The Role of Community Structure in Opinion Cluster Formation

Ryan J. Hammer, Thomas W. Moore, Patrick D. Finley, and Robert J. Glass

Complex Adaptive Systems of Systems (CASoS) Engineering
Sandia National Laboratories
Albuquerque, New Mexico USA
http://www.sandia.gov/CasosEngineering/
{rhammer, tmoore, pdfinle, rjglass}@sandia.gov

Abstract. Opinion clustering arises from the collective behavior of a social network. We apply an Opinion Dynamics model to investigate opinion cluster formation in the presence of community structure. Opinion clustering is influenced by the properties of individuals (nodes) and network topology. We determine the sensitivity of opinion cluster formation to changes in node tolerance levels through parameter sweeps. We investigate the effect of network community structure through rewiring the network to lower the community structure. Tolerance variation modifies the effects of community structure on opinion clustering: higher values of tolerance lead to less distinct opinion clustering. Community structure is found to inhibit network wide clusters from forming. We claim that advancing understanding of the role of community structure in social networks can help lead to more informed and effective public health policy.

Keywords: Social networks, community structure, complex networks, opinion dynamics, agent-based models, complex systems.

1 Introduction

Network science plays an important role in public health policy. Social, contact and organizational networks have been shown to affect various aspects of health. Public health researchers increasingly apply social network research to population health problems [1] in areas of infectious disease propagation [2], obesity [3], smoking [4], and even happiness [5].

Understanding the effects of social network topology is essential to crafting optimally effective public health policy. Social network research is increasingly focusing on the presence and effects of community structure within various networks. Community structure refers to heterogeneous degree distributions that result in groups of nodes more densely connected to each other than to the rest of the network [6]. This understanding of community structure can be traced back to Granovetter's research which found that two nodes having friendship relationships with a third node are more likely to be connected to each other [7]. Community structure is also a factor in homophily, where nodes with similar characteristics tend to be more highly connected to one another than nodes with dissimilar characteristics [8]. Community

2012-9938: Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

structure has been shown to play an important role in the dynamics of networks through social diffusion [9].

Opinion Dynamics (OD) is a powerful social-network modeling technique integrating research findings from sociology and statistical physics. Conceptually OD traces its origins to Cartwright and Harary's structural balance theory which posits that an individual's opinion regarding another person or idea is influenced by those with whom he/she shares positive social ties [10]. The mathematics and algorithms of OD are derived from the Ising spin model which captures spin alignment of adjacent particles within a lattice. OD models extend the particle interaction concepts of statistical physics to include structural balance theory to produce a generalizable method of modeling the flow of ideas, opinions and concepts through social networks [11].

Several different implementations of OD models have been proposed; for a comprehensive review, see Castellano, Fortunato and Loreto [11]. The basic assumptions of opinion dynamics can be extended to capture various facets of social dynamics. Some of these modifications include continuous-valued states, bounded confidence models using tolerance, and network moderated interactions to introduce social structure into the model. Opinion dynamics initially provided insight into the fundamental dynamics of information spread in networks and well-mixed populations. More recently, OD has been used to investigate the diffusion of agricultural practices in farming communities [12], the formation of extremist groups in larger communities [13], and the effects of influence-based interventions on differential social structures, including gendered networks [14].

Opinion dynamics is useful in studying the formation of opinion clusters. Opinion clusters are an emergent result of the collective interactions of people within a network influenced by their individual characteristics and the network structure. Various groups in nature display clustering behavior including flocks of birds [15], bacterial colonies [16], fish schools [17], and human behavior such as walking patterns [18]. Recent work that has highlighted the effect of opinion clusters on a system includes studies that show how clusters of unvaccinated individuals can lead to a dramatic increase in disease outbreak probability [19]. In a study of individual characteristics affecting opinion cluster formation, Schelling presented an exhaustive study using a spatial proximity model [20]. Studies on opinion cluster formation within adaptive networks [21] have elucidated network influences on opinion clusters.

In this paper, we address the role of individual node properties and the role of network topology on opinion cluster formation within and among communities. Section 2 documents the model formulation used in this investigation. Section 3 describes the experiments we performed and the results from these experiments. Section 4 presents a discussion of potential applications of our findings to public health.

2 Model Formulation

An agent-based model is used to investigate the influence of individual and network characteristics on the formation of opinion clusters. We use a modified version of the opinion dynamics model of Deffuant and Weisbuch to model the flow of opinions in a network of heterogeneous agents. The Deffuant-Weisbuch model simulates the spread of opinions in a well-mixed population [12]. We modify the original model by mapping it to a directed network of agents. This directed network represents relationship ties such as friendship nominations where the directionality indicates non-reciprocal nominations. Incorporating edge directionality is supported by studies showing that friendship networks are often directed [22]. In addition, empirical studies on network-based properties of tobacco use have identified correlations incorporating directionality [23].

Agents have an individual opinion, modeled as a continuous variable in the range [0.0, 1.0], and a tolerance value indicating how open an agent is to the opinions of others. The tolerance value constrains opinion-changing interactions to agents whose opinions are within a tolerance bound. If the difference in opinion between two agents is less than the tolerance value, the agents can influence one another, incrementally changing opinions to become more similar to each other. Henceforth, the term node shall refer to an agent. At each time step the opinion of each node is updated using the following equation:

$$x_i(t+1) = x_i(t) + \frac{1}{|S_i|} \sum_{j \in S_i} \mu_{ij} \left[x_j(t) - x_i(t) \right] . \qquad \textbf{(1)}$$

In equation (1), $x_i(t+1)$ represents the opinion of node i at the next time step. The opinion is updated by adding to node i's current opinion the average difference between node i's opinion and that of every one of its neighbors, $x_j(t)$, at the current time step t. If the difference in opinion between node i and a neighbor is greater than the tolerance bound, the two do not influence one another. An edge weight, μ_{ij}, allows for giving certain friendships more influence as might be the case for a family member or a close friend.

Tolerance is an important feature of the model to study opinion cluster formation. As the model execution proceeds, opinions of nodes in various portions of the network tend to converge to common local mean values with the number and average size of the clusters determined by tolerance [12]. Higher tolerance values generate fewer, larger clusters. Opinion clusters emerge from the network as a result of the constraint that tolerance places on the number of interactions that can take place. These clusters consist of groups of neighboring nodes with similar opinions. To be considered as part of a given opinion cluster, a node must both be reachable by every node in the cluster and hold an opinion within tolerance bounds of the most proximal nodes in the cluster to which it is connected.

Opinion cluster formation is also influenced by network structure: node interactions only occur if an edge exists. Random networks, such as Erdős-Rényi graphs, exhibit less distinct clustering properties than do scale free networks (Moore, et al, 2012 in review).

The algorithm employed for cluster detection is a variant of the DBSCAN algorithm [24]. We have modified the algorithm to map it onto a network. The algorithm operates by mapping nodes to clusters if they are within the tolerance bounds of the initial cluster node and reachable via an edge of the node. This algorithm also removes the need to know a priori the number of clusters involved, which is most suitable for our application.

We investigate the influence of tolerance-constrained interactions in a network containing community structure by generating 250-node networks comprised of five communities. We vary tolerance over the range [0, 0.5], and examine the formation of opinion clusters within and among communities. To investigate the contribution of community structure in the network, we decrease the community structure by increasing the number of edges between communities, adding random edges between nodes in different communities.

3 Experiments

We conducted two experiments to study the effects of individual-level constraints and network-level constraints on the formation of opinion clusters: one to study the effects of tolerance level and one to study the effects of network topology.

We use a similar network topology in both experiments. Our primary criterion for the network structure is that distinct community structure exists. We create our network by generating 5 communities using the Erdős–Rényi model [25]. Each community consists of 50 nodes connected using an edge probability $p = 0.1633$ resulting in 400 expected edges within each community.

Once constructed, we randomly connect each individual community to every other community with a specified number of edges depending on the experiment (Figure 1). We don't claim that this graph formation process generates networks representative of those in the real world, only that the generated networks contain distinct community structure, the condition under which we are interested in studying opinion clusters. Future studies will elicit the effects of using other network formation models more demonstrative of real-world topology.

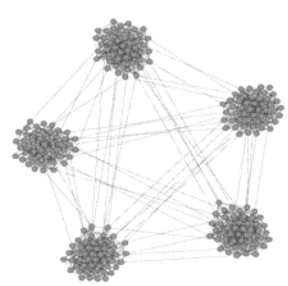

Fig. 1. An example network displaying community structure in which five densely connected communities are connected by more sparse inter-group edges.

3.1.1 Tolerance Experiment

The OD model is run using networks with community structure. The tolerance of each node is varied to determine the effect of tolerance on opinion cluster formation. For this experiment networks are generated with five distinct communities (Figure 1). Each community is connected to every other community by adding 25 edges at random between each community for a total of 250 additional edges. Using the modularity metric proposed by Newman and Girvan [26] to provide a measurement of community structure, this network produces a modularity of ~0.72 indicating a high level of community structure. Initial opinion values of the nodes are uniformly distributed at the outset of the model run. Tolerance is increased over the range [0, 0.5] in a series of simulations. For each simulation, a different stochastic realization of the example network is generated. For each tolerance value, the model is run with 100 stochastically-generated networks.

We investigate two things. First, as tolerance increases, at what tolerance value does each community form a single majority-opinion cluster? Second, at what tolerance value does the entire network form a single majority-opinion cluster? In this context, a single-majority cluster contains a large percentage of the nodes in the relevant community or network. It is rare that clusters are entirely defined by communities as network topology can cause certain nodes to be drawn into a different community.

3.1.2 Tolerance Results

As illustrated in Figure 2, at a tolerance value of 0.0 every node maintains its baseline opinion; no opinion adjustment can take place. At this value each node forms its own opinion cluster. As tolerance increases, the number of clusters decreases while the average cluster size increases.

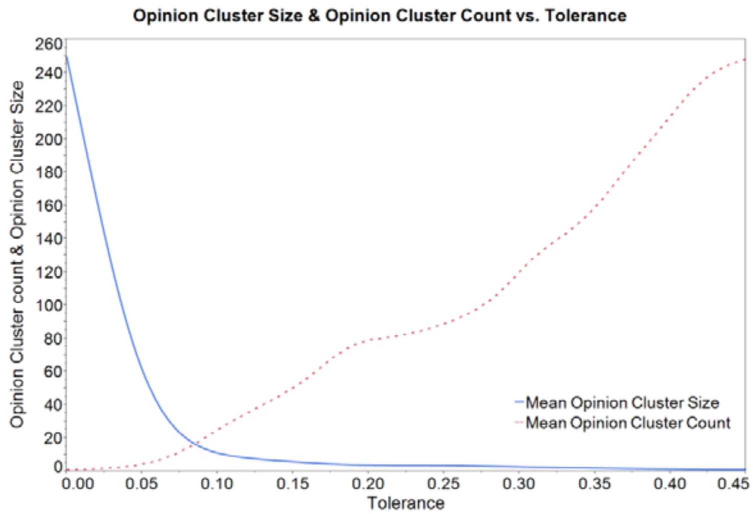

Fig. 2. Tolerance values affect both mean cluster count and cluster size. As tolerance increases the number of opinion clusters falls and the size of the clusters rises.

Once tolerance has increased to a certain level, each of the five communities emerges as a single opinion cluster, as seen in Figure 3. Individual node opinions for a single community are plotted relative to tolerance values. As tolerance increases, node opinions begin to draw closer together. Finally, at a tolerance value of ~0.27 the nodes in the community converge to a single majority-opinion cluster.

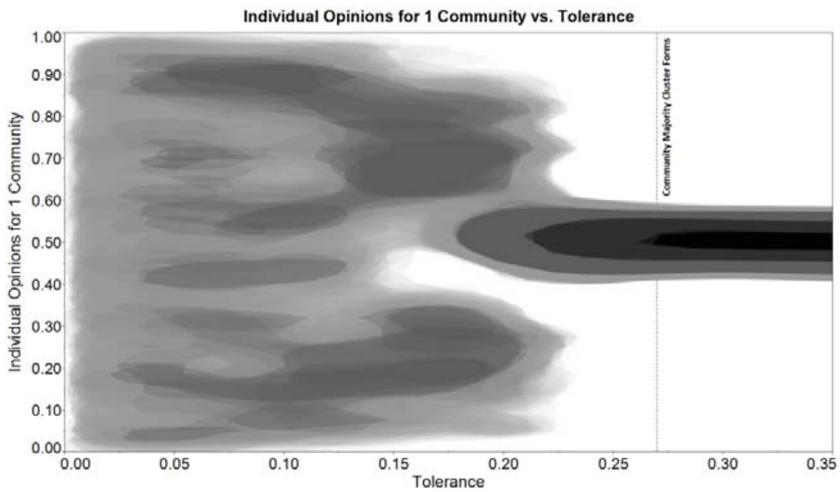

Fig. 3. Density plot shows the effect of tolerance on equilibrium distribution of opinion in a single community. Opinion distributions develop clustering patterns at a tolerance value of 0.05, become bimodally distributed at approximately 0.15 and coalesce to a single mean-value cluster at a tolerance of approximately 0.27.

Figure 4 shows a decrease in the standard deviation of node opinions for each community as tolerance increases. A sharp drop can be seen between tolerance values of ~0.14 to ~0.25, indicating minimal node interaction and opinion adjustment at low tolerance levels.

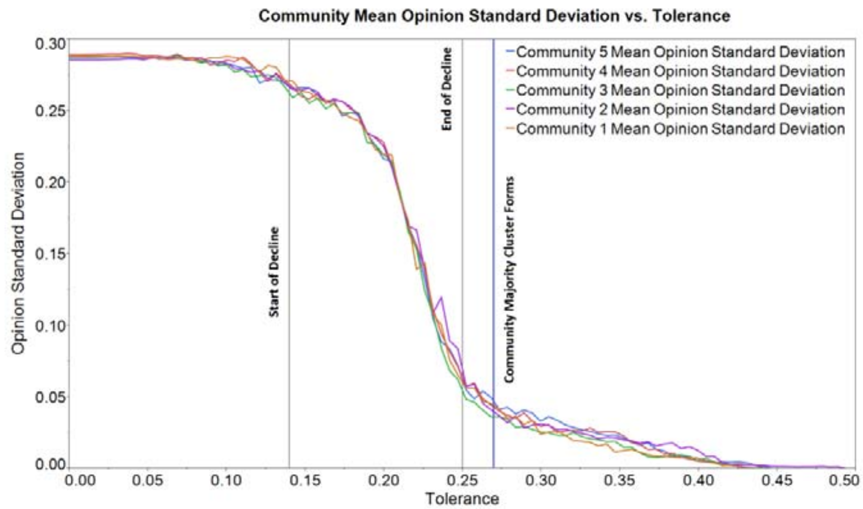

Fig. 4. Standard deviation of opinion distribution for each community. Note the transition in the rate of opinion deviation decline at 0.25. The tolerance value at which a community majority-opinion cluster forms (0.27) is highlighted by the blue line.

At progressively higher tolerance values, the clusters begin to merge together into larger clusters until eventually the entire network coalesces into a single-opinion cluster. Each community reaches a single majority-opinion cluster at a tolerance value of ~0.27, while coalescence into a single network-wide opinion cluster occurs at a tolerance value of ~0.45 (Figure 5). The top portion of the two-part figure plots each individual opinion in the entire network at tolerance values from 0.0 to 0.5. The bottom portion of the figure shows a plot of the standard deviation in opinion for the entire network across the same range of tolerance values. Two rough transitions can be seen in the top portion. A transition occurs at the tolerance value of ~0.27, the point at which communities form single-majority clusters as demonstrated in previous figures. The final transition to a single majority-opinion cluster occurs at a tolerance of ~0.45, a finding verified by the plot illustrated in the bottom half of Figure 5 in which the standard deviation reaches a minimum at a tolerance value of ~0.45.

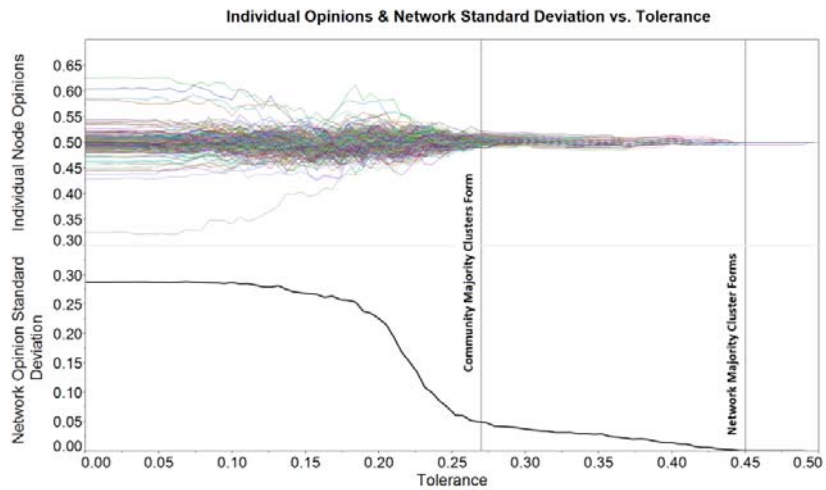

Fig. 5. Individual node opinions for the entire network plotted relative to tolerance values (above) are shown in the context of the standard deviation of opinion for the entire network (below). The individual node opinions come into consensus at tolerance ~0.45 indicating the formation of the single majority-opinion network cluster.

3.2.1 Topology Experiment

We examine the role of topology in opinion cluster formation, specifically in regard to the structure of communities. The same process of generating a network with five distinct communities is utilized. Based on our findings from the previous experiment, every node is assigned a constant tolerance value of 0.27, the value at which distinct community-wide opinion clusters form. Unlike the uniform distribution used in the previous experiment, here nodes in each community draw an initial opinion from one of five distinct opinion intervals ranging from low (0.00, 0.12) to high (0.88, 1.00). These intervals are defined in Table 1.

Table 1. Opinion ranges assigned to each community for topology experiment.

Community 1	[0.00, 0.12]
Community 2	[0.22, 0.34]
Community 3	[0.44, 0.56]
Community 4	[0.66, 0.78]
Community 5	[0.88, 1.00]

The heterogeneous assignment of community opinion among the different communities and a homogenous assignment within each community serves to illustrate model operation in community networks which more realistically reflect the homophily to be expected in social networks. Where communities hold distinctly separate opinion intervals, the effect of blurring the community structure can be seen more clearly. One side effect of community-specific opinion distribution is that the communities come to consensus very quickly within themselves.

The number of edges that connect the various communities is increased in a series of 100 runs starting with 0 edges and finishing with 250 edges. For each number of edge connections, 100 stochastic network realizations are modeled. Each edge increment lowers the amount of community structure until finally the network is a single densely connected community.

We examine edge ratios of between-community edges to within-community edges to identify the relationship at which communities will converge to a single majority-opinion cluster. That is, we ask to what degree community structure needs to be degraded to allow a majority cluster to form. Rather than varying tolerance to investigate cluster formation, we use network topology.

3.2.2 Topology Results

The modularity metric of Newman and Girvan is again used to measure the degree of degradation of community structure. A plot of modularity versus the ratio of between-community edges to within-community edges, presented in Figure 6, demonstrates that as the ratio increases the community structure decreases. In light of the degradation of community structure, both the modularity and the edge ratio can be analyzed at the point at which the single majority network cluster forms. Two sharp transitions in modularity can be seen in this figure: the first steep drop occurs in the edge ratio range of ~0.11 to ~0.20, the second in the edge ratio range of ~0.27 to ~0.31. This provides an understanding of what is happening to community structure as edges are added and at what edge ratio the community structure is obscured, giving us a clear picture of the modularity of the network when the communities converge.

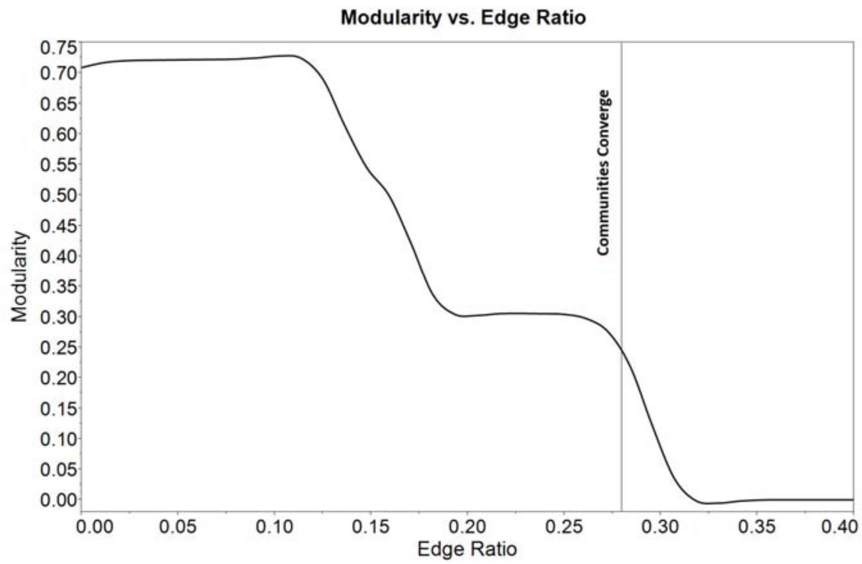

Fig. 6. Modularity plotted versus the ratio of between-community edges to within-community edges. As edges are added between the communities, the community structure fades. At the ratio of ~0.31, community structure has been eliminated.

A broader look at the changing relationships within each of the five communities can be seen in Figure 7 where the five communities' average opinions are plotted relative to edge ratios (top graph) as are the standard deviation of opinion for each community (lower graph). The mean opinion for each community starts to draw together at an edge ratio of ~0.065 and moves together more sharply at ~0.28. The standard deviation graph indicates that the communities are in consensus from the start based on the assigned tolerance value and the initial distribution range for each.

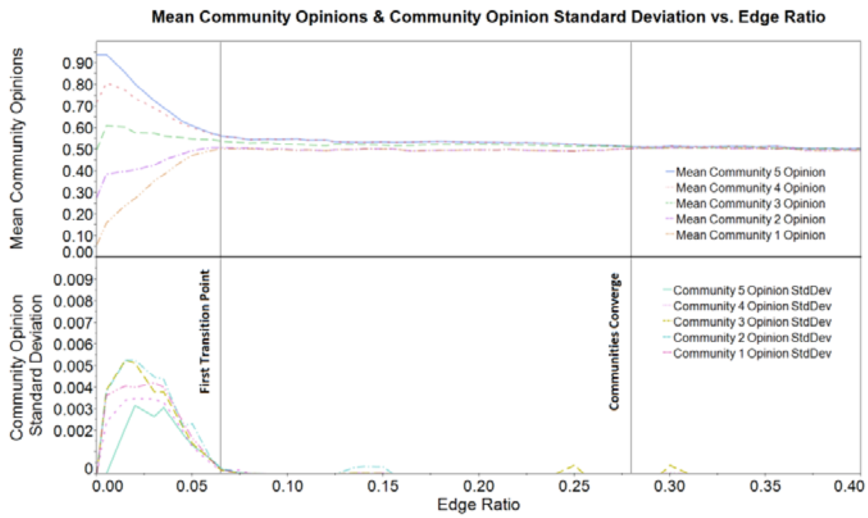

Fig. 7. Effect of between-community to within-community edge ratio on mean community opinion and standard deviation. The mean community opinion (upper) plot shows consensus among the five communities commencing at ~0.065 with near complete consensus at ~0.28. The standard deviation (lower) plot shows the communities at consensus amongst themselves at the onset.

The standard deviation in opinion for the network can be observed in Figure 8. Two transition points can be seen: the first occurs at the edge ratio of ~0.08 where the first steep drop ends; the second transition can be seen at the edge ratio of ~0.30 when the standard deviation finally levels out near 0. This provides further evidence that the communities have converged and the network has formed a majority cluster near an edge ratio of ~0.28. As seen earlier in Figure 6, the last steep drop of modularity concludes at the edge ratio of ~0.31 which highlights the fact that the network can converge once community structure has been obscured to a high enough degree. In this case, the convergence takes place when structure has been obscured to the point of an edge ratio of ~0.28 (as seen in Figure 7) and a modularity value near 0. In terms of this network, a ratio of 0.28 indicates that a total of 2000 edges exist within the communities and a total 560 edges connect the communities.

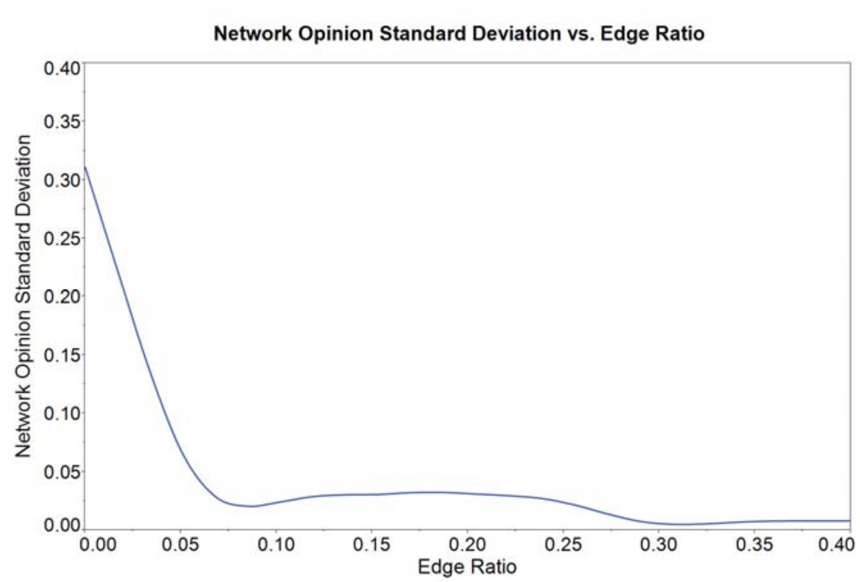

Fig. 8. Effect of in-community to inter-community edge ratio on overall network standard deviation.

The opinion cluster size should also level out at ~0.28. We observe this in Figure 9. Additionally it is of note in Figure 9 how quickly cluster size grows with just a small increase in the edge ratio. However, in order to achieve a majority cluster very close to the total number of nodes, an edge ratio ~0.28 must exist.

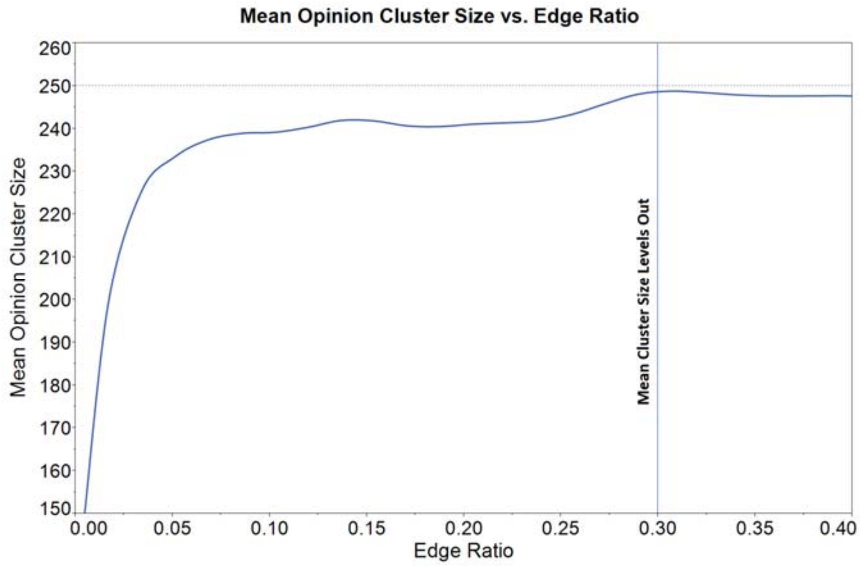

Fig. 9. Effect of between-community to within-community edge ratio on mean opinion cluster size. Mean cluster size can be seen to level out in the vicinity of an edge ratio of ~0.30, very near the point at which community opinions converged in the previous results (0.28).

These results indicate that community structure plays a role in the attainment of network-wide consensus. Even though communities are able to come to internal consensus, the community structure must be diminished to a certain degree in order to overcome the network-level constraint placed on opinion cluster formation. The results shown indicate that once an edge ratio of ~0.28 has been reached, the communities converge to form a single majority opinion cluster.

4 Discussion

We have presented an opinion dynamics model highlighting the role that community structure plays in the formation of opinion clusters. Using the individual constraint of tolerance, communities form single majority-opinion clusters at the relatively low tolerance value of ~0.27. A majority network-wide cluster forms at a tolerance value of ~0.45. We also found that community structure can inhibit the formation of a majority network opinion cluster when a constant tolerance value is imposed on a level which constrains the communities to single majority-opinion clusters. To overcome this network-level constraint, the community structure must be obscured to a certain degree by the establishment of inter-community relationships. In the case of the given topology, an edge ratio of between-community edges to within-community edges needed to be ~0.28 in order to overcome network constraints. This indicates that tolerance can be driven by means other than individual characteristics. Simply increasing the ratio of inter-community edges to intra-community edges promotes the formation of a majority opinion cluster.

Understanding the role that community structure plays in the spread of opinions or behaviors may be important to designing and implementing effective public health policies. The results we have presented can be used to gain insight into how communities in social networks will respond to policies. For example, consider the issue of implementing a public health policy in a high school. In high schools, community structure takes the form of cliques involving different groups of students. For a policy intervention to reach every group of students, the constraints implied by individual- and network-level characteristics must be taken into account. Students in marginalized communities often occupy peripheral positions on social networks and are excluded from participation in core communities by both individual differences and by the network of relationships connecting individuals. A more comprehensive understanding of these fundamental influences can help refine the design of field-based studies to elicit empirical data on dynamic network formation in high schools, and can ultimately contribute to more effective policy formation.

References

1. K. P. Smith and N. A. Christakis, "Social networks and health," *Annual review of sociology*, vol. 34, pp. 405–429, 2008.
2. L. M. Glass and R. J. Glass, "Social contact networks for the spread of pandemic influenza in children and teenagers," *BMC Public Health*, vol. 8, no. 1, p. 61, 2008.
3. N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, vol. 357, no. 4, p. 370, 2007.
4. T. W. Moore, P. D. Finley, J. M. Linebarger, A. V. Outkin, S. J. Verzi, N. S. Brodsky, D. C. Cannon, A. A. Zagonel, and R. J. Glass, "Extending Opinion Dynamics to Model Public Health Problems and Analyze Public Policy Interventions," in *Eighth International Conference on Complex Systems*, Quincy, MA, 2011.
5. J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study," *British Medical Journal*, vol. 337, no. dec04 2, p. a2338, 2008.
6. M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821 –7826, Jun. 2002.
7. M. Granovetter, "The strength of weak ties: A network theory revisited," Sociological theory, vol. 1, pp. 201–233, 1983.
8. M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, vol. 27, no. 1, pp. 415–444, 2001.
9. R. Colbaugh and K. Glass, "Predictive analysis for social diffusion: The role of network communities," *Arxiv preprint arXiv:0912.5242*, 2009.
10. D. Cartwright and F. Harary, "Structural Balance: A Generalization of Heider's Theory," *Psychological Review*, vol. 63, no. 5, 1956.
11. C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Reviews of modern physics*, vol. 81, no. 2, pp. 591–646, 2009.
12. G. Weisbuch, G. Deffuant, F. Amblard, and J. P. Nadal, "Meet, discuss, and segregate!," *Complexity*, vol. 7, no. 3, pp. 55–63, 2002.

13. G. Deffuant, “Comparing extremism propagation patterns in continuous opinion models,” *Journal of Artificial Societies and Social Simulation*, vol. 9, no. 3, 2006.
14. T. Moore, P. Finley, R. Hammer, and R. Glass, “Opinion dynamics in gendered social networks: an examination of female engagement teams in Afghanistan,” *Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 69–77, 2012.
15. M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, and A. Procaccini, “Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1232–1237, 2008.
16. E. Ben-Jacob, I. Cohen, and D. L. Gutnick, “Cooperative organization of bacterial colonies: from genotype to morphotype,” *Annual Reviews in Microbiology*, vol. 52, no. 1, pp. 779–806, 1998.
17. J. Gautrais, C. Jost, and G. Theraulaz, “Key behavioural factors in a self-organised fish school model,” in *Annales Zoologici Fennici*, 2008, vol. 45, pp. 415–428.
18. M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, “The walking behaviour of pedestrian social groups and its impact on crowd dynamics,” *PloS one*, vol. 5, no. 4, p. e10047, 2010.
19. M. Salathé and S. Bonhoeffer, “The effect of opinion clustering on disease outbreaks,” *Journal of The Royal Society Interface*, vol. 5, no. 29, pp. 1505–1508, 2008.
20. T. C. Schelling, “Dynamic models of segregation†,” *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.
21. B. Kozma and A. Barrat, “Consensus formation on adaptive networks,” *Phys. Rev. E*, vol. 77, no. 1, Jan. 2008.
22. J. Scott, *Social Network Analysis: A Handbook*. London: Sage Publications Ltd., 2000.
23. N. A. Christakis and J. H. Fowler, “The collective dynamics of smoking in a large social network,” *New England Journal of Medicine*, vol. 358, no. 21, p. 2249, 2008.
24. M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining*, 1996, vol. 1996, pp. 226–231.
25. P. Erdős and A. Rényi, “On the evolution of random graphs,” *Magyar Tud. Akad. Mat. Kutató Int. Közl*, vol. 5, pp. 17–61, 1960.
26. M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical review E*, vol. 69, no. 2, p. 026113, 2004.