PolicyFlow: Interpreting Policy Diffusion in Context

Yongsu Ahn
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, Pennsylvania 15260
yongsu.ahn@pitt.edu

Yu-Ru Lin
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, Pennsylvania 15260
yurulin@pitt.edu

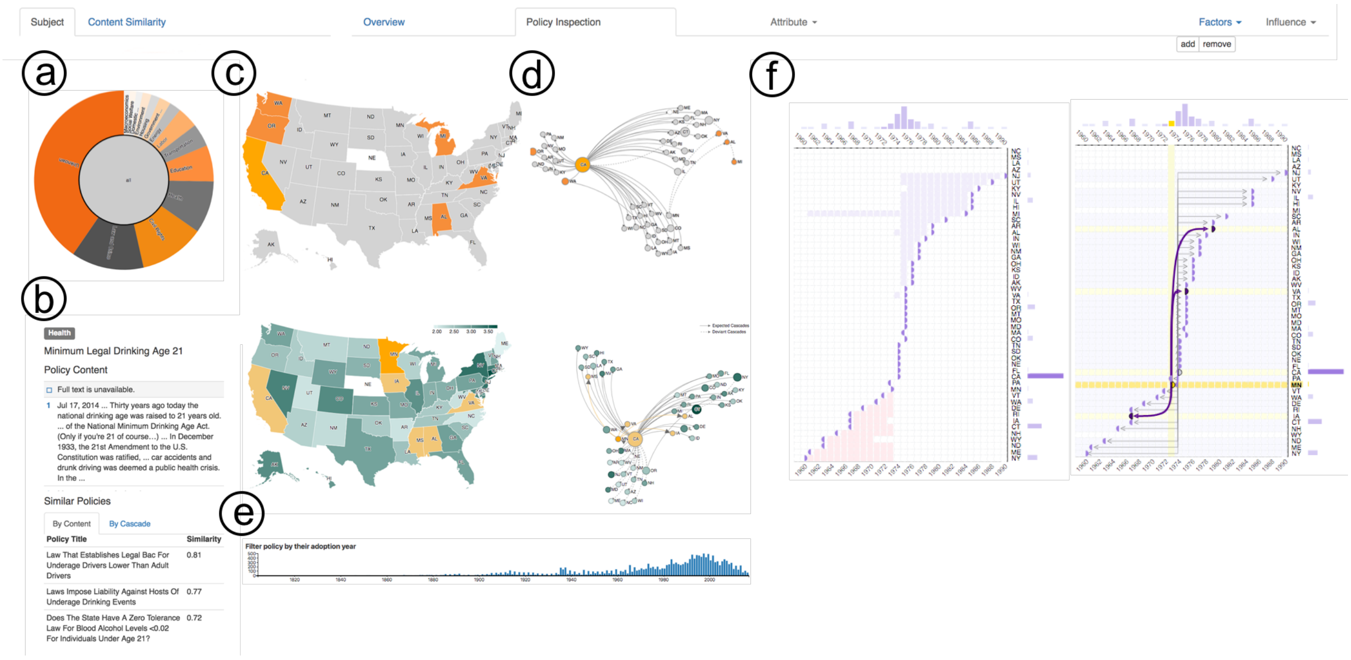

Figure 1: PolicyFlow is a visual interactive system for exploring the time-evolving patterns of policy adoption. The entire policy dataset can be browsed and filtered by content via the Subject Browser and Content Browser (a), by region via the Map View (c), and by time via the Timeline View (e). Details of the selected subset can be viewed in the Policy Detailed View (b). The Network View (d) shows underlying policy diffusion network of the selected subset, which is computed online using a network inference algorithm. The contextual and structural information of the diffusion network, such as their geographical distribution, the corresponding state-attributes (socio-economic covariates), and the network connectivity, can be viewed through the coordinated Map View (c) and Network View (d). The Policy Inspection View enables further inspecting a particular policy's adoption sequence over time and space, its relationship with the socio-economic covariates, and the conformity (of this particular policy) with the general policy diffusion pattern.

ABSTRACT

Stability in social, technical, and financial systems, as well as the capacity of organizations to work across borders, requires consistency in public policy across jurisdictions. The diffusion of laws and regulations across political boundaries can reduce the tension between innovation and consistency. Policy diffusion has been a topic of focus across the social sciences for several decades, but due to limitations of data and computational capacity, researchers have not taken a comprehensive and data-intensive look at cross-policy patterns of diffusion. This work combines visual analytics and text and network analyses to help understand how policies, as represented in digitized text, spread across states. As a result, our approach can quickly guide analysts to progressively gain profound insights into policy adoption data. We evaluate the effectiveness of our system via case studies with a real-world dataset and qualitative interviews with domain experts.

IDEA @ KDD'18, August 20th, 2018, London, United Kingdom

ACM ISBN .
https://doi.org/

KEYWORDS

Cascades, Diffusion Networks, Network Inference, Policy Diffusion

1 INTRODUCTION

Understanding the transmission of ideas, information, and resources among individuals and organizations has been a central theme in many fields, such as collective actions [10, 14], international cooperation [12], and economic development [4]. In the public policy domain, as the political actors (citizens, governments, or countries) repeatedly face similar political circumstances and uncertainty, the making and deployment of policies are often a *learning* or *diffusion* process where political actors look to each other when making policy choices. Identifying such policy diffusion pathways is crucial for developing innovative yet consistent policies to address new societal challenges. However, while the process involves dynamic connections among political actors, observing such connections is difficult. In this work, we present a visual analytics system that enables the discovery of persistent policy diffusion patterns.

Existing work. There has been abundant political science literature in studying *policy diffusion*, referred as general patterns of influence that the policies adopted in a given place and time are repeatedly influenced by prior policy choices made elsewhere. For example, the pioneering work by Walker analyzed and theorized how policies spread from the pioneering states to the rest of the American states [16]. Berry and Baybeck incorporated Geographic Information Systems (GIS) to analyze economic diffusion between contiguous states [4]. Most of these works were limited in studying a single policy. Recently, Desmarais et al. [9] combined machine learning algorithm and accumulation of policy data to characterize persistent policy diffusion patterns. They used the latent network inference algorithm, called *NetInf* [11], to infer policy diffusion networks connecting the American states over time. While these works have made great progress in learning a policy diffusion network from multiple policies, understanding such a network across different political settings is challenging.

Our proposed work. We propose a novel visual analytics system, called *PolicyFlow*, that allows for interpreting and examining policy diffusion *in context* – that is, across various spatial, temporal and multiple policy settings. We work closely with domain experts to design an interactive visualization system that helps answer relevant questions including: what would the underlying diffusion network look like? who are the leaders in the network? do the network and leaders change over time, regions, and across various topics of policies? what are possible factors associated with the diffusion patterns? to what extent the inferred patterns capture the observed data? In this work, we propose a suite of visual analytic tools to better explain and assess the results derived from the *black-box* network inference algorithm and aggregate policy adoption data.

Contributions. To summarize, our contributions include: (1) **System**: We propose the first visual analytics system that offers interpretable policy diffusion pattern discovery by seamlessly integrating network inference algorithm and complex political contexts. (2) **Visual analytics**: We provide a suite of analytics and visualizations that facilitate the explanation and examination of policy diffusion across different facets – geo-space, time, policy topics, similarities in political contexts, etc. (3) **Model interpretation**: We offer inspection that helps interpret and evaluate the inference results, e.g., how different policies or actors conform to or deviate from the generate diffusion patterns. (4) **Case study**: Moreover, we provide a case study that illustrates how our system can be used to discover insights into diffusion pathways in the American states using a large policy adoption data.

2 RELATED WORK

We describe related work in visual analysis of text corpus, and visualization of diffusion networks.

Visual Analysis of Text Corpus. The visual representation of text corpus is an effective way to summarize it, thereby lessening cognitive load. Among them, exploring temporal changes of a large corpus in different contexts have been salient over recent years. Many of them focused on visualizing a temporal stream of topics using stacked graph as visual representation with a slightly different focus. [7] is an earlier work of visualizing selected topics competing over time. The flow of streams in the system represented the merge and split of topics with varying amounts of texts. The following studies are featured as exploring hierarchically structured topics [8], anomalous spreading of information [17]. [15] introduced EvoRiver, a visual framework dealing with topics cooperating and competing for each other to attract opinion leaders.

Our system, on the other hand, focuses on inferring and visualizing diffusion pathways where political innovations are spreading over the US states. Text corpus provides a set of evidence, and we infer how states exert their political power onto others in the policy diffusion process using the policy metadata that consist of adoption sequences.

Visualization of Diffusion Networks. The information diffusion process involves various types of information. Many of studies have explored social media and micro-blogs in different contexts, but some studies focused on specific information type and phenomenon such as new diffusion [2], and rumor spreading [3] while others tried to convey the cascade pattern in general [1, 6]. [6] especially highlighted their study on how central users' influence propagate over other users to summarize the diffusion process in the hexagonal grid map. [5] integrated spatiotemporal contexts and provided a coordinated view. Retweet behaviors from a massive amount of tweets are monitored in real-time and represented in a visual framework that borrowed a metaphor from sunflower.

PolicyFlow coordinates with more various contexts. With a focus on providing the inferred diffusion pattern, such inference task can be derived from the combination of spatial, temporal, and topical context specified by users with interaction. We see that these functionality fit into users' need in the political domain in that policy-making process requires to explore the context and background, which our system can support the data-driven exploration on this.

3 DESIGN OVERVIEW

We design PolicyFlow for interpreting the latent policy diffusion network in context. Following a user-centric design process, we have worked closely with a group of domain experts – a team of political scientists that are specialized in the study of policy diffusion in American politics. We have scheduled a series of meetings over a year-long period to help us understand the system requirements and refine our prototype. Our discussions have centered on what kinds of patterns need to be captured and how to reveal and interpret them in various spatiotemporal contexts and political settings. We summarize the desired system requirements as follows.

R1 **Overview:** The system should offer a big picture of the general diffusion patterns learned from the historic policy diffusion dataset. It should allow users to identify the underlying diffusion network and leading states (i.e., influential nodes) in such a network.

R2 **Context:** The system should help reveal the diffusion patterns in heterogeneous contexts. It should allow users to explore questions such as (a) how do the network and leaders change over time, regions, and across policy topics? (b) how do the diffusion patterns associate with the socioeconomic context of states? and (c) what are the relationships between the diffusion and geo-proximity?

R3 **Structure:** The system should help reveal the structural details of the diffusion patterns. Specifically, (a) how does a particular state influences or is influenced by other states? (b) do policies of similar topics exhibit similar diffusion patterns? how to identify policies with similar diffusion patterns?

R4 **Inference assessment:** The system should enable users to interpret and assess the patterns derived from the network inference algorithm. In particular, it should clearly show the extent to which the inferred patterns apply to a particular policy, state, or a diffusion pathway.

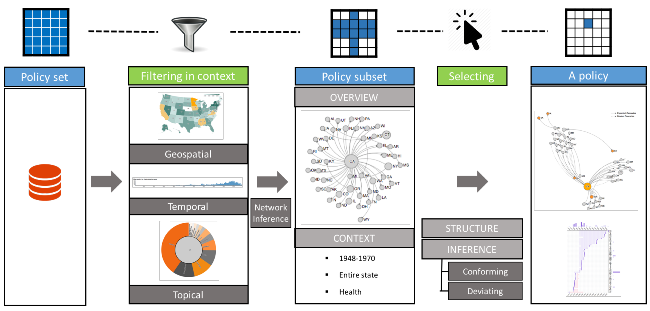

Figure 2: The system framework of PolicyFlow. The policy set can be specified in context with filtering and selection.

System overview. The proposed work, PolicyFlow, is designed and developed following the aforementioned system requirements. As shown in Fig. 2, the system allows users to interactively explore how the state-based policies diffused in the US. It allows users to browse and select a subset of policies from the entire policy database with the multifaceted (geospatial, temporal, and topical) context filters (**R2**). Based on the whole or user-selected subset of policy texts, the system computes the underlying policy diffusion network and generate an overview of the network (**R1**), with multiple coordinated views showing the geospatial and states' socioeconomic contexts (**R2**). Users can explore the structural details through interactions to reveal how states (nodes) are close to each other in terms of their network connectivity or policy adoption similarity (**R3**). Furthermore, the system allows users to inspect a policy's details and to evaluate how the inferred diffusion patterns conform to or deviate from a particular policy's actual adoption sequence (**R4**).

4 ANALYSIS OF POLICY ADOPTION

4.1 Policy Data

Here we use the policy adoption data collected by [9]. The scope of policies in the system includes 764 state-wised policies over 300 years ranging from 1691 to 2017. The dataset consists of the metadata of states, policies, and the set of policy adoption cases. Each policy is associated with a history of adoption that indicates which states have adopted this particular policy and the timeline of adoption. The additional information on policy adoption includes policy subject (e.g., health, education), start (the year of the first adoption), end (the year of the last adoption), and the number of states adopting the policy. In the policy data, the underlying diffusion network (i.e., who influenced whom in the adoption decision) is not observable. The process of inferring such an underlying network will be detailed in the next subsection.

Another information collected in our dataset is a set of state-related attributes that provide key socioeconomic context. These include Per Capita Income, Minority Diversity, Legislative Professionalism, Citizen Ideology, Total Population, and Population Density. Such attributes provide theoretically important covariates [9] as well as contextual information for users to reason the diffusion patterns. We provide analytical modules that may help hypothesize how each state's attributes correlate to its role in the propagation of political agendas as follower or influencer.

4.2 From Adoption to Diffusion Network

As mentioned above, the policy adoption data allows for observing how a policy was adopted over time and across states, the underlying influence or diffusion network, that is, who tends to lead and who tends to follow, is often unobservable [9]. The goal of network inference is thus to infer a *latent diffusion network* of political actors (i.e., states) based on observable data about the repeated adoption choices that those actors make.

We apply the latent network inference algorithm, NetInf [11], to infer policy diffusion networks. A policy's adoption history can be considered as multiple network *cascades*, where each cascade consists of a sequence of adoption cases called *contagions*. A contagion c, denoted as a tuple $(p, u, v, t_v)_c$, means that the adoption of a policy p has spread from state u to state v at time t_v. A directed edge $u \rightarrow v$ connecting a pair of state nodes was used to indicate that policies diffuse from *source* node u to *follower* node v. The set of cascades for a specific policy p can be obtained by grouping the contagions by policy, denoted as $\{(p', u, v, t_v)_c \mid p' = p\}$.

In practice, a contagion can only be observed through $(p, v, t_v)_c$ that describes the time t_v when node v got infected by the contagion c of a policy p. The NetInf algorithm aims to recover the unobserved directed network G^* , i.e., the policy diffusion network over which the contagions spread. The algorithm is based on a probabilistic model that models the probability of a contagion c between a pair of nodes u and v, and the probability that contagion c propagated in a particular cascade tree. Based on this, $P(C|G)$, the probability of a set of cascades C occurring in G, can be obtained, and the latent diffusion network G^* is approximated by $\hat{G} = \text{argmax}\ P(C|G)$ with a sparsity constraint on all possible G.

4.3 Relationship to Geopolitical Context

Our system supports the analytical task that enables users to examine the relationship between an inferred diffusion network and the geopolitical context across states. Specifically, we provide two correlation measures: (1) *influence*: the correlation between node influences and socioeconomic attributes, and (2) *connectivity*: the correlation between edge connection and geographical adjacency.

First, we calculate the correlation between each of socioeconomic attributes presented in Section 4.1, and five centrality measures as node influence including Outdegree, Page Rank, Betweenness, Hit, and Closeness. Since the centrality measures are often skewed, and we do not expect a linear relationship between variables in the correlations, we use Kendal's τ as a measure of correlations. Such correlation measures enable users to evaluate the ability of different socioeconomic attributes to explain the diffusion ties.

Second, we present how the geographical relationship of states can be related to the inferred diffusion ties that estimate the source-follower relationship in the decisions of policy adoptions. The correlation is measured as an overlap between the geospatial neighborhood and the network neighbors derived from the inferred diffusion network.

4.4 Cascade Pattern Comparison

For the context analysis of policy diffusion, we provide the similarity of policies in terms of cascade pattern and topics. A cascade of a policy is a set of policy adoptions represented as $c_p = \{(s,t),\ s \in S\}$, where S is a set of all states and t is adoption year. The comparison of cascade patterns cannot be obtained by using a standard correlation measure since the cascade sizes vary. To address this issue, we combine both the set overlap measure and a rank correlation measure. The similarity of two policies i and j is then given by: $\text{sim}_{ij} = J(c_i, c_j) \times K(c_i, c_j)$, where $J(\cdot)$ is the Jaccard index that measures the overlap between the two sets of states that adopted policies i and j (without considering the temporal ordering), and $K(\cdot)$ is the Kendall's τ rank correlation between the two sets of states that are adopted both policies and ordered by adoption years. The similarity score sim_{ij} ranges between 0 and 1. The higher the similarity, the more similar the two policies i and j are in terms of their cascade patterns.

We also provide the content similarity of policies. To compute this, we first retrieve the textual content of all policies using Google Search APIs and represent the set of policy texts using a word2vec model [13]. We then calculate the cosine similarity for all pairs of policies offline and store the five most similar policies of each policy. The most similar policies can be retrieved in the Policy Detailed View for users to discover some insights from the similarity in terms of adoption cascade and policy content.

4.5 Model Examination

Since the inferred diffusion network is estimated from the observed, repeated adoption choices made by states, there may be a disparity between the estimated diffusion patterns and the actual adoption sequences. Such disparity reflects (a) the quality of the network inference on a set of policies, and (b) the ability of the general inference to capture a particular set of policies or a specific policy. To capture such disparity, we introduce the notion of *expected* cascade and *deviant* cascade. Recall that a network is derived from a set of adoption cases, which indicates the influential relationship inferred from multiple policies as a whole over the given context. The adoption sequence of a single policy, on the other, is a set of recorded trajectories in the dataset which can differ from the whole network. In other words, some edges of a network direct from one state to another deviant from the time order in the selected policy's adoption sequence. Taking the example networks in the Fig. 3, the network on the left-hand side is an inferred network from the entire set of policies, and the network on the right-hand side is a network of the policy called "Mandated Coverage of Clinical Trials". The policy network consists of states involved in this policy as nodes, and their relationship as edges. When we look at a specific edge coming from CA to GA, it is represented as a deviant cascade with dotted line. It indicates that CA is likely to influence GA when we infer their relationship based on the given set of policies, but it's actually deviant in terms of the time order because GA actually adopted the policy name in 1998 which is earlier than CA did in 2000. These network edges reveal that the influential relationships between states in the context of policy adoption hold true in general, but it may not be the case for an individual policy where some follower states adopted earlier.

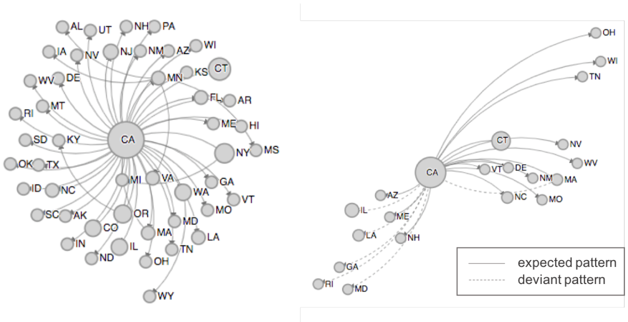

Figure 3: The whole network from all policies (on the left) and the network of the policy named 'Mandated Coverage of Clinical Trials' (on the right).

We quantify this pattern as a *conforming* score to provide a summary of how the diffusion pattern for a policy conforms to the whole network. For a specific policy, the score is calculated by the ratio of conforming edges among all edges. The higher score indicates that the adoption sequence of a policy better conforms to the inferred relationship. In the system, when a policy is selected, the conforming score of a policy is calculated and represented in the Policy Inspection View described in Section 5.4. Users can expect to identify the detailed expected/deviant pattern for a policy with the conforming score as a measure of how accurate the inference model is.

5 INTERACTIVE VISUALIZATION

5.1 Spatiotemporal Overview

The primary goal of PolicyFlow is to give users a comprehensive overview of the general diffusion pattern. Several components represent different aspects of the network, while the interface enables to explore the influential relationship from multiple perspectives.

Centered within the system is the Network View. A diffusion network is a directed network where an influencer state u and a follower state v is connected by an edge $e(u, v)$. When the system is initially loaded, it renders the general diffusion network inferred from the whole policies. With any interaction ending up with a set of policy adoptions or an individual policy selected, the size of nodes is adjusted by the influence. The five influence measures which are node centrality in the network show the most influential state. While the Network View visually provides the most intuitive representation, three components including the Timeline View, Map View, and Subject Browser, Content Browser convey temporal, spatial, and topical distribution of policies and adoptions. The Timeline View linearly shows the frequency of policy adoption throughout the entire policy adoption history. The Map View highlights the states hovered in the network, indicating where they locate. The Subject Browser and Content Browser represent how policies are topically and semantically distributed.

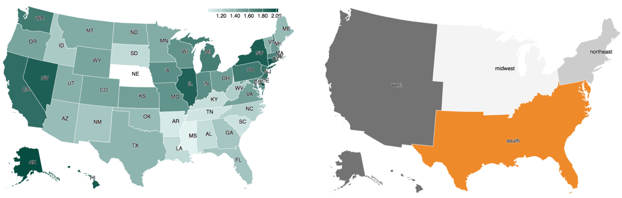

Figure 4: Two modes of the Map View. The state-wise and regional view provide different snapshots of spatial overview.

Even though those visual components serve as filters, they are basically dedicated to representing the distribution of policies and states in different perspectives. The Subject Browser and Content Browser effectively shows the proportion of topics as a pie chart, the Timeline View gives a trend of adoptions over time, and the Map View projects nodes into the map which provide the geo-proximity of states.

A whole network on the left and a network of the policy named 'Mandated Coverage of Clinical Trials' on the right on the Network View. The whole network is a network inferred from a set of policies with or without filtering. Once a policy is selected, the network is re-rendered as a policy network. The irrelevant nodes are filtered out, and edges are marked as one of two modes, which are the *expected* cascade with a solid line and the *deviant* cascade with a dashed line.

We also support the exploration of a specific policy. A major change on the layout when selecting a policy is that the inferred network in the Network View is re-rendered to represent the selected policy, as two networks in Fig. 3. Once a policy is selected, nodes are filtered such that only the states (nodes) that adopted the chosen policy appear in the network.

5.2 Diffusion Structural Details

We support exploring the similarity and connectivity of states and policies with interactive layouts by explicitly visualizing them. One of our requirements (R3) is to decompose the structural details of relationships between states or policies. Specifically, we try to help users figure out these questions: (a) How are states similar to each other in terms of their connections and similarities from the network? (b) What are the similar policies in the aspect of content and cascade?

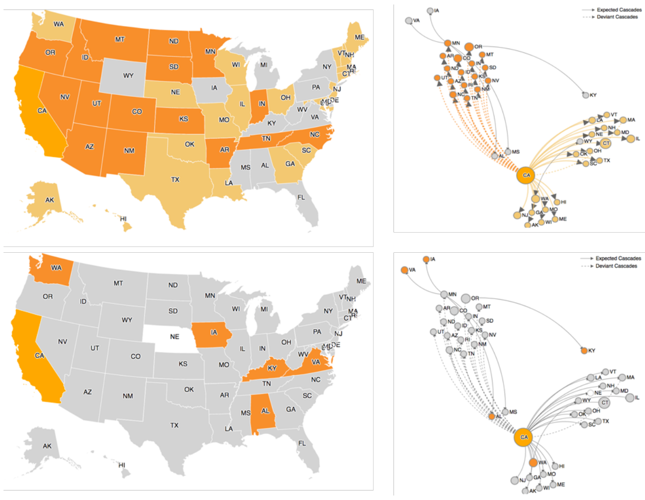

Figure 5: Two modes of examining structural details. The map and network pair renders the network of the policy 'Framework For Donation Of Organs Other Body Parts'. On a hover of the California node in the network, two hover modes are allowed to examine the structure centered around the selected node. The first pair of the map and network in this figure highlights the connected states with California (with lighter orange for expected cascades and darker orange for deviant cascades). The second one highlights the most top five similar nodes in terms of adoption sequences.

To answer the first question, we provide two hover modes for examining states: Connected and Similar. Users can select one of two modes by adjusting the slide bar, then hovering the cursor on any state on the map or any node on the network colors the most five similar states.

As mentioned and presented in Section 4.4, PolicyFlow derives the similarity of policies from two perspectives: Content and Cascade. The Policy Detailed View is dedicated to providing this information along with the detailed information of a policy. If a policy is selected, similar policies are listed with similarity score. Either of two lists is displayed when users adjust the tab interface.

5.3 Geo-political Context

The visual components in PolicyFlow not only serve to explore the overview but also to filter the dataset by context. In our system, the combination of three filters narrows down to a subset of policies. First, users can select a topic of their interest in the Subject Browser and Content Browser by clicking an arc of a topic. Second, Brushing through a period in the Timeline View automatically captures a subset of policies adopted within that time period. Third, the Map View is dedicated to picking specific states of user's interest. Any clicks end up counting in and out the selected state. These filtering behaviors lead to subsetting nodes and edges of the network in the Network View.

PolicyFlow also supports exploring two correlation measures presented in Section 4.3. First, our system computes the correlation between node centrality and socio-economic attributes in real time. When a policy is selected, the correlation measure correspondingly is updated in the dropdown menu. The correlation between the adjacency of states and the connectivity in the network, which is another correlation, is presented in the Policy Inspection View.

5.4 Policy Adoption Inspection

The Policy Inspection View visualizes how each of policies diffuses through states. When users select a policy, the system renders the Policy Inspection View in the Policy Inspection tab associated with the selected policy.

A matrix represented by this view is a two-mode matrix where each cell is a spatio-temporal incident of policy adoption (s, t) with a state s from the vertical axis and a year t from the horizontal axis. Each row is then a timeline of a state's policy adoption history as a set of linearly aligned cells. An example of the Policy Inspection View, which is a detailed adoption pattern of the policy called 'Bans Child Pornography', is presented Fig. 6. In the highlighted part on the left, a semicircle indicates a single event of adopting this policy by Florida in the corresponding year. Two different types of strip coming to a semicircle icon represents the *expected* or *deviant* pattern of policy adoption. A purple-colored strip is a *expected* pattern of adoption where the policy was adopted to the influential state and the follower state identified in the inferred network in chronological order. A pink-colored strip from the right to a semicircle on the left, on the other hand, is a *deviant* pattern of adoption where the follower state in the inferred network actually adopted the policy earlier than the influencer state, which means that the adoption sequence doesn't match the influential relationship we inferred based on a set of policy adoption cases. The semicircles head toward the direction where the influence comes from, so users can understand the aspect of patterns with the direction of semicircle heads and the color of strips.

The Policy Inspection View provides the connectivity between nodes within the general diffusion pattern. On hover of a semicircle component, the layout reveals the egocentric network of the state in a way that all the edges connected to it are represented. In Fig. 6, the hovered node is represented as a yellow node. The incoming and outgoing edges connect the node to its source and target nodes which are colored as light and dark purple. We also specify what other nodes are influenced by its source node. The gray edges coming from the source node indicate in the Fig. 6 that CA is a great influencer impacting on most of the other states.

6 CASE STUDY

In this section, we demonstrate the effectiveness and utility of PolicyFlow in helping users to gain a better understanding of policy diffusion throughout the history. We consider political analysts such as political scientists, policymakers and lobbyists, and members of NGOs, as the main users of our system. For them, understanding how the previously related public policies diffused across states and regions is important for planning and making a new policy, as well as for predicting how different aspects of a new policy may unfold in the near future. Here, based on our interview with a political scientist, we provide a use case scenario to show how a policy analyst can use PolicyFlow to gain insights from a policy dataset.

When the system was initially launched, the policy analyst explored the overview of diffusion pattern in the system. Without any filtering or selection, the system showed how state policies adopted throughout the history of United States. By looking at the Network View, the analyst immediately noticed that California, which is placed in the center of the network, was estimated and rendered the greatest influential state. He also found that policy adoptions have been in a noticeable increase after 1960, shown in the Timeline View. Looking at the Subject Browser, he realized that Law and Crime, Civil Rights, and Health were the most dominant subjects over the centuries.

At this time, he was interested in some influential policies in establishing abortion ban in the Northeast region in the late 1900s. As a preliminary analysis, he started exploring 'Civil Right' subject after the 1970s. When he adjusted the slider of the Map View to 'Region', all the nodes (states) from the Northeast area in the Network View were rendered as orange color. CT had especially a greater influence (outdegree) than other Northeastern states with the bigger size of the node. When he selected the closeness as another influence from the 'Influence' dropdown menu, CT had the second largest influence indicating that it is closer to all other nodes in the network in terms of the average of the shortest path.

Since the network had an intertwined connectivity between nodes, he wanted to further explore the structural details. The Map View and the Network View were helpful in a way that hovering a node in the network or a state in the map highlighted the relevant nodes to make it distinct based on the similarity and connectivity of nodes as shown in Fig. 5. With 'Similar' mode selected, the five most similar nodes were also highlighted in terms of cascade pattern. When the analyst changed the mode to 'Connected' with hovering over RI (Rhode Island), all the nodes connected to RI were highlighted. he found that RI is likely to influence KY, MD, NJ, and NT, and it is likely to be influenced by CA when it comes to policy adoptions within 'Civil Rights' after the 1970s.

The analyst decided to take a closer look at the policy 'Physicians Can Refuse To Do An Abortion' as the policy was the most spreading policies in this societal issue with dynamic diffusion pattern. When he selected the policy, Policy Detailed View displayed the policy contents, and also the list of similar policies. Interestingly, the most similar policy in terms of policy cascade was 'Require A Licensed Physician For Abortion' which is also related to the abortion issue. It makes sense that the adoption of this policy started and ended two years earlier than the policy according to the adoption

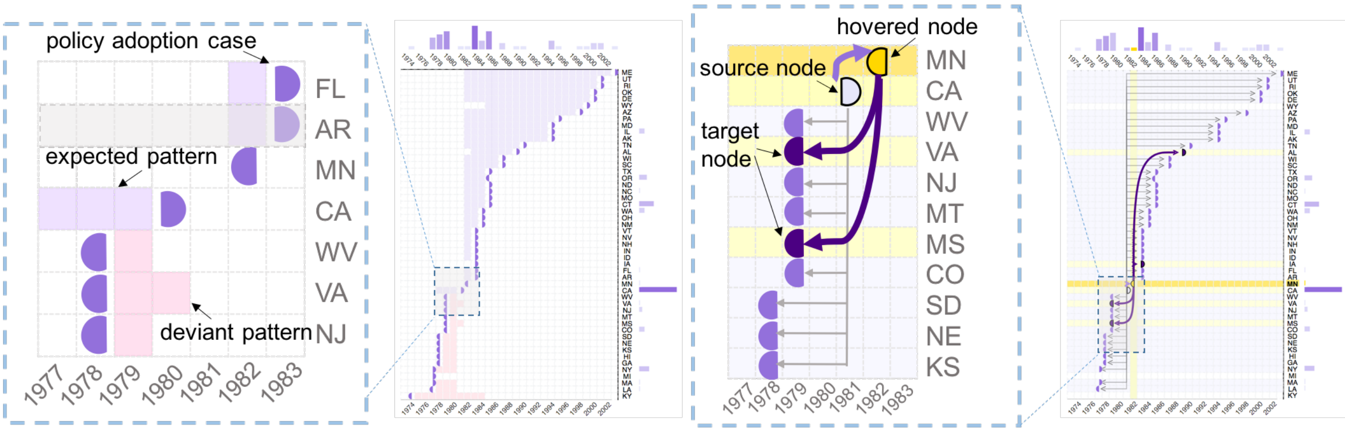

Figure 6: The Policy Inspection View conveys a detailed diffusion pathways of an individual policy.

period. It is also interesting that the five most similar policies in policy cascade all belonged to 'Civil Rights' subject.

The analyst moved on to the Policy Inspection View to identify the detailed diffusion pattern of the policy 'Physicians Can Refuse To Do An Abortion'. The pioneer states regarding this policy were NY and IL which adopted the policy in 1976 and 1978 respectively. The adoption started to burst out in 1984 and kept its pace until it was adopted by 45 states during ten years. Most of the edges showed the conformity to the general diffusion pattern of 'Civil Rights' policies in post-1970s with the conforming score 0.74.

After all these exploration tasks, the analyst explored how the socioeconomic state attributes correlate to this diffusion pattern. When he adjusted six different socioeconomic attributes from the 'Attribute' dropdown menu, states in the Map View and the Network View were colored by the selected attributes with the correlation score provided in the dropdown menu. He figured out that 'Total Population' was the most correlated to the diffusion pattern with the score of 0.68 when it comes to the selected policy.

7 DISCUSSION AND FUTURE WORK

In this paper, we presented PolicyFlow, an interactive framework for exploring diffusion pattern of policies in context. The policy-making process on public policy sector requires multifaceted exploration and careful inspection of the observable policy data within complex context. The visual components in our system not only provide an overview of diffusion patterns over states but also enable contextual filtering on spatial, temporal, and topical aspects. We made a step further by providing analytic components that allows for interactively inspecting the similarity of policies, and assessing the inference networks generated by the network inference algorithm. In the future, we plan to conduct more in-depth user study in two ways. First, we will interview domain experts to learn how our system can be embedded into the workflow of their research or policy-making process. For the usability and comprehensiveness of the system as a second one, we will conduct qualitative and quantitative study to validate the requirements of our system. We expect that identifying the usefulness and shortcomings of our system will help improve our system and bring users more meaningful findings in the exploratory tasks.

ACKNOWLEDGEMENT

We thank the anonymous referees for their useful suggestions. The authors would like to acknowledge the support from NSF Grants #1637067 and #1739413. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

REFERENCES

[1] 2011. Google+ Ripples: Revealing How Posts are Shared over Time. http://infosthetics.com/archives/2011/10/google_ripples_revealing_how_posts_are_shared_over_time.html. (2011). Accessed: 2018-05-09.
[2] 2011. NYTLabs Cascade: How Information Propagates through Social Media. http://nytlabs.com/projects/cascade.html. (2011). Accessed: 2018-05-09.
[3] 2016. How riot rumours spread on Twitter. https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter. (2016). Accessed: 2018-05-09.
[4] William D Berry and Brady Baybeck. 2005. Using geographic information systems to study interstate competition. *American Political Science Review* 99, 4 (2005), 505–519.
[5] Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. 2012. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2649–2658.
[6] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Xiaoru Yuan, Nan Cao, and Yadong Wu. 2016. D-Map: Visual analysis of ego-centric information diffusion patterns in social media. In *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on*. IEEE, 41–50.
[7] Weiwei Cui, Shixia Liu, Li Tan, Conglei Shi, Yangqiu Song, Zekai Gao, Huamin Qu, and Xin Tong. 2011. Textflow: Towards better understanding of evolving topics in text. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2412–2421.
[8] Weiwei Cui, Shixia Liu, Zhuofeng Wu, and Hao Wei. 2014. How hierarchical topics evolve in large text corpora. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2281–2290.
[9] Bruce A Desmarais, Jeffrey J Harden, and Frederick J Boehmke. 2015. Persistent policy pathways: Inferring diffusion networks in the American states. *American Political Science Review* 109, 2 (2015), 392–406.
[10] Glenn Ellison and Drew Fudenberg. 1995. Word-of-mouth communication and social learning. *The Quarterly Journal of Economics* 110, 1 (1995), 93–125.
[11] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. 2012. Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5, 4 (2012), 21.
[12] Erin R Graham, Charles R Shipan, and Craig Volden. 2013. The diffusion of policy diffusion research in political science. *British Journal of Political Science* 43, 3 (2013), 673–701.
[13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[14] W Russell Neuman, Lauren Guggenheim, S Mo Jang, and Soo Young Bae. 2014. The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication* 64, 2 (2014), 193–214.

[15] Guodao Sun, Yingcai Wu, Shixia Liu, Tai-Quan Peng, Jonathan JH Zhu, and Ronghua Liang. 2014. EvoRiver: Visual analysis of topic coopetition on social media. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1753–1762.

[16] Jack L Walker. 1969. The diffusion of innovations among the American states. *American political science review* 63, 3 (1969), 880–899.

[17] Jian Zhao, Nan Cao, Zhen Wen, Yale Song, Yu-Ru Lin, and Christopher Collins. 2014. # FluxFlow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1773–1782.