ONE-M: Modeling the Co-evolution of Opinions and Network Connections, *Proc. ECML-PKDD*, Dublin, Ireland, Sept 10-14, 2018, LNAI, Springer, Berlin, to appear.

ONE-M: Modeling the Co-evolution of Opinions and Network Connections

Aastha Nigam[1], Kijung Shin[2], Ashwin Bahulkar[3], Bryan Hooi[2], David Hachen[1], Boleslaw K. Szymanski[3], Christos Faloutsos[2], Nitesh V. Chawla[1]

[1] University of Notre Dame, Notre Dame, IN, USA
{anigam,dhachen,nchawla}@nd.edu,
[2] Carnegie Mellon University, Pittsburgh, PA, USA
{kijungs,christos}@cs.cmu.edu, bhooi@andrew.cmu.edu
[3] Rensselaer Polytechnic Institute, Troy, NY, USA
{bahula,szymab}@rpi.edu

Abstract. How do opinions of individuals on controversial issues such as marijuana and gay marriage and their underlying social network connections evolve over time? Do people alter their network to have more like-minded friends or do they change their own opinions? Does the society eventually develop echo chambers? In this paper, we study dynamically evolving networks and changing user opinions to answer these questions. Our contributions are as follows: (a) *Discovering Evolution of Polarization in Networks*: We present evidence of growing divide among users based on their opinions who eventually form homophilic groups (b) *Studying Opinion and Network Co-Evolution*: We present observations of how individuals change opinions and position themselves in dynamically changing networks (c) *Forecasting Persistence and Change in Opinions and Network*: We propose ONE-M to forecast individual beliefs and persistence or dissolution of social ties. Using a unique real-world network dataset including periodic user surveys, we show that ONE-M performs with high accuracy, while outperforming the baseline approaches.

1 Introduction

How do opinions of individuals on topics, specially controversial topics, and their network structures co-evolve? That is, we are interested in understanding and modeling how opinions develop and diversify, and their intricacies with the underlying network structure. We posit that it is not just the network effect that parlays diffusion or preponderance or change of an ego's opinion, but rather the opinions also dictate how the network of connections changes around an ego. When egos and alter-egos begin to evoke similar opinion, they form echo chambers, which are clique-like patterns.

Prior research [7–9, 13, 15, 17, 24, 25] has looked at polarization and opinion dynamics, but there is a paucity of work in studying how opinion and network co-evolve over time, and how it translates into persistence or change in opinions and network structure both. As such, the literature also lacks a principled model that brings together these characteristics to be able to forecast both spread of an opinion and change in the network structure (through persistence or dissolution of links).

Therefore, in this paper, we ask the following question:

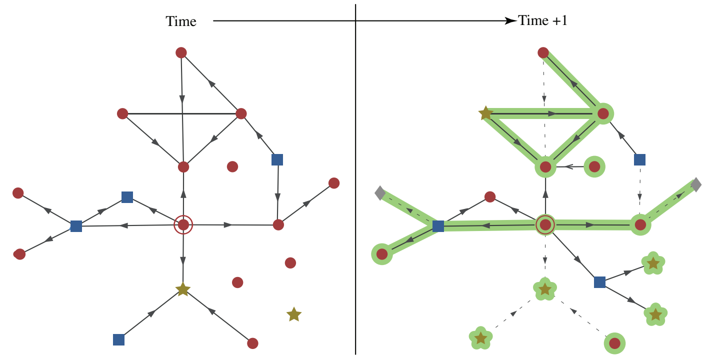


Fig. 1: **Effectiveness of ONE-M on real data set**: A real communication ego-network (central node markef by a boundary) snapshot from the NetSense study and the results obtained by using ONE-M. The nodes are color-coded by their opinions where red ($\circ$), yellow ($\star$) and blue ($\square$) indicate conservative, unsure and liberal opinion about political alignment, respectively. A grey($\Diamond$) node identifies an inactive node whereas the dotted edges indicate dissolved edges at $t+1$. Using opinion and network topology at t (left), ONE-M is able to correctly forecast the opinion and the persistence/dissolution of edges (as highlighted by green color) at $t+1$.

Informal Problem: *Given information about evolving user network interactions either derived from communication or collocation patterns, user opinions on diverse beliefs (e.g. gay marriage, marijuana, etc.) and external user attributes (e.g. age, gender, etc.), how can we model and leverage the two-way effect that opinions and networks have on each other, to forecast persistence and change in opinions and social ties?*

To address this question, we propose the ONE-M model (**O**pinion and **N**etwork co-**E**volution-**M**odel) to study and model how opinions and networks co-evolve. To jointly model the co-evolution of opinions and networks, ONE-M incorporates several data-driven features that capture opinions on beliefs at different time segments such as opinion pre-disposition and persistence at the ego and population level, as well as the network characteristics including strength of ties, reciprocity, and triads.

To evaluate our model, we use a longitudinal dataset that we have collected at the University of Notre Dame as part of the NetSense study. About 200 incoming freshmen in 2011 were enrolled in the NetSense study. Enrolling students on their arrival at the University gives us a unique opportunity to study the formation and evolution of their network, and also whether their opinions about beliefs changed or remained the same over time. We acknowledge that we used only this one dataset in the paper, but given the longitudinality and impressionable age of the participants (18 and older) it provides us with a novel opportunity to study the behavior as well as be able to forecast both opinion and link persistence or dissolution over time that also includes aspects of seasonality (such as holiday break, summer, etc.) and draw insights about a large audience.

Our work contributes to several novel discoveries as follows:

– **Discovering Evolution of Polarization in Networks**: We present evidence and showcase evolution of polarization among individuals across various controversial topics such as abortion and premarital sex.

Table 1: ONE-M is comprehensive compared to existing models of opinion and network dynamics.

Considerations	Das et al. [8]	Kim and Leskovec [21]	Durrett et al. [10]	Badev [2]	Bilò et al. [6]	ONE-M (Proposed)
Node Opinions	✓		✓		✓	✓
Node Attributes		✓		✓		✓
Edge Presence	✓	✓	✓	✓	✓	✓
Edge Weights	✓				✓	✓
Network Characteristics*		✓		✓		✓
Multiple Networks**						✓
Multiple Topics***						✓
Multiple Time Steps	✓	✓	✓	✓		✓

* reciprocity, transitivity, etc. ** co-location, communication, etc. *** euthanasia, death penalty, etc.

– **Studying Opinion and Network Co-Evolution**: Using ONE-M, we present three novel insights: a) majority of individuals are stubborn and resist opinion change; b) individuals who are unsure do not form cliques and tend to act as bridges; c) even if communication drops in summer months, the strongest ties survive.
– **Forecasting Persistence and Change in Opinions and Network**: Using ONE-M, we report high performance for both forecasting opinion and network change. Figure 1 highlights the results of ONE-M demonstrating both the co-evolution of opinion and network, and also the forecasting of opinion and persistence or dissolution of links on real-world data.

Reproducibility: Our code is available at `https://github.com/anigam/ONE-M`. The data used in the study is available from the NetSense Team at University of Notre Dame (`http://netsense.nd.edu/`).

2 Background and Related Work

In this section, we review existing models of opinion evolution and network evolution.

Evolution of Opinions. The interaction of individuals' opinions plays a role in nearly every social, economic, and political process. Understanding this interaction is particularly useful for political voting, public health campaigns, viral marketing, and information dissemination. A rich line of work in theoretical economics has studied mathematical models of opinion exchange; see [26] for a survey. The models are roughly divided into heuristic models and Bayesian models. In heuristic models, individuals follow simple update rules when interacting with each other. Notable examples of heuristic models are (a) averaging models [9, 13], where each individual's opinion is updated to the average opinion of its neighbors, (b) voter models [7,17], where each individual follows a randomly-chosen neighbor's opinion or the majority opinion in its neighbors, and (c) flocking models [15], where individuals give more weight to opinions that conform to theirs, and (d) the combination of these models [8]. On the other hand, in Bayesian models, rational individuals maximize their expected utilities that depend on their actions as well as randomness [24, 25]. In addition to this theoretical work, a rich body of empirical work in social and developmental psychology has studied how individuals update their opinions based on the opinions of those around them [1, 12].

Evolution of Networks. Social networks change over time by nature. Consequently, there has been great interest in dynamics of social networks, especially in underlying micro-mechanisms that result in macro-level evolution of networks. Models of how networks evolve are roughly divided into stochastic models and game-theoretic models. Stochastic models view each edge as a random variable whose probability of presence depends on many different effects; see [27] for a survey. Notable examples of such effects are (a) transitivity [16]: friends of friends tend to be connected, (b) popularity [4]: high-degree nodes tend to be connected with new nodes, (c) assortativity [23]: nodes with similar degrees tend to be connected, and (d) homophily [21]: nodes with similar sociodemographic characteristics tend to be connected. In game-theoretic models, nodes are rational individuals whose utilities depend on social network structures, and edges are formed at the discretion of the nodes [20, 30]; see [19] for a survey.

Co-evolution of Opinions and Networks. Relatively little attention has been given to the co-evolution of opinions and networks despite the considerable interest in the evolution of each of them. Recently, several models of opinion dynamics were extended so that nodes update their edges before updating their opinions [5, 6, 10, 14, 18]. In the extend models, nodes either follow simple heuristic rules (e.g., following a randomly chosen neighbor's opinion) [10,18] or optimize simple utility functions that are solely based on the agreement of opinion between neighboring nodes [5, 6]. In our model, however, diverse user attributes (age, gender, etc.) and network characteristics (reciprocity, transitivity, etc.) are taken into consideration as well as the agreement of opinion between neighboring nodes. Our work is also closely related to models of mutual interaction between networks and behaviors (e.g., drinking [29] and smoking [2]). In Table 1, we compare our model with several other models of opinion and network dynamics.

3 Proposed Method

In this section, we motivate the problem and provide intuition behind the proposed approach. Next, we define the notation used in the paper for easier comprehension followed by the details of the proposed model ONE-M.

3.1 Intuition & Problem

In human societies, opinions are guided by social interactions, and conversely, network connections are influenced by the opinions. For instance, all individuals supporting the conservative party might opt to maintain strong relationships and subsequently minimize communication with individuals who might be liberal — leading to echo chambers and polarization. Similarly, an individual might be against the use of marijuana, but the majority of his/her friends may support the usage — leading to the individual changing his/her opinion. On the contrary, some views might be more ingrained in an individual which he/she refuses to change irrespective of the network and would rather lose friends than change opinion.

While studying this two way effect might be challenging, other factors can also govern opinions and network connections. Personal factors such as age, gender, ethnicity and hometown could also impact the underlying network characteristics. For instance, an individual may choose to form same-gender friendships whereas another individual may select to conform his friendship among individuals of the same ethnic background.

Therefore, to truly understand and build a model around opinion and network co-evolution it is crucial to incorporate varied network interactions based on who calls whom or who spends time with whom, opinions on a diverse set of topics and personal attributes about the individuals. Intuitively, each action a user takes whether it be changing(or persisting) with their opinion and(or) altering their social connections, would be influenced by their surroundings and inherent characteristics. With this motivation, we build the ONE-M as described in the subsequent sections.

3.2 Notation

For a population of N individuals, we have a time-evolving collection of T networks across C modalities (represented by $\mu \in \{c, b, wb\}$ when $C = 3$): communication (c), collocation (b) and collocation over the weekends (wb). For each time-step, we record user networks G^t for each modality in μ where $G^t = \{g_c^t, g_b^t, g_{wb}^t\}$. Every network for each time step in each modalilty (g_μ^t) captures relationships between N individuals, with each node represented as u_i, is of size $N \times N$. Edges are directed and weighted by the strength of the connection represented as $w_c^t(i,j), w_b^t(i,j)$ and $w_{wb}^t(i,j)$ respectively. In the communication network, an edge $u_i \rightarrow u_j$ denotes if u_i chooses to call/message u_j. Further, in the collocation network, an edge $u_i \rightarrow u_j$ indicates if u_i selects to be physically present near u_j. Similarly, for collocation network over the weekend, an edge $u_i \rightarrow u_j$ denotes if u_i meets u_j over the weekends. Therefore, collectively, we have $G = \{G^1, G^2, ..., G^T\}$ which is a tensor of size $N \times N \times C \times T$.

In addition to the network information, for each u_i, we record their evolving opinions for a set of K beliefs. Typically, at each time t, for a belief a_k, u_i reports their opinion $a_{i,k}^t \in \{1, 2, 3\}$. Therefore, A of size $N \times K \times T$ captures opinion information for N users, K beliefs across T timesteps. Further, u_i is also associated with M external user attributes such as age, gender and ethnicity denoted by $X_i = \{X_{i1}, ..., X_{iM}\}$ which stay constant throughout the study for each individual. Thus, X capturing the user attributes is of size $N \times M$. Table 2 lists the symbols and their definitions.

In summary, the inputs to our setting are the following:

- Evolving network information: G, a tensor of size $N \times N \times C \times T$,
- Evolving user opinions: A, a tensor of size $N \times K \times T$,
- User attribute information: X, a matrix of size $N \times M$.

Recall that N is the number of users, C is the number of communication modalities (phonecall, bluetooth, etc), T is the count of time-ticks, K is the number of topics (marijuana, abortion, etc) and M is the number of demographic attributes (gender, etc).

3.3 Proposed Model: ONE-M

General Problem Definition *Given the networks (G^t), opinions (A^t) and user attributes (X) for individuals at time t, how can we model and forecast persistence and change in opinions(A^{t+1}) and network connections(G^{t+1}) at time $t+1$?*

To that end, we propose a model aimed at jointly capturing opinion and network co-evolution called ONE-M (**O**pinion and **N**etwork co-**E**volution-**M**odel). We define a

Table 2: Symbols and Definitions

Symbols	Definitions
$\mathcal{N}$	Set of all individuals
G	Set of networks across time $\{G^1, G^2, ..., G^T\}$
G^t	Collection of networks $\{g_c^t, g_b^t, g_{wb}^t\}$ at time t
g_μ^t	General representation of the desired network at time t where $\mu \in \{c, b, wb\}$
$w_\mu^t(i, j)$	Tie strength between u_i and u_j at time t in either networks where $\mu \in \{c, b, wb\}$
$e_\mu^t(i, j)$	0/1 if the tie is present u_i and u_j at time t in either networks where $\mu \in \{c, b, wb\}$
$\eta(i)$	Friends of i, who have a directed edge $i \to j$ and $j \to i$
A	Tensor of size $N \times K \times T$ capturing opinions on K beliefs for N individuals across T time-steps
A_i^t	Individual i's opinion for all beliefs at time t
$a_{i,k}^t$	Individual i's opinion for belief a_k at time t
$v(X_i)$	Probability of opinion a_i given the user attributes for individual i
X	Matrix of size $N \times M$ capturing M user attributes for N individuals
X_i	A vector of M user attributes (such as age, hometown, ethnicity and gender) for individual i

function for each individual over a set of 8 derived factors. Using information about G, A and X, we learn the importance/weights (β) for each of the factors. For each network-related factor, we learn weights corresponding to the type of the network as denoted by $\mu \in \{c, b, wb\}$. To build a more general system and have fewer parameters, we assume that the weight for each factor, stays constant over time and users. We propose that using 8 derived factors and their relative weights, we can capture the interplay between an individual's opinion and the corresponding network topology. Next, we present the definition and intuitive description for each of the factors.

1. Opinion Persistence: Given the controversial nature of many beliefs, this feature encapsulates the consistency of individuals in their opinions. For many, staying true to their opinions would be of priority over changing their opinions to conform to their surroundings. In order to incorporate for such persistence, we include opinion of user i at time t as shown below:

$$f_{stubborn} = a_{i,k}^t \tag{1}$$

2. Attribute Predisposition: Captures the phenomena that an individual can be predisposed to have a certain opinion based on their external attributes such as age, gender, hometown and ethnicity. For example, women might be more likely to support abortion or participants from a conservative leaning state would have a higher tendency to be conservative. In order to compute the dependency based on the individual's external attributes (X_i):

$$f_{predispose} = v(X_i) \sim P(a_{i,k}^t | X_i) \tag{2}$$

3. Population Belief: Measures the impact the global population can have on an individual's opinion. For instance, if all students around a participant i would indulge in marijuana usage, i is more likely to get influenced and conform to the surroundings. This phenomena can be measured as follows:

$$f_{population} = \frac{\sum_{j \in \mathcal{N} \setminus \{i\}} a_{j,k}^t}{N - 1} \tag{3}$$

4. Attribute similarity: Are we more likely to communicate or spend time with another student because they come from the same city or have the same gender as us?

This feature captures the strength of communication between user i and their friends based on how similar they are in terms of these external attributes such as ethnicity, hometown, age and gender. To account for this, we include the following feature:

$$f_{friend-sim} = sim(X_i, X_j) \tag{4}$$

5. Reciprocity Effect: While a student would like to be in harmony with their surroundings, they would be most influenced by their immediate friends whom they talk or spend time with. The strongest friendships would have the most impact on their opinion which can be measured as follows:

$$f_{friend-bias} = \frac{\sum_{j\in\eta(i)}(log(w^t_\mu(i,j)+1) + log(w^t_\mu(j,i)+1))a^t_{j,k}}{\sum_{j\in\eta(i)}(log(w^t_\mu(i,j)+1) + log(w^t_\mu(j,i)+1))} \tag{5}$$

6. Triads: Network and opinions could be impacted by the emergence of triads/-cliques. If u_i communicates or socializes with u_j and similarly u_j and u_k are friends, then u_i and u_k are more likely to communicate or influence each other beliefs either by agreeing or disagreeing. In order to capture the clique effect, we include the following:

$$f_{triads} = \sum_{q\neq j\in\eta(i)} (log(w^t_\mu(i,j)+1) + log(w^t_\mu(j,q)+1) + log(w^t_\mu(q,i)+1)) \tag{6}$$

7. Social Tie Persistence: Barring external shocks (such as disagreement or enrolling in same classes), typically, two friends tend to be consistent with their level of communication. In order, to account for such persistence between ties we include the strength of communication between two individuals:

$$f_{tie-persist} = log(w^t_\mu(i,j)+1) + log(w^t_\mu(j,i)+1) \tag{7}$$

8. Belief similarity: In principle, for a friendship to survive the two students should have similar beliefs to minimize conflict. We include this similarity between u_i and u_j based on their opinions across all beliefs. In this research, we use cosine similarity.

$$f_{same-beliefs} = sim(A^t_i, A^t_j) \tag{8}$$

We break down the problem definition into following sub-problems:

- **Sub-Problem 1 (Opinion Forecasting):** *Given the networks (G^t), opinions (A^t) and user attributes (X) for individuals at time t, how can we model and forecast persistence and change in opinions (A^{t+1}) at time $t+1$?*
- **Sub-Problem 2 (Edge Forecasting):** *Given the networks (G^t), opinions (A^t) and user attributes (X) for individuals at time t, how can we model and forecast*
  - **Sub-Problem 2.1 (Tie Strength)** *the strengths of ties and*
  - **Sub-Problem 2.2 (Persistence/Dissolution of Ties):** *the persistence and dissolution of ties in networks (G^{t+1}) at time $t+1$?*

Using ONE-M, in Section 5 we explore how the model can be used to study the opinion and network co-evolution and present novel insights about time-evolving networks. Further, we show its effectiveness to forecast opinions and changes of links.

4 Data

We use the NetSense data that we have collected at the University of Notre Dame. The dataset comprises of 199 freshmen students who joined in Fall 2011 and captures various facets of student opinions and interactions through periodic surveys and mobile phone monitoring for a 2 year period [28]. The participants were carefully selected to represent the general population. For our study, we leverage the survey, communication and collocation information as explained below:

1. Survey Data: Students were presented a survey at the beginning of each semester (summer included) beginning Fall 2011 to Summer 2013. A total of 6 surveys were conducted in the two year period where the students were requested to report their opinions on various beliefs such as premarital sex, euthanasia, death penalty, gay marriage, marijuana, political alignment towards the liberal party, abortion and homosexuality. In general, for each belief, the student could nominate as being against, unsure or in support of the belief. In addition to their views on such controversial topics, the students reported personal attributes such as age, gender, hometown and ethnicity. From among the 199 students, only 108 students participated in all 6 surveys and have been for the purpose of this study.

2. Communication Network (CN): During the two year period, as a part of the NetSense study, we also recorded the communication patterns among the students. Primarily, for all students in the study, we recorded call and message events. Using this information, we constructed a directed and weighted communication network, where each node is a student and the weight signifies the number of calls and messages sent from u_i to u_j aggregated across the semester.

3. Collocation Network (BN, WBN): In addition, to the call and message patterns, we also recorded bluetooth interactions between mobile devices of the participants. Each bluetooth interaction is associated with a signal strength, called the Received Signal Strength Indicator (RSSI). Higher RSSI value signify that the mobile devices were closer to each other which can safely be used as a proxy of in-person interactions [22]. To build a collocation network, we extracted the interactions which occurred more than 50 times to reduce the noise and ensure the quality of an edge [3]. The network was directed and weighted based on the number of times the two mobile devices were collocated. This network is referred to as the BN network. Next, to obtain a more representative and strong network of student interactions, we extracted the connections that occurred outside of school hours, typically outside of 9 AM and 5 PM on weekdays, and constructed a weekend-based network (WBN). We argue that WBN would be more representative of friendships over BN. BN could have edges that occur only due to students taking classes together, however a connection outside class hours indicates the students choosing to spend time together.

For all three networks, among the 108 students, we observe that only a subset of them are active in each semester. However, we observe a drop in the number of interactions for the summer semesters. We comment about the nature of these ties in Section 5.

5 Experiments

In this section we discuss ONE-M and its applicability for studying opinion and network co-evolution. We are primarily interested in answering the following questions:

Q1. **Discovering Evolution of Polarization in Networks**: How does polarization evolve? Are people changing their opinions or are they dropping unlike-minded friends?
Q2. **Studying Opinion and Network Co-Evolution**: Are people resistant to change in opinion? How do the unsure individuals differ from polarized population? Are the strong ties able to survive? or do they become in-active?
Q3. **Forecasting Persistence and Change in Opinions and Network**: How well does ONE-M work on real data for opinion forecasting and edge forecasting?

5.1 Discovering Evolution of Polarization in Networks

In order to study the presence/absence of polarization in dynamic networks, we study the communication patterns for the polarized users who are in-support and against a belief. The unsure/neutral individuals are dropped from this network as they are not polarized yet. A connection between individuals having the same opinion is called a homophilic edge. Conversely, a connection between students with different opinions is referred to as a cross edge.

We then conduct a test for polarization on each communication network snapshot (g_c^t). The test is based on the standard test for homophily in networks [11]. Let, p and q be the fraction of users against and for a belief, respectively. Then, the probability of a cross edge can be given by $\mu_0 = 2pq$ for a directed network. In order for polarization to exist, we expect to see the number of cross edges to be significantly lower than $2pq$ in the observed network (μ). Therefore, we test, $H_0 : \mu \geq \mu_0$ against $H_1 : \mu < \mu_0$. We present the results for polarization in Table 3.

Evidence of Polarization: In general, we witness *growing polarization* from Fall 2011 to Summer 2013 (as shown in Table 3). We also observe that topics such as abortion and premarital sex are most dividing across time. This makes sense because discussions about these topics are fervent at the Notre Dame campus which is a catholic university. Moreover, views on such topics can be governed by the cultural practices as observed in one's surrounding. For instance, an individual could more likely be in favor of marijuana usage if everyone around him/her are in favor of it. Further, we do not observe significant evidence of polarization for topics such as euthanasia and death penalty which could be defined as the "non-wedge" topics. Given the demographics of our participants, who are mostly aged between 17-19 years, we do not expect them to be discussing or have firm opinions on such topics.

Evolution of Polarization: While we do witness polarization in our data, we ask the next most natural question. Are people changing their belief over time causing the polarity of edges to flip (denoted as flipped cross edges) or are people simply dropping connections with people who do not share the same opinion as them (termed as dropped cross edges)? In order to answer this question, we present a breakdown of the evolving communication network in case of marijuana as shown in Table 4. On analyzing the evolution, we observe that the number of nodes in support of marijuana over time get much more active than their counterparts who consider marijuana not legal. Further, as communication transpires, we notice that at each time step we do not observe a lot of edges flipping the polarity, however we observe that most of the cross edges keep dropping in subsequent time intervals. We observe a similar pattern for the other beliefs too however due to space constraint we do not add those tables. Overall, we discover

Table 3: **Evidence of Polarization**: Results for each belief across 6 semesters using the communication network. ✓ indicates the presence of homophily at varying confidence levels: 99% (***), 95% (**), and 90%(*). We have color-coded based on significance for easy comprehension (darker shade indicate higher the confidence). × marks (or the absence of color) indicate that we do not have enough evidence to reject the null hypothesis (H_0).

	Fall 2011	**Spring 2012**	**Summer 2012**	**Fall 2012**	**Spring 2013**	**Summer 2013**
Premarital Sex	✓**	✓**	✓**	✓*	✓***	✓**
Gay Marriage	×	×	×	✓**	×	✓**
Marijuana	×	✓**	✓***	×	✓*	✓**
Political	✓*	✓*	×	✓**	✓**	✓***
Abortion	✓***	✓***	✓*	✓*	✓***	✓*
Homosexual	✓*	✓***	×	✓***	✓**	✓*
Euthanasia	×	×	×	×	×	×
Death Penalty	×	×	×	×	×	×

Table 4: **Evolution of Polarization**: Using marijuana usage as an example we break-down the evolution of the communication network among the students. In this illustration, we only consider the polarized individuals (Against Group: Not legal and For Group: Legal). Symbols mean the following: N: #nodes, E: #edges, N+/E+: #nodes/edges added at $t+1$, N-/E- :#nodes/edges dropped at $t+1$,

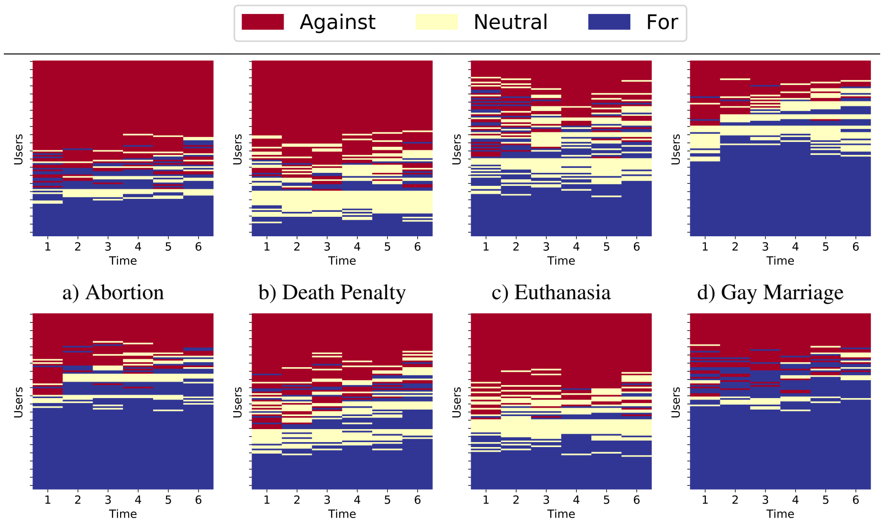
Time	**N**	**E**	**N+**	**N-**	**E+**	**E-**	**Against Nodes**	**For Nodes**	**Homo -philic Edges**	**Cross Edges**	**Flipped Cross Edges**	**Dropped Cross Edges**
Fall 2011	79	166	-	-	-	-	51	28	83	83	-	-
Spring 2012**	69	151	9	19	70	85	38	31	86	65	13	52
Summer 2012***	57	89	12	24	30	92	33	24	61	28	8	46
Fall 2012	70	122	21	8	60	27	37	33	56	66	0	13
Spring 2013*	67	124	12	15	49	47	31	36	69	55	12	28
Summer 2013**	54	100	6	19	34	58	19	35	64	36	0	34

that at the beginning of the study individuals start with casual friends but over time they drop ties (cross edges) and form more like-minded groups (or echo chambers).

5.2 Studying Opinion and Network Co-Evolution

We leverage ONE-M and explore opinion and network characteristics of individuals across the communication network and present the following key observations:

- **Stubbornness**: Given the controversial nature of the topics, we explore how resistant people are to changing their opinions. To that end, we leverage the Opinion Persistence factor of ONE-M. We present the persistent (and/or changing) opinions of individuals across the different beliefs in Figure 2. A user with same color coding for all semesters indicates that they did not change their opinion across the two year period. We observe that the polarized individuals (who are either against or for the belief) are more consistent with their opinions. Conversely, the unsure/neutral individuals frequently sway between the two extremes. Overall, we conclude that many individuals are stubborn about changing their opinions — being consistent with their own opinions is of more value than altering their opinion or network.
- **Bridges between Cliques**: We characterize the participants belonging in different opinion groups based on their tendency to form triads/cliques. In order to under-



a) Abortion b) Death Penalty c) Euthanasia d) Gay Marriage

e) Homosexuality f) Marijuana g) Liberal h) Premarital Sex

Fig. 2: **Stubbornness:** Figures a)-h) capture persistence and change in user opinions for 8 beliefs: abortion, death penalty, euthanasia, gay marriage, homosexuality, marijuana use, political alignment towards liberal group and premarital sex. For each figure, the 6 semesters are marked on the x-axis and each row represents a user. Red, yellow and blue indicate against, unsure/neutral and favor of the belief.

stand this phenomena, we leverage the Triads component of ONE-M. We compute a metric called *cliquishness* which captures the conditional probability of an individual forming a clique given their opinion of being against, unsure or in support of the belief. As shown in Figure 3, we observe that the tendency of forming cliques for both the extremes stay relatively constant across time. However, the unsure individuals not only have a lower probability of forming cliques but over time they are less likely to participate in cliques. We conclude that people with strong beliefs (i.e. individuals who are either for or against) tend to form cliques and follow their cliques over time. However, the unsure/neutral individuals have a lower tendency of participating in cliques and tend to act as bridges between the cliques.

– **Strong Ties Persist**: As reported earlier in Section 4, we observe that a lot of connections disappear in the summer months. While this is understandable as most students might not be present on campus, it is interesting to study the links that do persist. To that end, we study the Social Tie Persistence factor of ONE-M. As shown in Figure 4, we compare the strength of edges in the communication network between two consecutive semesters. As expected, we observe many edges to die in Figure 4b and e, however we observe that the links that do persist maintain the same strength of communication (falling close to the $45°$ line). We conclude that even though we observe many ties breaking between consecutive semesters, the ones that do survive are indeed the strong ties.

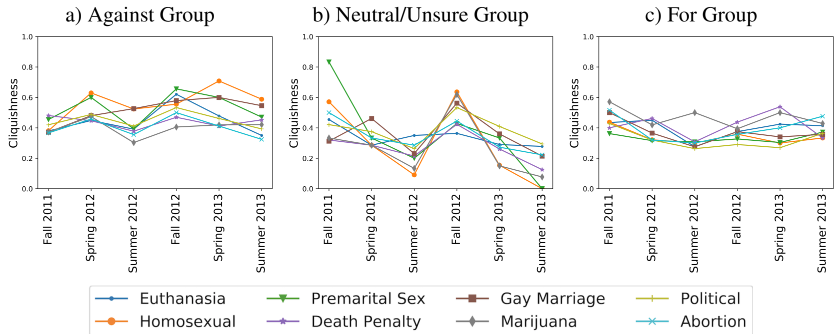


Fig. 3: **Bridges between Cliques**: People with strong beliefs have clique-ish ego networks ($\sim$ "echo chambers") whereas unsure people do not.

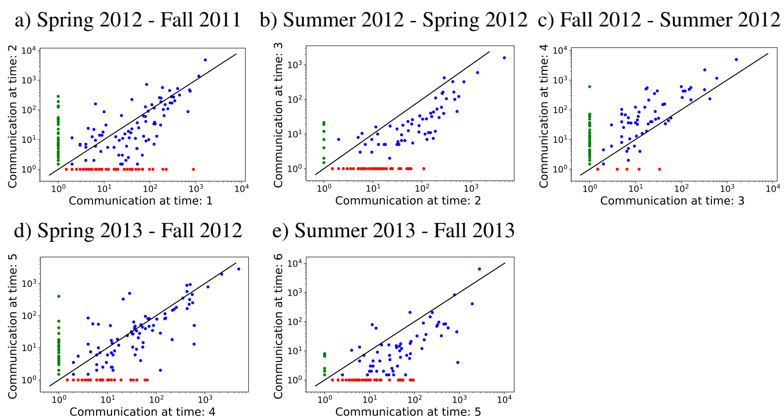


Fig. 4: **Strong Ties Persist**: Communication strength between $t+1$ (y-axis) and t (x-axis). Blue points: edges that persisted; Green points: new active edges; Red points: Dropped edges. Black line: $45°$. Observations: Communication drops in summer months, stays consistent between academic semesters.

5.3 Forecasting Persistence and Change in Opinions and Network

Sub-Problem 1: Opinion Forecasting We leverage ONE-M to forecast evolving opinions of individuals as described in Sub-Problem 1 (Section 3) — using networks (G^t), opinions (A^t) and user attributes (X) for individuals at time t, we forecast opinions (A^{t+1}) at time $t+1$. We address the problem as a classification task, where we apply ONE-M to extract features and predict $a_{k,i}^{t+1} \in \{1, 2, 3\}$. For opinion forecasting, we derive the following features: Opinion Persistence, Attribute Predisposition, Population Belief and Reciprocity Effect and learn their weights for the task. The weights for other features are set to 0.

We use ONE-M, to combine information across different network modalities, opinions and attributes, and study the following variations:

- **CN**: Extract features from the communication network (g_c^t) only.
- **BN**: Leverage the collocation network (g_b^t) only.
- **WBN**: Derive features from weekend-based collocation network (g_{wb}^t) only.

Table 5: **ONE-M forecasts opinions accurately**: the performance metric used is F1-score. We mark the best (bold and dark color) and the second best performing (underline and light color) model.

	ONE-M					Baselines		
	CN	**BN**	**WBN**	**CN+BN**	**CN+WBN**	**Yesterday**	**Majority**	**Random**
Premarital Sex	<u>0.79</u>	0.66	0.66	**0.81**	**0.81**	<u>0.79</u>	0.47	0.32
Gay Marriage	**0.65**	<u>0.54</u>	0.52	**0.65**	**0.65**	**0.65**	0.47	0.36
Marijuana	0.66	0.50	0.51	<u>0.69</u>	**0.70**	0.67	0.11	0.29
Political	<u>0.74</u>	0.71	0.71	<u>0.74</u>	**0.76**	0.56	0.23	0.36
Abortion	<u>0.57</u>	0.54	0.54	**0.59**	**0.59**	<u>0.57</u>	0.30	0.38
Homosexual	**0.77**	<u>0.65</u>	<u>0.65</u>	**0.77**	**0.77**	**0.77**	0.49	0.39
Euthanasia	0.56	**0.66**	0.45	<u>0.57</u>	0.53	0.53	0.25	0.32
Death penalty	**0.83**	0.68	0.71	<u>0.80</u>	0.65	0.62	0.46	0.44

- **CN+BN**: Derive features from CN (g_c^t) and BN (g_b^t).
- **CN + WBN**: Combine features from CN (g_c^t) and WBN (g_{wb}^t) networks.
- **Baselines**:
  - *Yesterday*: As we have previously seen, many individuals are stubborn and do not like to change their opinion. To incorporate for this effect, we propose a baseline which only includes Opinion Persistence.
  - *Majority*: The individual would select the opinion that was most popular in the previous time steps (training data).
  - *Random*: The individual would chose an opinion about a belief at random.

For the experiment setup, we train on data from G^1, G^2, G^3, and G^4 for time steps $1 \rightarrow 4$ and present results on performance for time step $5 \rightarrow 6$. We use logistic regression as the choice of classifier and F1-score as the performance metric.

Result: Based on the results in Table 5, we observe that for majority of the beliefs ONE-M achieves the best performance. We also observe that between the collocation (BN) and weekend-based collocation network (WBN), adding the latter provides a stronger signal and contributes more towards forecasting an individual's opinion. This makes sense because if two students are choosing to meet outside classes over the weekends they are more likely to influence each others opinions. We would also like to note the difference in the perception of some beliefs: while opinions about beliefs such as premarital sex might be more ingrained in an individual based on their religious value (explaining why the 'Yesterday' model based on opinion persistence performs better), opinions about marijuana usage and political views might evolve based on the cultural norms of the environment an individual is. Additionally, given that our population consists of students aged between 17-19 years, we do not expect them to discuss or have a firm opinion about the "non-wedge" issues such as death penalty and euthanasia (as also seen in Table 3). Overall, based on the performance of top two performing models (bold and underlined in Table 5), we conclude that ONE-M is able to forecast opinions and adding network features is of value for understanding evolving opinions. Moreover, we recommend using ONE-M with communication and collocation networks.

Sub-Problem 2: Edge Forecasting We evaluate the performance of ONE-M at forecasting information about the network connections as defined in sub-problem 2. To that end, we define experiments for both sub-problems:

Table 6: **ONE-M forecasts edges accurately**: for forecasting 1) tie strength 2) persistence/dissolution of a tie, we use mean-squared error (MSE) in experiment 1 (lower the better) and F1-score in experiment 2 (higher the better). We mark the best (bold and dark color) and the second-best (underline and light color) model.

	ONE-M					Baselines		
	CN	**BN**	**WBN**	**CN+BN**	**CN+WBN**	**Yesterday**	**Mean**	**Median**
Tie Strength (MSE)	2.50	3.59	3.34	2.48	**2.35**	<u>2.36</u>	4.21	4.00
	ONE-M					Baselines		
	CN	**BN**	**WBN**	**CN+BN**	**CN+WBN**	**Yesterday**	**Random**	
Persistence/Dissolution (F1-score)	**0.76**	0.61	0.65	0.72	<u>0.75</u>	<u>0.75</u>	0.52	

1. **Sub-Problem 2.1: Tie Strength:** Using features from G^t, A^t and X^t, forecast the strength of the communication between u_i and u_j at $t+1$, that is $log(w_c^{t+1}(i,j)+1)$. This can be treated as a standard regression problem.
2. **Sub-Problem 2.2: Persistence/Dissolution**: Using features from G^t, A^t and X^t, forecast whether an edge that exists between u_i and u_j at time t would persist or dissolve at time $t+1$ in the communication network (G^{t+1}). This can be considered as a classification problem with the target variable as $0/1$ class.

For both sub-problems, we derive the following features: Social Tie Persistence, Belief similarity, Attribute similarity and Triads and learn their weights for the task. The weights for other features are set to 0. For both experiments, we employ a series of ONE-M variations as described below:

- **CN**: Extract features from only the communication network (g_c^t).
- **BN**: Leverage only the collocation network (g_b^t).
- **WBN**: Derive features from weekend-based collocation network (g_{wb}^t) only.
- **CN+BN**: Derive features from CN (g_c^t) and BN (g_b^t).
- **CN+WBN**: Combine features from CN (g_c^t) and WBN (g_{wb}^t) networks.
- **Baseline**:
  - 'Yesterday': incorporates the persistence of the communication strength and edge by including only the Social Tie Persistence feature and excludes any network or opinion effects.
  - *For Tie Strength Forecasting*: we include two baselines. 'Mean' predicts the average tie strength, whereas 'Median' forecasts the median tie strength observed in the training data as the predicted strength.
  - *For Persistence/Dissolution Forecasting*: we include a 'Random' baseline for the edge persistence/dissolution experiment, which randomly assigns if a link persists or dissolves in the consecutive time step.

For both experiments, we train on data from G^1, G^2, G^3, and G^4 for time steps $1 \rightarrow 4$ and present results on performance for time step $5 \rightarrow 6$. For tie strength, we employ linear regression with mean squared error (MSE) as the performance measure. For persistence/dissolution, we use logistic regression with F1-score as the performance metric. The results for the both the experiments are listed in Table 6.

Result. Based on results in Table 6, we observe that ONE-M performs better for tie strength and persistence/dissolution experiment as compared to baselines. We observe that for predicting tie strength between two nodes, ONE-M using communication and collocation network based on weekends combined obtains the lowest error. For persistence/dissolution, we observe that CN wins. Between the two collocation networks, we again observe a strong signal from the WBN (as also seen from opinion forecasting experiment). Our ONE-M is able to leverage the connections between friends who are spending time with each other outside classes rather than random meetings for communication strength forecasting. We also notice that using simple social tie persistence ('Yesterday' model) is not enough to understand how connections persist and dissolve in a polarizing environment. Again, based on the results, we recommend using ONE-M with the communication and strong collocation networks that capture friendships for edge forecasting.

6 Conclusions

In summary, the contributions of our work are as follows:

- **Discovering Evolution of Polarization in Networks**: We discover (a) polarization indeed happens, for several topics and (b) people prefer to severe ties with disagreeing contacts, than change opinion.
- **Studying Opinion and Network Co-Evolution**: With our proposed ONE-M model, we made additional discoveries: "stubborness" (people keep their opinions); "bridges between cliques" (strong-opinion people tend to belong to clique-like groups); "strong ties persist" (despite summer breaks).
- **Forecasting Persistence and Change in Opinions and Network**: Our proposed model, ONE-M, outperforms baselines on forecasting accuracy, for beliefs as well as network changes.

7 Acknowledgements

This work is supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053 and by the National Science Foundation (NSF) Grant IIS-1447795.

References

1. Allen, J.P., Porter, M.R., McFarland, F.C.: Leaders and followers in adolescent close friendships: Susceptibility to peer influence as a predictor of risky behavior, friendship instability, and depression. Development and psychopathology 18(1), 155–172 (2006)
2. Badev, A.: Discrete games in endogenous networks: Equilibria and policy. arXiv preprint arXiv:1705.03137 (2017)
3. Bahulkar, A., et al.: Coevolution of a multilayer node-aligned network whose layers represent different social relations. Computational Social Networks 4(1), 11
4. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. science 286(5439), 509–512 (1999)
5. Bhawalkar, K., Gollapudi, S., Munagala, K.: Coevolutionary opinion formation games. In: STOC. pp. 41–50. ACM (2013)

6. Bilò, V., Fanelli, A., Moscardelli, L.: Opinion formation games with dynamic social influences. In: WINE. pp. 444–458. Springer (2016)
7. Clifford, P., Sudbury, A.: A model for spatial conflict. Biometrika 60(3), 581–588 (1973)
8. Das, A., Gollapudi, S., Munagala, K.: Modeling opinion dynamics in social networks. In: WSDM. pp. 403–412. ACM (2014)
9. DeGroot, M.H.: Reaching a consensus. Journal of the American Statistical Association 69(345), 118–121 (1974)
10. Durrett, R., Gleeson, J.P., Lloyd, A.L., Mucha, P.J., Shi, F., Sivakoff, D., Socolar, J.E., Varghese, C.: Graph fission in an evolving voter model. PNAS 109(10), 3682–3687 (2012)
11. Easley, D., Kleinberg, J.: Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press (2010)
12. Evans, W.N., Oates, W.E., Schwab, R.M.: Measuring peer group effects: A study of teenage behavior. Journal of Political Economy 100(5), 966–991 (1992)
13. Friedkin, N.E., Johnsen, E.C.: Social positions in influence networks. Social Networks 19(3), 209–222 (1997)
14. Gu, Y., Sun, Y., Gao, J.: The co-evolution model for social network evolving and opinion migration. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 175–184. ACM (2017)
15. Hegselmann, R., Krause, U., et al.: Opinion dynamics and bounded confidence models, analysis, and simulation. Journal of artificial societies and social simulation 5(3) (2002)
16. Holland, P.W., Leinhardt, S.: A dynamic model for social networks. Journal of Mathematical Sociology 5(1), 5–20 (1977)
17. Holley, R.A., Liggett, T.M.: Ergodic theorems for weakly interacting infinite systems and the voter model. The annals of probability pp. 643–663 (1975)
18. Holme, P., Newman, M.E.: Nonequilibrium phase transition in the coevolution of networks and opinions. Physical Review E 74(5), 056108 (2006)
19. Jackson, M.O.: A survey of network formation models: stability and efficiency. Group Formation in Economics: Networks, Clubs, and Coalitions pp. 11–49 (2005)
20. Jackson, M.O., Wolinsky, A.: A strategic model of social and economic networks. Journal of economic theory 71(1), 44–74 (1996)
21. Kim, M., Leskovec, J.: Multiplicative attribute graph model of real-world networks. Internet mathematics 8(1-2), 113–160 (2012)
22. Liu, S., Jiang, Y., Striegel, A.: Face-to-face proximity estimation using bluetooth on smartphones. TMC 13(4), 811–823 (2014)
23. Morris, M., Kretzschmar, M.: Concurrent partnerships and transmission dynamics in networks. Social Networks 17(3-4), 299–318 (1995)
24. Mossel, E., Sly, A., Tamuz, O.: Asymptotic learning on bayesian social networks. Probability Theory and Related Fields 158(1-2), 127–157 (2014)
25. Mossel, E., Sly, A., Tamuz, O.: Strategic learning and the topology of social networks. Econometrica 83(5), 1755–1794 (2015)
26. Mossel, E., Tamuz, O.: Opinion exchange dynamics. Probability Surveys 14, 155–204 (2017)
27. Snijders, T.A., Van de Bunt, G.G., Steglich, C.E.: Introduction to stochastic actor-based models for network dynamics. Social networks 32(1), 44–60 (2010)
28. Striegel, A., et al: Lessons learned from the netsense smartphone study. SIGCOMM Comput. Commun. Rev. 43(4), 51–56 (2013)
29. Wang, C., Hachen, D.S., Lizardo, O.: The co-evolution of communication networks and drinking behaviors. In: Proc AAAI Fall Symposium Series (2013)
30. Watts, A.: A dynamic model of network formation. Games and Economic Behavior 34(2), 331–341 (2001)