9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Characterizing Polarization in Social Networks using Archetypal Analysis

Keywords: polarization, social networks, link prediction, latent space modeling, embeddings

Extended Abstract

Ideological polarization refers to the significant differences in policy preferences expressed by elite groups, while affective polarization describes the emotional intensity of voter attitudes toward different policy positions. With media usually presenting these extreme positions as existential threats, an "us-versus-them" mentality is created. To better understand ideological polarization, we turn to signed networks, which capture positive, negative, and neutral relationships between nodes and can characterize social polarization more accurately than unsigned networks. In this paper, we focus on extreme positions and argue that the "us-versus-them" multipolarity, reinforced by homophily and influence, can be represented by a latent position model such as the latent distance model (LDM) (1). This model is applied to networks that are confined to a social space formed by a polytope, which we refer to as a "sociotope." The corners of the sociotope represent distinct aspects/poles formed by polarized network tendencies, where positive ties reinforce homophily among similar individuals, and negative ties repel dissimilar individuals to opposing poles. These multiple poles are important for defining the corners of the sociotope and revealing the different aspects of the social network. For the modeling of signed networks and for the characterization of polarization, we present the Signed Relational Latent Distance Model (SLIM). A generative model utilizing a likelihood function for weighted signed links based on the Skellam distribution (3). The Skellam distribution is the discrete probability distribution of the difference between two independent Poisson random variables. The SLIM model is used to characterize the latent social space in terms of extreme positions forming polytopes inspired by archetypal analysis (AA) (2) enabling archetypal analysis for relational data. We apply SLIM on four real signed networks believed to reflect polarization and demonstrate how SLIM uncovers prominent distinct positions (poles).

We introduce the Signed Latent relational dIstance Latent Model (SLIM) by defining a relational archetypal analysis approach. The generative model has endowed a parameterization akin to archetypal analysis in order to efficiently extract a K-dimensional polytope from relational data defined by signed networks. Specifically, for a given signed network $\mathcal{G} = (\mathcal{V}, \mathcal{Y})$ where $\mathcal{V} = \{1, \ldots, N\}$ denotes the set of vertices, the SLIM generative process is:

$$
\begin{aligned}
\gamma_i &\sim Normal(\mu_\gamma, \sigma_\gamma^2) && \forall i \in \mathcal{V}, \\
\delta_i &\sim Normal(\mu_\delta, \sigma_\delta^2) && \forall i \in \mathcal{V}, \\
\mathbf{a}_k &\sim Normal(\boldsymbol{\mu}_A, \sigma_A^2 \mathbf{I}) && \forall k \in \{1, \ldots, K\}, \\
\mathbf{z}_i &\sim Dirichlet(\boldsymbol{\alpha}) && \forall i \in \mathcal{V}, \\
\lambda_{ij}^{+} &= \exp\left(\gamma_i + \gamma_j - \|\mathbf{A}(\mathbf{z}_i - \mathbf{z}_j)\|_2\right), \\
\lambda_{ij}^{-} &= \exp\left(\delta_i + \delta_j + \|\mathbf{A}(\mathbf{z}_i - \mathbf{z}_j)\|_2\right), \\
y_{ij} &\sim Skellam(\lambda_{ij}^{+}, \lambda_{ij}^{-}) && \forall (i, j) \in \mathcal{V}^2,
\end{aligned}
$$

where λ_{ij}^{+} is the rate responsible for generating positive interactions and λ_{ij}^{-} negative. The set $\{\gamma_i, \delta_i\}_{i\in\mathcal{V}}$ denote the node-specific random effect terms, and $||\cdot||_2$ is the Euclidean distance function. More specifically, γ_i, γ_j represent the "social" effects/reach of a node and the tendency to form positive interactions, expressing positive degree heterogeneity . In contrast, δ_i, δ_j provide the "anti-social" effect/reach of a node to form negative connections, and thus models negative degree heterogeneity. In addition, $\{\mathbf{z}\}_{i\in\mathcal{V}}$ are the (shared across rates) latent representations which according to AA lie in the standard simplex set Δ^K, and we further assume that they follow a Dirichlet distribution. Lastly, matrix $\boldsymbol{A}^{K\times K}$ denotes the location of extreme positions (i.e., corners of the polytope regarded as archetypes).

Table 1: Area Under Curve Precision-Recall scores for representation size of $K = 8$.

	WikiElec			*WikiRfa*			*Twitter*			*Reddit*		
Task	*p@n*	*p@z*	*n@z*	*p@n*	*p@z*	*n@z*	*p@n*	*p@z*	*n@z*	*p@n*	*p@z*	*n@z*
POLE	.929	.922	.544	.927	.937	.779	.998	.932	.668	x	x	x
SLF	**.964**	.926	**.787**	**.983**	.922	.881	.994	.870	.740	**.966**	.956	**.850**
SiGAT	.960	.724	.439	.969	.646	.497	**.999**	.861	.582	.965	.692	.232
SIDE	.907	.779	.608	.920	.806	.739	.974	.831	.469	.957	.820	.614
SigNet	.944	.670	.298	.950	.572	.417	.998	.647	.248	.956	.510	.083
SLIM (ours)	.953	**.956**	.785	.973	**.969**	**.907**	**.999**	**.962**	**.813**	.958	**.960**	**.850**

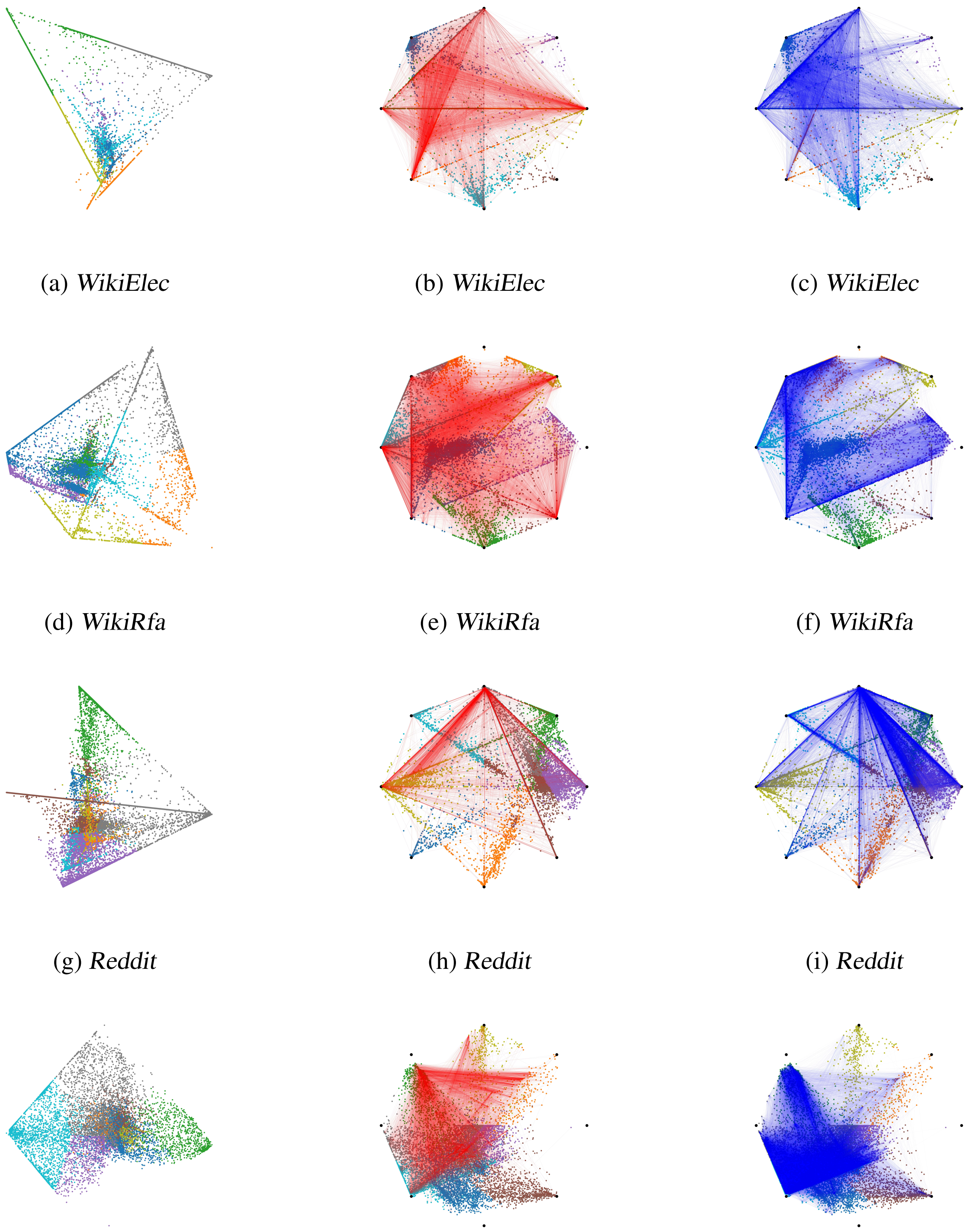
Experiments: We employ four networks, describing electoral voting records and opinions. We benchmark the performance of our proposed frameworks against five prominent signed graph representation learning methods, including random-walk-based methods and graph neural networks. We create a test set by removing 20% of the total network links while preserving connectivity on the residual network. *Link sign prediction (p@n):* in this setting, we utilize the link test set containing the negative/positive cases of removed connections. We then ask the models to predict the sign of the removed links. *Signed link prediction:* a more challenging task is to predict removed links against disconnected pairs of the network, as well as, infer the sign of each link correctly. For that, the test set is split into two subsets positive/disconnected and negative/disconnected. We then evaluate the performance of each model on those subsets. The tasks of signed link prediction between positive and zero samples are denoted as $p@z$ while the negative against zero is $n@z$. Results provided in Table 1 show mostly favorable or on-par results. *Visualization of latent space:* in Figure 1 we observe how the polytope successfully uncovers extreme positional nodes, discovering "dislike", "like" and "controversial" hubs.
Conclusion: SLIM provides easily interpretable network visualizations with favorable performance in the link prediction tasks for signed networks. Endowing the model with a space constrained to polytopes enabled us to characterize distinct aspects in terms of extreme positions in the social networks akin to archetypal analysis but for graph-structured data.

References

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *JASA*, 97(460):1090–1098, 2002.

M. Mørup and L. Kai Hansen. Archetypal analysis for machine learning. In *2010 IEEE International Workshop on Machine Learning for Signal Processing*, pages 172–177, 2010.

J. G. Skellam. The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society.*, 1946.



(a) *WikiElec* (b) *WikiElec* (c) *WikiElec*

(d) *WikiRfa* (e) *WikiRfa* (f) *WikiRfa*

(g) *Reddit* (h) *Reddit* (i) *Reddit*

(j) *Twitter* (k) *Twitter* (l) *Twitter*

Figure 1: Inferred polytope visualizations for various networks. The first column showcases the $K = 8$ dimensional sociotope projected on the first two principal components (PCA) — second and third columns provide circular plots of the sociotope enriched with the negative (red) and positive (blue) links, respectively.