

Functional and structural clustering of social relationship layers among college students for link prediction with applications to perceived drinking networks

Eun Lee¹, Ovidia Stanoi², Xie He³, Yoona Kang⁴, Mia Jovanova⁵, Amanda L. McGowan⁶, David M. Lydon-Staley⁶, Dani S. Bassett⁷, Kevin N. Ochsner², Zachary M. Boyd⁸, Emily B. Falk⁶, Peter J. Mucha³

¹Department of Scientific Computing, Pukyong National University, Busan, South Korea

²Department of Psychology, Columbia University, New York, USA

³Department of Mathematics, Dartmouth College, USA

⁴Department of Psychology, Rutgers University, USA

⁵School of Medicine, University of St. Gallen, Switzerland

⁶Annenberg School for Communication, University of Pennsylvania, USA

⁷Department of Bioengineering, University of Pennsylvania, USA

⁸Department of Mathematics, Brigham Young University, USA

Keywords: multilayer networks, social networks, link prediction, structural similarity, alcohol behavior

Extended Abstract

Motivation. Peer relationships play a central role in shaping behaviors such as alcohol consumption among college students. Most prior studies model peer influence using undifferentiated friendship network, even though social interactions involve multiple types of relationships, including friendship, emotional support, and leadership. Understanding how these different relationship types interact may provide deeper insight into the social mechanisms shaping health-risk behaviors, including how individuals perceive and infer drinking behavior within their peer networks. In this work, we analyze multilayer social networks collected from college student groups to examine structural similarities among different types of social relationships and their role in predicting perceived drinking behavior.

Approach and Methodology. We analyze peer nomination data from nine college student groups (351 participants), where participants nominated peers across ten questions capturing different types of social relationships, including friendship, emotional support, leadership, and perceived drinking. These nominations were used to construct ten directed network layers for each group. To quantify structural similarity between layers, we analyze networks at three levels: (1) Node-level similarity; We compute centrality measures (in-degree, eigenvector, and betweenness) for each node and evaluate correlations across layers. (2) Link-level similarity; Edge overlap between layers is measured using the Jaccard similarity coefficient. (3) Triad-level similarity; Higher-order structural patterns are examined using triad census analysis and Triad Significance Profiles.

These similarity measures are combined and analyzed using Principal Component Analysis (PCA) and unsupervised k-means clustering to identify groups of structurally similar layers. In addition, layers are projected onto a conceptual axis ranging from hierarchical relationships (e.g., leadership) to horizontal relationships (e.g., friendship). Finally, we evaluate whether structurally similar relationship layers can help predict links in perceived drinking networks, examining whether the structural features of these

differentiated social ties provide useful information for inferring drinking-related perceptions using link prediction with stochastic block models.

Results. Across node-, link-, and triad-level analyses, social relationship layers consistently group into three functional clusters: (1) Affiliation cluster, including friendship and interaction-based relationships (i.e., Like, Closest, Interact), (2) Emotional support cluster, including relationships related to sharing news or providing support (i.e., Good News, Bad News, Offer Support), and (3) Leadership cluster (i.e., Leaders, Influence) layers.

Drinking-related layers form a distinct group but appear structurally closer to leadership, and emotional support layers than to affiliation layers, suggesting that perceived heavy drinking may be associated with socially prominent individuals rather than simply close friends. Triadic motif analysis further shows that affiliation networks exhibit strong reciprocity and transitivity, reflecting cohesive relationships, whereas leadership and drinking layers display more heterogeneous and hierarchical structures. These findings suggest that drinking behaviour is more likely to be better predicted from peers who are leaders rather than from peers who are close friends. Link prediction experiments demonstrate that structurally related layers can effectively infer missing links in drinking networks, suggesting that the structural patterns of different social relationships contain useful information for predicting perceived drinking behavior. In particular, emotional support layers substantially improve prediction accuracy, achieving performance close to models that incorporate all network layers. This indicates that structurally coherent subsets of layers capture much of the relevant information needed for network inference.

Conclusions and Outlook. This study demonstrates that multilayer social networks capture meaningful structural regularities across different types of relationships. By integrating node-, link-, and triad-level analyses, we identify coherent clusters of social layers and show that these clusters improve the prediction of missing ties in perceived drinking-related networks. Our findings underscore the importance of considering multiple types of social relationships when examining perceived drinking-behaviors in peer groups. Beyond theoretical implications, this approach can offer practical insights for designing interventions aimed at reducing risky health-related behaviors.

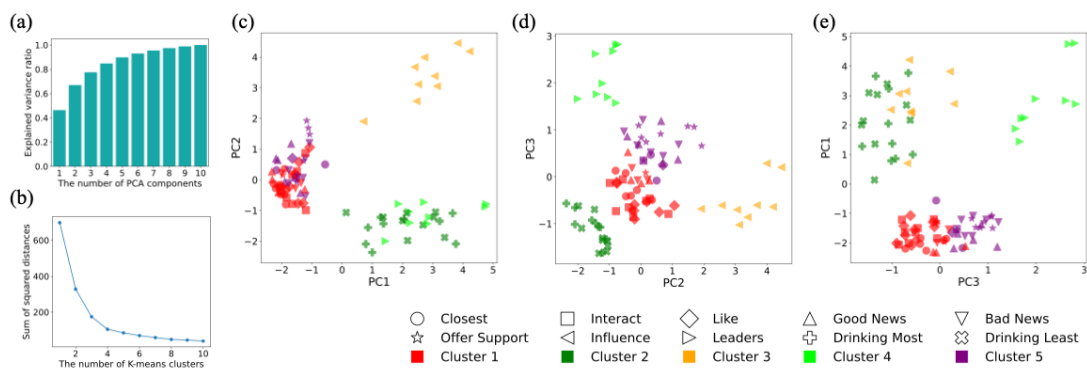


Figure 1. Clustering of social network layers based on multi-level structural similarity using PCA and k-means. Unsupervised k -means clustering is applied to classify 10 different social network layers in each of the nine student groups. Node-, link-, and triad-level similarities are averaged for each pair of layers within a student group, separately. Principal Component Analysis (PCA) is used to reduce the 10 resulting similarity dimensions to three components. The scree plot in (a) visualizes the amount of variance explained over different numbers of components. Unsupervised k -means clustering scores on these three PCA components with different numbers of clusters are plotted in (b). The results of clustering into five groups are shown in (c–e), projected onto different combinations of the first three PCA axes.