

RESEARCH

Open Access



# Dyads, triads, and tetrads: a multivariate simulation approach to uncovering network motifs in social graphs

Diane Felmlee<sup>1\*</sup> , Cassie McMillan<sup>2</sup> and Roger Whitaker<sup>3</sup>

\*Correspondence:  
dhf12@psu.edu

<sup>1</sup> Department of Sociology  
and Criminology,  
Pennsylvania State University,  
University Park, PA, USA  
Full list of author information  
is available at the end of the  
article

## Abstract

Motifs represent local subgraphs that are overrepresented in networks. Several disciplines document multiple instances in which motifs appear in graphs and provide insight into the structure and processes of these networks. In the current paper, we focus on social networks and examine the prevalence of dyad, triad, and symmetric tetrad motifs among 24 networks that represent six types of social interactions: friendship, legislative co-sponsorship, Twitter messages, advice seeking, email communication, and terrorist collusion. Given that the correct control distribution for detecting motifs is a matter of continuous debate, we propose a novel approach that compares the local patterns of observed networks to random graphs simulated from exponential random graph models. Our proposed technique can produce conditional distributions that control for multiple, lower-level structural patterns simultaneously. We find evidence for five motifs using our approach, including the reciprocated dyad, three triads, and one symmetric tetrad. Results highlight the importance of mutuality, hierarchy, and clustering across multiple social interactions, and provide evidence of “structural signatures” within different genres of graph. Similarities also emerge between our findings and those in other disciplines, such as the preponderance of transitive triads.

**Keywords:** Social networks, Network motifs, Dyads, Triads, Tetrads, Conditional distributions

## Introduction

Complex networks that arise in nature, such as those from biochemistry, neurobiology, ecology, and engineering, exhibit some of the same, simple structures that occur with greater than expected frequency, known as “network motifs” (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, typically through the overrepresentation of a particular substructure. 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 (Alon 2007; Shenn-Orr et al. 2002) and food-web structures (Stouffer and Bascompte 2010).

In recent years, a veritable industry of research has arisen in which commonly occurring network motifs form the focus of study. Motif research, per se, often concentrates on interconnections that are biological or physical, with less attention to social science methods and theory until recently (e.g., Cunningham et al. 2013; Gu et al. 2019; Michienzi et al. 2021; Zignani et al. 2018). Nevertheless, certain motifs may be universal across a variety of domains, but this remains unclear. The purpose of this research, therefore, is to examine motifs within multiple types of social networks, and focus on the relative frequency of two, three, and four node subgraphs, known as dyads, triads, and tetrads, respectively. Although decades of statistical studies of social graphs examine local, network structure, such as dyads and triads (e.g., Holland and Leinhardt 1971, 1978), little work explicitly adopts a “motif” framework. Given that disciplinary boundaries sometimes generate silos of methods and interests (Brandes et al. 2013), such a framework facilitates comparisons across possible divisions.

To identify these overrepresented graphlets, it is necessary to compare observed networks to sets of comparable, randomly-generated graphs. Yet several conditional distributions exist that control for differing aspects of network structure, such as degree sequence or the dyad distribution, and the correct choice remains the subject of debate (Gotelli and Graves 1996; Yaveroglu et al. 2015). Another goal of this research is to extend the state-of-the-art by applying a multivariate exponential random graph model (ERGM) framework to generate random graphs for the identification of network motifs. This approach reduces the scope for misleading results by controlling for multiple, potential correlates in the same set of random models.

We employ our methodological approach in a case study in which we analyze a total of 24 social networks, with four networks representing each of six, broad genres: (1) Friendship, (2) Legislative co-sponsorship, (3) Twitter, (4) Advice, (5) Email communication, and (6) Terrorism. Additionally, we examine the degree to which these social graphs exhibit unique local structural, motif patterns, or “signatures” that differentiate one from the other (e.g., Faust and Skvoretz 2002; Welser et al. 2007). We end by comparing our findings to those from more traditional approaches to motif detection and note key discrepancies.

### **Network motifs**

Motifs identify small, local subgraphs that occur more often than would be expected in random graphs with comparable characteristics. In the biological domain, motifs, or rather their underlying subgraphs, are viewed as naturally occurring entities that capture atomic functions, rather than artificial constructs engineered for analysis. For example, motifs between networks of genes and proteins act as the circuitry for local decision-making on feature activation (Shoval and Alon 2010). Unlike engineered networks (Alon 2003), small, recurring, biological subgraphs are products of evolution, where they contribute to high degrees of network modularity (Faust and Skvoretz 2002). Motifs aid in understanding various network characteristics and functions (Jain et al. 2019), such as social mobility patterns (Schneider et al. 2013), triad closure in authors (Braines et al. 2018), inter-firm connections (Ohnishi et al. 2010), and resilience in power-grids (Dey et al. 2019).

## Subgraphs of interest

### Dyads

The simplest subgraphs capture reciprocity, or the lack of reciprocity (asymmetry), between a pair of nodes. Reciprocity refers to the tendency for pairs of nodes to develop mutual connections. Dating back to early empirical and theoretical work (Gouldner 1960; Moreno 1934), reciprocity receives extensive attention within both the social and the physical network literature (e.g., Felmlee and Faris 2016; Whitaker et al. 2016). Mutual ties powerfully influence network growth and higher-order network structures [e.g., triads (Faust 2010)] for a host of directed networks (Holland and Leinhardt 1971; Wasserman and Faust 1994).

### Triads

Triads, or three node subgraphs, are often considered to be the structural foundation of social networks (Holland and Leinhardt 1971; Wasserman and Faust 1994). Triads have a long history of scholarly attention within the social sciences, dating back to theoretical work by Simmel (1902) and multiple empirical applications (e.g., Davis 1970; Hallinan 1974; Holland and Leinhardt 1971). Through the study of three person groups, we can examine a variety of network patterns, 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  $k$ , a concept informed by Heider's balance theory (Heider 1946).

Following the labelling system developed by Holland and Leinhardt (1971), a quintessential example of transitivity can be observed in the 030 T triad, where the first number in the label refers to the number of mutual dyadic ties in the triad, the second indicates the number of asymmetric edges, and the third represents the number of null, or non-existent, ties. The T refers to a transitive relationship among the asymmetric ties and distinguishes it from other arrangements of three asymmetric edges, such as the 030C triad, in which the asymmetric edges are cyclical, rather than transitive.

The 030 T transitive triad represents a dominant network motif in biological networks, where it is referred to as the "feed-forward loop." The feed-forward loop provides the architecture for logical "AND/OR" gates (Mangan and Alon 2003) through which a range of signal processing occurs in gene networks (Ingram et al. 2006). Previous work finds that the subgraph is prominent in key "super families" of biological networks (Milo et al. 2004) and appears in hundreds of gene systems (Shoval and Alon 2010), ranging from bacterial networks (Shenn-Orr et al. 2002) to protein interactions (Yeager-Lotem et al. 2004) and networks in yeast (Lee et al. 2002). In a biological setting, the feed-forward loop operates by an arrangement of three genes where one gene can influence another both directly and indirectly (i.e., via a third party). Within community ecology, moreover, the presence of omnivory represents a literal feed-forward pattern, whereby humans eat rabbits, rabbits eat carrots, and humans eat carrots, for instance (Stouffer and Bascompte 2010). In a social context, these tendencies map exactly to a pattern of triad transitivity, resulting in a form of triadic closure.

### Tetrads

Research points to four node network motifs as additional patterns of interest in biological and electronic networks (Hale and Arteconi 2008; Kashtan and Alon 2005; Milo et al.

2004). With some exceptions (e.g., Bearman et al. 2004; Coleman 1988; Marcum et al. 2016; McMillan and Felmlee 2020), tetrads receive less attention than dyads or triads in the social science literature, but patterns formed by small groups of four nodes, or tetrads, also have likely ramifications for the study of social interaction in the context of larger organizational structures.

We argue that tetrad subgraphs can be used to identify the presence of key social structures in network data, such as clustering, “anti-clustering,” and bridging. For example, the completely connected 4-node clique, in which all four actors extend ties to each other, represents an instance of local clustering. Given the tendency of human groups to reciprocate interchanges (Gouldner 1960), and for homophilous subgroups to emerge in social interaction (McPherson et al. 2001), we expect the clique tetrad will constitute a motif in our data sets.

Other four-node subgraphs can capture elements of hierarchy and “anti-clustering” (Krumov et al. 2011), where anti-clustering refers to a local subgraph pattern that lacks the complete connectivity of a clique. The “box” tetrad (i.e.,  $a-b$ ,  $b-c$ ,  $c-d$ ,  $d-a$ ) is an example of an anti-clustering subgraph and is one of the few types of four actor configurations that garners consideration in the social sciences. For example, Coleman (1988) argues that two higher status actors (e.g., parents) in certain groups of four can reinforce one another’s guidelines and sanctions for each of their two respective, lower status ties (e.g., their children). At the same time, previous work documents a pattern of avoidance of “four-cycles,” or the “box” tetrad, in adolescent sexual dating networks (Bearman et al. 2004).

Certain types of tetrads also signify the presence of “bridging” or weak ties among actors in a group. The “kite” tetrad (i.e.,  $a-b$ ,  $b-c$ ,  $b-d$ ,  $c-d$ ), or the “1-Edge-Triangle” (Wang et al. 2009), for example, indicates that one actor in a group of three connected actors has a sole connection to a fourth. The link that extends to the actor represents a local “bridge,” or “weak tie,” between two locations in the network. Several theories argue that network bridges play a crucial role in the dissemination and diffusion of novel information and the transmission of resources (Granovetter 1973). We expect the kite tetrad to represent social interactions in which individuals cross “structural holes” (Burt 2004) to connect disparate groups.

### Control distributions

The wider literature can leave the impression that network motifs are an “absolute” concept, in the sense that motifs are uniquely specified on a graph as a consequence of comparison against similar random graphs. However, this is not the case: random graphs may take many forms, and therefore motifs are defined relative to the type of random graph with which assessments are made. Thus, to benchmark the relative over (or under) representation of subgraphs and determine the presence of motifs, it is necessary to draw comparisons against an appropriate sample of random graphs (Wasserman and Faust 1994; Honey et al. 2007) referred to as the “null model”. Discussion continues regarding the most suitable model, and there is debate over the strengths and weaknesses of using marginal tests versus statistical modeling approaches for motif research (Gotelli and Graves 1996; Yaveroglu et al. 2015).

Within the network field, the choice of a null model varies, with many studies using a uniform conditional distribution that either randomizes the distribution of edges (Ashford et al. 2019; Milo et al. 2002), or controls for the indegree and outdegree of vertices (Milo et al. 2004; Shenn-Orr et al. 2002). Controlling for the number of observed edges and nodes represents the most rudimentary approach, wherein observed networks are compared to random graphs that share their same size and density. Previous work shows that utilizing a conditional distribution that solely controls for the number of observed edges and nodes can produce biased estimates of subgraph presence, in the case of highly skewed degree sequences (Artzy-Randrup et al. 2004). As a result, research in the natural and physical sciences often compares the patterns of local structures to random networks that are conditional on observed degree distributions (e.g., Kashtan and Alon 2005; Shenn-Orr et al. 2002; Yeager-Lotem et al. 2004).

Many other foundational analyses of local network structure apply a random, directed graph distribution known as the U|MAN distribution (Faust 2010; Holland and Leinhardt 1978; Wasserman and Faust 1994) to the study of triads. This uniform distribution conditions on three elements, the numbers of mutual, asymmetric, and null dyads in a directed graph. It 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.

Recent approaches to studying local patterns use multivariate statistical modeling techniques to evaluate the over and under representation of various micro-level structures (e.g., Yaveroğlu et al. 2015). Since relying on a single structural control can yield biased findings, multivariate methods, such as exponential random graph models (ERGMs), hold promise because they can account for various network properties simultaneously. ERGMs represent a statistical network method that can compare the patterns observed in actual networks to what would be expected to occur randomly (Hunter et al. 2008; Robins et al. 2007; Snijders et al. 2006; Wasserman and Pattison 1996). Through the application of such methods, motif researchers can control for competing structural processes that underlie the formation of network edges, such as tendencies towards reciprocity and skewed degree distributions.

### **ERGM-based simulation technique**

Traditional, univariate, conditional distributions used to identify motifs are limited in the extent to which they can incorporate underlying structural patterns. While multivariate statistical techniques can address these limitations, issues of near degeneracy, model instability, and sensitivity can make it challenging to estimate multivariate statistical models, such as ERGMs, particularly across diverse samples of empirical networks (Schweinberger 2011; Yaveroğlu et al. 2015). For instance, ERGMs may fail to converge when structural parameters control for subgraph patterns that are nested within one another, or concentrated in local pockets in a network (Hunter 2007). These problems are especially prevalent when models simultaneously include parameters that count the occurrence of multiple dyad, triad, and tetrad configurations (Yaveroğlu et al. 2015).

Given the limitations and challenges associated with current marginal tests and statistical approaches, we employ an alternative technique that extends the ERGM modeling approach by using the statistical method to simulate random graphs. More specifically,

our approach examines the extent to which motifs occur in a sample of social networks with the use of conditional distributions that are generated using an ERGM framework. ERGMs, or  $p^*$ , models are far from new to network science (Hunter et al. 2008; Robins et al. 2007; Wasserman and Pattison 1996), and numerous studies use this approach in statistical analyses of network data (e.g., Faris et al. 2020; Goodreau et al. 2009; Lusher et al. 2013; McMillan 2019). What is novel here, however, is the application of ERGMs to formulate comparable conditional graphs used to identify network motifs. This new framework has the advantage that it enables flexibility to incorporate the varying strengths of different distributions, as well as account for multiple, endogenous structural processes in the same model (e.g., reciprocity and transitivity). Additionally, since there does not exist a “general” ERGM parameterization that can be applied across all networks (Yaveroğlu et al. 2015), our approach holds promise for comparative network research as it enables researchers to apply more parsimonious tests.

### Hypotheses

In our analyses, we examine the relative frequency of patterns of dyads, triads, and tetrads in our sample of networks, with a view to understanding how these different size substructures provide insight into network characteristics both within and between network genres. We simulate three sets of random networks to examine the prevalence of individual, subgraphs that are conditioned on various structural features of the observed networks (e.g., density, reciprocity, transitivity). First, we begin by examining patterns of dyads, and here *we expect to find reciprocated dyads to be overrepresented in social data sets* (Hypothesis 1). Next, we examine subgraph ratio profiles for each network to uncover both three and four node motifs (Milo et al. 2004). In a robustness analysis, we also note several differences between our findings and those that utilize either the degree distribution or U|MAN. Given prior theoretical and empirical work across various disciplines, *we predict that transitive triads, such as the 030 T triad, will be overrepresented in our data sets* (Hypothesis 2). Due to social clustering, *we expect four-node cliques to be overrepresented* (Hypothesis 3).

Finally, we investigate the extent to which the six network genres exhibit motifs that differ from those of the other types of networks by comparing patterns between all individual networks. That is, we address the following question: *Does each network genre exhibit a triad and tetrad significance pattern that represents a “structural signature”* (Skvoretz and Faust 2002) *that is distinctive, and varies across types of network relations?*

### 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. These data sets each represent anonymized network structures that were collected outside of the current study and are used here for purposes of secondary data analysis. Within each of the six social network genres, we consider four distinct networks, which yields a total sample of 24 graphs. A descriptive summary of all included networks is included in Table 1.

**Table 1** Summary of networks and descriptive statistics

Network	Number of nodes	Directed?	Density	Proportion reciprocated	Clustering coefficient
Friendship 1	1408	Yes	0.0028	0.3966	0.1931
Friendship 2	757	Yes	0.0069	0.4191	0.1867
Friendship 3	729	Yes	0.0068	0.3793	0.1598
Friendship 4	987	Yes	0.0037	0.2726	0.1132
Cosponsorship 1	104	Yes	0.3910	0.5148	0.5667
Cosponsorship 2	101	Yes	0.5629	0.6315	0.6955
Cosponsorship 3	100	Yes	0.5041	0.5931	0.6379
Cosponsorship 4	102	Yes	0.6227	0.7040	0.7552
Twitter 1	1994	Yes	0.0004	0.1640	0.5363
Twitter 2	2399	Yes	0.0003	0.0878	0.4895
Twitter 3	2095	Yes	0.0004	0.0090	0.4682
Twitter 4	2469	Yes	0.0003	0.0076	0.3069
Advice 1	46	Yes	0.4237	0.7457	0.6932
Advice 2	21	Yes	0.4524	0.4737	0.6640
Advice 3	71	Yes	0.1795	0.3924	0.4433
Advice 4	56	Yes	0.0776	0.3013	0.3464
Email 1	309	Yes	0.0318	0.7212	0.5158
Email 2	162	Yes	0.0679	0.8205	0.7118
Email 3	142	Yes	0.0687	0.7884	0.7001
Email 4	180	Yes	0.0800	0.5943	0.3507
Terrorism 1	34	No	0.1034		0.5000
Terrorism 2	27	No	0.3191		0.5647
Terrorism 3	27	No	0.4444		0.7224
Terrorism 4	27	No	0.2336		0.5863

For our adolescent friendship data, we randomly selected four school-based networks from the in-school survey collected during the first wave of the National Study of Adolescent to Adult Health (Add Health) (Harris et al. 2009). During this wave, entire student bodies from a nationally representative sample of U.S. middle and high schools were surveyed and respondents were asked to nominate up to ten of their closest school friends. From these friendship nominations, we constructed directed networks in which 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 interactions 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  $a$  co-sponsored at least one piece of legislation for which senator  $b$  was the primary sponsor.

The Twitter networks used in our sample were previously gathered outside of this study (Felmlee et al. 2020) via the Twitter API using a keyword search function, based on a period of one week at the end of February 2017. Tweets were collected that contained aggressive, harmful terms (i.e., curse words) that targeted women and minorities. Networks of online interaction were created by downloading connected messages based

on retweets and “likes.” Each of the four networks represents a network of retweets and likes that evolved over the week in response to an initiating tweet that included a racial or gendered slur. Each network contained several negative, conflictual ties but also positive ties of support. The networks used in the current study were in an anonymous form.

Our four advice networks were collected from surveys administered to employees in four different workplaces, including a law firm (Lazega 2001), high-tech company (Krackhardt 1987), IT department of a Fortune 500 company (Almquist 2014), and consulting company (Cross and Parker 2004). In each survey, employees were asked to nominate the coworkers whom they would ask for professional advice. From their nominations, we were 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 include three networks from different administrative departments in the European Union (Leskovec et al. 2007) and one from the company ENRON (Klimt and Yang 2004). For all four networks, we consider email-sending patterns over an eighteen-month period. We construct a directed network where nodes represent individual employees, and 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 and ARTIS Transnational Terrorism Database (2009). Each network is focused on 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.<sup>1</sup>

### Analytical approach

Our novel approach to identify those dyads, triads, and tetrads that occur more frequently than expected consists of three steps. First, we estimate an exponential random graph model (ERGM) on each of the observed networks that includes controls for the structural processes of interest. Second, we use the coefficient values of the estimated ERGM to simulate 1000 conditional graphs. Finally, we compare the prevalence of subgraphs across the observed and random graphs by calculating subgraph ratio profiles (SRPs).

#### Step 1. Exponential random graph models (ERGMs)

To simulate conditional graphs, it is first necessary to estimate ERGMs that account for structural processes of interest on each of the observed networks. Here, we define  $\mathbf{Y}$  as

---

<sup>1</sup> We note that the original data sets from which we extract network ties were collected using a variety of methods (e.g., survey versus direct observation, varying nomination limits). For instance, all four advice networks originate from different sources. However, despite these variations, our results suggest that all the advice networks are substantially more similar to one another in their local structural foundations than they are to any of the other types of networks in our sample. This finding increases our confidence that the subgraph patterns we observe are primarily the result of social processes, as opposed to data collection strategies, not unlike the results from other analyses of multiple data sets (Skvoretz and Faust 2002).



an  $n \times n$  matrix (where  $n$  equals the number of actors) where the  $(i, j)$  entry of this matrix is equal to 1 if there is a relational tie between actors  $i$  and  $j$  or 0 if no such tie exists. The ERGM specifies the probability that a set of relational ties  $\mathbf{Y}$  will form given a set of individuals:

$$P(\mathbf{Y} = y | \mathbf{X}) = \frac{\exp[\theta^T g(y)]}{k(\theta)}$$

The  $\mathbf{X}$  term represents a matrix of covariates and  $\theta$  represents a vector of all network coefficients hypothesized to relate to the probability of the observed network's formation. A vector of network statistics,  $g(y)$ , is calculated using the observed adjacency matrix and  $k(\theta)$  is a normalizing factor that ensures the equation predicts a legitimate probability distribution.

In the current project, we estimate three sets of ERGMs that include different specifications according to whether they will be utilized to simulate graphs that assess the overrepresentation of dyads, triads, or symmetric tetrads. To identify those *directed dyads* that occur more frequently than expected, we first estimate an ERGM on each of the twenty directed graphs that produces random graphs that are tantamount to an Erdős-Rényi or Bernoulli model (Goodreau et al. 2008). These ERGMs represent the simplest possible model and only include a control for the *base log odds of a tie*, which accounts for the likelihood that an edge will exist between any two actors in the network. To identify overrepresented *directed triads*, we estimate a second set of ERGMs on each of the directed networks in our sample that generate random graphs that are identical to those that condition on the dyad distribution of the observed network. This second set of ERGMs includes two controls: one for *reciprocity* and one for the base log odds of a tie. The reciprocity parameter accounts for actors' tendencies to send ties to those actors from whom they also receive social ties.

Finally, we estimate a set of ERGMs that enable us to uncover the *symmetric tetrads* that are overrepresented in our sample of networks. Following the precedent of previous motif studies (e.g., Hale and Arteconi 2008), we focus on tetrads in which all four nodes are connected. We consider symmetric tetrads, rather than directed tetrads, moreover, because it reduces the number of tetrads in which each node is adjacent to at least one edge from 199 to 6.

This third set of ERGMs introduces *open triad* and *closed triad* variables, which control for tendencies towards transitivity. The open triad variable accounts for the necessary preconditions, while the closed triad term measures tendencies towards the structural phenomenon itself (Hunter 2007). When we estimate this third set of ERGMs on the directed networks in our sample, we include controls for the base log odds of a tie and reciprocity. Among the symmetric networks in our sample (i.e., the terrorism networks), the ERGMs only include the triad-level variables and a term for the base log odds of a tie. Across these various models, the base log odds of a tie term refers to the edges parameter, reciprocity refers to the mutual parameter, open triads is the geometrically

weighted dyad-wise parameter, and closed triads is the geometrically weighted edge-wise parameter.<sup>2</sup>

**Step 2. ERGM simulations**

After estimating each ERGM (see discussion of convergence and goodness of fit in the Additional file 1), we use the models’ coefficients to simulate samples of random networks. This enables us to generate random networks that are conditional on the structural phenomena parameterized in the model (for additional details, see Hunter et al. 2008). For each of the directed networks, we generate 1000 random networks according to each of the three sets of ERGMs, resulting in 60,000 simulated networks (20 observed networks × 3 ERGM specifications × 1000 simulations). To uncover which symmetric tetrads occur more frequently than expected, we weakly symmetrize our observed networks and the corresponding simulated networks (i.e., a tie exists between node *a* and node *b* if either node *a* or node *b* send ties to one another). Because most networks in our sample are relatively sparse, this method of symmetrizing is useful. To make comparisons among the symmetric terrorist networks, we generate 1000 random networks based on the specification of the symmetric ERGM that controls for triadic patterns (4 observed networks × 1000 simulations). Taken together, these analyses result in a final sample of 64,000 random, conditional networks.

**Step 3. Subgraph Ratio Profiles (SRPs)**

To consider the prevalence of dyads across our networks, we compare the proportion of reciprocated dyads in our observed networks to the average proportion across their associated conditional networks. We calculate the subgraph ratio profile (SRP) for each network’s distribution of triads and tetrads (Milo et al. 2004). This ratio compares what we see in the observed graph to a set of randomly simulated graphs, and it forms the basis for motif analysis across the wider literature. For each subgraph *i* in an individual graph, we start by calculating the following:

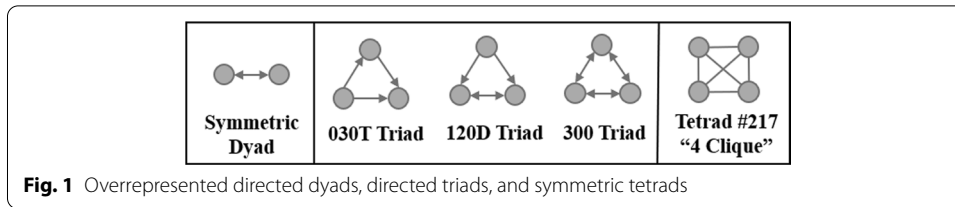
$$\Delta_i = \frac{N_{observed_i} - \langle N_{random_i} \rangle}{N_{observed_i} + \langle N_{random_i} \rangle + \epsilon}$$

where  $N_{observed_i}$  is the number of observed *i* subgraphs in the network and  $\langle N_{random_i} \rangle$  is the mean number of such subgraphs in the random conditional graphs. The  $\epsilon$  is an error term to make sure that  $\Delta_i$  is not too large when the subgraph rarely appears in both the observed and random conditional graphs. The term is set to three for triads, and four for tetrads (Milo et al. 2004). We use these values to calculate the SRP, or a normalized vector of all  $\Delta_i$ :

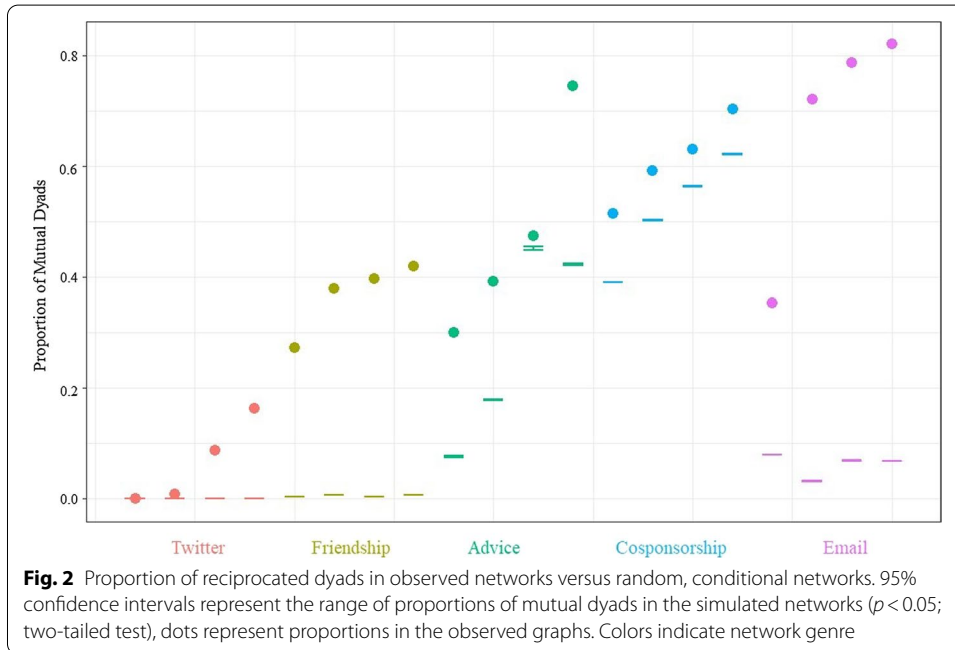
$$SRP_i = \Delta_i / \sqrt{\sum \Delta_i^2}$$

---

<sup>2</sup> In an attempt to evaluate these subgraph patterns through a multivariate modeling approach, we began by estimating preliminary ERGMs with parameters that measure the count of specific three- and four-node configurations (e.g., the 030 T triangle, the kite subgraph). Unfortunately, most of these models did not converge because of problems with degeneracy, and as a result, we adopt the alternative simulation approach outlined here.



**Fig. 1** Overrepresented directed dyads, directed triads, and symmetric tetrads



**Fig. 2** Proportion of reciprocated dyads in observed networks versus random, conditional networks. 95% confidence intervals represent the range of proportions of mutual dyads in the simulated networks ( $p < 0.05$ ; two-tailed test), dots represent proportions in the observed graphs. Colors indicate network genre

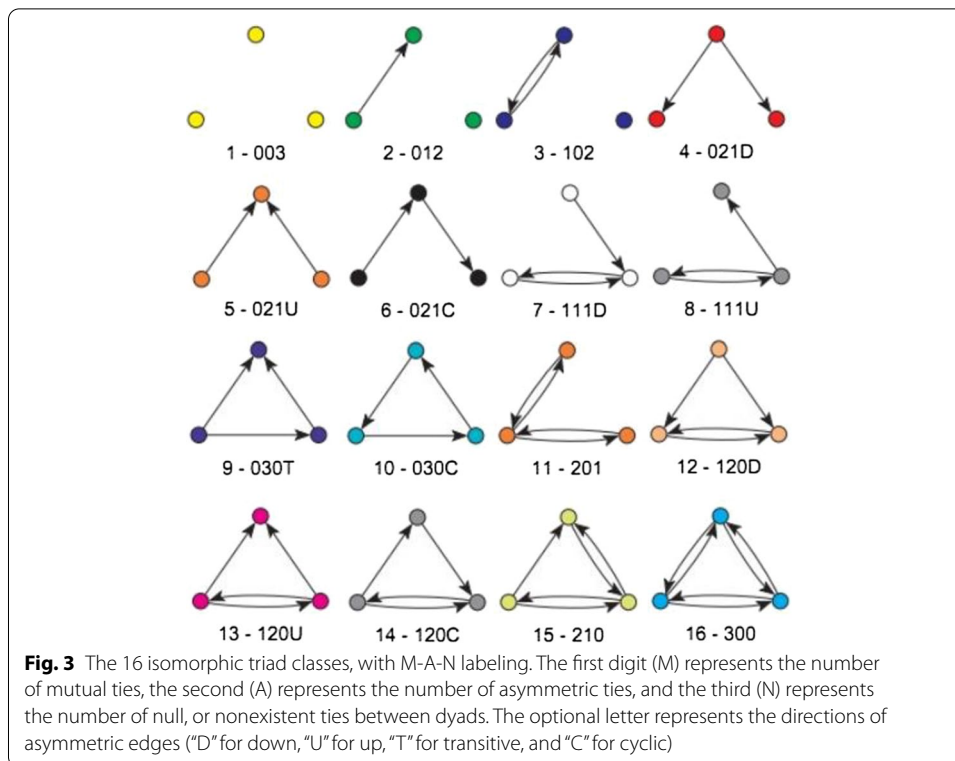
For each subgraph and each network, we calculate a vector of SRPs with a length equal to the number of potential isomorphic subgraphs. A positive  $SRP_i$  indicates that subgraph  $i$  is more likely to occur in the observed graph when compared to the conditional distribution, while a negative value suggests that subgraph  $i$  is less common than would be expected. We define motifs as those subgraphs  $i$  where the  $SRP_i$  for each individual network is greater than or equal to zero. Furthermore, since SRPs are normalized according to the number of nodes in the graph, we can compare  $SRP_i$ s across networks.

**Results: ERGM simulation approach**

When we use ERGMs to simulate multivariate, conditional graphs, we identify five motifs, including one dyad, three triads, and one symmetric tetrad. These motifs appear in Fig. 1.

**Dyads**

For all five genres of directed social networks, reciprocated dyads consistently appear more frequently than expected when compared to random, simulated graphs that are conditioned on the number of observed nodes and edges (see Fig. 2). The proportion of reciprocated dyads is as high as 0.82 in some of the denser email networks. For our



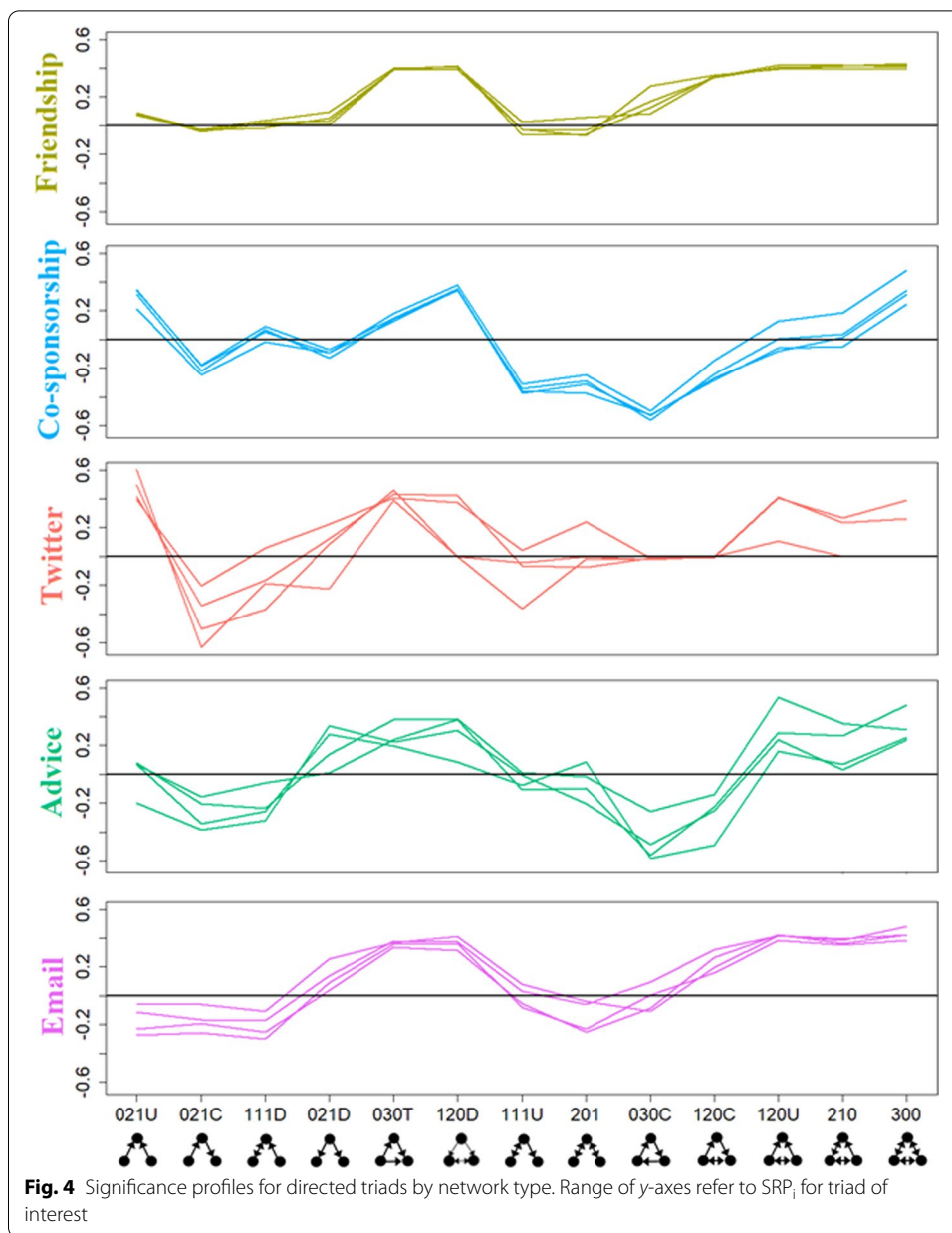
sparser networks, such as the Twitter graphs, the proportion of reciprocated dyads reaches levels as low as 0.001. However, the proportion of reciprocated dyads occurring in all the observed networks remains higher than what we would expect ( $p < 0.001$ ), even in those graphs with low density.<sup>3</sup>

### Triads

While there exists a total of 16 possible isomorphic directed triads (see Fig. 3), we only consider those 13 triads in which all three nodes in the subgraph are connected, following previous research (e.g., Benson et al. 2016), and because we report dyad patterns above. We compare the prevalence of these 13 triads across simulated graphs that are conditional on the observed networks’ tendencies towards reciprocity and counts of nodes and edges.

Average SRPs for the five directed network genres show an overrepresentation of the triad consisting of all three mutual ties (i.e., the “300” triad, following Holland and Leinhardt 1971) (see Fig. 4). The “300” triads are particularly prominent in friendship (mean  $SRP_i = 0.418$ ) and email networks (mean  $SRP_i = 0.427$ ) (see Table 2), implying that small, densely interconnected cliques play a key role in the structure of these networks. The classic transitive triad (i.e., “030 T” triad), also is overrepresented, and it signifies a second triad motif. This graphlet is particularly prominent in networks of friendship,

<sup>3</sup> Note that the mutuality coefficients included in our second and third sets of ERGMs were consistently positive and statistically significant. This finding further highlights the overrepresentation of mutual ties across our sample of networks.



Twitter, and email. Finally, the “120D” triad, which consists of one mutual tie and two asymmetric ties directed towards the mutual tie, also represents a motif. The “120D” triad is also transitive and suggests simultaneous patterns of clustering and hierarchy; both actors involved in the mutual dyad receive unreciprocated ties from the third actor.

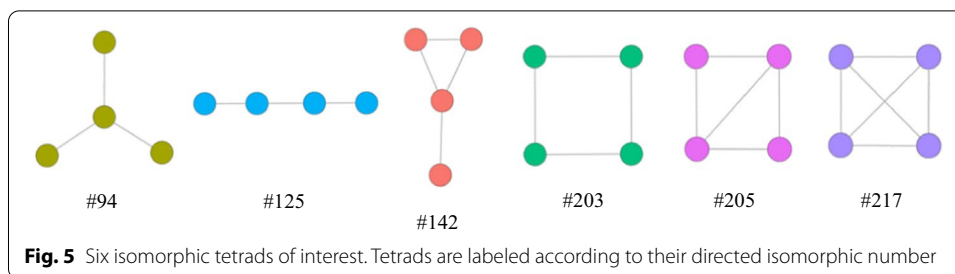
**Tetrads**

Of the six possible symmetric tetrads (see Fig. 5), one can be characterized as a motif in our data (see Table 2). The “four clique” (i.e., four-node subgraph #217) is more common than we would expect across all network genres (see Fig. 6), as hypothesized. Note that the four clique represents a motif, even after we control for reciprocity and triad

**Table 2** Average SRP<sub>i</sub> for directed triads and symmetric tetrads by network genre

	Friendship		Cosponsorship		Twitter		Advice		Email		Terrorism	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
<i>Triads</i>												
021U	0.084	0.008	0.304	0.062	0.476	0.095	0.004	0.134	-0.170	0.100		
021C	-0.035	0.008	-0.207	0.034	-0.420	0.186	-0.270	0.109	-0.170	0.082		
111D	0.014	0.021	0.049	0.045	-0.167	0.175	-0.218	0.113	-0.208	0.086		
021D	0.046	0.038	-0.097	0.025	0.053	0.193	0.190	0.145	0.131	0.093		
030T	0.392	0.004	0.151	0.023	0.422	0.031	0.260	0.082	0.360	0.020		
120D	0.409	0.010	0.354	0.017	0.199	0.231	0.286	0.140	0.366	0.040		
111U	-0.023	0.036	-0.346	0.029	-0.107	0.178	-0.044	0.055	-0.006	0.075		
201	-0.026	0.058	-0.305	0.055	0.037	0.138	-0.058	0.122	-0.145	0.113		
030C	0.166	0.082	-0.529	0.026	-0.016	0.007	-0.472	0.149	-0.023	0.094		
120C	0.344	0.008	-0.236	0.063	-0.003	0.003	-0.276	0.150	0.236	0.073		
120U	0.407	0.010	-0.001	0.093	0.231	0.210	0.306	0.160	0.409	0.018		
210	0.413	0.011	0.049	0.099	0.125	0.145	0.179	0.154	0.374	0.020		
300	0.418	0.014	0.347	0.100	0.162	0.194	0.322	0.111	0.427	0.041		
<i>Tetrads</i>												
94	0.168	0.119	0.096	0.138	0.538	0.066	0.153	0.323	0.109	0.222	0.238	0.096
125	0.117	0.099	-0.254	0.121	0.196	0.095	-0.301	0.165	-0.338	0.184	-0.161	0.119
142	0.271	0.061	-0.074	0.038	0.400	0.019	0.040	0.125	0.004	0.260	0.184	0.139
203	0.591	0.045	-0.831	0.102	0.523	0.064	-0.443	0.308	-0.570	0.330	-0.400	0.443
205	0.377	0.081	-0.147	0.120	0.406	0.076	0.158	0.198	0.038	0.330	0.207	0.335
217	0.603	0.053	0.384	0.146	0.166	0.196	0.633	0.263	0.488	0.209	0.613	0.236

Labeling for triads based on the labeling scheme suggested in Holland and Leinhardt (1971). Labeling for tetrads refers to isomorphic number SD, Standard deviation



patterns in our ERGMs. This higher-level clustering is particularly common in the terrorism and workplace email networks, both of which represent social interaction that requires high levels of cooperation to achieve group goals.<sup>4</sup>

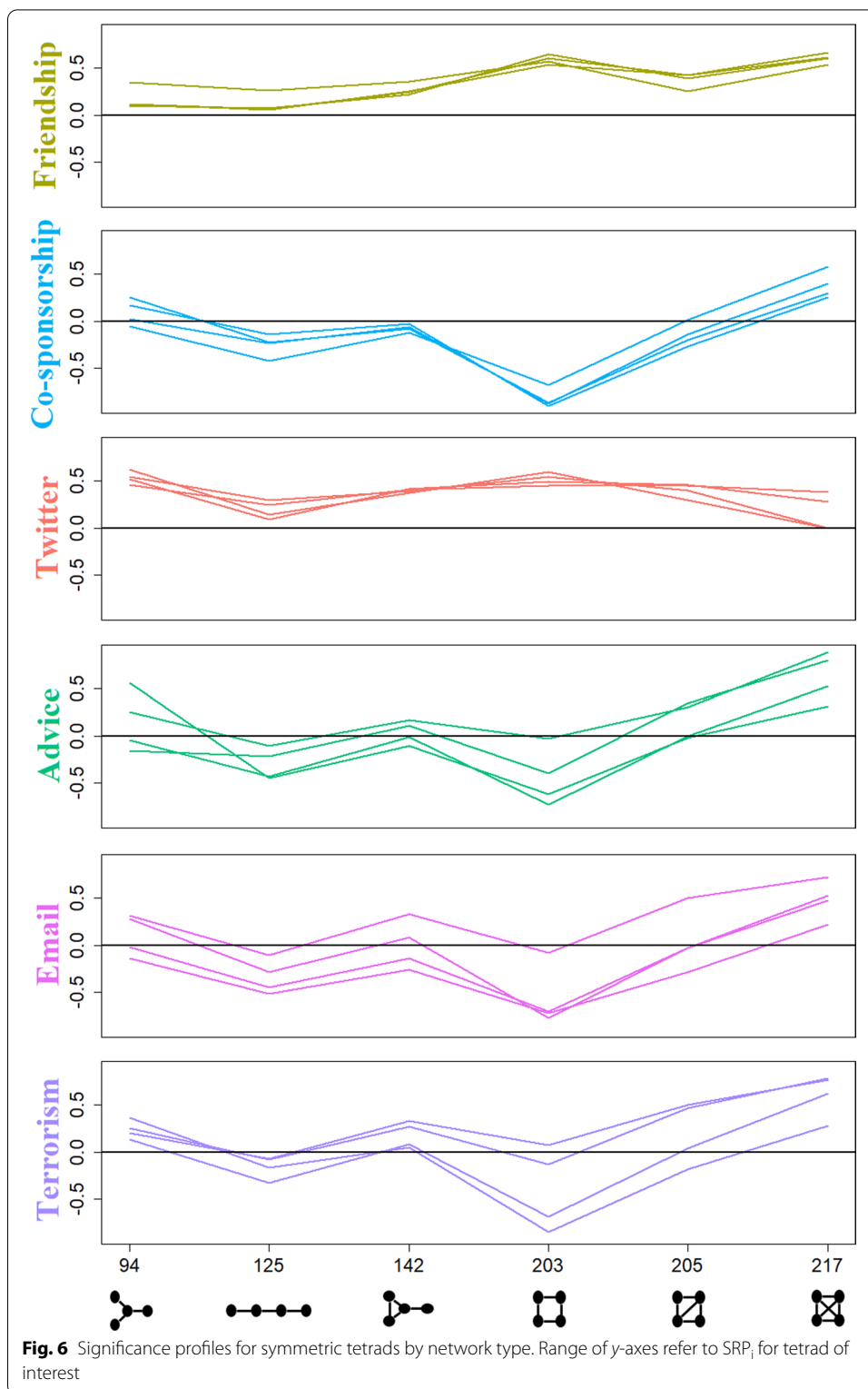
#### Variation within and between network genres

While several motifs are more common than expected among all the social networks in our sample, there are also key differences. From calculating the correlations between vectors of the directed triad SRPs and symmetric tetrad SRPs for each pair of networks, it is apparent that SRPs tend to be more similar within network genres than between genres, as confirmed by a two-sample  $t$ -test ( $p < 0.001$ ) (see Fig. 7). Within network genres, the average correlation among SRP vectors is relatively high (0.863), whereas the average correlation between SRPs of different types of networks is more moderate (0.392). Overall, we conclude that each network genre tends to have a highly correlated structural pattern, that is, a unique “fingerprint,” based on the significance profiles for triads and tetrads.

Comparing correlations among SRP vectors suggests that certain subgraphs are prevalent within some network genres, but not others. For example, triad “120C” is an intransitive triad that includes a mutual dyad and a cyclic pattern between asymmetric ties. The “120C” triad is less common than expected according to the null model among the co-sponsorship and advice networks (mean SRP<sub>i</sub>s are  $-0.236$  and  $-0.276$ , respectively), more common than expected in the friendship and email data (mean SRP<sub>i</sub>s are  $0.344$  and  $0.236$ , respectively), and close to zero in Twitter (mean SRP<sub>i</sub> is  $-0.003$ ). Friendship and email networks exhibit particularly high levels of triads that build towards cliques (e.g., 210 and 300), and the “120C” triad likely represents a step in that direction. See Fig. 8 for a visual comparison of the relative prevalence of the 120C triad in each type of network with one sample graph for each type.

In another interesting example, the prevalence of the kite tetrad (i.e., tetrad #142) also varies across network genre. Compared to our null model, the kite tetrad occurs more frequently than expected in our friendship, Twitter, and terrorism networks (mean SRP<sub>i</sub>s are  $0.271$ ,  $0.400$ , and  $0.184$ , respectively). The kite tetrad appears about as frequently as expected across the advice and email networks (mean SRP<sub>i</sub>s are  $0.040$  and  $0.004$ , respectively), but it is less common within co-sponsorship networks (mean SRP<sub>i</sub> =  $-0.074$ ).

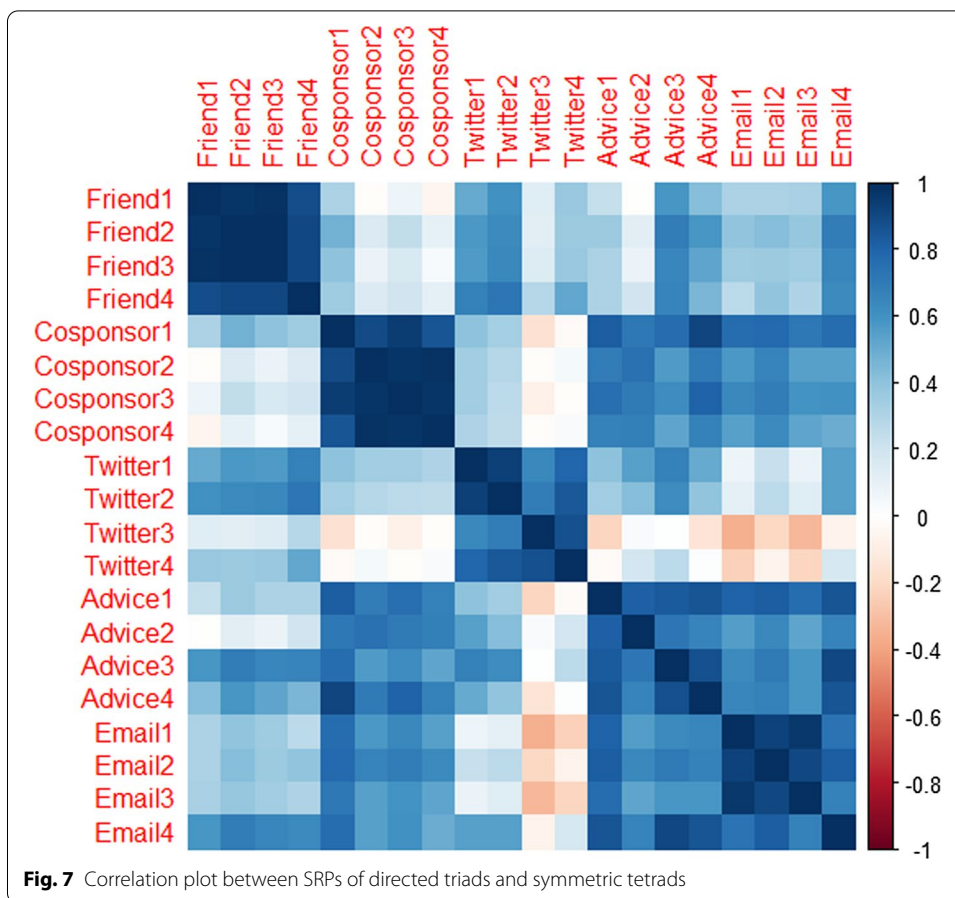
<sup>4</sup> It is important to note that Tetrads #203 and #205 are embedded within the structures that the geometrically weighted open and closed triad ERGMs parameters model. While the decay parameters help ensure that these tetrads are not perfectly controlled, we exercise caution when interpreting the results for these two symmetric tetrads.



**Comparing with results from degree and dyad conditional distributions**

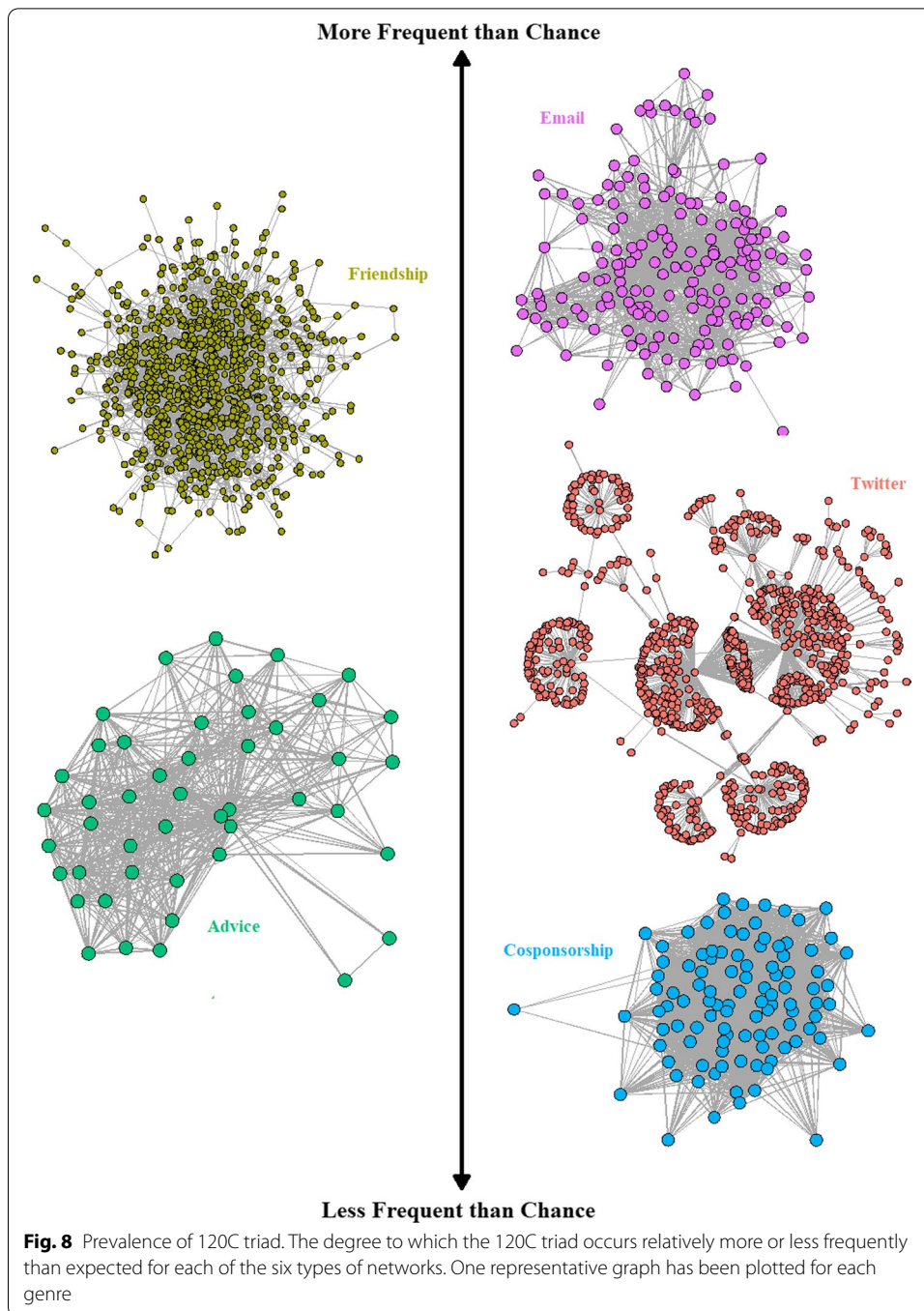
Next, we examine the occurrence of graphlets across our data using two other distributions: one that controls for the conditional dyad distribution (U|MAN) and another





that uses degree sequence. As expected, we uncover multiple variations in motif patterns, depending on the approach used to generate random graphs (see Tables 3, 4 in the Appendix). One of the most notable differences concerns the well-known 030 T, transitive triad. When we use either our multivariate, ERGM approach or the dyad conditional distribution, this triad is identified as a common motif, whereas when we only control for degree sequence, this triad tends to occur less frequently than we would expect (mean within-genre  $SRP_i$  reaches a minimum of  $-0.348$ ). The 030 T triad is not over-represented under the latter set of controls because this subgraph lacks reciprocal ties, which are exceptionally common among social graphs. Since our ERGM technique and the U|MAN distribution control for the number of asymmetric dyads and there are only two ways in which three asymmetric ties can be arranged, the 030 T triad is more common than expected with these latter controls.

Another key difference concerns the three-star tetrads (tetrad #94), which are over-represented in 19 of our 24 networks when we compare the observed data to networks generated using our multivariate, simulation framework. Interestingly, these tetrads instead are underrepresented in 10 of our observed networks when random graphs are generated according to the U|MAN dyad distribution. Most notably, our ERGM approach finds that three-stars occur more frequently than expected across the email networks, while conditioning comparison graphs solely on the dyad-distribution



leaves the impression that these tetrads are underrepresented (average  $SRP_i$ s are 0.109 and  $-0.127$ , respectively). When comparison graphs only take the degree distribution into account, three-star tetrads also are underrepresented across all (but one) of the 24 networks in our sample, demonstrating key discrepancies in findings between the various approaches.

## Discussion

Using a novel, multivariate approach to identify motifs, we find extensive evidence of recurring local patterns of network structure within a relatively large sample of social networks. The five motifs we uncover point to the relevance of fundamental, micro-level processes within the social world that occur with some commonality across diverse forms of activity and interaction. Certain motifs emerge in our graphs that are common in various disciplines, such as the four-clique, which predominates in graphs of protein structures and electrical power grids (e.g., Milo et al. 2004). In other words, clustering among four nodes crosses scientific domains. The 030 T transitive triad, or “feed-forward loop,” also is overrepresented in biology and in our sample of networks. Not only do patterns of omnivory in ecology and genes in biology display triadic closure, in other words, but humans also extend a diverse array of social ties in this manner. Although early social network research documents transitivity in largely face-to-face, friendly, interactions (Holland and Leinhardt 1971; Hallinan 1974), here we find evidence of this phenomenon in multiple human behaviors, such as positive and negative online communications and legislative actions. What remains especially remarkable, however, is the cross-disciplinary predominance of this type of network subgraph, underlining the prominence of local hierarchy processes across a diversity of biological and social structures. To the best of our knowledge, comparisons of local network patterns across the social, natural, and engineering sciences receive limited explicit attention.

Not unlike their counterparts in biology and ecology, social graph motifs also signify functional aspects of networks. For example, symmetric dyads were more likely to occur in all the observed networks when compared to randomly simulated networks that condition on density. Reciprocity of ties, thus, represents a basic element of many social interactions, including those consisting of face-to-face or electronic interchanges, bonds of affinity, strategic, political connections, as well as informal or formal bases of organization. Taken together, our results provide ample support for the “norm of reciprocity,” which is associated with the maintenance of stable social systems and the collective benefit of cooperation (Gouldner 1960).

The motifs identified here also point to two latent yet fundamental social processes, those of clustering and hierarchy. The frequent appearance of completely connected dyads, triads, and tetrads in this sample of social graphs highlights that actors often form clusters of ties, and the overabundance of transitive triads with asymmetric ties (i.e., 030 T and 120D) implies the occurrence of social hierarchy. Both clustering and hierarchy represent fundamental social processes in formative theories of group processes (Holland and Leinhardt 1971; Homans 1961), and we find evidence that these mechanisms are interconnected. Clusters tend to emerge among strong ties in our sample of networks, but in addition, weak, intransitive ties are present that bridge these clusters in most graphs. Finally, although there remain exceptions, network genres tend to be much more similar than different in their triad and tetrad significance patterns. These results suggest a tendency towards network structure signatures, or fingerprints, that emerge based solely on two small substructures, that is, triads and tetrads.

Notable differences between genres of networks also arise. For example, the kite tetrad was overrepresented in several networks genres, but less common in the co-sponsorship networks of the U.S. Senate. Given that the kite tetrad portends bridging

across structural holes (Burt 2004), the prevalence of kite tetrads suggests that many of our graphs contain both strong ties that lead to clustering and weak ties that result in bridging. Co-sponsorship networks, on the other hand, are likely to be characterized by less bridging and contain fewer weak ties, which may be an artifact of the highly partisan, two-party political system in the United States. In other words, only rarely does a U.S. legislator attempt to push through legislation by bridging a gap across the political divide.

Finally, our results highlight the importance of selecting appropriate, theoretically-driven conditional distributions for baseline comparison within motif research. We argue that a multivariate ERGM approach for generating random comparison graphs has utility for future research that attempts to identify recurrent, local structural patterns, due to its ability to incorporate various lower-level, structural controls simultaneously. In the current project, we show that univariate distributions, such as the degree distribution and U|MAN distribution, can yield differing conclusions from those using our multivariate method. For example, the 030 T triad tends to occur *more* frequently than expected using our approach but *less* frequently than expected when our observed networks are compared to random graphs conditional on the degree distribution. This three-node configuration represents a transitive triad, which theories of balance predict to be more likely in social graphs than would be expected (e.g., Heider 1946). The fact that our results align with theoretical predictions and trends from other disciplines (e.g., the feed-forward-loop), provides support for the strengths of our ERGM simulation methodology.

In another example, three-star tetrads are overrepresented in the majority of networks with our multivariate simulation approach, while they are often underrepresented when using either of the two univariate distributions to produce random graphs. These differences highlight the importance of accounting for triad-level properties in a multivariate model when considering patterns with four or more nodes. Given that all our networks are defined by transitivity, four-node motifs that consist of many intransitive triads (e.g., the three-star subgraph includes three intransitive triples) will be estimated to occur *less* frequently than chance without such a control. Moreover, it is reasonable to expect a three-star pattern to emerge frequently in several social interactions, such as in email, where one administrative person could send messages to multiple employees, without expectations of triad closure. Our multivariate approach, unlike those of some traditional distributions, identifies this type of key, three-star regularity.

Our research is not without limitations, nevertheless. Although our sample of networks is relatively large, it is not random and does not represent many forms of social interaction. Findings could differ noticeably, depending on the type of human behavior studied, as well as the sample of networks and data collection methods. In addition, we examine networks at a cross-section. Future research could benefit from an examination of change in micro-level, network structures over time. Moreover, the conclusions drawn from any multivariate model depend on the specific control variables, an issue that becomes especially pertinent with highly endogenous network data (see Duxbury 2021). We conditioned on both mutuality and triad tendencies because patterns of reciprocity and transitivity were overrepresented in all our graphs, and because the inclusion

of these measures was informed by well-established theory. Yet alternative control variables could shape findings, and additional research is required to address other models. The multivariate approach we use here holds promise for such investigations.

Future research also needs to continue to explore which type of local structures drive network formation over time. Typical controls for structural interdependence in motif research assume that lower-level graph properties could account for higher-level graph properties, with density, degree, and reciprocity, for example, helping to explain triad or tetrad properties. Yet we usually do not know which property precedes the other. Perhaps tendencies to form 030 T triads condition the development of network degree, or reciprocity, instead of the other way around. Theories of balance (Cartwright and Harary 1956; Heider 1946) suggest that tendencies towards transitivity would supersede other social phenomena, like the propensity for well-connected individuals to become increasingly central over time. Applying temporal motifs (Paranjape et al. 2017) to investigate sequential, communication patterns (Michienzi et al 2021; Zignani et al 2018), for example, could be helpful in addressing these issues.

In sum, our work builds on a long tradition in which researchers investigate patterns among network graphlets, or in our case, “motifs” (e.g., Moreno 1934; Holland and Leinhardt 1971, 1978; Milo et al. 2002; Wasserman and Pattison 1996). Several advantages exist in further extending the focus on overrepresented subgraphs. For example, patterns of dyads, triads, and tetrad motifs could aid in determining which local structural measures to include in multivariate network models. Note also that motifs contain the seeds of change. Social networks characterized heavily by reciprocity would be expected to turn asymmetric links into those that are symmetric over time, and those composed of high levels of transitivity would be predicted to close open triads.

By understanding how network motifs are similar and vary across a range of social network types, our findings also have the potential to inform social theories. For instance, our results show that all our networks display overrepresented levels of clustering at the triad and tetrad levels. When crafting research questions, therefore, theories of tie formation that apply to one of these seemingly disparate categories of social relationships may be informative for the others. Our results identify network motifs that are prevalent not only in social networks, but also in other scientific disciplines, which raises the potential for scholarly implications that cross disciplinary boundaries. Finally, our unique methodological approach, in which we employ multivariate, exponential random graph models to generate comparable random graphs for identifying local structural patterns, increases confidence in the robustness of our findings.

## Appendix

See Tables 3 and 4.

**Table 3** Average SRP<sub>i</sub> for directed triads and symmetric tetrads by network genre according to the UJMAN distribution

	Friendship		Cosponsorship		Twitter		Advice		Email		Terrorism	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
<i>Triads</i>												
021U	0.081	0.015	0.300	0.062	0.430	0.045	0.001	0.133	-0.170	0.100		
021C	-0.044	0.016	-0.209	0.033	-0.324	0.123	-0.266	0.110	-0.170	0.082		
111D	-0.044	0.105	0.049	0.044	-0.224	0.191	-0.214	0.112	-0.208	0.085		
021D	0.028	0.051	-0.100	0.026	0.046	0.202	0.188	0.145	0.129	0.094		
030T	0.395	0.010	0.147	0.023	0.435	0.025	0.257	0.077	0.362	0.020		
120D	0.404	0.013	0.353	0.017	0.279	0.191	0.287	0.139	0.366	0.040		
111U	-0.095	0.079	-0.345	0.030	-0.204	0.199	-0.047	0.057	-0.007	0.075		
201	-0.127	0.147	-0.301	0.057	0.043	0.122	-0.053	0.121	-0.145	0.114		
030C	0.160	0.054	-0.530	0.026	-0.022	0.016	-0.475	0.146	-0.025	0.102		
120C	0.321	0.058	-0.235	0.065	-0.002	0.002	-0.274	0.146	0.236	0.071		
120U	0.402	0.006	-0.002	0.093	0.315	0.131	0.301	0.154	0.408	0.016		
210	0.390	0.042	0.052	0.100	0.136	0.158	0.183	0.157	0.372	0.021		
300	0.405	0.035	0.354	0.100	0.168	0.199	0.335	0.111	0.426	0.042		
<i>Tetrads</i>												
94	-0.031	0.063	0.086	0.151	0.462	0.030	0.057	0.285	-0.127	0.274	0.046	0.088
125	-0.036	0.069	-0.253	0.115	0.189	0.182	-0.297	0.153	-0.325	0.173	-0.184	0.141
142	0.494	0.015	-0.070	0.029	0.459	0.025	0.077	0.185	0.315	0.155	0.351	0.128
203	0.323	0.099	-0.821	0.118	0.450	0.037	-0.469	0.296	-0.175	0.258	-0.308	0.309
205	0.564	0.027	-0.137	0.139	0.450	0.015	0.288	0.299	0.527	0.057	0.486	0.112
217	0.562	0.026	0.397	0.164	0.282	0.193	0.604	0.197	0.564	0.069	0.627	0.055

Labeling for triads based on the labeling scheme suggested in Holland and Leinhardt (1971). Labeling for tetrads refers to isomorphic number SD, Standard deviation

**Table 4** Average  $SRP_i$  for directed triads and symmetric tetrads by network genre according to the degree distribution

	Friendship		Cosponsorship		Twitter		Advice		Email		Terrorism	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
<i>Triads</i>												
021U	-0.209	0.106	-0.147	0.015	-0.006	0.004	-0.148	0.116	-0.375	0.025		
021C	-0.262	0.077	-0.267	0.015	-0.321	0.133	-0.295	0.103	-0.358	0.025		
111D	0.042	0.205	0.118	0.020	-0.020	0.349	0.000	0.114	-0.203	0.071		
021D	-0.202	0.084	-0.175	0.046	-0.112	0.088	-0.107	0.081	-0.303	0.031		
030T	-0.026	0.191	-0.348	0.020	0.454	0.261	-0.167	0.105	-0.338	0.037		
120D	0.238	0.157	-0.050	0.017	0.321	0.217	0.116	0.160	-0.084	0.086		
111U	0.073	0.122	0.119	0.042	0.121	0.251	0.083	0.088	-0.096	0.021		
201	0.354	0.090	0.134	0.035	0.158	0.194	0.320	0.136	0.206	0.038		
030C	-0.261	0.073	-0.711	0.013	-0.053	0.089	-0.499	0.113	-0.412	0.067		
120C	0.0200	0.195	-0.321	0.037	-0.047	0.080	-0.333	0.217	-0.247	0.018		
120U	0.265	0.137	-0.001	0.023	0.327	0.218	0.079	0.246	-0.003	0.116		
210	0.367	0.058	-0.038	0.069	0.136	0.158	0.168	0.216	0.186	0.094		
300	0.466	0.075	0.285	0.086	0.182	0.214	0.346	0.109	0.354	0.024		
<i>Tetrads</i>												
94	-0.125	0.03	-0.726	0.112	-0.001	0.001	-0.048	0.503	-0.333	0.115	-0.282	0.038
125	-0.100	0.027	0.181	0.099	-0.583	0.077	-0.202	0.019	-0.425	0.106	-0.311	0.045
142	0.362	0.016	0.101	0.06	-0.188	0.296	-0.016	0.08	-0.025	0.097	-0.018	0.002
203	0.197	0.037	-0.584	0.122	0.205	0.195	-0.296	0.672	-0.448	0.081	-0.743	0.065
205	0.588	0.014	-0.177	0.019	0.273	0.443	0.088	0.212	0.315	0.089	-0.040	0.110
217	0.675	0.007	0.140	0.100	0.090	0.582	0.393	0.431	0.601	0.118	0.490	0.142

Labeling for triads based on the labeling scheme suggested in Holland and Leinhardt (1971). Labeling for tetrads refers to isomorphic number SD, Standard deviation

### Abbreviations

ERGMs: Exponential random graph models; U|JMAN: Uniform distribution conditional on a fixed dyad census; Add Health: National Study of Adolescent and Adult Health; API: Application programming interface; SRP: Subgraph ratio profile; SD: Standard deviation.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-021-00403-5>.

**Additional file 1.** ERGM Goodness of Fit.

### Acknowledgements

One portion of the study uses Add Health data. The Add Health data collection is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01 AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I–V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

### Authors' contributions

DF conceptualized and coordinated the study, acquired funding and data, interpreted findings, drafted and revised the manuscript. CM performed data analysis, contributed to conceptualization, acquired data, interpreted findings, drafted and revised the manuscript. RW contributed to conceptualization, drafted and revised the manuscript. All authors read and approved the final manuscript.

### Funding

This research received assistance from the Penn State Population Research Institute, which is supported by an infrastructure grant by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P2CHD041025) under the of the National Institutes of Health under award number 2P2CHD041025. This research was sponsored in part by the DAIS-ITA, the U.S. Army Research Laboratory, and the U.K. Ministry of Defense 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 Defense or the U.K. Government. Partial support was provided by Pennsylvania State University and the National Science Foundation under an IGERT award # DGE-1144860, Big Data Social Science.

### Availability of data and materials

The data sets analyzed in the current study are available from the corresponding author on reasonable request. In one exception, legal restrictions prohibit the sharing of the friendship data (National Study of Adolescent and Adult Health – Add Health). Interested readers can obtain information regarding data acquisition at (<https://addhealth.cpc.unc.edu/>).

### Declarations

#### Ethics approval and consent to participate

We use secondary data sources for our analyses and complied with the terms of service for all websites and databases from which we collected the data. This research was approved by the Pennsylvania State University IRB (STUDY00008769; STUDY00008773).

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Sociology and Criminology, Pennsylvania State University, University Park, PA, USA. <sup>2</sup>Department of Sociology and Anthropology, School of Criminology and Criminal Justice, Northeastern University, Boston, MA, USA. <sup>3</sup>Department of Computer Science and Informatics, Cardiff University, Cardiff, Wales, UK.

Received: 17 May 2021 Accepted: 26 July 2021

Published online: 28 August 2021

### References

- Almquist ZW (2014) networkdata: Lin Freeman's network data collection. R package version 0.01. <https://github.com/Z-co/networkdata>
- Alon U (2003) Biological networks: the tinkerer as an engineer. *Science* 301:1866–1867
- Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8:450–461
- Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L (2004) Comment on "Network motifs: simple building blocks of complex networks" and "Superfamilies of evolved and designed networks." *Science* 305:1107
- Ashford JR, Turner LD, Whitaker RM, Preece A, Felmlee D, Towsley D (2019) Understanding the signature of controversial Wikipedia articles through motifs in editor revision networks. In: WWW '19: companion proceedings of the 2019. World Wide Web conference, pp 1180–1187. <https://doi.org/10.1145/3308560.3316754>



- Bearman PS, Moody J, Stovel K (2004) Chains of affection: the structure of adolescent romantic and sexual networks. *Am J Sociol* 110:44–91
- Benson AR, Gleich DF, Leskovec J (2016) Higher-order organization of complex networks. *Science* 33:163–166
- Braines D, Felmlee D, Towsley D, Tu K, Whitaker RM, Turner LD (2018) The role of motifs in understanding behavior in social and engineered networks. In: Proceedings of SPIE, vol 10653, Next-generation analyst VI, 106530W. <https://doi.org/10.1117/12.2309471>
- Brandes U, Robins G, McCranie A, Wasserman S (2013) What is network science? *Netw Sci* 1:1–15
- Burt RS (2004) Structural holes and good ideas. *Am J Sociol* 110:349–399
- Cartwright D, Harary F (1956) Structural balance: a generalization of Heider's theory. *Psychol Rev* 63:277–293
- Coleman JS (1988) Social capital in the creation of human capital. *Am J Sociol* 94:595–5120
- Cross R, Parker A (2004) The hidden power of social networks. Harvard Business School, Cambridge
- Cunningham P, Harrigan M, Wu G, O'Callaghan D (2013) Characterizing ego-networks using motifs. *Netw Sci* 1:170–190
- Davis JA (1970) Clustering and hierarchy in interpersonal relations: testing two graph theoretical models on 742 socio-matrices. *Am Sociol Rev* 35:843–851
- Dey AK, Gel YR, Poor HV (2019) What network motifs tell us about resilience and reliability of complex networks. *Proc Natl Acad Sci* 116:19368–19373
- Duxbury SW (2021) Diagnosing multicollinearity in exponential random graph models. *Sociol Method Res* 50:491–530
- Faris R, Felmlee D, McMillan C (2020) With friends like these: aggression from amity and equivalence. *Am J Sociol* 126:673–713
- Faust K (2010) A puzzle concerning triads in networks: transitivity and homophily in strong-tie relations. *Soc Netw* 22:221–233
- Faust K, Skvoretz J (2002) Comparing networks across space and time, size and species. *Sociol Methodol* 32:267–299
- Felmlee D, Faris R (2016) Toxic ties: networks of friendship, dating, and cyber victimization. *Soc Psychol Quart* 79:243–262
- Felmlee D, DellaPosta D, Inara Rodis P, Matthews SA (2020) Can social media anti-abuse policies work? A quasi-experimental study of online sexist and racist slurs. *Socius* 6:1–6. <https://doi.org/10.1177/2378023120948711>
- Fowler JH (2006) Connecting the congress: a study of cosponsorship. *Polit Anal* 14:456–487
- Gotelli NJ, Graves GR (1996) Null models in ecology. Smithsonian Institution, Washington, DC
- Gouldner AW (1960) The norm of reciprocity: a preliminary statement. *Am Sociol Rev* 25:161–178
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78:1360–1380
- Gu W, Luo J, Liu J (2019) Exploring small-world network with an elite-clique: bringing embeddedness theory into the dynamic evolution of a venture capital network. *Soc Netw* 57:70–81
- Hallinan MT (1974) The structure of positive sentiment. Elsevier, Netherlands
- Harris KM, Halpern CT, Whitsel E, Hussey J, Tabor J, Entzel P, Udry JR (2009) The national longitudinal study of adolescent to adult health: research design. <http://www.cpc.unc.edu/projects/addhealth/design>
- Heider F (1946) Attitudes and cognitive organization. *J Psychol* 21:107–112
- Holland PW, Leinhardt S (1971) Transitivity in structural models of small groups. *Comp Group Stud* 2:107–124
- Holland PW, Leinhardt S (1978) An omnibus test for social structures using triads. *Sociol Methods Res* 7:227–256
- Homans G (1961) Social behavior: its elementary forms. Harcourt, Brace and World, New York
- Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci* 104:10240–10245
- Hunter DR (2007) Curved exponential family models for social networks. *Soc Netw* 29:216–230
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M (2008) ERGM: a package to fit, simulate, and diagnose exponential-family models for networks. *J Stat Softw* 24:1–29
- Ingram PJ, Stumpf MP, Stark J (2006) Network motifs: structure does not determine function. *BMC Genomics* 7:108
- Jain D, Patgiri R (2019) Network motifs: a survey. In: Singh M, Gupta P, Tyagi V, Flusser J, Ören T, Kashyap R (eds) Advances in computing and data sciences. ICACDS 2019. *Commun Comput Inf Sci*, vol 1046. Springer, Singapore, pp 80–91
- John Jay & ARTIS transnational terrorism database (2009) <http://doitapps.jjay.cuny.edu/jjatt/index.php>
- Kashtan N, Alon U (2005) Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci* 102:13773–13778
- Klimt B, Yang Y (2004) Introducing the Enron corpus. In: Proceedings of the first conference on email and anti-spam, CEAS
- Krackhardt D (1987) Cognitive social structures. *Soc Netw* 9:109–134
- Krumov L, Freeter C, Müller-Hannemann M, Weihe K, Hütt MT (2011) Motifs in co-authorship networks and their relation to the impact of scientific publications. *Eur Phys J B* 84:535–540
- Lazega E (2001) The collegial phenomenon: the social mechanisms of cooperation among peers in a corporate law partnership. Oxford University, New York
- Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK et al (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799–804
- Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: densification and shrinking diameters. *ACM Trans Knowl Discov Data* DOI 10(1145/1217299):1217301
- Lusher D, Koskinen J, Robins G (2013) Exponential random graph models for social networks: theory, methods, and applications. Cambridge University, New York
- Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci* 100:11980–11985
- Marcum CS, Lin J, Koehly L (2016) Growing-up and coming-out: are 4-cycles present in adult hetero/gay hook-ups? *Netw Sci* 4:400–405
- McMillan C (2019) Tied together: adolescent friendship networks, immigrant status, and health outcomes. *Demography* 56:1075–1103
- McMillan C, Felmlee D (2020) Beyond dyads and triads: a comparison of tetrads in twenty social networks. *Soc Psychol Quart* 83:383–404
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27:415–444

- Michienzi A, Guidi B, Ricci L, De Salve A (2021) Incremental communication patterns in online social groups. *Knowl Inf Syst* 63:1339–1364
- Milo R, Shen-Orr SS, 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, Alon U (2004) Superfamilies of evolved and designed networks. *Science* 303:1538–1542
- Moreno JL (1934) Who shall survive? A new approach to the problem of human interrelations. In: *Nervous and mental disease*, Washington, DC
- Ohnishi T, Takayasu H, Takayasu M (2010) Network motifs in an inter-firm network. *J Econ Interact Coord* 5:171–180
- Paranjape A, Benson AR, Leskovec J (2017) Motifs in temporal networks. In: *Proceedings of the tenth ACM international conference on web search and data mining*, pp 601–610. <https://doi.org/10.1145/3018661.3018731>
- Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph ( $p^*$ ) models for social networks. *Soc Netw* 29:173–191
- Schneider CM, Belik V, Couronné T, Smoreda Z, González MC (2013) Unravelling daily human mobility motifs. *J R Soc Interface* 10:20130246
- Schweinberger M (2011) Instability, sensitivity, and degeneracy of discrete exponential families. *J Am Stat Assoc* 106:1361–1370
- Shenn-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31:64–68
- Shoval O, Alon U (2010) SnapShot: network motifs. *Cell* 143:326–326
- Simmel G (1902) The number of members as determining the sociological form of the group. *Am J Sociol* 8:1–6
- Skvoretz J, Faust K (2002) Relations, species, and network structure. *J Soc Struct* 3:39–145
- Snijders TAB, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. *Sociol Methodol* 36:99–153
- Stouffer DB, Bascompte J (2010) Understanding food-web persistence from local to global scales. *Ecol Lett* 13:154–161
- Wang P, Robins G, Pattison P (2009) PNet: program for the simulation and estimation of exponential random graph models. Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*. Cambridge University, Cambridge
- Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and  $p^*$ . *Psychometrika* 61:401–425
- Welser HT, Gleave E, Fisher D (2007) Visualizing the signatures of social roles in online discussion groups. *J Soc Struct* 8:1–32
- Whitaker RM, Colombo GB, Allen SM, Dunbar RIM (2016) A dominant social comparison heuristic unites alternative mechanisms for the evolution of indirect reciprocity. *Sci Rep* 6:31459
- Yaveroglu ON, Fitzhugh SM, Kurant M, Markopoulou A, Butts CT, Przulj N (2015) ergm.graphlets: a package for ERG modeling based on graphlet statistics. *J Stat Softw* 65
- Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY et al (2004) Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proc Natl Acad Sci* 101:5934–5939
- Zignani M, Quadri C, Del Vicario M, Gaito S, Rossi GP (2018) Temporal communication motifs in mobile cohesive groups. In: Cherifi C, Cherifi H, Karsai M, Musolesi M (eds) *Complex networks and their applications VI*, vol 689. Springer, Cham, pp 490–501

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.