communities (over 30 communities), a multi-level hierarchical community timeline (based on hierarchical parallel coordinates [FWR99]) could be used with the lowest level showing individuals and the highest level showing collectives of communities. Another limitation stems from the one-community-at-a-time affiliation model. In complex social system such as human societies, an individual is more likely to affiliate with multiple social groups simultaneously. To depict this phenomenon, the individual's thread could split into branches that join different communities.

The feedback we received during the user study was positive, however there is a need for further evaluation and assessment. We intend to continue working with domain scientists who use dynamic social network analysis in their research to refine the visualization and carry out further evaluation. In particular, we intend to follow up with a long-term study similar to [PS08] to evaluate how domain scientists use the visualization over an extended period of time. We also plan to perform a controlled perceptual evaluation of the community timeline and compare it against stacked and animated graphs using a task-oriented study.

8. Conclusions

Visualization has been a central theme in the field of social network analysis since its inception. While graphs have been viewed as the de facto method for visualizing social networks, they have inherent limitations when applied to dynamic networks. We introduced a visualization methodology for dynamic social networks in an attempt to provide a new perspective on this problem. The visualization depicts the evolution of communities using a simple visual representation which employs a commonly used visual metaphor. Our technique enables the exploration of the network at the finest-temporal resolution, while maintaining good scalability along the number of individuals and timesteps in the network. We applied our method to visualize political social networks, yielding interesting observations into the dynamic of the US House of Representatives. A user study also provides evidence that the technique is useful to other domain scientists, especially when the visualization is enriched with additional domain data.

Acknowledgements

This publication is based on work supported in part by the National Science Foundation (NSF) awards CNS-0821121 and OCI-0943559. We would also like to thank Dan Rubenstein, Siva Sundaresan, and Ilya Fischhoff for access to their data and for their help in evaluating the visualization.

References

[ADM*04] AHMED A., DWYER T., MURRAY C., SONG L., XIN WU Y.: Wilmascope graph visualisation. *IEEE Symp. on Information Visualization (InfoVis 2004)* (2004), r4. 2

[And99] ANDERSON R. M.: Epidemic models and social networks. *Math. Scientists 42* (1999), 128–147. 1

[BdM06] BENDER-DEMOLL S., MCFARLAND D. A.: The art and science of dynamic network visualization. *Journal of Social Structure 1*, 2 (2006). 2

[BETT99] BATTISTA G. D., EADES P., TAMASSIA R., TOLLIS I. G.: *Graph drawing: algorithms for the visualization of graphs*. Prentice Hall, 1999. 2

[BHKL06] BACKSTROM L., HUTTENLOCHER D., KLEINBERG J., LAN X.: Group formation in large social networks: Membership, growth, and evolution. In *Proc. SIGKDD '06: 12th Int'l Conf. on Knowledge Discovery and Data Mining* (2006). 3, 4

[Car03] CARLEY K. M.: Dynamic network analysis. In *National Research Council workshop on Social Network Modeling and Analysis* (2003), Breiger R., Carley K. M., (Eds.), The National Academies Press. 1

[Cle] House of representatives, office of the clerk. http://clerk.house.gov/. Retrieved 11-27-2010. 6

[FBS06] FALKOWSKI T., BARTELHEIMER J., SPILIOPOULOU M.: Mining and visualizing the evolution of subgroups in social networks. In *Proc. of 2006 Int'l Conf. on Web Intelligence* (2006), IEEE. 3

[FE02] FRIEDRICH C., EADES P.: Graph drawing in motion. *Journal of Graph Algorithms and Applications 6*, 3 (2002), 353–370. 2

[Fre93] FREEMAN L. C.: On the sociological concept of "group": a empirical test of two models. *American Journal of Sociology 98* (1993), 152–166. 2, 3, 4

[Fre00] FREEMAN L. C.: Visualizing social networks. *Journal of Social Structure 1*, 4 (2000). 1

[Fre04] FREEMAN L. C.: *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004. 1

[FWR99] FUA W.-H., WARD M. O., RUNDENSTEINER E. A.: Hierarchical parallel coordinates for exploration of large datasets. In *Proc. Visualization '99* (1999). 9

[GHW09] GROH G., HANSTEIN H., WÖRNDL W.: Interactively visualizing dynamic social networks with dyson. In *Proc. of the IUI'09 Workshop on Visual Interfaces to the Social and the Semantic Web* (2009). 2

[Het00] HETHCOTE H. W.: The mathematics of infectious diseases. *SIAM Review 42* (2000), 599–643. 1

[Hul10] HULSE C.: House rejects plan to leave afghanistan by year's end. *New York Times March*, 10 (2010). 7

[Ins09] INSELBERG A.: *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, 2009. 3

[Kar03] KARRAK M.: The space-time cube revisited from a geovisualization perspective. In *Proc. of the 21st Int'l Cartographic Conference* (2003). 8

[KG06] KUMAR G., GARLAND M.: Visual exploration of complex time-varying graphs. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 805–812. 2

[KLJ09] KRAUSE J., LUSSEAU D., JAMES R.: Animal social networks: an introduction. *Behavioral Ecology and Sociobiology 63*, 7 (2009), 967–973. 1

[MB06] MCGRATH C., BLYTHE J.: Do you see what i want you to see? the effects of motion and spatial layout on viewers' perceptions of graph structure. *Journal of Social Structure 5*, 2 (2006). 2

[MMBd05] MOODY J., MCFARLAND D., BENDER-DEMOLL S.: Dynamic network visualization. *American Journal of Sociology 110*, 4 (2005), 1206–1241. 2

[Mun] MUNROE R.: Movie narrative charts. http://xkcd.com/657/. Retrieved 11-27-2010. 5

[NBW06] NEWMAN M., BARABÁSI A.-L., , WATTS D. J.: *The Structure and Dynamics of Networks*. Princeton University Press, 2006. 2, 3

[OMB*07] OGAWA M., MA K.-L., BIRD C., DEVANBU P., GOURLEY A.: Visualizing social interaction in open source software projects. In *Proc. 6th Int'l Asian-Pacific Symp. on Visualization* (2007), pp. 25–32. 3

[PHG07] PURCHASE H. C., HOGGAN E., GÖRG C.: How important is the "mental map"? an empirical investigation of dynamic graph layout algorithm. In *Proc. GD '14: 14th Int'l Symp. on Graph Drawing* (2007), vol. 4372 of *Lecture Notes in Computer Science*, Springer. 2

[PS08] PERER A., SHNEIDERMAN B.: Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. In *Proc. of the 26th annual SIGCHI conference on Human factors in computing systems* (2008), ACM, pp. 265–274. 9

[RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1325–1332. 2

[Sch02] SCHONLAU M.: The clustergram: A graph for visualizing hierarchical and non-hierarchical cluster analyses. *The Stata Journal 3* (2002), 316–327. 3

[SM08] SHEN Z., MA K.-L.: Mobivis: A visualization system for exploring mobile data. In *Proc. PacificVis '08* (2008), IEEE, pp. 175–182. 3

[SP08] SAFFREY P., PURCHASE H.: The "mental map" versus "static aesthetic" compromise in dynamic graphs: a user study. In *Proc. AUIC '08: 9th conference on Australasian user interface* (2008), vol. 76. 2

[TBW09] TANTIPATHANANANDH C., BERGER-WOLF T.: Constant-factor approximation algorithms for identifying dynamic communities. In *Proc. SIGKDD '09: 15th Int'l. Conf. on Knowledge Discovery and Data mining* (2009), ACM, pp. 827–836. 4, 8

[UC03] U. B., CORMAN S. R.: Visual unrolling of network evolution and the analysis of dynamic discourse. *Information Visualization 2* (2003), 40–50. 2

[VWD04] VIÉGAS F. B., WATTENBERG M., DAVE K.: Studying cooperation and conflict between authors with history flow visualizations. In *Proc. of the SIGCHI conference on Human factors in computing systems* (2004), CHI '04, ACM, pp. 575–582. 3

[WF94] WASSERMAN S., FAUST K.: *Social Network Analysis: Methods and Applications (Structural Analysis in Social Science)*. Cambridge University Press, 1994. 1

[WKCH10] WOOK KIM N., CARD S. K., HEER J.: Tracing genealogical data with timenets. In *Proc. AVI '10: Int'l Conf. on Advanced Visual Interfaces* (2010), ACM, pp. 241–248. 3

[ZFT*08] ZHANG Y., FRIEND A. J., TRAUD A. L., PORTER M. A., FOWLER J. H., MUCHAM P. J.: Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications 387*, 7 (2008), 1705–1712. 1