Using Four-Node Network Motifs and Anti-Motifs to Understand the Local Structure of Populations

Cassie McMillan, Diane Felmlee, Don Towsley, Roger Whittaker

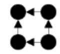
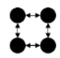
Abstract (150 word max): Previous research in the social sciences analyzes the micro-level features of larger networks by considering how relational ties connect small subsets of two to three actors (i.e., dyads and triads). However, significantly less work considers how tetrads, or configurations of four nodes, pattern our social world. The current project addresses this gap in the literature by comparing the prevalence of 199 directed tetrads across 20 social networks drawn from five unique populations. By comparing our observed networks to randomly generated networks, we specifically look for those tetrads that occur more frequently than expected, or network motifs, and those that are less likely to occur, or anti-motifs. Across all of our networks, we find evidence for 20 network motifs and one anti-motif, as well as key differences between the motif patterns of different populations. We argue that theoretical insight can be gained by studying four-node motifs in social interaction.

Introduction

Data on our social interactions are is becoming increasingly large, detailed, and complex. Yet, perhaps paradoxically, we can make broad and general conclusions about these social networks by analyzing simple, small-scale patterns of interaction. Previous research in the social sciences has analyzed the micro-level features of larger networks by considering how relational ties connect small subsets of two to three actors, or what are commonly referred to as dyads and triads (e.g., Faust and Skvoretz, 2002; Faust 2010; Holland and Leinhardt, 1971; Holland and Leinhardt, 1976; Wasserman and Faust, 1994). The occurrence of certain dyads and triads can inform our understanding of broader social processes, such as hierarchy, clustering, and reciprocity (Homans, 1961; Holland and Leinhardt, 1971).

However, significantly less work considers how tetrads, or configurations of four-nodes, pattern our social world. The current project addresses this gap in the literature by comparing the prevalence of 199 directed tetrads across 20 social networks drawn from five unique populations. Specifically, we look for those tetrads that occur more frequently than expected, or network motifs, and those that are less likely to occur, or anti-motifs (Alun, 2007; Milo et al., 2002; Milo et al., 2004; Shenn-Orr et al., 2002). Network motifs refer to recurring, small-scale patterns of interaction between sets of nodes, while anti-motifs refer to those configurations that actors typically avoid. Taken together, network motifs and anti-motifs represent the essential building blocks of larger structures. Identifying these local network patterns of subgraphs, including tetrads, has provided insight into the functioning of biological networks (e.g., Alun, 2007; Shenn-Orr et al., 2002), however, researchers are only beginning to consider the motif patterns of social networks (e.g, Felmlee et al. 2018). By taking a comparative network approach, we look for four-node network motifs both within and across the five populations in our sample: (1) friendship, (2) legislative cosponsorship, (3) Twitter, (4) advice, and (5) email communication. Overall, we believe that new theoretical insight can be gained by studying patterns of four-node subgraphs in social interaction.

Social Network Tetrads: Theory and Empirical Findings

While four-node motifs have largely been overlooked in the social science literature, they have received moderate attention in the analysis of biological and engineering networks (e.g., Milo et al. 2004). For example, particular four-node configurations, commonly labeled the “bifan,” i.e., , and the “biparallel” motifs, i.e., , have been identified in neuronal networks (Kashtan and Alon 2005) and electronic circuits (Milo et al. 2002). There also exists limited theoretical work devoted to tetrads, especially when compared to dyads and triads. Nevertheless, a couple of conceptual arguments exist regarding the properties of specific types of network arrangements of four actors. For example, in his classic treatment of social capital, Coleman (1988) argues that four actors in networks as diverse as wholesale markets or an intergenerational families can monitor and guide actors’ behavior most effectively when that network consists of a “box” tetrad, i.e., . In this particular four node network, two higher status actors (e.g., parents) can reinforce one another’s guidelines and sanctions for each of their respective, lower status ties (e.g., their children). This form of network closure is noteworthy, furthermore, not only because it facilitates the development of effective norms, but also because it encourages the creation of trustworthiness among actors. Taken together, this line of reasoning suggests that we would expect closed, box tetrads to be more common than their open counterparts in certain types of social data.

Bearman, Moody, and Stovel (2004) also discuss the key role of four node social network subgraphs. In a longitudinal study of the romantic, sexual network of a high school, they document a pattern of avoidance of “four-cycles,” in which an adolescent dates the former partner of their current partner’s former partner. This results in a configuration that is similar to the box tetrad suggested by Coleman (1988). According to the authors, this aversion to a set of four mutual, heterosexual dating or sexual ties arises because choosing to become involved with the ex of a current partner’s ex would represent a public loss of status.

In our analysis, we examine the relative frequency of patterns of 199 directed tetrads in our sample of networks. We simulate random networks to examine the prevalence of individual, network subgraphs that control for the number of edges, nodes, and distribution of dyads in the observed graph. Based on Milo et al. (2004), we examine subgraph ratio profiles for each network and first look for those subgraphs are more or as likely to occur by chance across all of the networks in our sample, or what we define as network motifs (Research Question 1). Next, we consider whether there are any tetrads that occur less likely than would be expected by chance, or anti-motifs, across all networks in our sample (Research Question 2). Finally, we compare motif patterns between population types and note when they differ (Research Question 3).

Methods

Data

We examine network motifs in 20 social networks, including those of adolescent friendship, U.S. senate bill co-sponsorship, Twitter online messaging, advice seeking, and email communication. Within each of the five social network genres, we consider four distinct networks, which yields a total sample of 20 network graphs. Examples of the graphs analyzed in our study are presented in Figure 1.

For our adolescent friendship data, we randomly select four school-based networks from the in-school survey collected during the first wave of the National Longitudinal Study of Adolescent to Adult Health (Add Health). During this wave, entire student bodies at over 100 U.S. middle and high schools were surveyed and respondents were asked to nominate up to ten of their closest within school friends. From these friendship nominations, we are able to construct directed networks where nodes represent individual adolescents and a tie from node *a* to node *b* indicates that adolescent *a* nominated adolescent *b* as a friend.

Our four co-sponsorship networks were constructed from data on US Senate co-sponsorship patterns during the 1995, 2000, 2005, and 2010 congressional terms (Fowler 2006). Each node represents an individual senator and edges are directed. A tie from node *a* to node *b* indicates that during the congressional term of interest, senator *a* cosponsored at least one piece of legislation for which senator *b* was the primary sponsor.

The Twitter online messaging data was collected by searching instances of the use of aggressive, harmful terms and downloading connected messages, in the form of retweets. Two of our networks represent cyberbullying instances that surrounded the use of either a racial or gendered slur. The other two networks consist of cyberbullying attacks that either originated from, or targeted, a well-known celebrity.

Our four advice networks were collected from surveys administered to employees in four different workplaces including a law firm (Lazega, 2001), a high-tech company (Krackhardt, 1987), a consulting firm, and manufacturing company (Cross and Parker, 2004). In each survey, employees and/or managers were asked to nominate the coworkers whom they went to for professional advice. From their nominations we are able to construct directed networks where nodes represent individual employees. A tie from node *a* to node *b* indicates that employee *a* seeks advice from employee *b*.

Our four email communication networks are also collected from workplace environments. Three of the networks are from different bureaucratic departments in the European Union (EU) and one is from the company ENRON. For all four networks, we consider email sending patterns over an eighteen month period. From this information, we are able to construct a directed network where nodes represent individual employees and all networks are directed. A directed edge from node *a* to node *b* indicates that employee *a* sent at least one email to employee *b* during the period of interest.

Plan of Analysis

For the current project, ware interested in finding motifs and anti-motifs that occur among directed tetrads (four-node subgraphs). We compare the census of tetrads in our observed networks to those of 100 randomly simulated graphs. In analyses not shown here, we used larger sets of randomly simulated graphs for a subset of our models, but our substantive conclusions remain the same. All randomly simulate graphs that have the same number of nodes and edges as the observed networks and that are conditional on the observed dyad distribution (e.g., the UMAN distribution, see Holland and Leinhardt, 1976). Given that mutual, or reciprocal, dyads tend to be quite common in social networks (e.g., Jiang, Zhang, and Towsley, 2015), we believe it is important to control for this elemental aspect of social networks in our analyses.

Following Milo et al. (2004), we calculate the subgraph ratio profile (SRP) for each network for each subgraph type. This ratio compares what we see in the observed graph to a set of randomly simulated graphs and has been shown to be uncorrelated with network size. For each *i* motif in each individual graph, we start by calculating the following:

$$\Delta_i = \frac{N_{observed_i} - < N_{random_i} >}{N_{observed_i} + < N_{random_i} > + \varepsilon}$$

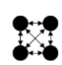
where $N_{observed_i}$ is the number of observed *i* motifs in the graph and $< N_{random_i} >$ is the mean number of such motifs in the random graphs. Following Milo et al. (2004), ε is an error term that we set to 4 to ensure that Δ_i is not too large when the motif rarely appears in both the observed and random graphs. We use these values to calculate the SRP, or a normalized vector of all Δ_i:

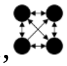
$$SRP_i = \Delta_i / \sqrt{\sum \Delta_i^{\,2}}.$$

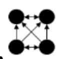
For each network, we are able to calculate a vector of SRPs with a length of 199, the total number of isomorphic tetrads. For each SRP_i in the vector, a positive value indicates that subgraph *i* is more likely to occur in the observed graph than expected by random chance (i.e., a motif), while a negative value suggests that subgraph *i* is less common than would be expected (i.e., an anti-motif). Furthermore, since the SRPs are normalized according to the number of nodes in the graph, we can compare SRP_i's across the different networks. If network *a* has a larger SRP_i than network *b*, this indicates that motif *i* is more frequent in network *a* than it is in network *b*.

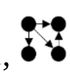
Results

<u>*Motifs*</u>

Of the 199 possible directed tetrads, 20 appear more frequently or as frequently as we would expect among the networks in our sample (see Table 1 for a graphical depiction of each tetrad). We define these 20 subgraphs as motifs across the variety of different populations we consider. Here, we only discuss a selection of these motifs from which we can make interesting substantive conclusions. First, directed tetrad #217 (i.e.,  where all possible ties are present and all ties are reciprocated, a directed four-clique) is more common than we would expect by random chance across all network genres. Tetrad #217 is especially frequent in networks of cosponsorship (mean SRP = 0.138) and email (mean SRP = 0.103).

Furthermore, we argue that several additional motifs can be understood as transition stages towards this four-clique because they contain the necessary preconditions for the subgraph to form. For example, we also find that tetrad #216 is a motif, which includes five mutual dyads and one asymmetric dyad, i.e.,. In other words, tetrad #216 is only missing one of the necessary edges to be classified as a four-clique. As time progresses, we suspect that this asymmetric tie will become reciprocated and the subgraph would resemble to four-clique. Overall, the high frequency of the four-clique tetrad, as well as that of several transition subgraphs, suggests that, in our sample of networks, there exists a high degree of clustering that occurs on a larger scale than the level of the triad.

Another key motif in our sample includes two dyads connected by mutual edges. Both nodes in one pair send ties to each of the members in the other pair, but these ties are not reciprocated (i.e., tetrad #196, i.e.,). This motif suggests that both clustering and hierarchy are simultaneously occurring within our sample of social networks, complementing the conclusions of Holland and Leinhardt (1971). Nodes tend to cluster together and these clusters are hierarchically ordered so that certain groups receive more directed ties than others, which may mean that they have more influence, are more desirable, or are the object of more communication. The #196 tetrad is especially likely to occur in networks of cosponsorship (mean SRP = 0.097), advice (mean SRP = 0.085), and email (mean SRP = 0.088).

Finally, another motif that occurs in our sample of networks is characterized by a transitive triad in which one of the nodes is reciprocally tied to a fourth node. The fourth node has no ties to either of the other two nodes in the transitive triad (i.e., tetrad #77, i.e., ). Particularly, tetrad #77 is more likely to occur than predicted by random chance in networks of friendship (mean SRP = 0.090) Twitter (mean SRP = 0.153), and email (mean SRP = 0.088).This configuration suggests that the bridging of gaps in ties, or “structural holes,” occurs in our sample more than would be expected by random chance (Burt, 2004). These bridges likely represent weaker ties that play a key role in diffusing information or ideas across the network

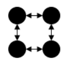
(Granovetter, 1973). Overall, this motif suggests that the networks in our sample are made of both strong ties that lead to clustering and weaker ties that result in bridging.

Anti-Motifs

We only find evidence for one anti-motif, tetrad #29, which is less likely to occur than expected by chance across all of the networks in our sample. This anti-motif is characterized by a string of four nodes where each actor sends a tie to one of the others, but there is no mutuality or clustering, i.e., . Overall, this anti-motif suggests a strict, highly segmented hierarchy that is characterized by no clustering or reciprocity. Tetrad #29 is especially unlikely to occur in our Twitter networks (mean SRP = -0.178).

Variations across Populations

The SRPs of individual networks tend to be more highly correlated between graphs from the same network genre than those across network genres (see Figure 1). The average SRP correlation for networks from the same population is 0.80, while the average correlation for those from different populations is 0.38. However, some populations are more alike than others. For instance, the average correlation between email and friendship networks is 0.67, while the average correlation between Twitter and advice networks is only 0.23. We further demonstrate the variation in subgraph prevalence by plotting SRPs of each network for a selected of tetrads (see Figure 2). Patterns of SRPs tend to be more alike within each type of population than they are across populations

From comparing SRPs across multiple types of populations, it is clear that some subgraphs are more likely to be found within certain social networks, but not others. For instance, tetrads that are referred to as "bifans" (i.e., tetrad #89, ) are more likely to occur than expected in the friendship (mean SRP = 0.087), cosponsorship (mean SRP = 0.058), and Twitter networks (mean SRP = 0.207), but are less likely to be observed in certain advice and email networks than would be expected by random chance. Similarly, the box tetrad that we mentioned previously (i.e., tetrad #203, is slightly more likely to occur in the friendship networks (mean SRP = 0.020). Among all other networks, the box subgraph is less likely to occur by random chance.

Conclusion

Human interaction remains rooted within numerous overlapping social networks that heavily influence, and are influenced by, individual, group, and societal outcomes. We argue that, in order to fully understand how micro-level processes shape the social networks in which we are situated, it is necessary to move beyond studying dyads and triads to additionally focus on subgraphs of four nodes, or tetrads. Furthermore, by specifically looking for motifs and anti-motifs, our current study finds extensive evidence of recurring and absent four-node configurations within a number of networks drawn from diverse populations. Across our five populations, we find evidence for 20 motifs and one anti-motif, suggesting there are fundamental elements of interaction within the social sphere. Such patterns do not limit themselves to the biological and the physical, despite the difficulty inherent in capturing and measuring often messy and complex human behavior.

Additionally, there are other key differences between the motif patterns of the populations we study, suggesting that processes such as hierarchy and clustering operate differently across the populations in our sample. Certain types of hierarchy were more common in Twitter, friendship, and cosponsorship networks, as demonstrated in patterns bifans. Furthermore, symmetric four-cycles were more common than expected in the friendship, but not the remaining networks. Note that these results suggest that the contrasting theories regarding tetrads discussed earlier are supported in specific types of social interactions, but not others.

References

Alun, U. (2007). Network motifs: Theory and experimental approaches. *Nature, 3*: 450-461.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1), 44-91.

Burt, R. S. 2004. Structural holes and good ideas. *American Journal of Sociology* 110: 349-399.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology,* 94, S95-S120.

Cross, R., & Parker, A. (2004). *The hidden power of social networks.* Harvard Business School Press, Boston, MA.

Faust, K., & Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology, 32*(1), 267-299.

Faust, K. (2010). A puzzle concerning triads in networks: Transitivity and homophily in strong-tie relations. *Social Networks, 22,* 221-233.

Felmlee, D. H., McMillan, C., Towsley, D., & Whitaker R. (2018). Social network motifs: A comparison of building blocks across multiple social networks. *Presented at the 2018 Annual Meeting of the American Sociological Association, Philadelphia, PA.*

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship. *Political Analysis 14*(4), 456-487.

Granovetter, M. S. (1973). The strength of weak ries. *American Journal of Sociology* 78(6), 1360-1380.

Jiang, B., Zhang, Z., & Towsley, D. (2015). Reciprocity in social networks with capacity constraints. *Proc. KDD 2015*.

Kashtan, N., & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13773-13778.

Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9(2), 109-134.

Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, *2*(2), 107-124.

Holland, P. W., & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, *7*, 1-45.

Homans, G. (1961). *Social behavior: Its elementary forms*. New York: Harcourt, Brace and World.

Lazega, E. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press.

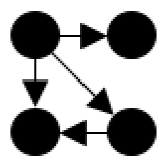
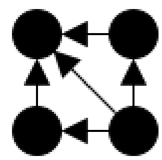
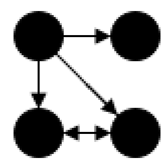
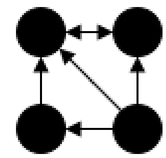
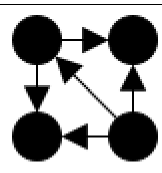
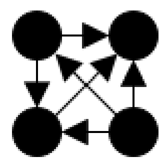
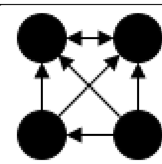
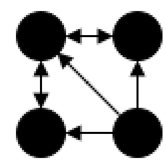
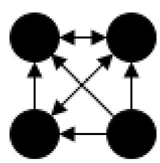
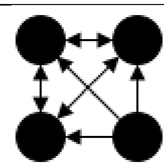
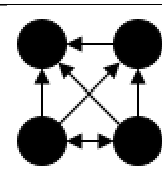
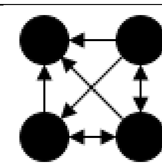
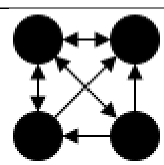
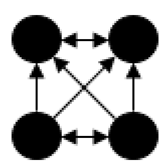
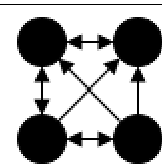
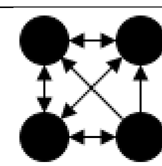
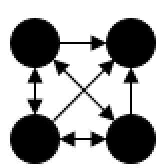
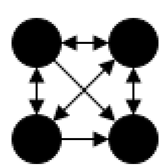
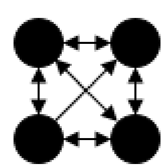
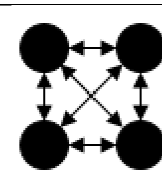
Milo, R., Shenn-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science,* 298, 824-827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., & Alon, U. (2004). Superfamilies of evolved and designed networks. *Science*, 303, 1538-42.

Shenn-Orr, S., Milo, R., Mangan, S., & Alon, U. (2002). Network Motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31, 64-68.

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and Applications*. Cambridge University Press: New York.

Table 1. Tetrad motifs by isomorphic number (e.g., significance profile is greater or equal to 0 for all networks in sample)

77	78	79	81
			
82	83	84	86
			
90	91	184	186
			
187	196	197	198
			
213	214	216	217
			

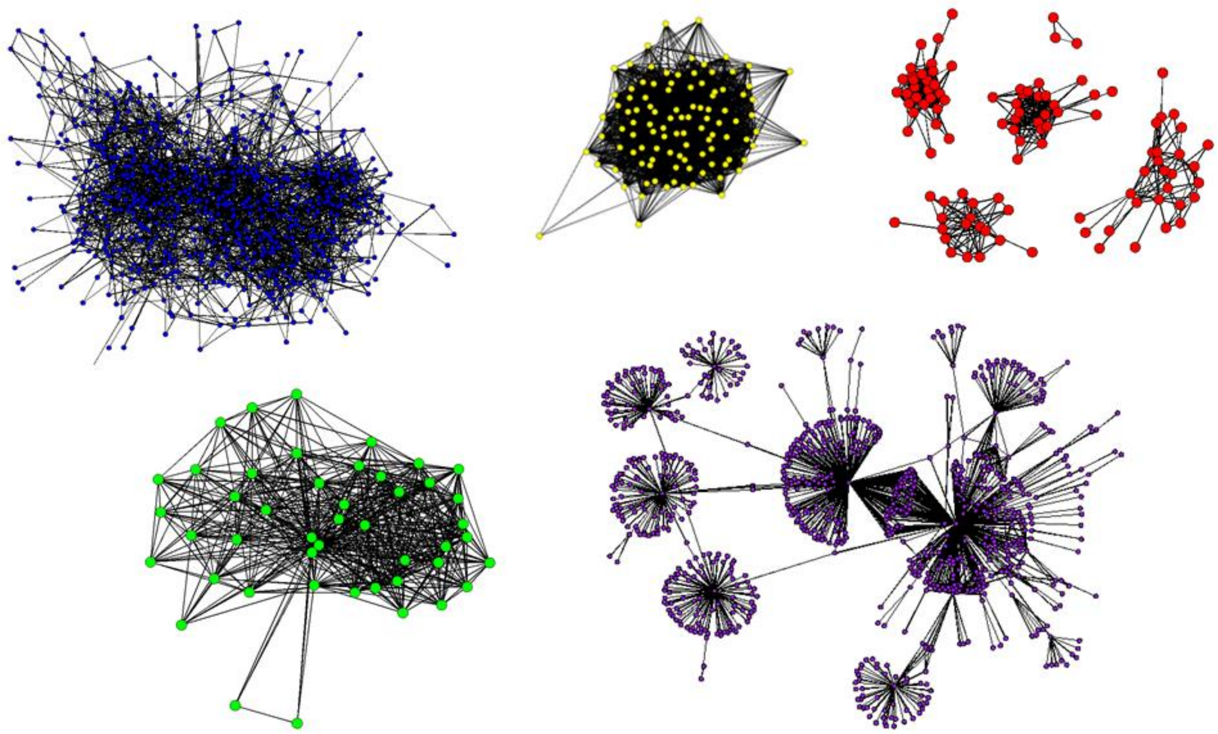


Figure 1. Example network graphs from different populations. Blue represents friendship, yellow is cosponsorship, red is email, green is advice, and purple is Twitter.

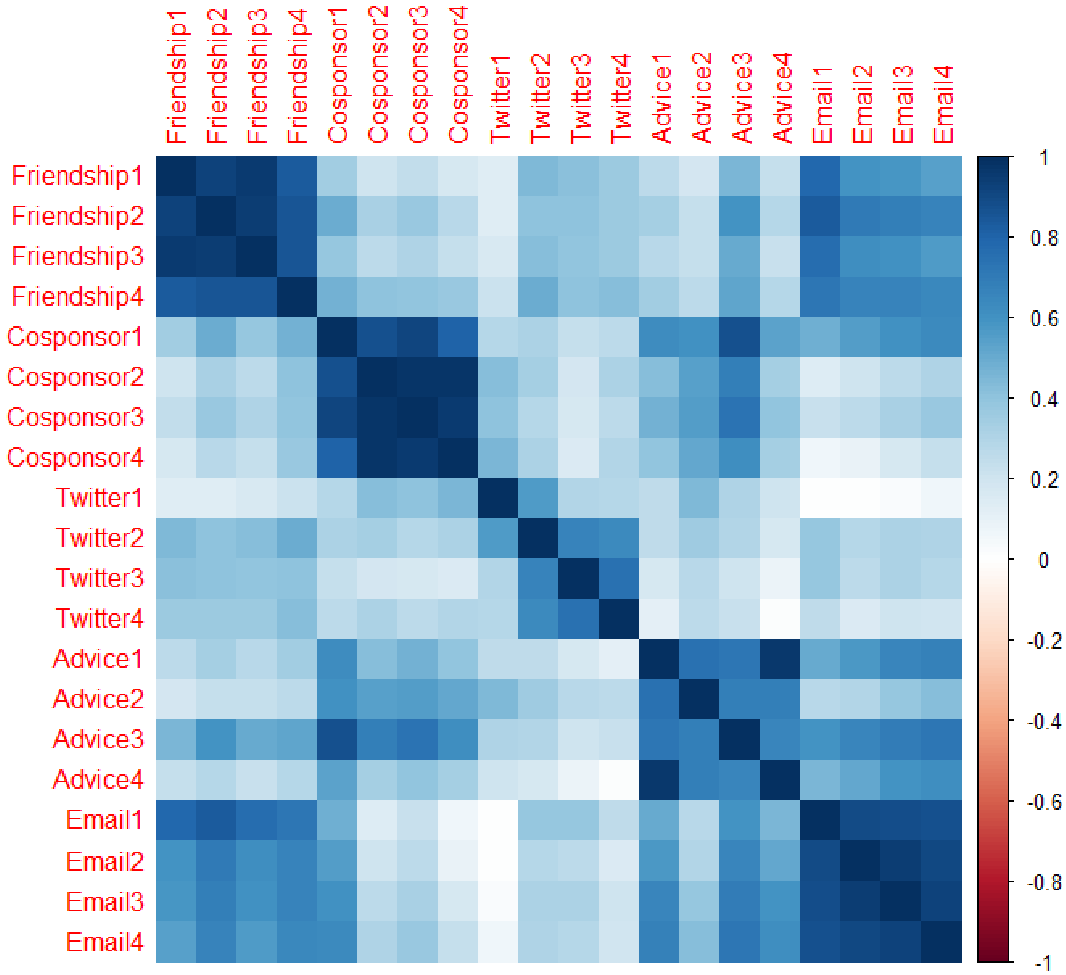


Figure 2. Correlation plot for tetrad significance ratio profiles

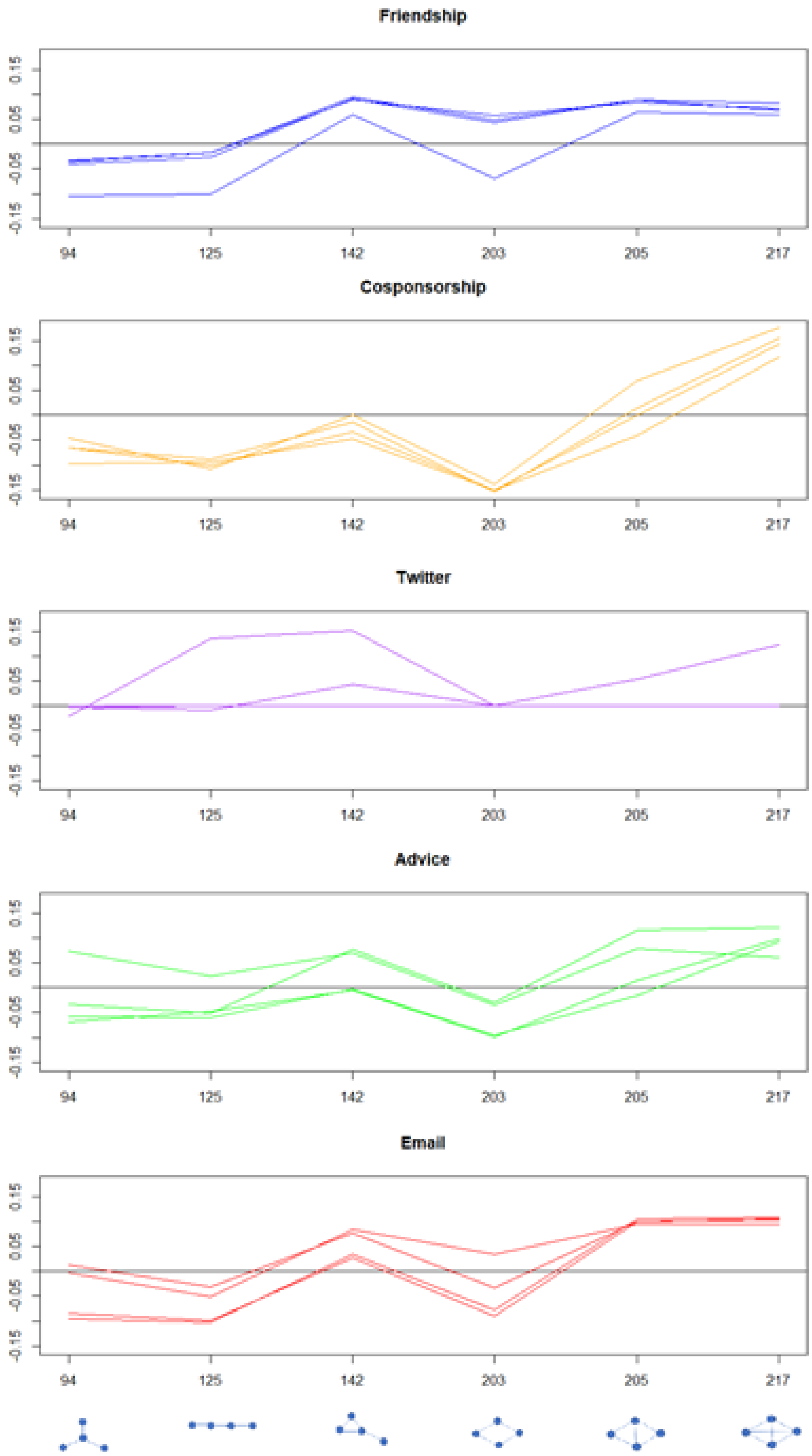


Figure 3. Significance profiles for select tetrads by network type