Social Network Motifs: A Comparison of Building Blocks across Multiple Social Networks

Diane Felmlee, Cassie McMillan
Department of Sociology
Pennsylvania State University
State College, PA, USA

Don Towsley
Department of Computer Science
University of Massachusetts - Amherst
Amherst, MA, USA

Roger Whittaker
Department of Computer Science
Cardiff University
Cardiff, Wales, UK

***Abstract*— Network motifs represent local subgraphs, such as dyads, triads, and tetrads, that occur frequently in networks. The biological and physical sciences document multiple instances in which motifs appear in graphs that provide insight into the structure and processes of these networks. Yet, little work has studied motifs within social graphs. In this research, we examine the prevalence of dyad, triad, and tetrad motifs among six types of social interactions, including friendship, advice seeking, email communication, twitter messages, terrorist ties, and legislative cosponsorship. We use four networks of each type, for a total sample of 24 social networks. One contribution of our work is the use of the UMAN distribution, which controls for the number of mutual, asymmetric, and null dyads, when determining the significance profile of triad and tetrad frequencies. We argue that this distributional control has advantages over other controls used in previous research, because mutual dyads are extremely common in social graphs. We find important commonalities among the six types of networks in our sample, suggesting that there are specific motifs that characterize multiple genres of social network data. Reciprocity of directed ties occurs more frequently than expected by chance in all of our graphs. Completely connected triads and tetrads (i.e., four-node subgraphs) also occur more often than expected, which highlights the tendency of actors to form clusters of ties. We also identify motifs that reflect patterns of hierarchy. In addition, one intriguing tetrad motif points to the bridging of gaps, or "structural holes," between nodes and implies that the networks in our sample are made of both strong ties that lead to clustering, and weaker ones that result in bridging. Furthermore, we find that certain motifs are specific to some genres of social networks, but not others. For instance, a common motif among biological networks, known as a "bifan" tetrad (i.e., nodes i and j each send non-reciprocated ties to nodes k and l), is more likely to occur among the Twitter, friendship, and advice networks in our sample, but not in others.**

Keywords—social networks, network motifs, dyads, triads, tetrads, hierarchy, clustering, bridging

I. Introduction

Complex networks that occur in nature, such as those from biochemistry, neurobiology, ecology, and engineering, exhibit some of the same, simple, network structural properties that occur with greater than expected frequency, that is, "network motifs" (e.g., Milo et al., 2002). Network motifs refer to recurring, significant patterns of interaction between sets of nodes, or actors, and they represent the basic, building blocks of graphs. In particular they are useful in examining complex networks, such as those characterized by significant scale or temporal change. Common network motifs include repeated patterns that occur among three nodes, for example, such as the classic case in which a directed tie extends from node A to node B, from node B to node C, and then also from A to C. This type of three node pattern is referred to as a "transitive triad" in the social sciences, a "feed-forward loop" in biology, and when involving symmetric graphs in network science, it is often labelled "triadic closure." Identifying such local network patterns among small numbers of graph nodes, or "graphlets," has been particularly useful in providing insight into the functioning of basic, biological networks, including those in bacteria and yeasts (e.g., Alun, 2007; Shenn-Orr, 2002).

In recent years, research considering networks that are social, rather than biological or physical, has expanded dramatically (e.g., Felmlee & Sinclair, In Press). Studies repeatedly establish the importance of social network characteristics for a wide range of interaction processes, such as those in social organizations (e.g., Feld 1981), close relationships (e.g., Felmlee, 2001; Felmlee and Faris, 2016), political connections (e.g., Fowler 2006), crime and deviance (Kreager and Staff, 2009), social media (e.g., Noë et al, 2016), as well as in war and terrorist activity (e.g., Everton, 2012; Krebs, 2002). However, research on “network motifs” seldom investigates graphs that capture the ties and interactions among these social networks, and with a few exceptions (e.g., the World Wide Web (Milo et al., 2004)), instead focuses largely on interconnections that are biological or physical. Yet it is likely that certain network motifs proliferate not only in the biological and physical sciences, but also in graphs based on social ties, and some network subgraphs may be universal across domains. The main purpose of this research, therefore, is to examine network motifs that occur within multiple types of social networks, and to what extent these social networks exhibit similarity or differences through network motifs.

An investigation of patterns of “graphlets” in social networks is insightful for a number of reasons. First, given that network motifs appear to align with the functions of networks that represent biological interactions (e.g., Lee, 2002), the same may be true for motifs that occur within networks that are socially based. Network subgraphs also represent the micro mechanisms that produce particular structural patterns within the larger network. A better understanding of motifs within a network can provide insight into the development of a network’s structural characteristics, such as clustering and hierarchy. Furthermore, since motifs signify “overrepresented” graph patterns, they can be used to identify where future network changes are most expected to occur. For example, as networks grow and evolve, subgraphs are more likely to develop into motif arrangements over time than would be predicted by chance.

In this research we examine motif patterns among a total of 24 social networks, with four networks representing each of six, broad genres: 1) Friendship, 2) Terrorist, 3) Twitter, 4) Advice, 5) Legislative Co-Sponsorship, and 6) E-mail Communication. These networks differ on several key dimensions that are likely to influence network structure, such as whether they represent positive or negative behaviors, face-to-face or electronic interaction, and informal or formal bases of organization. Most of the networks contain asymmetric, or directed ties, but one of the networks consists of symmetric, or undirected edges (i.e., Terrorist networks). The multiple functions of these six types of networks vary, as well, and include the coordination of plans and actions (e.g., terrorist, advice), the dissemination of information (e.g., Twitter, e-mail communication), bonds of affinity (e.g., friendship), strategic, political connections (e.g., co-sponsorship), and expressions of positive and negative sentiment (e.g., Twitter). Motif patterns may differ substantially, depending upon the type of behavior and the function, of these various networks.

II. BACKGROUND

Motifs in networks are basic representations of small subsets of vertices and edges in a graph. They identify small, local networks that occur more often than would be expected in random graphs with the same number of nodes and edges. In this research, we examine the relative frequency of subgraphs, or graphlets, within dyads (2 nodes), triads (3 nodes), and tetrads (4 nodes).

Diagrams of several illustrations of basic, network motifs for dyads, triads, and tetrads appear in Figure 1.

Figure 1: Examples of Network Motifs

Dyads and Reciprocity. The simplest motifs capture reciprocity, or the lack of reciprocity (asymmetry), between a dyad, or a pair of two nodes. Reciprocity refers to the tendency of pairs of nodes to develop mutual connections. Dating back to early theoretical work (e.g., Gouldner, 1960; Simmel), reciprocity has received extensive attention within both the social and the physical network literature (e.g., Whitaker et al., 2016). Reciprocal ties powerfully influence network growth and higher-order network motifs (e.g., triadic motifs, Faust 2010) for a host of directed networks (e.g., Holland and Leinhardt, 1975; Wasserman & Faust, 1994).

Triads/Triangles. Triads, or subsets of three actors, are often considered to be the structural foundation of social networks (Holland and Leinhardt, 1975; Wasserman & Faust, 1994). Through studying triads we can better understand a variety of network phenomena, including transitivity, or the tendency for actor i to be tied to actor k if a tie exists between actor i and actor j and between actor j and actor k.

Triads have a long history of scholarly attention, dating back to the classic, work of Cartwright and Harary (1956), who developed a graph theory model of structural balance based on triads. The structural balance model was extended and elaborated by several others over the years (e.g., Davis and Leinhardt 1967; Hallinan, 1972) and eventually led Holland and Leinhardt (1970) to develop the structural concept of “transitivity” and the triad census. This body of research demonstrates that triadic patterns have implications for the overall structure of a network. For example, transitive graphs in which there are no asymmetric dyads are clusterable, with mutual dyads within clusters and null dyads occurring between clusters (Wasserman and Faust 1994). Furthermore, digraphs that consist only of transitive triads, composed of unsigned, mutual, asymmetric and null ties, produce graphs that consist of sets of hierarchically ranked clusters, with asymmetric ties occurring between clusters (Holland and Leinhardt 1970). Given that patterns of transitivity predict both graph clusterability and hierarchy, several researchers maintain that transitivity can account for the bulk of the structure in an overall network, (Holland and Leinhardt, 1971; Wasserman and Faust, 1994; Wellman and Berkowitz, 1988). A good deal of current statistical, network research, therefore, controls for patterns of transitivity, or "triad closure," and other "triangle" network tendencies in models in which the main focus is on other network processes, such as homophily (e.g., Hunter et al. 2008; Morris, Hancock, and Hunter, 2008; Steglich, Snijders, and Pearson 2010).

Tetrads. Recent research points to four node network motifs as common recurring patterns in biological networks (e.g., Milo et al. 2004). For example, particular four-node configurations, commonly labelled the "bifan" and the "biparallel" motifs have been identified in neuronal networks (Kashtan and Alon 2005) and electronic circuits (Milo et al. 2002). Yet, seldom does research devote substantial, empirical attention to four-node subgraphs within networks that are based on social ties.

Compared with the literature on triads within the social sciences, there also is little theoretical work devoted to tetrads. 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 social networks and relationships exhibit structures that facilitate particular forms of social capital. He maintains that certain forms of interpersonal networks aid in providing social capital, or social resources, to network actors, in the sense that the structure enables multiple actors to effectively impose negative sanctions on another. Gossip, for example, can be used as a collective sanction when it is applied to a shared acquaintance. Coleman argues that four actors in networks as diverse as wholesale markets or an intergenerational family can monitor and guide actors' behavior most effectively when that network consists of a "rectangle" tetrad. 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, rectangular 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. A "four-cycle," refers to the same closed rectangle discussed by Coleman, but one that consists of asymmetric, rather than symmetric edges. 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. They note, too, that antipathy to this type of four-node subgraph generates a specific, overall network structure that represents a spanning tree. The spanning tree consists of an extremely large component connecting over half of the students, with an absence of cycles, or redundant paths, and very long, chains of ties. On the other hand, when four-cycles are common in a graph, the broader network tends to be characterized by a core-periphery structure, rather than a spanning tree.

In our analysis, we examine the relative frequency of patterns of dyads, triads, and tetrads in our sample of networks. We simulate random networks to examine the prevalence of individual, network subgraphs, using comparisons against random networks with the same number of edges (Milo et al., 2002), while including other controls in additional analyses. In one set of comparisons, we symmetrize all of our networks. In another set of analyses, we compare patterns for the directed networks and treat the symmetric terrorist network as a set of directed (reciprocated) edges. Based on Milo et al. (2004), we examine subgraph ratio profiles for each network for both three and four node motifs. We use correlation plots to compare patterns between types of networks for both the symmetrized and directed significance profiles.

UMAN Distribution. One of the most frequently used random, directed graph distributions in social network analysis is the classic, UMAN distribution (Holland and Leinhardt 1975). It is a uniform distribution that conditions on three elements, the numbers of mutual, asymmetric, and null dyads in a directed graph. This uniform distribution assigns equal probability to all diagraphs of n nodes that have the same number of mutual, asymmetric and null dyads, and where conditioning on two of the three counts, will mean that the third is fixed. The UMAN distribution is useful because it conditions on the lower-level properties of the complete dyad census, such as the tendency towards reciprocity, and, by default, conditions on the density of the observed graph. This distribution has been used in multiple studies that investigate the properties of triads in social graphs (e.g., Faust 2010, Levina and Hilmann 2012, Shizuka and McDonald 2012). Furthermore, 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. Thus, we will examine the extent to which three and four motif patterns occur, over and above trends towards mutuality among dyads (and trends towards asymmetric and null dyads).

One common, alternative conditional distributions used within the natural and physical sciences to investigate network motifs is to control for degree sequence (e.g., Milo et al., 2004). Another typical distributional control (e.g., Milo, 2002;), especially used in early motif studies, is to control for degree, or network size. In analyses not shown here we explored the use of these distributions in our investigation, and with some exceptions, the findings were similar to those presented here. One of the main exceptions concerned the 030T triad, which consists of 3, directed edges, in a pattern of transitivity, or triad closure. When we use the UMAN conditional distribution, this triad tends to be a common motif in our data, whereas when we use a simple control for network size, or when we control for degree sequence, this triad tends to be less common than we'd expect in the random distribution. The UMAN, unlike the other distributions, controls for the number of asymmetric dyads. There are two ways in which three asymmetric ties can be arranged in a triad, and the 030T arrangement appears to be more common than expected under the UMAN. Another exception is the 201 triad, which consists of two mutual dyads alone. This triad is not a motif under the UMAN distribution, but it is a motif when using the other types of distributional controls. Since the UMAN controls for the number of mutual dyads already, our finding means that it is less likely than expected to find triads with two mutual dyads, out of three possible, in our data. Note, too that both these triad patterns align with social psychological theories of triadic formation. The 030T triad represents a transitive, or balanced, and closed, triad, and that type of pattern is predicted to be more likely in social graphs than would be expected. The 201 triad, on the other hand, represents an intransitive, or imbalanced triad with no triadic closure. Such a subgraph should be less frequent than expected, based on theories of transitivity and balance. Findings from the UMAN distribution, in other words, aligned best with predictions of social science theories, such as those of transitivity and balance, which is why we focus here on this distribution.

Hypotheses. We expect to find reciprocated dyads to be over-represented in social data sets, with the possible exception of Twitter data. Given transitivity theory and previous empirical research, we predict that transitive triads, especially the "300" triad, which consists of 3 mutual dyads, will be overrepresented in most of our data sets. We anticipate that certain networks are likely to exhibit rectangular four-cycle motifs, whereas others, similar to the adolescent sexual network, may avoid such patterns. Finally, we expect that the 4-node clique will be highly prevalent in our data.

III. Data and Methods

Data. We examine network motifs in 24 social networks, including those of adolescent friendship, U.S. senate bill co-sponsorship, Twitter online messaging, advice seeking, email communication, and terrorist organizations. Within each of the six social network genres, we consider four distinct networks, which yields a total sample of 24 graphs.

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 Study of Adolescent and 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.

Finally, our four terrorist networks were randomly selected from the John Jay & ARTIS Transnational Terrorism Database (JJATT). Each network is focused around a specific terrorist act (e.g., 2002 Bali bombings) and nodes represent individual terrorists. Ties indicate whether a social relationship existed between the two terrorists in the year of interest (e.g., the pair was acquainted, roommates, operational collaborators, etc.). Due to the symmetric nature of these relationships, all four terrorist networks are characterized by undirected edges.

Plan of Analysis. We are interested in finding motifs that occur among dyads (two node subgraphs), triads (three node subgraphs), and tetrads (four node subgraphs). Specifically, we consider patterns of directed motifs that occur within our sample of asymmetric graphs. To better compare our findings to the symmetric terrorist networks, we treat these graphs as directed in our analyses.

We compare the census of dyads, triads, and 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 the conclusions remained the same. Our confidence intervals tend to be so small that the use of larger samples sizes of random graphs does not alter the findings. We 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.

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. 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 to make sure that Δ_i is not too large when the motif rarely appears in both the observed and random graphs. As recommended by Milo et al. (2004), the term is set at 2 for dyads, 3 for triads, and 4 for tetrads. 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 the terrorist networks that have been modified to compare as asymmetric networks, we set the SRP values for all “impossible” i asymmetric subgraph arrangements to zero. As a result, for each subgraph type (e.g., directed triads, symmetric tetrads) and each network, we are able to calculate a vector of SRPs with a length equal to the number of potential isomorphic subgraphs. 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, while a negative value suggests that subgraph i is less common than would be expected. 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.

IV. RESULTS

Dyads. For all six types of directed social networks, reciprocated dyads (i.e., node a nominates node b and node b nominates node a) consistently appeared more frequently than would be expected by random chance. The proportion of reciprocated dyads was as high as 0.66 in some of the denser advice networks. For our sparser networks, such as the Twitter graphs, the proportion of reciprocated dyads reached levels as low as 0.004. However, the proportion of reciprocated dyads present in all of the observed networks remains higher than what we would expect by random chance ($p < 0.001$), even in those with low density. On average, the observed proportion of reciprocated dyads is 38.6 times larger than would be expected by random chance. These findings suggest that among all of the directed dyads in our sample, there are strong tendencies towards reciprocation.

Triads. While there exists a total of 16 possible isomorphic directed triads (see Figure 2), we only consider those 13 triads in which all three nodes in the subgraph are connected, following previous motif research (e.g., Milo, 2004).

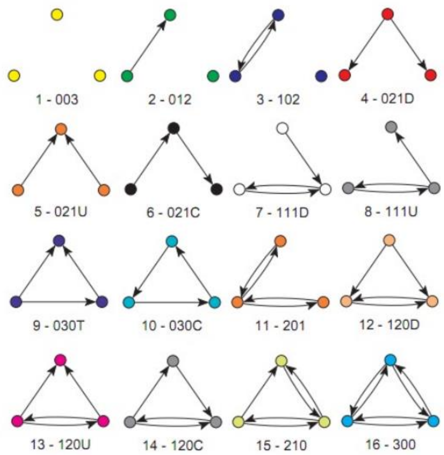

Figure 2: The 16 isomorphic triad classes, with M-A-N labeling. The first digit (M) represents the number of mutual ties, the second (A) represents the number of asymmetric ties, and the third (N) represents the number of null, or nonextistent ties between dyads. The optional letter represents the directions of asymmetric edges ("D" for down, "U" for up, "T" for transitive, and "C" for cyclic.

SRPs for the six network genres all show an overrepresentation of three types of triads, in particular. For instance, the triad consisting of all three mutual ties (i.e., the "300" triad) is more common than expected by chance across all networks in our sample (see Figure 3). The "300" triads were especially more common in friendship (mean SRP = 0.405), email (mean SRP =0.426), and terrorism networks (mean SRP = 0.990), suggesting that small, densely interconnected cliques play a key role in the structure of these networks. The classic transitive triad (i.e., "030T" triad), or feed forward loop, is also overrepresented in our sample, particularly in networks of friendship (mean SRP = 0.395), Twitter (mean SRP = 0.435), and email (mean SRP = 0.362). Finally, the "120D" triad, which consists of one mutual tie and two asymmetric ties directed towards the mutual tie (i.e., a → b, b→ a, c → a, c → b) appears more common than expected by chance across all of networks in our sample. The "120D" triads, which are characterized by patterns of mutuality, or clustering, and hierarchy, are most frequent in friendship (mean SRP =0.404) and email networks (mean SRP = 0.366).

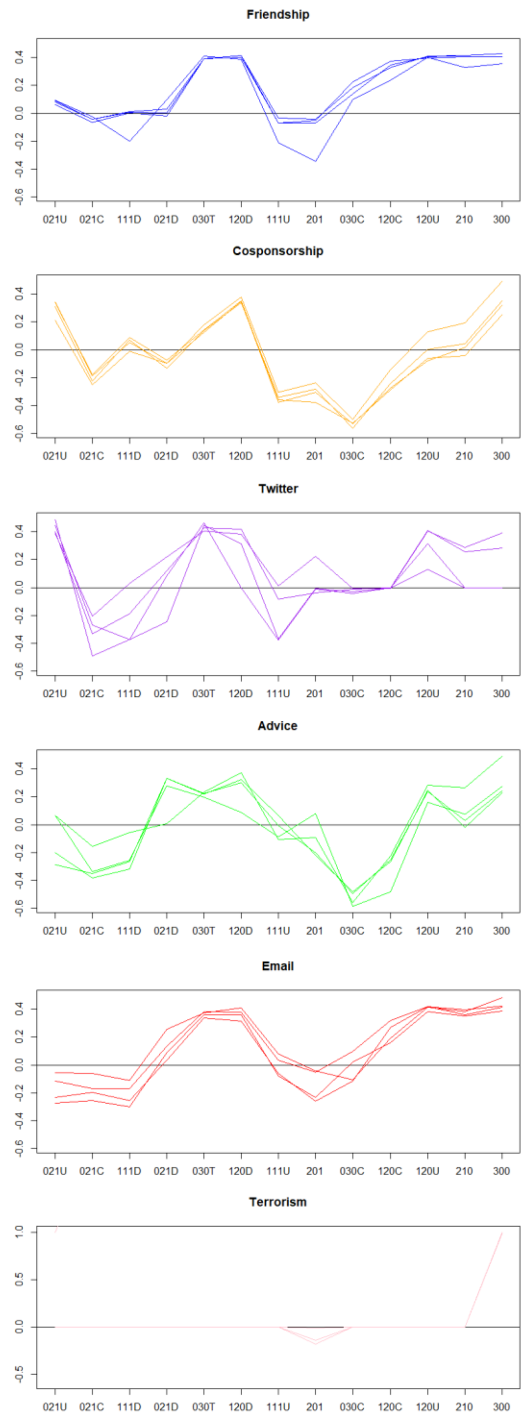

Figure 3: Significance profiles for directed triads by network type.

While there are directed triad motifs that are consistently more common than expected among all of the social networks in our sample, there are also key differences between the different types of networks. From calculating the correlations between the directed triad SRP vectors for each pair of networks, it is apparent that some graphs are more alike than others (see Figure 4). Overall, SRPs tend to be more similar within network genres than between network genres. For example, each of the friendship networks is more similar to the other friendship networks (correlation coefficients r > 0.90), than it is to any of the cosponsorship networks (r = 0.37 to 0.70). This finding reflects the tendency for specific three node subgraphs to be a prevalent characteristic within certain network genres, but not others. For example, triad “021U” is a “vacuously transitive” triad that suggests a specific type of hierarchy characterized by non-reciprocity. The “021U” triad is less common than chance among the email networks (mean SRP = -170), but more common than chance in the Twitter and cosponsorship networks (mean SRPs are 0.300 and 0.4350 respectively).

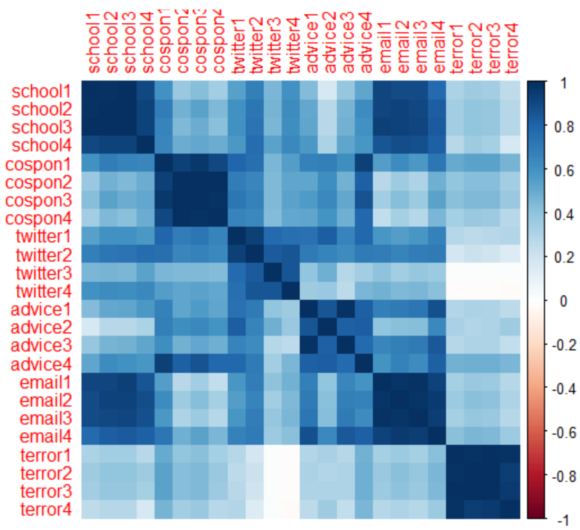

Figure 4. Correlation plot for directed triad significance profiles.

Tetrads. There is a total of 199 possible, directed tetrads, 21 of which appear more frequently or as frequently as we would expect among all of the networks in our sample. We define these 21 subgraphs as motifs across all of the network genres in our sample. Here, we only discuss a selection of these motifs from which we can make interesting substantive conclusions. First, directed four-node motif #217 (i.e., tetrad where all possible ties are present and all ties are reciprocated, a directed four-clique, see Table 1) is more common than we would expect by random chance across all network genres. Motif #217 is especially frequent in networks of cosponsorship (mean SRP = 0.138) and terrorism (mean SRP = 0.612) (see Figure 5).Taken together these findings further highlights that there is substantial clustering within our sample that tends to occur on a larger scale than the level of the triad, particularly in networks of cosponsorship and terror.

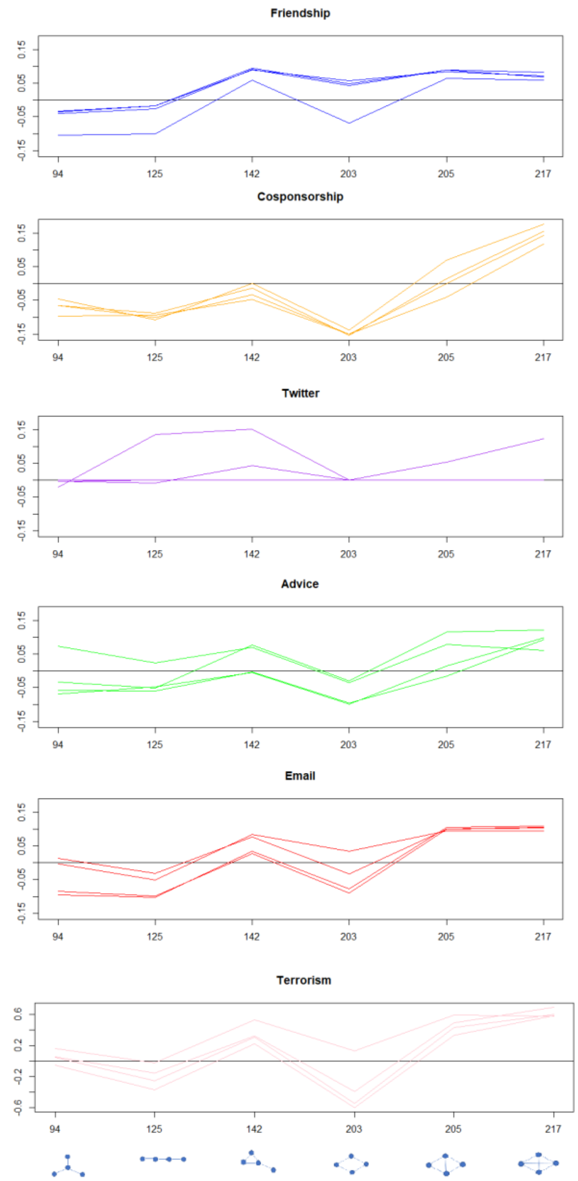

Figure 5. Significance profiles for select tetrads by network type.

A second 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., directed four-node motif #196). This motif suggests that both clustering and hierarchy are simultaneously occurring within all the social networks in our sample, 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 motif 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, a third motif that occurred in our sample 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., directed four-node motif #77). Particularly, motif #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.

Again, 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 6). This is because some subgraphs are more likely to be found for certain relational patterns, but not others. For instance, tetrads that are referred to as "bifans" (i.e., a → c, a → d, b → c, b → d) were 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 less likely to be observed in certain advice and email networks.

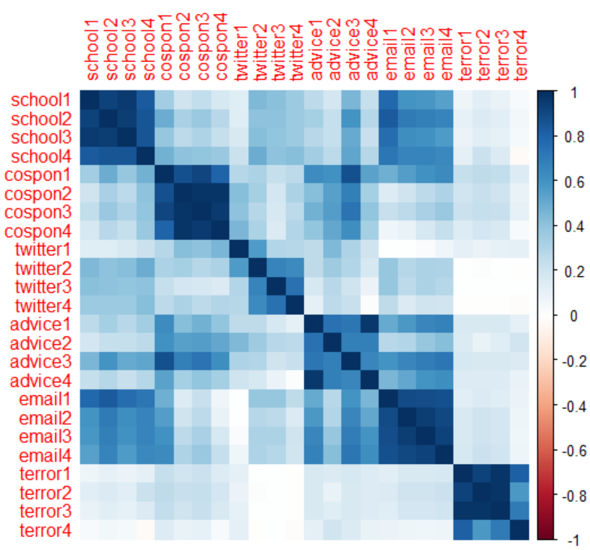

Figure 6. Correlation plot for directed tetrad (4-node) significance profiles.

V. Conclusions

Human interaction remains rooted within numerous overlapping social networks that heavily influence, and are influenced by, individual, group, and societal outcomes. We find extensive evidence of recurring local patterns of network structure within a number of these noteworthy social networks, which points to 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. Certain motifs emerged in social graphs that are common among their biological counterparts, as well, such as the "feed-forward loop", or the transitive triad, in friendship and several other networks, and the "bi-fan" tetrad in Twitter.

Not unlike their biological counterparts, social graph motifs also signify functional aspects of the networks. For example, in all of the graphs studied here symmetric dyads were more likely than expected by random. Reciprocity of ties, thus, represents a basic element of many social interactions, including those consisting of either negative or positive interactions, face-to-face or electronic interchanges, as well as formal and informal ties between individuals. The motifs identified here also point to two additional, fundamental social processes, those of clustering and hierarchy. The frequent appearance of completely connected dyads, triads, and tetrads in this sample of social graphs suggests that actors often form clusters of ties, whereas the overabundance of triads with asymmetric ties in particular arrangements, implies the occurrence of social hierarchy. Both clustering and hierarchy are identified as fundamental social processes in formative social theories of group processes (e.g., Homans 1961, Holland and Leinhardt 1971). Finally, we find evidence that both of these processes of clustering and hierarchy often are interconnected in our data. Clusters tend to emerge among strong ties in our graphs, but in addition, weak, intransitive ties develop that bridge these clusters.

At the same time, notable differences between genres of social networks arise in our analyses. For example, as confirmed by the particularly high presence of "120D" triads, patterns of clustering and hierarchy were particularly frequent in friendship and Twitter networks. Certain types of hierarchy were more common in Twitter and cosponsorship networks, as demonstrated in patterns of 021U triads and bifans. Furthermore, symmetric four-cycles were more common than expected in the friendship and twitter graphs, but not in cosponsorship, advice, email or terrorist 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.

A number of advantages exist in identifying motifs in graphs. For example, network motifs can be used in building simulations of templates for various types of social networks. Motifs also contain the seeds of dynamic change. Networks characterized heavily by reciprocity would be expected to turn asymmetric links into those that are symmetric in the future, and networks composed of high levels of triangulation and transitivity would be predicted to close open triads. There also is an inherent interdependence between the composition of basic, network motifs and the pattern of overall structure in a network, furthermore, especially regarding its level of hierarchy and subgroup formation. Finally, by understanding how network motifs are similar and varied across a range of social network types, our findings have the potential to inform the creation and application of social theories. For instance, our results show that electronic communication, friendship, and terrorist networks are all characterized by high degrees of transitivity, or triad closure. When crafting particular research questions, therefore, theories of tie formation that apply to one of these categories of social relationships may be informative for the others.

VI. Acknowledgement

This research was sponsored in part by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. 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 U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] Alun, Uri. 2007. Network Motifs: Theory and Experimental Approaches. *Nature* 3: 450-461.

[2] Bearman, Peter S., Moody, James, and Katherine Stovel. 2004. Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks. *American Journal of Sociology* 110(1): 44-91.

[3] Burt, Ronald S. 2004. Structural Holes and Good Ideas. *American Journal of Sociology* 110: 349-399.

[4] Cross, Rob, and Andrew Parker, 2004. *The Hidden Power of Social Networks.* Harvard Business School Press, Boston, MA.

[5] Everton, Sean. 2012. Disrupting Dark Networks. Cambridge University Press.

[6] Krebs, Valdis E. 2002. "Mapping Networks of Terrorist Cells." *Connections* 24(3): 43-52.

[7] Faris, Robert and Diane Felmlee. 2014. "Casualties of Social Combat: School Networks of Peer Victimization and Their Consequences. *American Sociological Review* 79: 228-57.

[8] Faust, Kathleen. 2010. A Puzzle Concerning Triads in Networks: Transitivity and Homophily in Strong-Tie Relations. *Social Networks* 22: 221-233.

[9] Feld, Scott. 1981. The Focused Organization of Social Ties. *American Journal of Sociology* 85(6): 1015-35.

[10] Felmlee, Diane and Rober Faris. 2016. Toxic Ties: Networks of Friendship, Dating, and Cyber Victimization. *Social Psychology Quarterly* 79.

[11] Felmlee, Diane and Colleen H. Sinclair. Forthcoming. Social Networks and Personal Relationships. In the *Cambridge Handbook of Personal Relationships*, 2nd Ed. edited by Anita Vangelisti and Daniel Perlman

[12] Fowler, James H. 2006. Connecting the Congress: A Study of Cosponsorship. *Political Analysis* 14(4): 456-487.

[13] Granovetter, Mark S. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78(6): 1360-1380.

[14] Holland, Paul W. and Samuel Leinhardt. 1971. Transitivity in Structural Models of Small Groups. *Comparative Group Studies*.

[15] Holland, P. and Leinhardt, S. 1975. Pp. 1-45 in *Sociological Methodology* (Ed). Heise, D.; Jossey-Bass: San Francisco.

[16] Holland, Paul W. and Samuel Leinhardt. 1981. An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association* 76(373): 33-50.

[17] Homans, George. 1961. *Social Behavior: Its Elementary Forms*. New York: Harcourt, Brace and World.

[18] Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Marina Morris. 2008. "Ergm: A Package to Fit, Stimulate, and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24:1-29.

[19] Jiang, Bo, Zhang, Zhi-Li, and Don Towsley. 2015. Reciprocity in Social Networks with Capacity Constraints. Proc. KDD 2015, August 2015.

[20] Krackhardt, David. 1987. "Cognitive Social Structures." *Social Networks* 9(2): 109-134.

[21] Lazega, Emmanuel. 2001. The Collegial Phenomenon: The Social Mechanisms of Cooperation among Peers in a Corporate Law Partnership. Oxford University Press.

[22] Lungeanu, A. & Contractor, N. S. 2015. The effects of diversity and network ties on innovations: The emergence of a new scientific field. American Behavioral Scientist, 59(5), 548-564.

[23] Kashtan, N. and 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.

[24] Kreager, Derek A. and Jeremy Staff. 2009. The Sexual Double Standard and Adolescent Peer Acceptance. *Social Psychology Quarterly* 72: 143-64.

[25] Milo, R., et al. 2002. Network Motifs: Simple Building Blocks of Complex Networks. Science 298: 824-827.

[26] Milo, Ron, Itzkovitz, Shlaev, Kashtan, Nadav, Levitt, Reuven, Shen-Orr, Shai, Ayzenshtat, Inbal, Sheffer, Michal, and Uri Alon. 2004. Superfamilies of Evolved and Designed Networks. *Science* 303: 1538-42.

[27] Morris, Martina, Mark S. Hancock, and David R. Hunter. 2008. Specification of Exponential –Family Random Graph Models: Terms and Computational Aspects. Journal of Statistical Software 24: 1548-7660.

[28] Noë, N. Whitaker, R.M. Chorley, M.J. and Pollet, T.V., 2016. Birds of a feather locate together? Foursquare checkins and personality homophily. *Computers in Human Behavior*, *58*: 343-353.

[29] Shenn-Orr, S. S., Milo, R., Mangan, S. and alon, U. Network Motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31: 64-68.

[30] Squartini, T., F. Picciolo, F. Ruzzenenti, and D. Garlaschelli. 2013. Reciprocity of Weighted Networks. *Scientific Reports* 3: 2729.

[31] Steglich, Christian, Snijders, Tom A. B., and Michael Pearson. 2010. Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology* 40(1): 329-393.

[32] Wasserman, S. and Faust, K. 1994. *Social Network Analysis*. Cambridge University Press: New York.

[33] Wasserman, Stanley and Philippa Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: An Introduction to Markov Graphs and p*." *Psychometrika* 61: 401-25.

[34] Whitaker, R. et al., 2016. A Dominant Social Comparison Heuristic Unites Alternative Mechnisms for the Evolution of Indirect Reciprocity. *Scientific Reports* 6: 31459.