5th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Demo Session, Anchorage, AL, August 2019

Modeling and predicting human social behavior

Ashwin Bahulkar*, Boleslaw K. Szymanski*, Kevin Chan† and Omar Lizardo‡
*Department of Computer Science and Network Science and Technology Center
Rensselaer Polytechnic Institute, 110 8th Street, Troy NY 12180, USA
Email: bahula@rpi.edu, szymab@rpi.edu
†US Army Research Laboratory, Adelphi, MD 20783, USA
Email: kevin.s.chan.civ@mail.mil
‡Department of Sociology, UC Los Angeles, Los Angeles, CA 90095
Email: olizardo@ucla.soc.edu

I. Project goals and team

The team consists of Dr. Omar Lizardo of the Department of Sociology at University of California, Los Angeles, Dr. Kevin Chan of Army Research Laboratory as well as Dr. Boleslaw Szymanski and Ashwin Bahulkar of the Department of Computer Science at Rensselaer Polytechnic Institute, but included collaboration with teams lead by Dr. Nitesh Chawla of the Department of Computer Science at the Notre Dame University, and Dr. Christos Faloutsos of Carnegie Mellon School of Computer Science. The overall project goal is to model evolving attribute-rich social systems. Accordingly the most of the work of the team has been focused on modeling and predicting change in social and human networks. Previous work establishes how node attributes influence edge formation and dissolution over time [4] [3] [5] and how interactions patterns can predict the lifetime of edges [2]. Current work focuses on modeling the formation of social interacting groups and exploring how the formation of these social groups is related to opinion change of persons [1].

II. Introduction

Previous work on human social networks has established a number of well-document regularities. First, most of human social interaction occurs in the context of a set of relatively small face-to-face groups and not in isolated dyads [9], [14], [16], [18]. Even the rise of telecommunications technology allowing for connectivity without face-to-face interaction has not altered this. Instead, interaction via such channels like texting, online chatting, social media complements rather than replaces face-to-face interaction in small groups, as the great majority of “online” ties ultimately become (or emerge from) offline face-to-face ties [10].

Second, interaction in small groups has implications for the society-wide distribution of beliefs and opinions [12], [14]. In particular, individuals are more likely to belong to groups in which other individuals share their beliefs and opinions [9]. Dyadic ties within the context of small groups tend to be “homophilous” with respect to a variety of individual attributes, such as race, diffusion, religion, place of residence, and the like [11]. These structural commonalities then lead to individuals adopting similar beliefs and practices, via processes of structurally channeled diffusion, such as the spread of an opinion among members of a given racial group or education stratum [13], coupled with dynamics of opinion adjustment and social influence when people interact with similar others in the context of small groups [7], [16]. It follows that cross-individuals differences and commonalities in opinions are largely a function of the groups they belong to, and dedicate the bulk of their time and attention. In this way, group based interaction contributes to both processes of cultural convergence, and most importantly, cultural differentiation, including polarization and segregation, at the level of society [12], [13].

Yet, is is also well established that, rather than being relatively static, social connections within face-to-face groups are dynamic, not static [17]. That is, rather than being “locked-in” into a small set of groups (as are some social non-human primates), humans have the capacity to choose which groups to join, and are also likely to choose when to leave a group, especially in the context of contemporary liberal societies. This means that the group-based nature of social interaction may not only be a key factor explaining individual differences of opinion and belief in the cross-section, but may also be crucial to accounting for temporal patterns of opinion change. The opinions that an individual holds at any one time, in their turn, may also figure prominently in their decisions to either join a new group or exit an existing one. More specifically, these decisions may be driven by the extent to which an individual faces a trade-off between adjusting their current opinion to the (possibly changing) opinions of the others in a group to which they are strongly committed, or decides to keep their opinion and join a group of like-minded others.

Keeping this interdependence in mind, we introduce a model which can predict both the formation and dissolution of memberships in social groups, as well as prediction of opinion changes made by the people to improve their level of satisfaction they achieve from group membership. The novelty of the model is that it can predict all the different kinds of interrelated changes people make in their social lives. Accordingly, for each person, we predict the following possible changes: joining a new group, leaving one of the groups to which this person belongs, and changing an opinion on one of the issues import to this person. The basis for the model prediction is the level of social benefits that a person derives from group membership. These benefits are measured in terms of the amount of time that a person spends with the group, and the level of similarity (homophily) of opinions on issues of importance to this person to opinions of other members of the group. A person deriving a low social benefit from group interactions might decide to either leave the group, or perhaps change her own opinion in order to improve her benefits from group membership. Our model learns the specific weights of these and other metrics using the records of past changes made by people as recorded in the training set. After the model is trained it can be used to predicts future changes.

The utility of this work is in finding how people with various opinions form stable groups in organizations and how these groups can influence the opinions of their members over a period of time. While this study has been performed only on a certain section of society within a particular age group, the methodology developed by our research can be applied to different demographics to help policy makers identify social barriers which might exist within organizations.

III. Data, model and results

We use the NetSense data that have been collected at the University of Notre Dame. The dataset contains records of nearly 200 students who joined in Fall 2011 as freshmen. This data set contains students' opinions on various subjects collected via periodic surveys and records of mobile phone activities collected over a two year period [6]. The attempt was made to select for the study a representative sample of the general student population. In our study, we use this data set in the following ways:

Survey data: Students were presented a survey at the beginning of each semester beginning Fall 2011 to Spring 2013 (summers were excluded due to small number participants). A total of four surveys were collected in the two year period in which the students reported their opinions on various beliefs such as gay marriages, marijuana legalization, political orientation, abortion and homosexuality.

Communication network: During the two year period, as a part of the NetSense study, the communication patterns among the students were collected. Call and message events were recorded for all students in the study. Using this information, we constructed a directed and unweighted communication network, in which each student was represented as a node and the edge was drawn between students when the number of calls and messages between those students aggregated over the semester exceeded a small threshold.

Collocation network: In addition to the call and message patterns, the data contains also records of bluetooth interactions between mobile devices of the participants. 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. The network was directed and weighted based on the number of times the two mobile devices were collocated.

Using the collocation network based on Bluetooth interactions between students, we identify social groups within a particular time frame using the hierarchical clustering method. This method was introduced by us in [1] and applied to the same Bluetooth interactions. The method identifies connected components of nodes over short periods of time, ranging from 10 minutes to one hour. These connected components are further merged using hierarchical clustering with additional constraints. The constraints are based on two postulates (axioms) for a well-formed group: sufficient *intersection* of potential members at all meetings, and sufficient *attendance* of each potential member in the meetings. Intersection postulate requires that two groups are merged into one if and only if the number of intersecting members of the two groups is above a certain threshold. Similarly, attendance postulate requires that each member in a group needs to have a certain level of attendance, exceeding another threshold, to be considered a part of the group. Using this method, we extract stable social groups formed in each semester. This is achieved by tuning up the thresholds for both postulates to achieve the average size of groups and the average number of groups to which a person belongs consistent with those reported in the sociological studies of groups. With the selected values for these thresholds, we observe that the average number of groups to which a node belongs is around four to five, while majority of groups include from five to six members.

Our goal is to predict which all of many potential changes each person in the network will make. In our model, people can make the following types of changes: changing opinion on an issue, joining a new group, and leaving a group to which the person currently belongs. We propose a benefit driven model in which the set of changes predicted to be made maximizes the benefits that a person derives from groups membership. We define benefit as a function of the interactions the person has with the group, the similarity the person has with other group members on various opinions and the number of groups the person is part of, where weights between these change options are defined by model parameters. The optimal values of model parameters are computed by minimizing penalty for changes predicted by the model but not made in reality and for changes not predicted as made by the model but made by the node in reality. There is no penalty for changes predicted correctly or changes not made and predicted as such. As usual, we use part of the NetSence dataset as the training set to learn optimal values of the model parameters. This is done by computing the penalty for every possible set of changes for each person and then choosing the one with the lowest penalty for prediction. We further propose a machine learning based model which uses many more parameters to predict if a set of changes will be made by a person or not.

Using the analytical model, we are able to predict the changes made by persons with a good accuracy, obtaining a good recall and precision as well. The machine learning model performs better than the purely analytical solution with only one or two parameters.

IV. Visualization

We have recently completed the development of this model and obtained results, and now we have started to develop visualizations. We plan to show visually how people are faced with multiple decisions on making changes to their social groups and opinions. We would then show how the model developed by us learns from the decisions people make to predict changes made by people in the future. We would also be showing some visualizations on how groups are identified from Bluetooth data using the hierarchical clustering method developed by us.

V. Acknowledgement

This work was supported in part by the Army Research Laboratory (ARL) under Cooperative Agreement Number W911NF-09-2-0053 (NS-CTA), and by the Office of Naval Research (ONR) Grant No. N00014-15-1-2640.

References

[1] A. Bahulkar, B. K. Szymanski, K. Chan and O. Lizardo, "Impact of Attributes on Group Formation," *Proc. IEEE/ACM ASONAM*, Barcelona (2018), pp. 1250-1257. doi: 10.1109/ASONAM.2018.8508658 URL: https://www.cs.rpi.edu/ szymansk/papers/SI-ASONAM-Groups.pdf

[2] A. Bahulkar, B. K. Szymanski, "Interaction Patterns in a Multilayer Social Network," *Proc IEEE ICCCN*, Hangzou, China, July 30 - Aug. 2 (2018), pp. 1-8. URL: https://www.cs.rpi.edu/ szymansk/papers/ICCCN-Patterns.pdf

[3] Bahulkar, A., Szymanski, B. K., Chen, K., Lizardo, O., "Co-evolution of two networks representing different social relations in NetSense," *Proc. of the International Workshop on Complex Networks and their Applications*, Studies in Computational Intelligence Series, vol. 693, Springer, pp. 423–434 (2016). URL: https://www.cs.rpi.edu/ szymansk/papers/cons17.pdf

[4] Bahulkar, A., Szymanski, B. K., Chen, K., Lizardo, O., "Influence of Personal Preferences on Link Dynamics in Social Networks," *Complexity* vol. 2017 (2017), https://doi.org/10.1155/2017/4543563 URL: https://www.cs.rpi.edu/ szymansk/papers/4543563.pdf

[5] Nigam, A., Shin, K., Bahulkar, A., Hooi, B, Hachen, D., Szymanski, B. K., Faloutsos, C., Chawla, N. V., ONE-M: "Modeling the Co-evolution of Opinions and Network Connections," *Proc. ECML-PKDD*, Dublin, Ireland, Sept 10-14, 2018, Lecture Notes on Artificial Intelligence, vol. 11053, pp. 122-140, Springer, Berlin, (2019). URL: https://www.cs.rpi.edu/ szymansk/papers/ECML-PKDD.pdf

[6] Striegel, A., Liu, S., Meng, L., Striegel, A., Hachen, D., Lizardo, O., "Lessons learned from the netsense smartphone study," *SIGCOMM Comput. Commun. Rev.* vol. 43(4), pp. 5156 (2013).

[7] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, **34**, 1990, pp. 193–206.

[8] Hofstra, B., Corten, R., Van Tubergen, F. and N.B. Ellison, "Sources of segregation in social networks: A novel approach using Facebook" *American Sociological Review*, **82(3)**, 2017, pp.625-656.

[9] J.M. McPherson and L. Smith-Lovin, "Homophily in voluntary organizations: Status distance and the composition of face-to-face groups," *American Sociological Review*, **52(3)**, 1987, pp. 370-379.

[10] Reich, S.M., Subrahmanyam, K. and G. Espinoza, "Friending, IMing, and hanging out face-to-face: Overlap in adolescents' online and offline social networks." *Developmental psychology*, **48(2)**, 2012, p.356.

[11] McPherson, M., Smith-Lovin, L., and J.M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, **27(1)**, 2001, pp.415-444.

[12] Centola, D., Gonzalez-Avella, J.C., Eguiluz, V.M., and M. San Miguel, "Homophily, cultural drift, and the co-evolution of cultural groups" *Journal of Conflict Resolution*, **51(6)**, 2007, pp.905-929.

[13] DellaPosta, D., Shi, Y. and M, Macy, "Why do liberals drink lattes?," *American Journal of Sociology*, **120(5)**, 2015, pp.1473-1511.

[14] Carley, K., "A theory of group stability," *American sociological review*, 1991, pp.331-354.

[15] Shi, Y., Dokshin, F.A., Genkin, M. and M.E Brashears, "A member saved is a member earned? The recruitment-retention trade-off and organizational strategies for membership growth," American Sociological Review, 82(2), 2017, pp.407-434.

[16] G.A. Fine and B. Harrington, "Tiny publics: Small groups and civil society." *Sociological Theory*, 2004, pp. 341-356.

[17] M.T. Rivera, S.B. Soderstrom and B. Uzzi, "Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms." *Annual Review of Sociology*, 2010, pp.91-115.

[18] L .C. Freeman, "The sociological concept of" group": An empirical test of two models." *American Journal of Sociology*, **98** 1992, pp. 152–166.