Socio-Cultural Cognitive Mapping

Kathleen M. Carley, Geoffrey P. Morgan and Joel Levine*

August 31, 2017
CMU-ISR-17-115

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

This work was supported in part by the Office of Naval Research (ONR) N000141512797 Minerva award for Dynamic Statistical Network Informatics, and the Center for Computational Analysis of Social and Organization Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the U.S. government.

*Joel Levine is a Professor of Mathematical Social Sciences at Dartmouth College

Keywords: socio-cultural cognitive mapping, social-networks

Abstract

Socio-cultural cognitive mapping is a new empirical procedure for inferring social networks based on non-network data. The socio-cultural cognitive map (SCM) is the best fit network model given a set of attributes of for the nodes. At one level, this can be viewed as a case of reduced dimensionality mapping. At another level, this can be viewed as an approach for inferring network links such that the links themselves, and the positions of the nodes in space, carry meaning about the nature of the similarity and difference among the nodes. Herein, the advances that have been made in developing and testing this technique are described and illustrated.

1 Objectives

The objective of this research is to develop, test and employ a new statistical-network method to improve socio-cultural assessment. This method is referred to as Socio-cultural Cognitive Mapping (SCM). A secondary objective is to understand the strengths and limitations of this methodology, the types of data for which it is robust, and its sensitivity to alternative forms of data manipulation. The third objective is to demonstrate the use of this methodology for assessing & predicting resiliency and change in groups, based on diverse data sets.

2 Methods

Two alternative procedures are described – one based on a statistical frequentist approach (LSS) and the second based on a visual analytics approach (MVS). Both approaches begin with a set of frequencies with which the nodes in question are linked and then place these nodes into a space (defined using a Minkowski geometry) such that the positions of the nodes in this space generates the best fit to the underlying frequencies. The Levine Statistical Segregation method (LSS) has three key features: First, it allows the user to set the power and attenuation used in constructing the resulting network geometry. Second, it controls node prominence in the frequency matrix via row and column multipliers. Third, it calculates a fitted frequency and compares this against the original frequency. In contrast, the Morgan Visual Segregation (MVS), needs the user to only set the Minkowski power. Second, it identifies the maximum frequency and then uses that to generate an inverse frequency matrix, and it is this against which the nodal geometry is fit. Finally, it calculates distance based on the geometry and then compares this against inverse frequency directly. No row and column multipliers are used.

For both approaches, a greedy optimization procedure using a combination of near and long distance jumps is used to move the nodes about the space. For LSS initial node placement is done using a random uniform placement within one unit centered at the origin. Whereas, for MVS although nodes are still placed randomly, they are placed between the origin and the maximum frequency. The optimizer working to minimize the Chi-statistic – which is used simply as a measure of nearness. For LSS, the Chi statistic is the sum of the Chi-squares for each node defined on the difference between the observed and the fitted frequency. For MVS, the Chi-statistic is the sum of the Chi-squares for each node defined in terms of the difference between the inverted frequency and the distance. To compare the fit of SCM's created for networks of varying sizes, the Wilson-Hilferty F metric is used. The network that results from the SCM process is one in which node position in the dimensional space is meaningful and the distance between nodes interpretable; which, is not the case for standard network diagrams.

3 Results

SCM has been tested on several data sets ranging in size, data collection strategy, and the nature of the underlying data (e.g., categorical, continuous, or binary). To illustrate the

SCM data on the Hatfield and McCoy feud is used. Characteristics of this data are shown in the table. This data is binary. The best known network fit to the data is shown in Figure 1. As can be seen, the Hatfields are on the top and the McCoys on the bottom, while the men are on the right and the women on the left.

Attribute	*Percent with Attribute*
Man	75.8%
Woman	24.2%
Hatfield	45.5%
McCoy	54.5%
Devil Anse (Hatfield) Family	18.2%
Randolph (McCoy) Family	24.2%
Intermarried	10.6%
Harmed Hatfield	7.6%
Harmed McCoy	6.1%
Killed in Feud	16.7%

Notes: Table 1 showing the characteristics of the Hatfield and McCoy data.

Fig.1. Best known network. The Minkowski power is .7 and the attenuation 3. The Chi score is 38.88.

Both LSS (Figure 2) and MVS variants of SCM were run on this data. Only that for LSS is shown. Several points are worth noting. First, the hand-curated image in Figure 1 took days to identify; in contrast, that in Figure 2 took less than an hour. The automated system is faster. Second, as with the hand-curated image, the Hatfields and McCoys are on opposite sides of the figure as are the men and women. The visualization is essentially the same with a 90 degree rotation. Third, the chi-square in the hand-curated version is a little lower. That means it is a somewhat better fit to the underlying data.

A characteristic of optimizing on the Chi metric is that the optimizer moves to plateaus where minor moves can be used to make improvements. However, the time to find the improvement in that plateau can be exorbitant. Thus additional iterations of the optimizer will push the Chi score down. Future works needs to examine alternative optimizers that support rapid movement in these plateaus.

Due to fewer calculations, the MVS is faster than LSS. LSS appears slightly better suited to frequentist data and MVS to binary data; however, with the current optimizer the LSS get's lost in local minima more than the MVS and is more sensitive to minor changes in the data. Further studies are underway to identify improved optimization procedures and alternative factors that influence the robustness of the results.

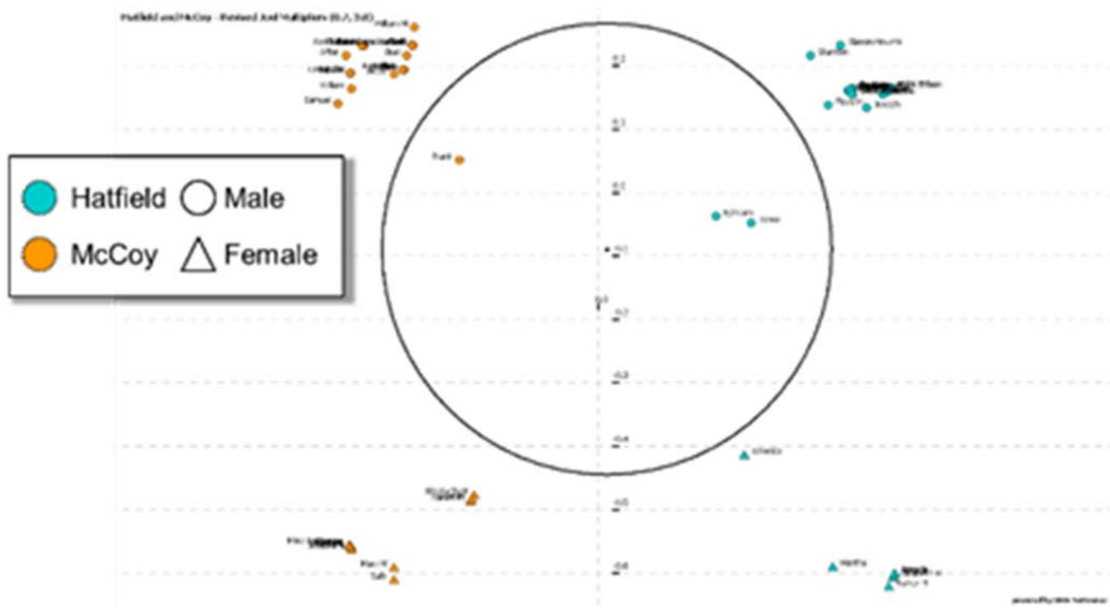

Fig.2. LSS generated network. The Minkowski power is .7 and the attenuation 3. The Chi score is 54.09. This is similar to the best known network, but rotated 90 degrees.

SCM has been operationalized and tested. It is available wthin ORA to other researchers and the DoD. Additional details on the method are available in Levine & Carley, 2016[1]; Morgan and Carley, 2017[2].

[1] Joel H. Levine and Kathleen M. Carley, 2016, SCM System, Carnegie Mellon University, School for Computer Science, Institute for Software Research, Pittsburgh, Pennsylvania, Technical Report CMU-ISR-16-108.

[2] Geoffrey P Morgan and Kathleen M. Carley, 2017, "Socio-cultural Cognitive Modeling." In Proceedings of the International Conference SBP-BRiMS 2017, Dongwon Lee, YuRu Lin, Robert Thompson and Nathaniel Osgood (Eds.) July 5-8, 2017 Washington DC, Springer.

Scientifically, SCM represents a new approach for inferring network relations from non-network data. A key feature of the SCM approach is that it supports meaningful comparison of networks between a set of nodes derived from alternative data sources; thus, providing an empirical metric of how divergent or similar two data sets are. The approach also supports combining alternative data sets to create a comprehensive view. The key advantage of SCM from a naval perspective is that it supports improved social informatics. By creating and SCM, the commander can use the resulting network to rapidly identify the relative power of different groups, and can use the SCM to assess how the power structure is likely to change in response to changes in the features used to infer the network.