5^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2019, University of Amsterdam, The Netherlands

Relating Values and Social Network Structure

Neda Paryab, Alexander Sachs, Andrew Li,
Meiyappan Nagappan and Jesse Hoey

{n2paryab, alexander.sachs, acli, mei.nagappan, jhoey}@uwaterloo.ca
David R. Cheriton School of Computer Science
University of Waterloo.

Keywords: Social Structure, Group Dynamics, SYMLOG, Emotion, GitHub

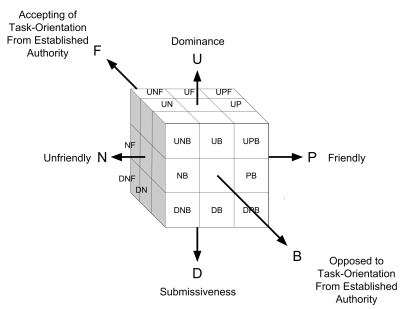
Extended Abstract

This paper investigates correlations between a social-psychological method for the analysis of group behaviour (Bales' SYMLOG [1]) with three network structural properties related to important properties of group structures as discussed by Martin [2]. We manually annotate pull request discussions in a project on the GitHub online network yielding a three dimensional personality profile for each of 39 developers. We then compute three properties of these same developers according to the network formed from their pull request comments: PageRank, reciprocity, and clustering coefficient. Analysis of the correlations and trends between these two three-dimensional spaces show promise for an automated method for group analysis based on the SYMLOG method using network structures automatically computed from GitHub data.

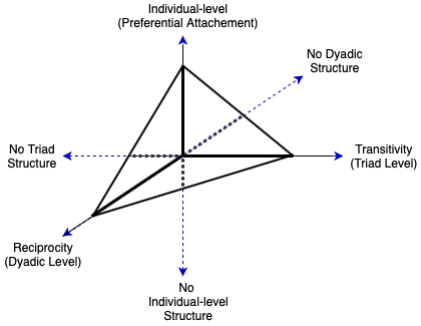
Systematic Multiple Level Observation of Groups (SYMLOG [1]) is a method of small-group analysis which assesses individuals using a 26-item survey tool measuring personal values, and places each individual in a three dimensional space shown in Figure 1a. Based on these measures, personality conflicts can be detected and addressed in order to increase group cohesion and minimize loss of productivity. However, the SYMLOG process requires interviews that may be difficult to administer reliably in distributed online communities like GitHub.

Martin's social structure model [2] defines the connectivity of each individual based on three factors: assertiveness (individual-level), direct interaction with the known counterpart (reciprocity), and interaction with strangers (transitivity). These three structural aspects are proposed as heuristics, associated with values and preferences, that guide individual behaviour (see Figure 1b). For example, a developer might prefer to be assertive and extend her relations, by communicating with new people (individual-level), or she might prefer to converse with people she knows (reciprocity), or try to connect with some small sets of developers that mostly work together to form a clique (transitivity). Here, we propose three simple network structural properties to represent these three dimensions and seek a correlation with the SYMLOG dimensions that will allow a mapping to be computed from one to the other.

A single sample GitHub project, called "onivim/oni" was chosen with 1676 pull request discussions and we: (1) annotated 3-dimensional SYMLOG values as an aggregation of the manual annotations of the raw comments; and (2) computed 3-dimensional network structure characteristics pulled directly from the raw data extracted through the GitHub API. In total, we have 39 members in the chosen repository. We only considered active developers with at least one conversation, in which there were more than two comments by different people.



(a) SYMLOG Vectors [1]



(b) Heuristics of social structure [2].

	Coefficient	P-Val
	P/N	
PageRank	-0.009	0.90
Reciprocity	0.159	0.30
Transitivity	-0.155	0.35
	U/D	
PageRank	0.02	0.87
Reciprocity	0.04	0.79
Transitivity	-0.13	0.42
	F/B	
PageRank	-0.12	0.4
Reciprocity	-0.31	0.049*
Transitivity	-0.14	0.4

(c) Correlation results

Figure 1: (a) SYMLOG has 6 poles (**U**pward, **D**ownward, **P**ositive, **N**egative, **F**orward, and **B**ackward) that are combined to form 26 subspaces defininining individual personalities [1]. (b) Martin [2] proposes three dimensions of social structure that do not vary independently: a simplex. (c) Spearman Correlations and p-values between Bales' dimensions and the network structural properties.

Data annotation was done by the second author by reading each discussion, and filling out a binary version of the 26-item values questionnaire [1] based on his impression of each individual. Each value is marked as being present or not present (1 or 0). One questionnaire was filled out for each individual that participated in the pull request discussion, and the results are averaged across discussions and projected to the three dimensional space in Figure 1a. The final results have a range from -18 to +18 for each dimension. A second judge re-annotated a separate subset of data, and we found an intraclass correlation ICC(3,2) [3] of 0.54, 0.49, and 0.64 for the P/N, F/B, and U/D dimensions respectively. The questionnaire was filled for each developer by a second party (the annotator) looking at their conversations, not directly by developers as in the conventional SYMLOG annotation process. This potential limitation is mitigated in two ways. First, the remote annotation process is more manageable and less time consuming than reaching out to developers to fill out the values questionnaire. Second, extracting personality from conversation text is commonly used by IBM Watson [4] and similar academic research projects [5, 6, 7]. Recent research looks at the comparability of face-to-face and online communications [8], and at the idea of remotely annotating data for SYMLOG values [9].

The set of pull requests form a directed graph where nodes are members and edges are from the person commenting to the pull request starter. We compute three measures of this graph that we propose are related to the three dimensions in Figure 1b: PageRank [10] for the individual level; reciprocity (the ratio of the number of edges in both directions to the total number of edges [11]) for the dyadic level; and the clustering coefficient [12] for the triad level.

Correlations between the manual annotations and the network properties are shown in Figure 1c. Note that p-values for Spearman coefficients must be taken somewhat lightly due to the number of datapoints with exactly the same values (e.g. zeros in the Bales' scores). We find one significant moderate (negative) correlation between the F/B dimension and reciprocity, indicating that more task-oriented and instrumentally controlled developers (F) are less likely to engage in reciprocal interactions, while developers with a higher emotional engagement (B) are more likely to develop these reciprocal connections. We also see some trends (non-significant weak correlations) between F/B and both other network structures, showing that task-orientation is negatively correlated with clustering and PageRank. Further weak correlations exist between the P direction and reciprocity, an indication that more positive developers may be more likely to reciprocate, and between the N and D directions and clustering, indicating that more negative and submissive developers may be involved in more triadic relationships.

While our results were mostly trends, our methods lead the way to further investigation and the potential for a fully automated SYMLOG-like group analysis. Future work involves further network structure analyses and dataset annotations, supplemented by more "bottom-up" analyses of natural language and sentiment [13]. We will seek to model less active group members, and consider actors playing multiple roles. We will investigate whether remote third party annotation of SYMLOG values correlates with developer questionnaire responses. The combination

of these methods will yield more predictive algorithms for automated group analysis.

Acknowledgments

This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and Social Sciences and Humanities Research Council of Canada (SSHRC), through the Trans-Atlantic Platform's Digging into Data Program.

References

[1] R. F. Bales, *Social Interaction Systems: Theory and Measurement*. New York, NY, USA: Routledge, 2001.

[2] J. L. Martin, *Social structures*. Princeton University Press, 2009.

[3] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability." *Psychological bulletin*, vol. 86, no. 2, p. 420, 1979.

[4] "Personality insights," https://www.ibm.com/watson/services/personality-insights, accessed: 2019-03-30.

[5] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 2011, pp. 149–156.

[6] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2. IEEE, 2012, pp. 386–393.

[7] F. Mairesse, M. Walker *et al.*, "Words mark the nerds: Computational models of personality recognition through language," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 28, no. 28, 2006.

[8] P. J. Fahy, "Online and face-to-face group interaction processes compared using bales' interaction process analysis (ipa)," *European Journal of open, distance and e-learning*, vol. 9, no. 1, 2006.

[9] F. D. Berdun, M. G. Armentano, L. S. Berdun, and M. Cincunegui, "Building symlog profiles with an online collaborative game," *International Journal of Human-Computer Studies*, 2018.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[11] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.

[12] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, p. 440, 1998.

[13] J. Hoey, T. Schröder, J. Morgan, K. B. Rogers, D. Rishi, and M. Nagappan, "Artificial intelligence and social simulation: Studying group dynamics on a massive scale," *Small Group Research*, vol. 49, no. 6, pp. 647–683, 2018.