

# Geodesic Cycle Length Distributions in Delusional and Other Social Networks

Alex Stivala<sup>1</sup>

Accepted manuscript to appear in *Journal of Social Structure* (2020). License: CC BY-NC 4.0

## Abstract

A recently published paper [Martin (2017) JoSS 18(1):1-21] investigates the structure of an unusual set of social networks, those of the alternate personalities described by a patient undergoing therapy for multiple personality disorder (now known as dissociative identity disorder). The structure of these networks is modeled using the  $dk$ -series, a sequence of nested network distributions of increasing complexity. Martin finds that the first of these networks contains a striking feature of a large “hollow ring”; a cycle with no shortcuts, so that the shortest path between any two nodes in the cycle is along the cycle (in more precise graph theory terms, this is a geodesic cycle). However the subsequent networks have much smaller largest cycles, smaller than those expected by the models. In this work I re-analyze these delusional social networks using exponential random graph models (ERGMs) and investigate the distribution of the lengths of geodesic cycles. I also conduct similar investigations for some other social networks, both fictional and empirical, and show that the geodesic cycle length distribution is a macro-level structure that can arise naturally from the micro-level processes modeled by the ERGM.

**Keywords**— Geodesic cycle, Exponential random graph model, ERGM,  $dk$ -series random graphs, Social networks, Fictional networks, Dissociative identity disorder

## Introduction

In order to investigate the structure of ideas, or schemata, of social networks, Martin (2017) investigated a very unusual set of three social networks. These are delusional social networks of alternative personalities described by a patient undergoing therapy for multiple personality disorder (David et al., 1996), now known as dissociative identity disorder (Kihlstrom, 2005). In order to do this, Martin (2017) uses random graphs, specifically the  $dk$ -series model (Mahadevan et al., 2006; Orsini et al., 2015), to generate random networks that fit different aspects of these delusional social networks, and then compares other (not explicitly fit by the model) structures of the random networks with those of the observed networks.

Martin (2017) finds that one of the networks contains what he describes as a large “hollow ring” — a cycle with no shortcuts so that the shortest path is along the cycle. In this network the size of this largest hollow ring is much larger than expected under the  $dk$ -series model. However in the other two networks, the size of the largest such hollow ring is smaller than expected under the model.

He concludes that the logic of these networks is spatial, with long path lengths indicating a “large world”, and that the bias against large hollow rings is indicative of a belief that a chain of relationships that “goes away” in space is not going to return. In summary, Martin (2017) concludes with the hypothesis that the root schema for social networks is local and spatial.

In this work I will re-examine these networks using a different random graph model, the exponential random graph model (ERGM), which allows some more flexibility in the structures it can model, and also allows use of additional data (such as nodal covariates) not available in the purely structural  $dk$ -series model. In addition, I will make precise the “hollow ring” definition and put it in the context of mathematics and computer science research in graph theory, where it is known as a geodesic cycle. Similarly to Martin (2017) I will use random graph models (both ERGM and  $dk$ -series) to generate random networks. However rather than examining the size of just the largest geodesic cycle, I will compare the distribution of geodesic cycle lengths in the observed network to that in the generated networks.

In addition, I will repeat this exercise with several other empirical networks (human and animal social networks), and a fictional character network, and compare the results with those for the delusional social networks.

---

<sup>1</sup>Università della Svizzera italiana, Via Giuseppe Buffi 13, 6900 Lugano, Switzerland. Email: alexander.stivala@usi.ch

## Cycles, chordless cycles, and geodesic cycles

Martin (2017) describes a “hollow ring” of degree ten as “a cycle containing ten nodes, no pair of which have a distance in the graph lower than that in the cycle”, so that there are no “short cuts” between nodes in the ring (Martin, 2017, p. 16). This definition makes intuitive sense, and seems precise enough for us to understand it unambiguously (and for Martin (2017) to count such structures), however I believe it is useful to relate it to existing definitions in graph theory literature, in which this idea has already been defined precisely.

First, I prefer the term “cycle” rather than “ring”, as the former is well-known in graph theory, while “ring” is generally not used in graph theory, being better-known as an algebraic structure. In the following definitions, all graphs are undirected, and all graphs considered in this work are undirected.

A *walk* is a sequence of edges joining a sequence of vertices; a *trail* is a walk in which the edges are distinct; a *circuit* is a non-empty trail in which the first and last vertices are repeated. Then a *cycle* (or *simple circuit*) is a circuit where only the first and last vertices are repeated. A *Hamiltonian cycle* is a cycle that visits each node exactly once.

An *induced subgraph* of a graph is a subgraph formed by a subset of vertices of the graph and all the edges connecting vertices in that subset. An *induced path* in a graph  $G$  is a path that is an induced subgraph of  $G$ . That is, any two adjacent vertices in the path are connected by an edge in  $G$  and any two non-adjacent vertices in the path are not connected by an edge in  $G$ .

A *chord* is an edge joining two non-adjacent nodes in a cycle. A *chordless cycle* is a cycle in which no two vertices of the cycle are connected by an edge that is not itself in the cycle; i.e. it is a cycle with no chords. A chordless cycle is also known as an *induced cycle*, as (just as for an induced path) in a chordless cycle in a graph  $G$ , any two adjacent vertices in the cycle are connected by an edge in  $G$  and any two non-adjacent vertices in the cycle are not connected by an edge in  $G$ . A chordless cycle is sometimes also called a *hole*.

The preceding definitions are all standard and well-known in graph theory, and can be found in any discrete mathematics or computer science textbook, e.g. Roman (1989), or Wolfram MathWorld, e.g. Weisstein (2020). The following definitions, however, are not so well-known.

A *geodesic cycle* (Li and Shi, 2018) or *isometric cycle* (Lokshtanov, 2009) is a cycle where the length of the shortest path between any pair of vertices along the cycle is equal to the length of the shortest path between them in the graph:

$$d_G(u, v) = d_C(u, v)$$

where  $d_G(u, v)$  is the distance (length of shortest path, i.e. geodesic) between  $u$  and  $v$  in  $G$ . This is not such a well-known (textbook) concept, but this definition of “geodesic cycle” goes back to at least Negami and Xu (1986). Every geodesic cycle is chordless, but not every chordless cycle is geodesic.

Note that although finding the largest cycle and largest chordless cycle in a graph are both  $\mathcal{NP}$ -complete problems (Garey and Johnson, 1979), finding the largest geodesic (isometric) cycle is in  $\mathcal{P}$  (Lokshtanov, 2009). In simple terms, this means that finding the largest cycle or largest chordless cycle is a computationally intractable problem<sup>1</sup> (it is in the same class as the famous “traveling salesman problem”), but that finding the largest geodesic cycle is not. This means that not only are geodesic cycles a fairly intuitive idea, perhaps because of our spatial ideas of (social) networks (the idea of the “shortcut” being inherently spatial), but it is also computationally convenient to use them rather than cycles or chordless cycles.

To illustrate these definitions, consider the graph shown in Figure 1. This graph has seven cycles. Four of them are chordless (1–4–5–9–3–2–1, 4–6–5–4, 5–6–7–8–9–5, and 1–4–6–7–8–9–3–2–1), however only three of the chordless cycles are geodesic: the largest chordless cycle is not geodesic as the paths 4–5–9 and 6–5–9 are “shortcuts” across the cycle.

A cycle  $C$  of a graph  $G$  is (*geodetically*) *convex* if for any pair of distinct vertices  $u, v \in V(C)$  (where  $V(G)$  denotes the vertex set of a graph  $G$ ):

$$d_C(u, v) < d_{G-C}(u, v)$$

i.e. the distance along the cycle between any pair of vertices in the cycle is less than the distance between them in the graph excluding the cycle. This is not such a well-known concept, but is described in Hellmuth et al. (2014). Geodetic (or geodesic) convexity in graphs (more generally, not necessarily of cycles) had earlier been discussed in Batten (1983); Farber and Jamison (1986); Farber (1987). Note that these papers on “isometric”, “geodesic”, or “convex” cycles do not cite each other or mention any equivalent definitions with different names, although Hellmuth et al. (2014) also discusses isometric subgraphs and isometric cycles:

- A subgraph  $H$  of  $G$  is *isometric* if  $d_H(u, v) = d_G(u, v)$  holds for all  $u, v \in V(H)$ .
- $H$  is a (*geodetically*) *convex* subgraph of  $G$  if and only if for all  $u, v \in V(H)$ , all shortest  $uv$ -paths  $P \in \mathbb{P}_G[u, v]$  are contained in  $H$ .
- Convex implies isometric (Hellmuth et al., 2014, p. 125).

<sup>1</sup>Unless  $\mathcal{P} = \mathcal{NP}$ .

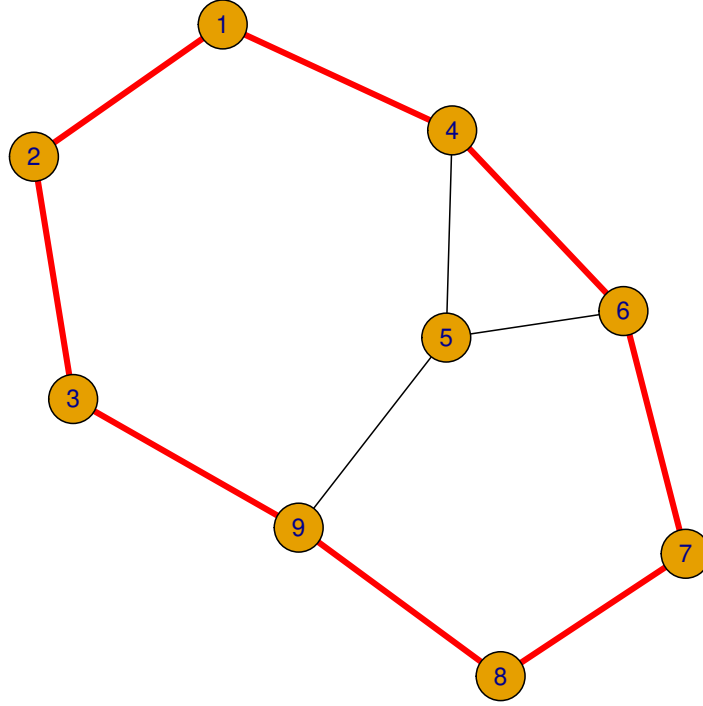


Figure 1: This graph has seven cycles (one is Hamiltonian). Four are chordless and three of those are also geodesic (and, in addition, convex). The cycle in red (around the “outside” of the graph) is chordless, but not geodesic.

An *atomic cycle* is defined by Gashler and Martinez (2012) as a generalization of a chordless cycle:

- An  $n$ -*chord* is a path of length  $n$  connecting two vertices in a cycle, where  $n$  is less than the length of the shortest path in the cycle between the vertices.
- An *atomic cycle* is a cycle with no  $n$ -chords.

We can see that this coincides with the definition of a geodesic cycle, and also with the description of a “hollow ring” by Martin (2017). I prefer the term “geodesic cycle” as it makes sense mathematically given its definition, without confusing it with other properties such as atomicity, and it dates back at least as far as Negami and Xu (1986).

## Counting cycles

For an Erdős–Rényi random graph, the expected number of cycles of a given length can be calculated (Erdős and Rényi, 1960; Takács, 1988), as can the probability of a chordless cycle (Łuczak, 1991), and the expected length of the largest chordless cycle (Łuczak, 1993). Geodesic cycles in random graphs were studied in Benjamini et al. (2011), and the length of the longest geodesic cycles in random graphs in Li and Shi (2018).

For more complex families of random graphs, such as the  $dk$ -series or ERGMs, however, cycle length distributions can only be estimated from simulations. In this work geodesic cycles are counted using the `find_large_atomic_cycle` algorithm of Gashler and Martinez (2012).

## Exponential random graph models and $dk$ -series random graphs

ERGMs provide a way of modeling network ties based on structure and attributes. Given an observed network, we estimate parameters for local effects, such as closure (clustering), activity (greater tendency to have ties), homophily, and so on. The sign (positive for the effect occurring more than by chance, negative for less than by chance) and significance tell us about these processes, taking dependency into account. That is, the parameter tells us about the process occurring significantly more or less than by chance, given all the other effects in the model occurring simultaneously. ERGMs are widely used in the social sciences to model social networks (Robins et al., 2007a; Lusher et al., 2013; Amati et al., 2018).

An ERGM is a probability distribution with the form:

$$\Pr(X = x) = \frac{1}{\kappa} \exp \left( \sum_A \theta_A z_A(x) \right)$$

where:

- $X = [X_{ij}]$  is a 0-1 matrix of random tie variables,
- $x$  is a realization of  $X$ ,
- $A$  is a “configuration”, or motif, a (small) set of nodes and a subset of ties between them,
- $z_A(x)$  is the network statistic for configuration  $A$ ,
- $\theta_A$  is a model parameter corresponding to configuration  $A$ ,
- $\kappa$  is a normalizing constant to ensure a proper distribution.

Which configurations  $A$  are allowed depends on the assumptions as to which ties are independent. Here I will use the *social circuit dependence* assumption (Snijders et al., 2006; Robins et al., 2007b), under which two potential ties are conditionally dependent exactly when, if they were observed, they would form a four-cycle in the network (Robins et al., 2007b).

So given the observed network  $x$ , the problem is to estimate the parameter vector  $\theta$  which maximizes the probability of  $x$  under the model. A variety of algorithms, generally based on Markov chain Monte Carlo (MCMC) techniques, have been used to do this (Corander et al., 1998, 2002; Snijders, 2002; Hunter and Handcock, 2006; Robins et al., 2007b; Caimo and Friel, 2011; Hummel et al., 2012; Hunter et al., 2012; Byshkin et al., 2016; Amati et al., 2018; Byshkin et al., 2018; Borisenko et al., 2020), and are implemented in different software packages (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b; Wang et al., 2009; Stivala et al., 2020).

Rather than estimating model parameters to fit an observed network and then, if an ensemble of random networks is required, simulating random networks from the model as required by ERGM,  $dk$ -random graphs are simulated so that particular statistics of the observed network are fixed. Therefore no parameter estimation is required to obtain the appropriate simulated networks. The  $dk$ -series defines a sequence of nested network distributions of increasing complexity, labelled by the value of  $d$ . Specifically first the density is fit ( $d = 0$ ), then degree distribution ( $d = 1$ ), degree homophily ( $d = 2$ ), average local clustering ( $d = 2.1$ ), and clustering by degree ( $d = 2.5$ ) (Orsini et al., 2015). As noted by Orsini et al. (2015), the  $d \in \{0, 1, 2\}$  distributions are well-known random graph models, and in particular, the  $0k$  random graph distribution is just the Erdős–Rényi random graph model. However the  $2.1k$  and  $2.5k$  random graphs are introduced to model clustering, as an important and ubiquitous feature of many empirical networks (Orsini et al., 2015).

Martin (2017) mentions that a difficulty with ERGMs, along with other techniques such as block modeling (White et al., 1976), is that they require strong structural assumptions. He further notes that, for more flexible modeling approaches, the difficulty in obtaining a converged model for theoretically interesting effects means that model convergence itself becomes the criterion for the model selected (Martin, 2017, p. 6). For these reasons, Martin (2017) instead uses the  $dk$ -series.

I contend, however, that rather than necessarily being a difficulty, the strong structural assumptions made by the use of ERGM is in some ways an advantage. The social circuit dependence assumption means that statistical significance of parameters tells us something about the corresponding local processes generating the observed global structure of the network. The configurations (and their corresponding parameters) in the model can therefore chosen not purely to fit local structure (such as degree homophily and local clustering) but to test theoretically relevant hypotheses about local processes.

A potential disadvantage of the  $dk$ -series model is that it is purely structural, unlike ERGMs, which can incorporate node (and other) attributes. As noted by Martin (2017, p. 6) if extra information such as nodal covariates is unavailable or unimportant, this is no disadvantage at all. However in cases where such information is available, it is potentially problematic. Martin (2017) does not consider node attributes in the “Patricia” data, finding little evidence of structural implications (Martin, 2017, p. 8). I will show that this is largely true, although ERGM models that incorporate node attributes do have a better fit than those that do not. However for some of the other social networks I consider, it is vital to incorporate node attributes to obtain a reasonable model.

As for the potential difficulty in obtaining a converged ERGM model due to problems with degeneracy, for example, it is true that this can be a challenging task. However advances in model specifications such as the use of the “alternating” or “geometrically weighted” model terms (Hunter and Handcock, 2006; Snijders et al., 2006; Robins et al., 2007b), curved ERGMs (Hunter and Handcock, 2006; Hunter, 2007), and alternative estimation algorithms (Hummel et al., 2012; Byshkin et al., 2018) mean these degeneracy problems can generally be overcome (Schweinberger et al., 2019).

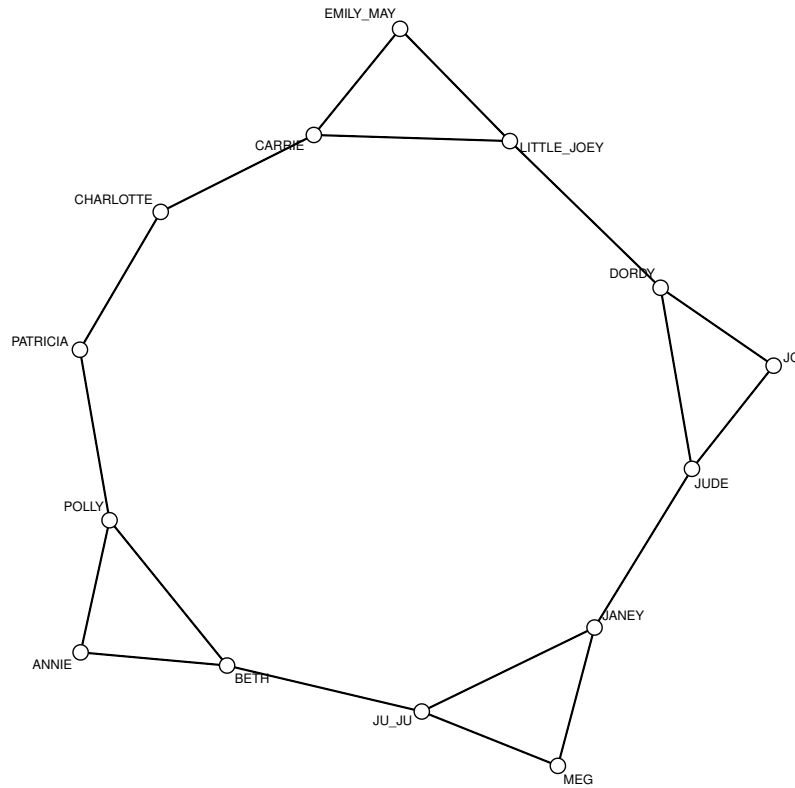


Figure 2: Patricia’s 1990 network. Visualization created using the network R package (Butts, 2008, 2015).

## Data

As described in David et al. (1996) and used in Martin (2017), the patient “Patricia” drew maps in which she represented her personalities as nodes and relationships between them as edges. I manually coded the networks from Patricia’s hand drawings as reproduced in David et al. (1996). Three such drawings are available, for the years 1990, 1992, and 1993. Network visualizations of these three delusional social networks are shown in Figures 2, 3, and 4, respectively.

The 1990 network has no nodal attributes other than names. However the 1992 and 1993 networks have some nodal attributes. In the original drawings, shaded nodes are labeled as “Integrated alters” and I coded these as a binary attribute *Integrated*. “Christian alters” are marked with an asterisk in the original drawings, and I coded these as a binary attribute *Christian*. Some nodes are annotated with a number (which may correspond to age, although this is not clear) and these are coded as a numeric attribute (or NA if missing). Some nodes are included in a region labeled the “Sphere of the blue flame” and I coded this with a binary attribute *Sphere*. In the 1993 network there is in addition a graph component labeled “Behind” on the drawing, and I coded this as a binary attribute *Behind*.

Martin (2017) includes a detailed discussion of the structure and evolution of these networks, however due to their unusual nature, some more discussion of what they might mean is warranted. According to David et al. (1996),

It would be hard to conceive of a single mind capable of sustaining dozens of other minds, lives, and relationships, leaving aside the question of whether or not such a feat is intended.

(David et al., 1996, pp. 139-140), and

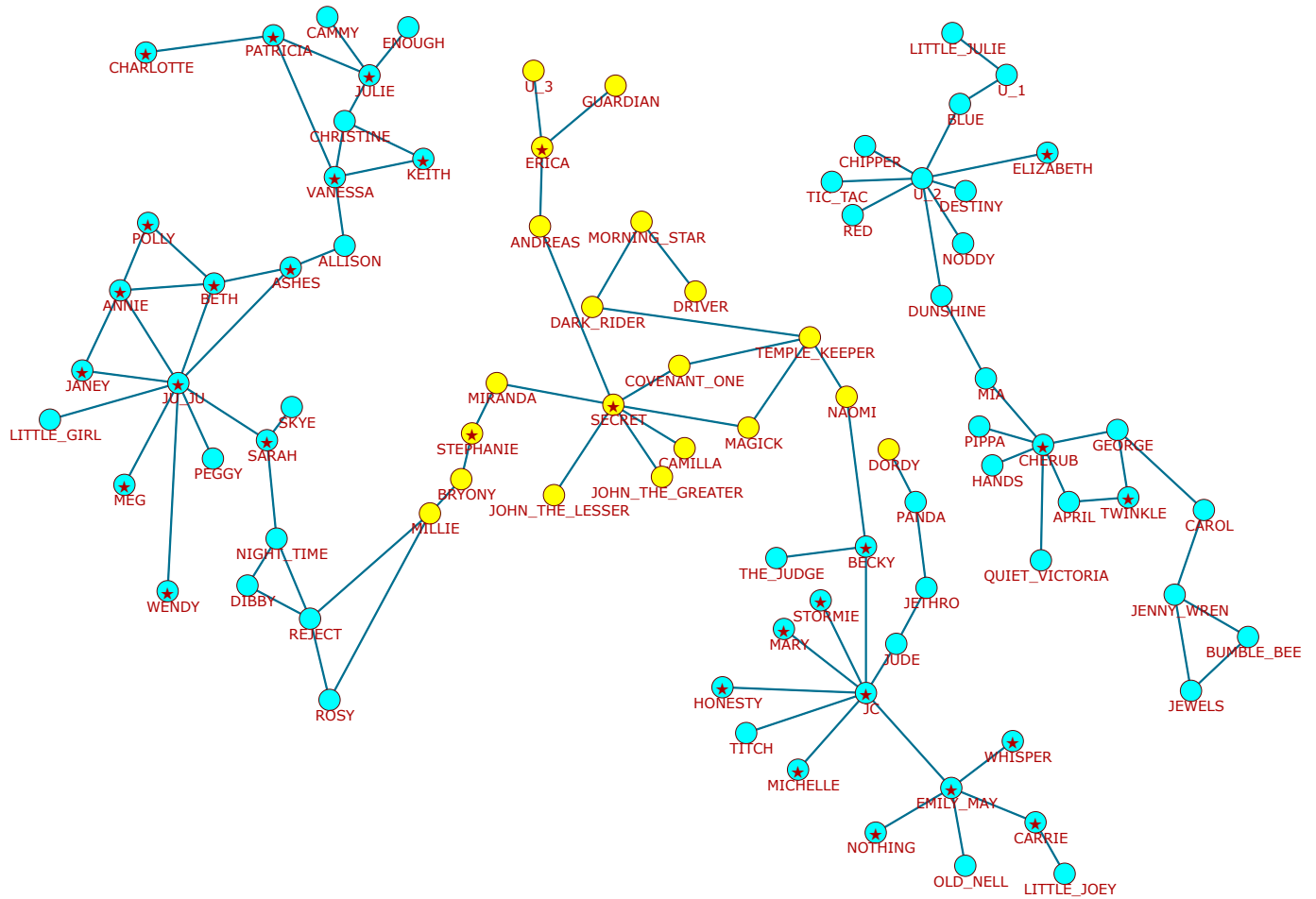


Figure 3: Patricia's 1992 network. Nodes marked with a star are marked as "Christian Alters" in Patricia's original diagram, and nodes colored yellow are included inside the "Sphere of the Blue Flame" in Patricia's original diagram. Visualization created with Pajek (Batagelj and Mrvar, 2004; Mrvar and Batagelj, 2016).

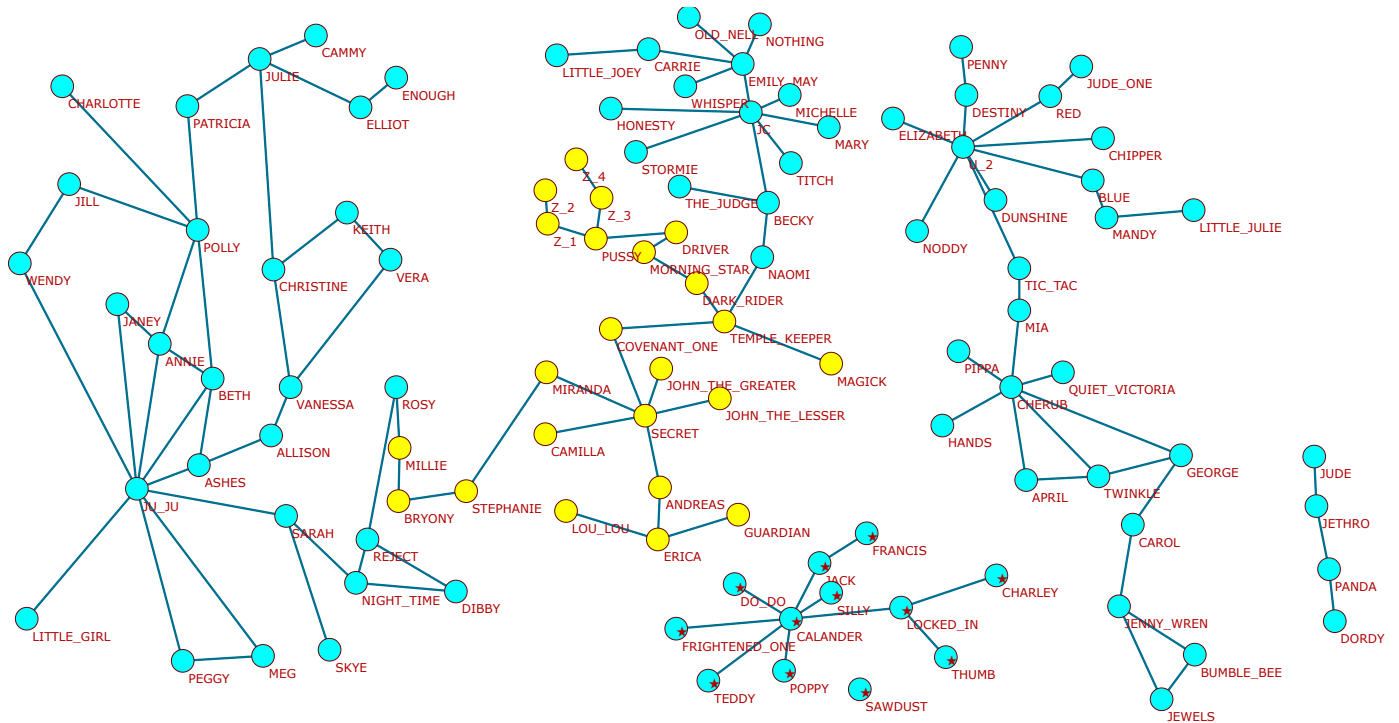


Figure 4: Patricia's 1993 network. Nodes marked with a star are from the box labelled "(Behind)" in Patricia's original diagram, and nodes colored yellow are included inside the "Sphere of the Blue Flame" in Patricia's original diagram. Visualization created with Pajek (Batagelj and Mrvar, 2004; Mrvar and Batagelj, 2016)

The illusion of unity of consciousness is not exposed in MPD [multiple personality disorder] but is repeated over and over again. One illusory consciousness gives way (or joins) another. Each is a coherent autonomous homunculus [*sic*]. It is like an illiterate forger passing off dud bank notes of different denominations but always with the word "pound" mis-spelled.

(David et al., 1996, p. 143).

Without referring to Freud, David et al. (1996) refers to MPD as "a means of dealing with social and interpersonal conflict" that "becomes fossilised and embellished..." (David et al., 1996, p. 143). Freud, famously, had the view, that, in relation to paranoia,

In every instance the delusional idea is maintained with the same energy with which another, intolerably distressing, idea is fended off from the ego. *Thus they love their delusions as they love themselves.* That is the secret.

(Freud, 1950, pp. 211–212).

These theories of how such delusions may arise, however, give us no insight into what the lines in Patricia's diagrams (network edges) might represent. According to Martin (2017),

In many ways, this returns us to core intuitions of early network analysts: that we are interested in some sort of intertwining of lives, but we might be unable to specify a single content to the nature of the ties involved. ... Patricia's maps presumably are indicating a more fundamental connection, one perhaps as obscure as our own strong feelings.

(Martin, 2017, p. 5).

However, is not "a single mind capable of sustaining dozens of other minds, lives, and relationships" (David et al., 1996, pp. 139–140) just (part of) what a novelist, or screenwriter, or any other fiction writer must have, to some degree? The characters ("coherent autonomous homunculi") in a work of fiction are individual and coherent, but they perhaps have a common hallmark, from the author. So, leaving aside any consideration as to how "real" the alter are, it makes sense to consider Patricia's alters just as we might consider fictional characters. They and their relationships are not completely arbitrary or random, but follow some schema which makes them meaningful.

In addition to the delusional social networks, I also use six other networks from a variety of domains.

The first is a fictional character network. The *Grey's Anatomy* network (Lind, 2012; Weissman, 2019; Leavitt and Clark, 2014) is a sexual contact network of characters on the ABC television medical drama "Grey's Anatomy". There are nodal covariates for sex, birth year, race, and position in the hospital. This network is illustrated in Figure A1. ERGM models are also based on those described in Lind (2012); Leavitt and Clark (2014).

The *Dolphins* social network (Lusseau et al., 2003), downloaded from <http://www-personal.umich.edu/~mejn/netdata/> (accessed 2 January 2018), is a social network of a small closed population of bottlenose dolphins, with ties representing significantly more frequent co-occurrence of individuals within schools (Lusseau et al., 2003). This network is illustrated in Figure A2.

The *Lazega law firm* (Lazega and Pattison, 1999; Lazega, 2001; Snijders et al., 2006) network downloaded from <http://moreno.ss.uci.edu/data.html> (accessed 2 November 2017) is a friendship network of lawyers in a New England law firm. There are nodal covariates for gender, law school, office, practice, status, age, and seniority. This network is illustrated in Figure A3.

The *Zachary karate club* network (Zachary, 1977), obtained via statnet (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b) is the well-known network of social relationships among members of a karate club. This network is illustrated in Figure A4.

The *Kapferer tailor shop* (Kapferer, 1972) network, obtained via statnet, is a network of social interactions between workers in a tailor shop in Zambia. This network is illustrated in Figure A5. ERGM models are replications of those in Hummel et al. (2012).

The *High school friendship* network (Mastrandrea et al., 2015) downloaded from the SocioPatterns website (<http://www.sociopatterns.org/datasets/high-school-contact-and-friendship-networks/>, accessed 25 September 2019) is a network of reported friendships among students at a high school in France. The directed network of reported friendship is treated as undirected here (reciprocity in the directed network was .78). There are nodal covariates for age and school class. This network is illustrated in Figure A6.

Summary statistics of all the networks considered are shown in Table 1.

## Methods

For each of the networks considered, ERGM models are estimated using the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Unless otherwise noted, the default MCMCMLE (Markov chain Monte Carlo maximum likelihood estimation) estimation method is used; sometimes the "Stepping" algorithm (Hummel et al., 2012) is used instead. Where a curved ERGM is used, this is noted by the estimation of the corresponding decay parameter  $\alpha$  in the model, otherwise the fixed value of  $\alpha$  is specified. The ERGM parameters used are shown in Table B1.



Table 1: Summary statistics of the networks.

Network	N	Components	Mean degree	Density	Clustering coefficient	Assortativity coefficient	Mean path length
Patricia 1990	14	1	2.57	0.19780	0.40000	-0.25000	2.96
Patricia 1992	85	2	2.21	0.02633	0.10345	-0.46399	8.41
Patricia 1993	107	5	2.13	0.02010	0.10266	-0.37400	8.88
Grey’s Anatomy	44	4	2.09	0.04863	0.00000	-0.22567	3.49
Dolphins	62	1	5.13	0.08408	0.30878	-0.04359	3.36
Zachary karate club	34	1	4.59	0.13904	0.25568	-0.47561	2.41
Kapferer tailor shop	39	1	8.10	0.21323	0.38506	-0.18269	2.04
Law firm friendship	71	3	11.24	0.16056	0.44862	0.07948	2.19
High school friendship	134	3	6.06	0.04556	0.47540	0.28718	4.02

Note: All networks are undirected. The statistics were computed using the igraph (Csárdi and Nepusz, 2006) package in R (R Core Team, 2016).

Only models that show acceptable convergence and goodness-of-fit according to statnet diagnostics are included in the results. For each network, the “best” model, according to AIC, BIC, and goodness-of-fit plots was selected, and 100 networks simulated from that model using statnet.

Random graphs from the  $dk$ -series distributions are generated with the RandNetGen software (Mahadevan et al., 2006; Colomer-de Simón et al., 2013; Colomer-de Simón and Boguñá, 2014; Orsini et al., 2015). 100 networks from each of the  $dk$ -series distributions  $d = 1$ ,  $d = 2$ ,  $d = 2.1$  and  $d = 2.5$ , are simulated using RandNetGen to do the appropriate  $dk$ -randomization for each empirical network.

The geodesic cycle length distributions are then computed from these simulated networks, using the `findLargeAtomicCycle` algorithm (Gashler and Martinez, 2012) as implemented in the Waffles machine learning toolkit (Gashler, 2011) (software was written to call this toolkit routine in order to find geodesic cycles of all possible lengths). Cycle and chordless cycle distributions from the simulated networks are computed using the CYPATH program (Uno and Satoh, 2014).

Box plots were generated using the ggplot2 R package (Wickham, 2009) and follow the default convention that the lower and upper hinges show the first and third quartiles, with the whiskers extending to the largest or smallest value no further than 1.5 times the interquartile range from the corresponding hinge.

Code and scripts written for this work are available from [https://sites.google.com/site/alexdstivala/home/geodesic\\_cycles](https://sites.google.com/site/alexdstivala/home/geodesic_cycles).

## Results

### Models of the delusional social networks

Table 2 shows ERGM models for Patricia’s 1990 network. As is apparent from Figure 2 and discussed in Martin (2017), this network hardly resembles a typical social network at all, having only nodes of degree two or three, a large geodesic cycle (“hollow ring”) of length 10, and in fact resembles a connected caveman graph (Watts, 1999). There is therefore little to be gained by attempting to interpret these model parameters, except to note that we can indeed obtain converged ERGM models which fit this network adequately, and that Model 3 indicates an over-representation of nodes of degree 3.

Table 3 shows ERGM models of Patricia’s 1992 network. As is clear from Figure 3, and discussed in Martin (2017), this network is far larger and very different from the 1990 network, and contains node attributes. We can conclude from these models that there is a significant tendency against preferential attachment on degree (the GWDEGREE parameter is positive and significant). This is not what we might have expected from the  $dk$ -series models as discussed in Martin (2017) where this network is described as having degree heterophily or a hub-spoke structure (Martin, 2017, p. 12). The positive and significant GWESP parameter indicates a significant tendency for clustering (transitive closure). This is a very typical result in social networks, and also a feature noted in this network by Martin (2017, p. 16).

Models 2 and 3 incorporate nodal attributes as well as structural information, which is not possible with the  $dk$ -series models used by Martin (2017). There appears to be no significant effect of the “Christian” and “Integrated” attributes, consistent with Martin (2017, p. 8) finding little evidence of structural implications. However Model 3 shows that nodes in the “Sphere of the Blue Flame”

Table 2: ERGM models of Patricia's 1990 network.

Effect	Model 1	Model 2	Model 3
Edges	-2.023 (0.507)***	-2.030 (0.503)***	-2.391 (0.669)***
GWESP	0.697 (0.341)*	0.710 (0.355)*	0.861 (0.505)
Degree 2	1.029 (0.573)		
Degree 2-3		1.006 (0.568)	
Degree 3			1.501 (0.584)*
AIC	90.62	90.61	84.12
BIC	98.16	98.14	91.65

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. The GWESP  $\alpha$  parameter is fixed at 0. \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

Table 3: ERGM models of Patricia's 1992 network.

Effect	Model 1	Model 2	Model 3
Edges	-4.588 (0.311)***	-6.047 (0.561)***	-8.392 (0.775)***
GWDEGREE	1.397 (0.503)**	1.908 (0.560)***	1.838 (0.572)**
GWESP	0.793 (0.173)***	0.764 (0.178)***	0.608 (0.177)***
Activity Christian		0.734 (0.196)***	0.772 (0.217)***
Homophily Christian		0.377 (0.205)	0.305 (0.215)
Activity Integrated		0.232 (0.310)	0.150 (0.331)
Homophily Integrated		0.427 (0.350)	0.290 (0.380)
Activity Sphere			0.646 (0.200)**
Homophily Sphere			2.744 (0.513)***
AIC	854.40	843.20	782.60
BIC	872.90	886.40	838.20

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.5 and the GWESP  $\alpha$  parameter is fixed at 0. \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

Table 4: ERGM models of Patricia’s 1993 network.

Effect	Model 1	Model 2	Model 3	Model 4
Edges	−4.941 (0.282)***	−6.609 (0.470)***	−9.914 (0.869)***	−10.169 (0.901)***
GWDEGREE	1.501 (0.443)***	2.384 (0.547)***	2.314 (0.550)***	2.249 (0.538)***
GWESP	0.870 (0.163)***	0.835 (0.168)***	0.656 (0.169)***	0.643 (0.172)***
Activity Christian		0.708 (0.190)***	0.747 (0.203)***	0.789 (0.203)***
Activity Integrated		0.540 (0.243)*	0.507 (0.258)*	0.580 (0.247)*
Homophily Christian		0.473 (0.203)*	0.453 (0.205)*	0.512 (0.214)*
Homophily Integrated		0.725 (0.252)**	0.670 (0.260)**	0.828 (0.271)**
Activity Sphere			0.660 (0.192)***	0.729 (0.194)***
Homophily Sphere			3.611 (0.730)***	3.600 (0.744)***
Activity Behind				0.631 (0.377)
AIC	1094.00	1062.00	975.60	975.00
BIC	1114.00	1108.00	1035.00	1041.00

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.5 and the GWESP  $\alpha$  parameter is fixed at 0. \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

are both more likely to form ties than other nodes, and in addition to that, are more likely to form ties to other nodes within the Sphere than to nodes outside it. This, of course, is entirely consistent with the Sphere resembling a network community (see Figure A7), and with the spatial and planar nature of the network noted by Martin (2017).

Table 4 shows ERGM models for Patricia’s 1993 network (Figure 4). As discussed in Martin (2017) this network is not very different from the 1992 network, and this is reflected in the ERGM models. One difference, however, is that the Christian and Integrated attributes, which had no statistically significant effects in the models for 1992 network (Table 3), now both have positive significant effects for activity and homophily.

ERGM models of the other networks are shown in Appendix B.

## Geodesic cycle length distributions

Figure 5, Figure 6, and Figure 7 show the distribution of geodesic cycle lengths and of the largest geodesic cycle length for Patricia’s 1990, 1992, and 1993 networks, respectively. In agreement with the results described by Martin (2017) using the  $dk$ -series models, we find that the geodesic cycle of length 10 in the 1990 network is extremely unlikely under the ERGM model (under which a much shorter longest geodesic cycle is expected), while the fact that the largest geodesic cycle for the 1992 network is only of length 4 is extremely unlikely (larger maximum geodesic cycle lengths are expected). The situation for the 1993 network is not as clear, with the largest geodesic cycle (of length 8) being on the lower hinge of the boxplot for both the  $dk$ -series 2.5 $k$  distribution and the ERGM (Figure 7). Further, the distribution of geodesic cycle lengths is not a very good fit in the 1990 and 1992 networks (Figures 5 and 6), although for the 1993 networks (Figure 7) it is acceptable in the case of the ERGM.

For the 1990 and 1992 networks, the ERGM does not fit the maximum geodesic cycle length or geodesic cycle distribution any better (or worse) than the  $dk$ -series 2.5 $k$  distribution, despite the ERGM making use of nodal attributes for the 1992 network, which the  $dk$ -series cannot. However for the 1993 network, the ERGM has a better fit to the geodesic cycle length distribution than the  $dk$ -series 2.5 $k$  distribution (Figure 7). Note that all four ERGM models fit the maximum geodesic cycle length as well or better than the  $dk$ -series 2.5 $k$  distribution, including ERGM Model 1, which does not include any nodal attributes and has only three estimated parameters (Table 4). Figure 8 shows the geodesic cycle length distributions for the  $dk$ -series 1 $k$  distribution (which shows the best fit to maximum geodesic cycle length in Figure 7) and for ERGM Model 1. The ERGM shows a better fit to the geodesic cycle length distribution in this network than any of the  $dk$ -series distributions, even when node attribute information is not included in the model.

Figures 9 – 14 show the corresponding results for the other networks. Unlike Patricia’s networks, in all of these networks the ERGM shows a good fit to the largest geodesic cycle sizes in the observed networks, and the fit to the overall geodesic cycle length distributions is also all acceptable for all of these networks.

With two exceptions, the  $dk$ -series 2.5 $k$  distribution shows a similarly good fit to the largest geodesic cycle sizes and geodesic cycle length distributions. The exceptions are the Grey’s Anatomy network and the high school friendship network. In the Grey’s Anatomy network (Figure 9), although both  $dk$ -series 2.5 $k$  and ERGM fit the maximum geodesic cycle length, the ERGM also fits

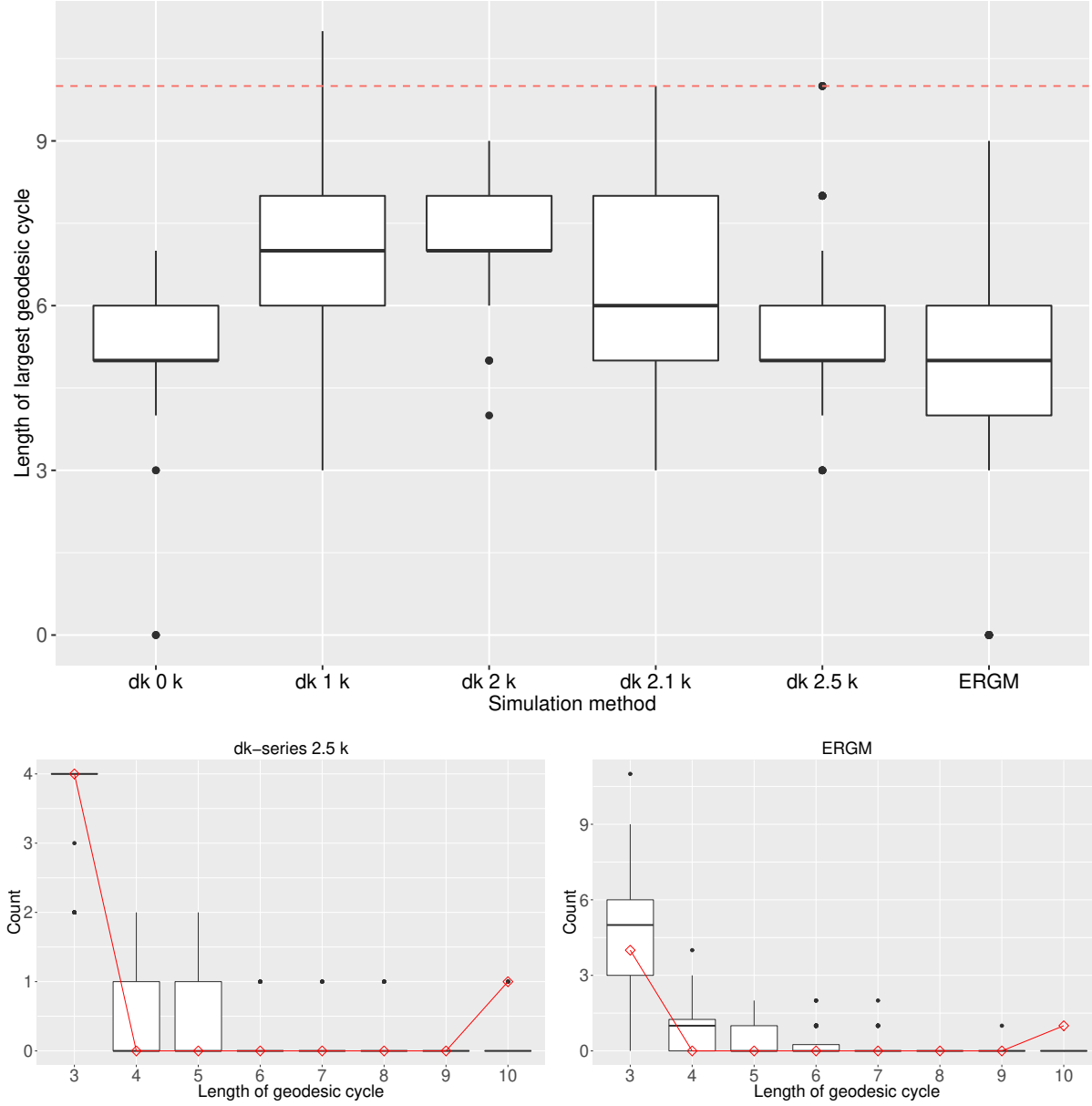


Figure 5: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for Patricia's 1990 network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 3 in Table 2.

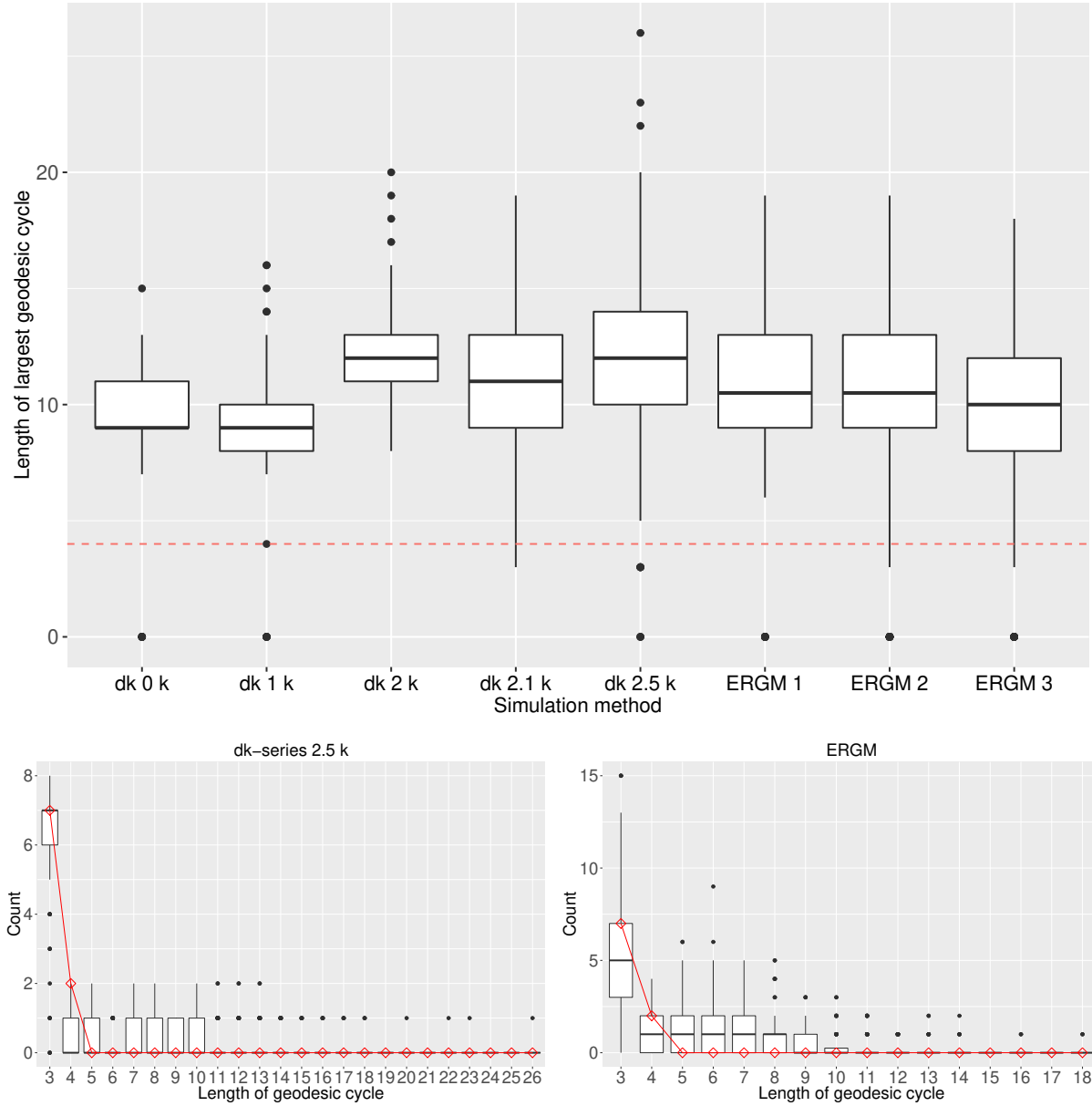


Figure 6: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for Patricia's 1992 network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM Model 3 (right). ERGM model numbers refer to those in Table 3.

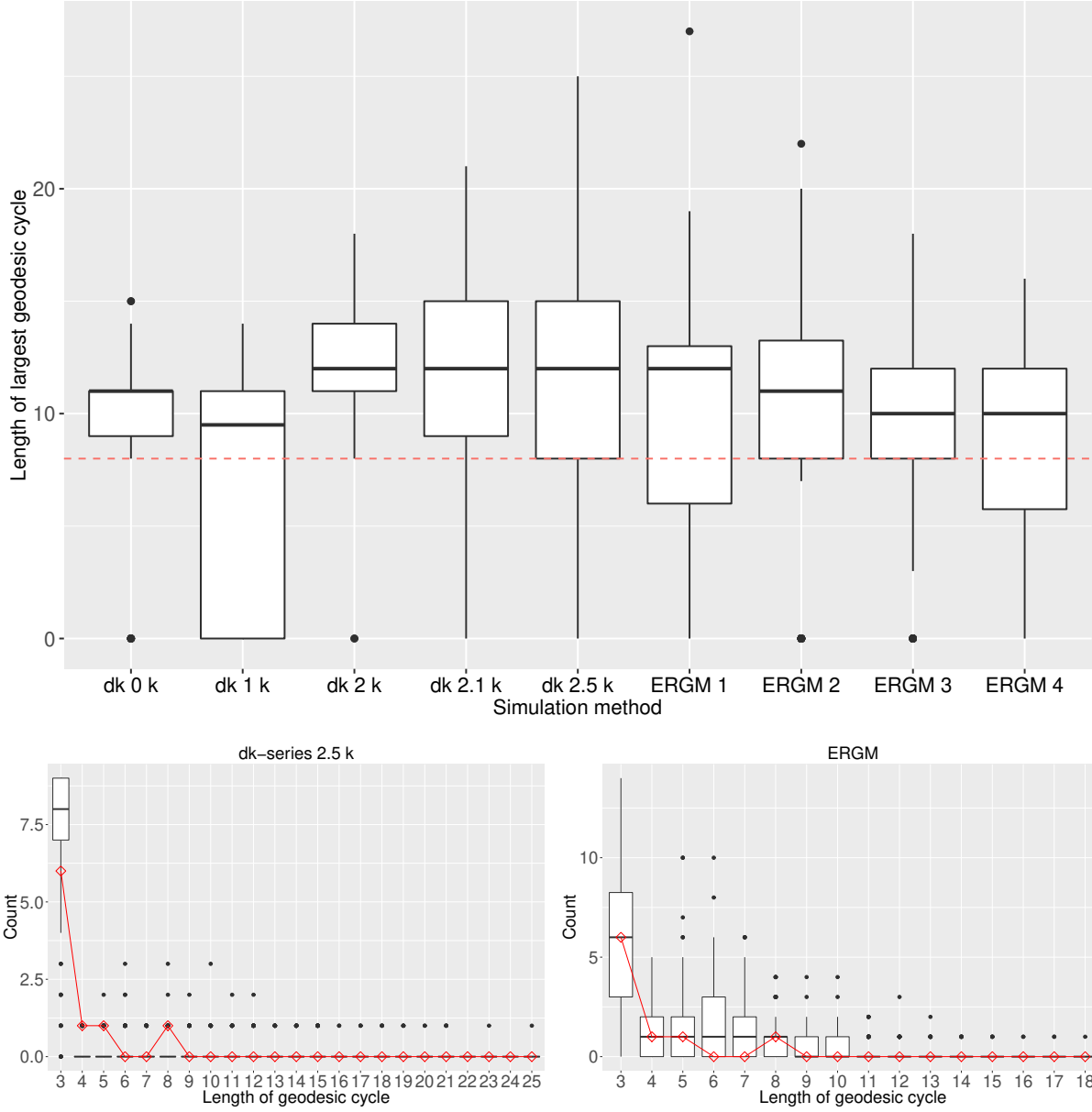


Figure 7: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for Patricia's 1993 network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM Model 3 (right). ERGM model numbers refer to those in Table 4.

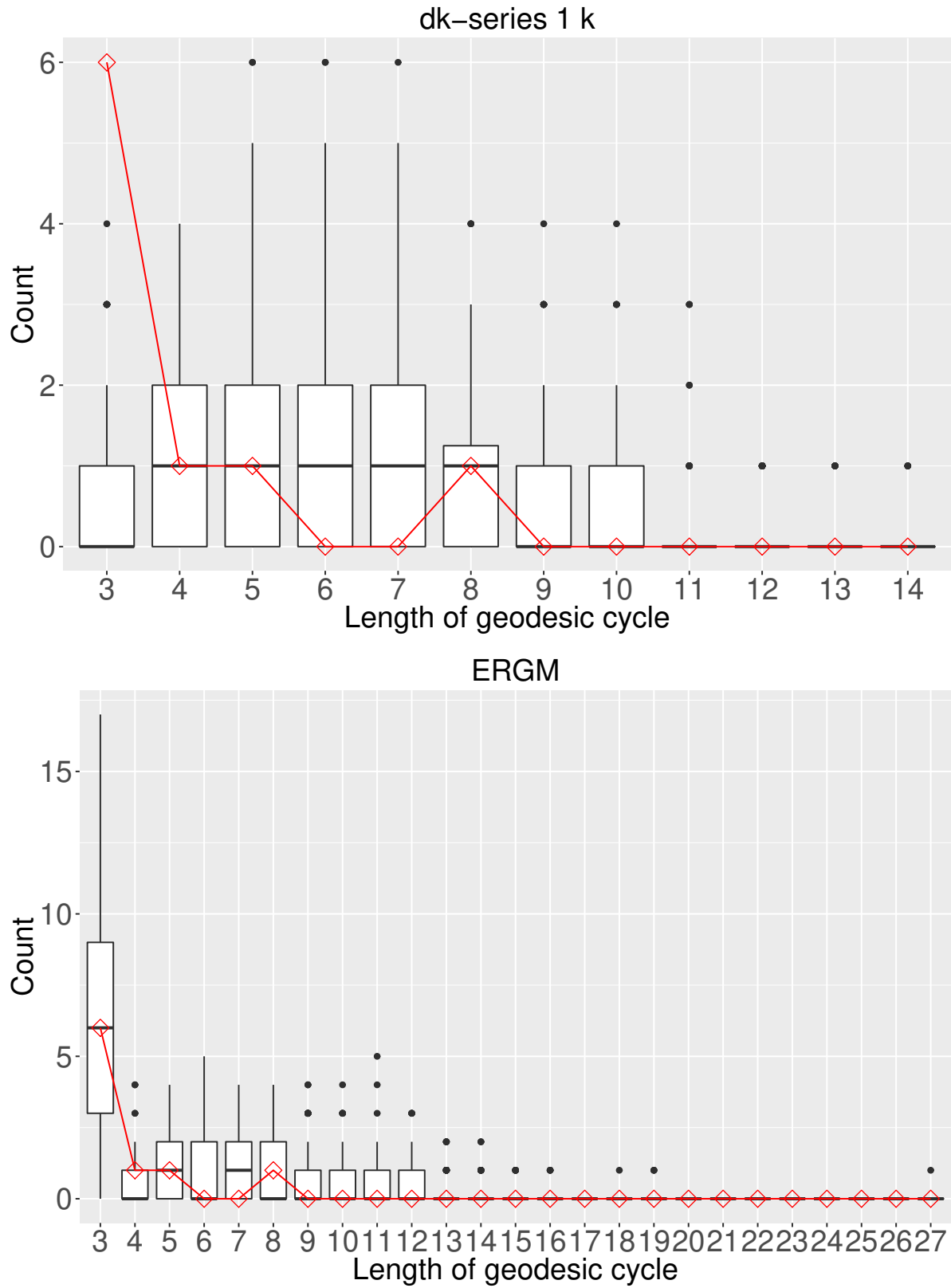


Figure 8: Distribution of geodesic cycle sizes (bottom) for Patricia’s 1993 network. The points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the *dk*-series 1*k* distribution (top) and ERGM Model 1 (bottom). ERGM model numbers refer to those in Table 4.

the geodesic cycle distribution well, while the  $dk$ -series distribution does not. Specifically, the observed network has significantly more four-cycles, and fewer five-cycles, than expected under  $dk$ -series  $2.5k$  distribution. (The importance of four-cycles in sexual contact networks is further discussed in the Conclusions). That the ERGM is able to fit cycle length distributions on this network while the  $dk$ -series cannot is due to the ERGM being able to make use of the node attribute information the sex of actors. Most of the relationships in the network are heterosexual, and as a consequence there are no odd-numbered cycles in the network: the ERGM can model this using the sex of the actors while the  $dk$ -series cannot.

In the high school friendship network (Figure 14), the ERGM fits the maximum geodesic cycle length better than the  $dk$ -series  $2.5k$  distribution does. Again, this could be a consequence of the ERGM making use of node attribute information, while the  $dk$ -series cannot. Note that we cannot test this directly on this network by fitting an ERGM without using node attributes, as such models do not converge: a parameter for homophily on class needs to be included in the model to fit the community structure induced by the school classes.

## Conclusions and further work

Neither ERGM nor  $dk$ -series models (up to  $2.5k$ ) of Patricia’s delusional social networks (for 1990 and 1992; the 1993 case is less clear) show good fits to geodesic cycle length distributions, in particular for the largest geodesic cycle length. However for the other networks examined, including empirical social networks, and a fictional character network, the ERGM models fit geodesic cycle length distributions well, as do (with two exceptions, which can be explained by the ERGM taking account of node attributes which the  $dk$ -series does not) the  $dk$ -series  $2.5k$  distributions. This would appear to indicate that there is something “special” about Patricia’s networks in this regard. Perhaps, as discussed by Martin (2017), this is because Patricia’s networks are the product of a single mind (leaving aside the reality or otherwise of multiple personalities) and reflect her spatial idea of what a social network “should” look like. In contrast, the other networks are either empirical, or in the case of the sole fictional network, the Grey’s Anatomy network, the product of multiple creators (over an extended period).

This work suggests that geodesic cycle length distributions could be useful goodness-of-fit diagnostics for social network models, in addition to the usual ones of degree distribution, maximum geodesic distance, edge-wise and dyad-wise shared partners, and triad census, as used in statnet. In the case of empirical networks, that ERGMs, explicitly based on their local (social circuit) dependence assumption, reproduce well the (non-local) geodesic cycle size distribution is an encouraging confirmation that purely local interactions really can produce the observed global structures.

The high school dating network from the Add Health research program described in Bearman et al. (2004) also notably contains a large geodesic cycle (“hollow ring”). Bearman et al. (2004) propose a normative proscription against four-cycles (“don’t date your old partner’s current partner’s old partner”), based on low number of four-cycles in their data relative to simulated networks. However Rolls et al. (2015) find in a more sophisticated ERGM model of this data with acceptable goodness-of-fit that small numbers of four-cycles are generated. The (much smaller) fictional Grey’s Anatomy sexual contact network contains seven four-cycles (of which all seven are chordless and five are geodesic) — violating this proposed proscription could be because it makes compelling entertainment (Lind, 2012). I could not find a converged ERGM with a four-cycle term for this network to explicitly test for over- or under-representation (neither could Lind (2012)). But note the ERGM (but not the  $dk$ -series) fits geodesic cycles of length four (and also cycles and chordless cycles of length four) well, similar to the case for four-cycles in Rolls et al. (2015) for the Bearman et al. (2004) network. It would also be interesting to test the geodesic cycle length distribution of this network. Although Rolls et al. (2015) have an ERGM model of this, it is quite involved (multilevel, fixing various ties), and the data are not publicly available (in Rolls et al. (2015) the network was recoded manually from the figure in Bearman et al. (2004)).

It could be informative to repeat the modeling and geodesic cycle size comparisons with other fictional networks, specifically those with a single author, to test the hypothesis that such cases would also produce anomalous geodesic cycle length distributions like Patricia’s. However fitting ERGMs to these is difficult. There is usually a main character, who is linked to most of the other characters, creating a single very strong “hub” node, which causes problems for ERGM model fitting. This problem can potentially be overcome by fixing the ties for that character, but I have not been able to get good converged ERGMs for the publicly available data sets I have tried (Les Misérables, David Copperfield, Anna Karenina, Huckleberry Finn). Note that this is not the case with Patricia’s networks: “Patricia” is not, as it were, the star of her own story in relation to her alters. I was able to fit all three Patricia networks well with ERGM, as well as the other six networks described earlier. ERGMs also have the advantage of allowing node attributes, while the  $dk$ -series models are purely structural.

This difficulty with ERGMs (having to only use data for which a converged model with good fit can be obtained) is one reason Martin (2017) gives for using  $dk$ -series instead. The present work confirms that the  $dk$ -series approach can work as well as ERGM for the purposes of simulating appropriately constrained random graphs (rather than making statistical inferences on model parameters), at least when node attributes are not available or not structurally relevant.



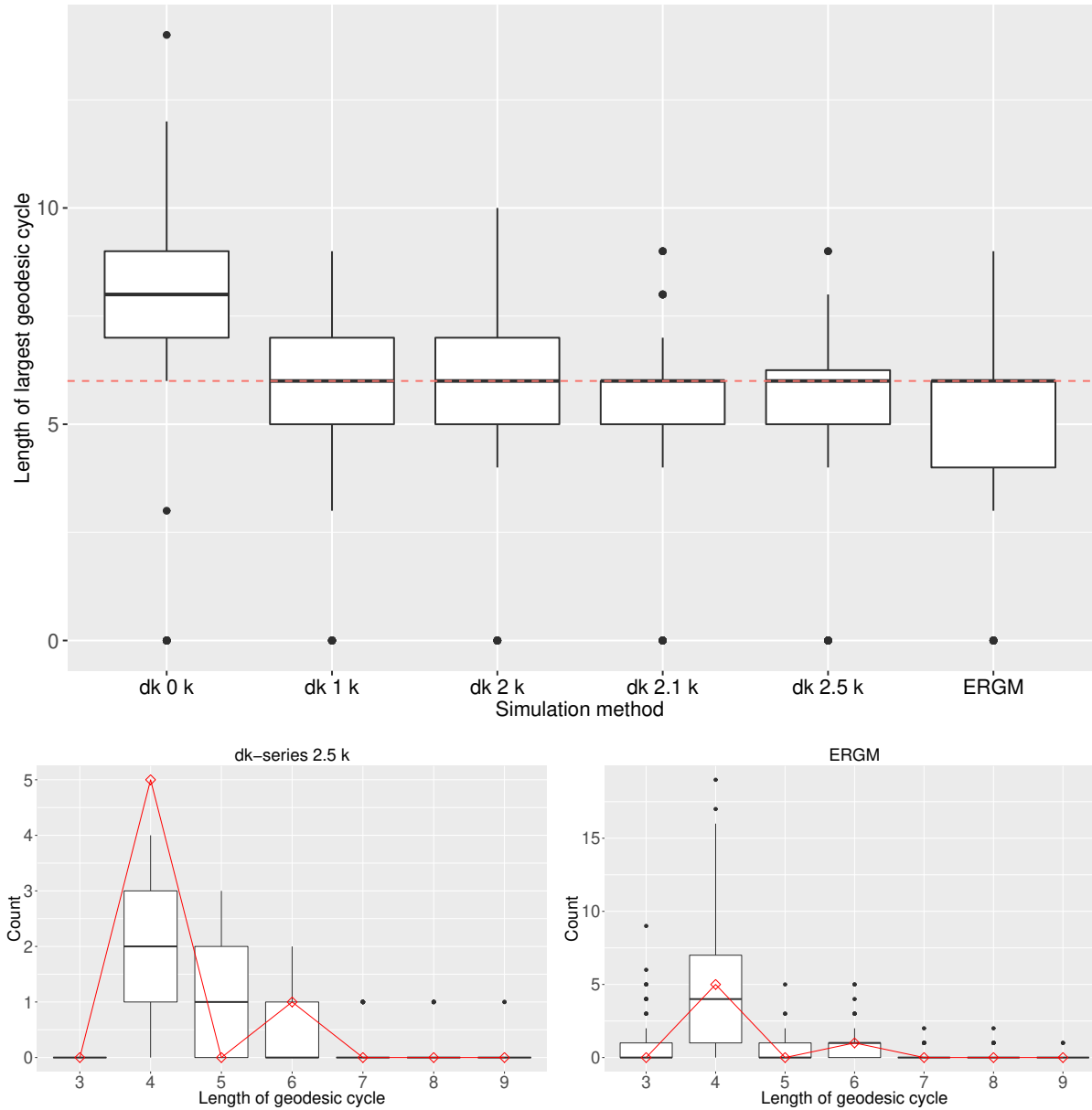


Figure 9: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the Grey's Anatomy sexual contact network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 3 in Table B2.

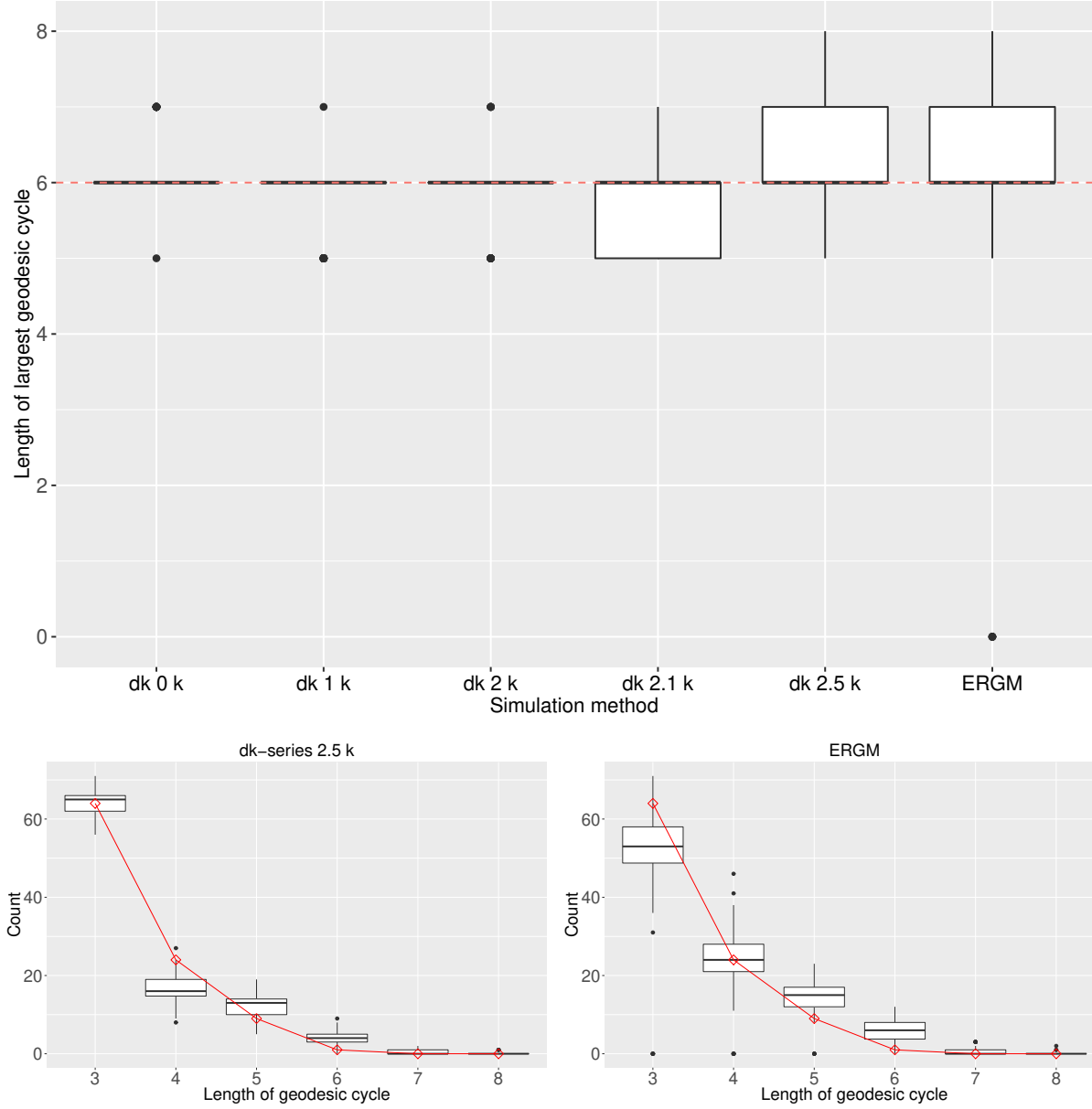


Figure 10: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the dolphin social network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 1 in Table B3.

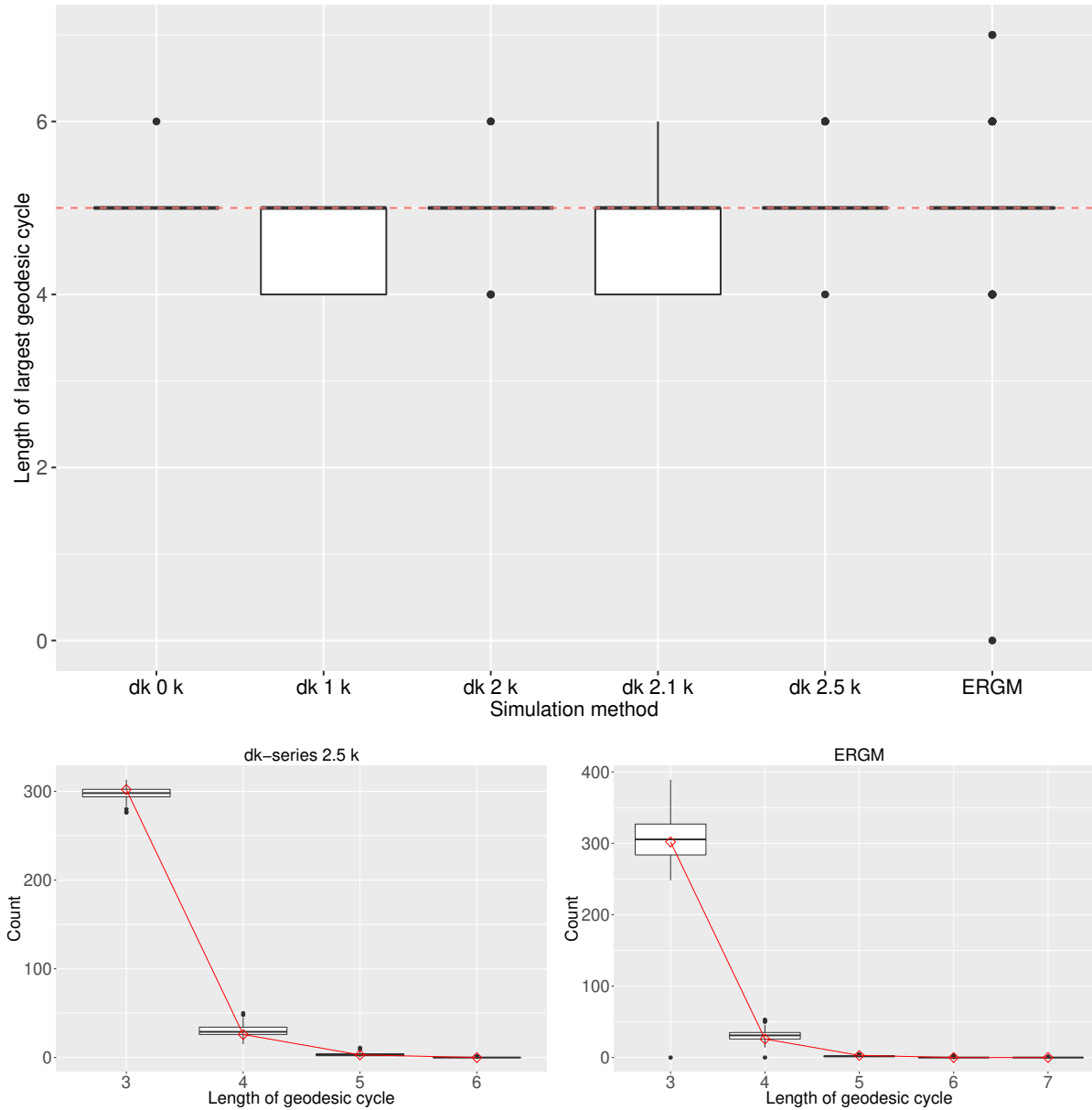


Figure 11: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the Lazega law firm friendship network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 1 in Table B4.

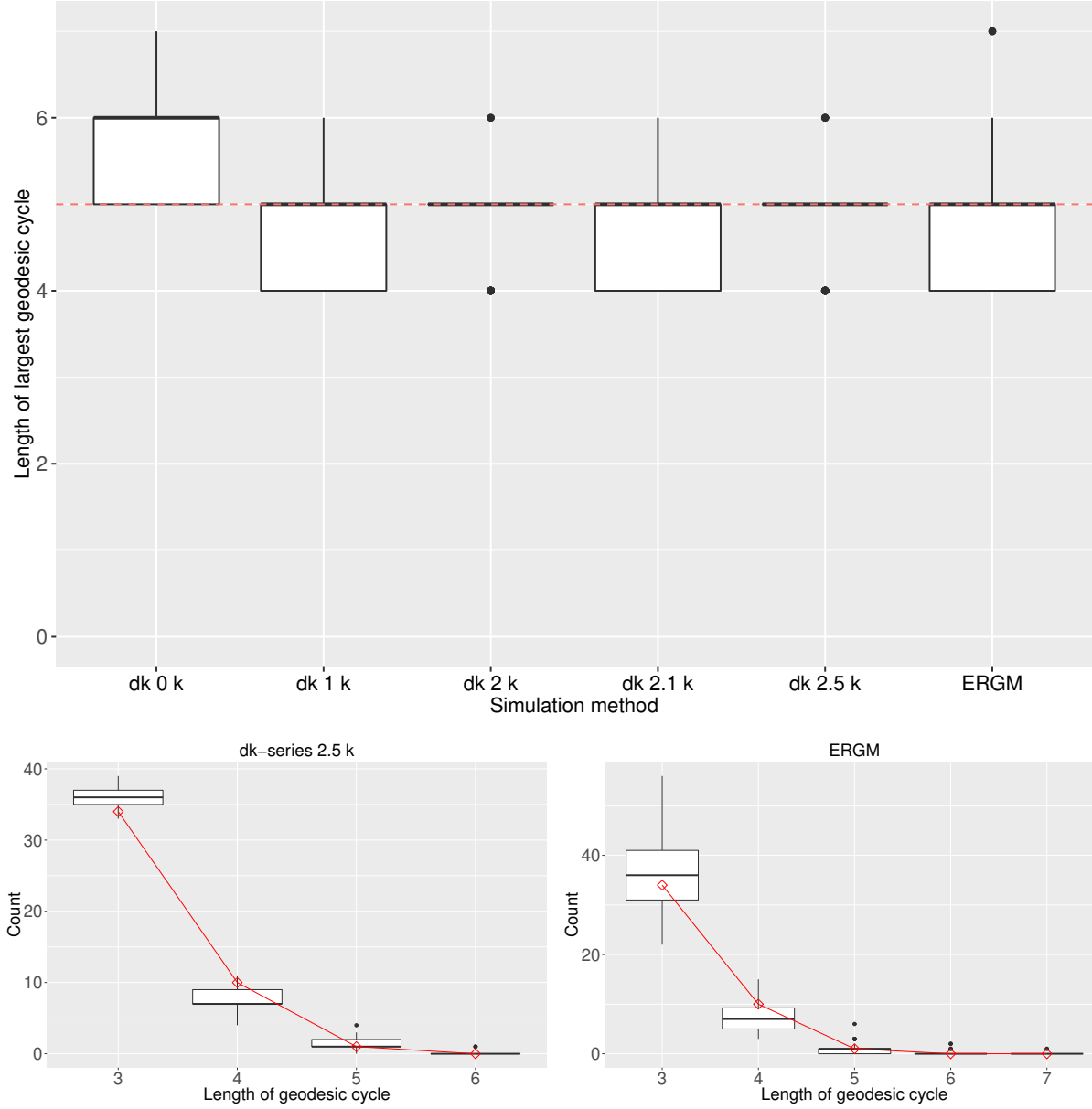


Figure 12: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the Zachary karate club network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 2 in Table B5.

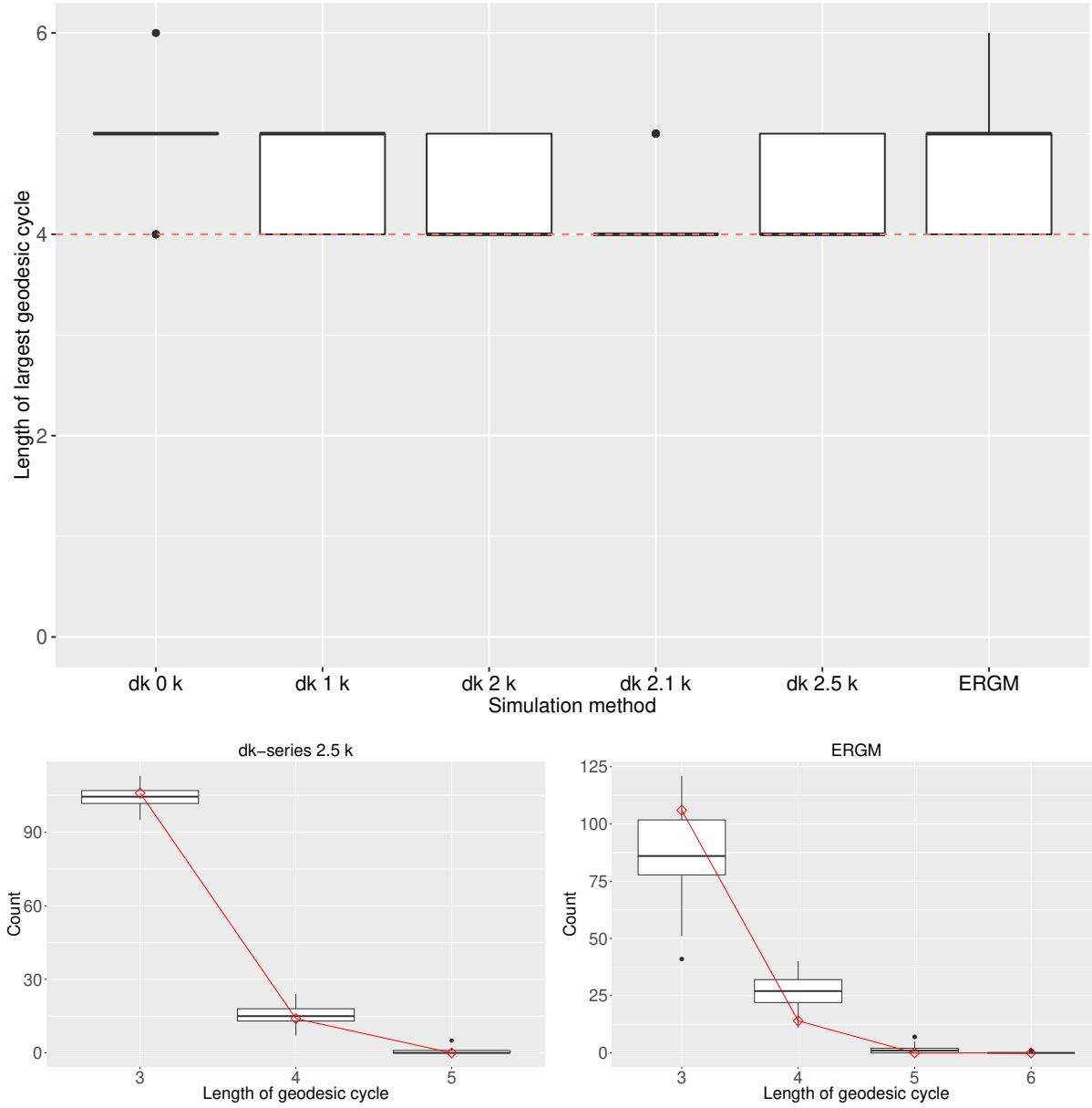


Figure 13: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the Kapferer tailor shop network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 2 in Table B6.

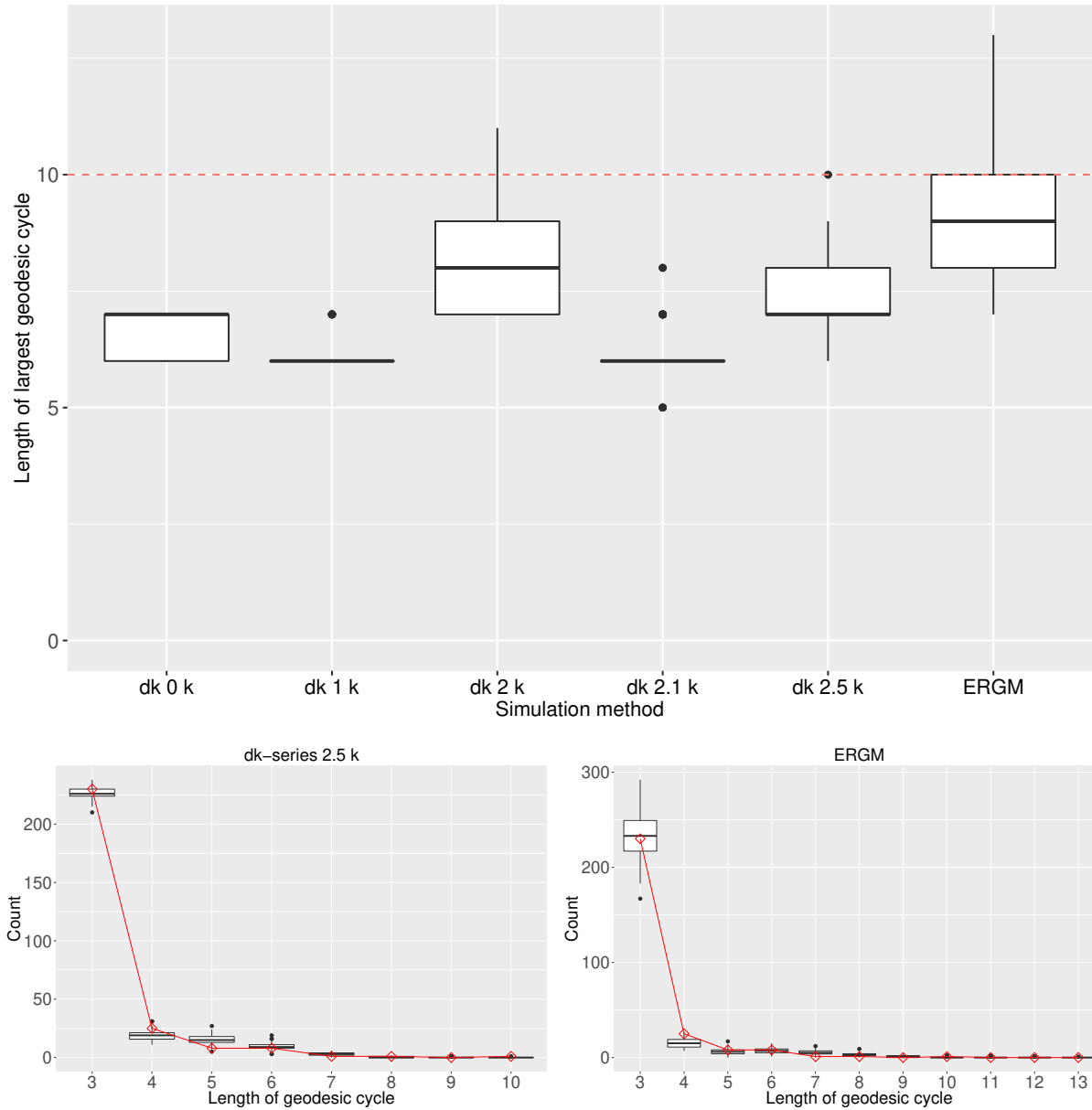


Figure 14: Largest geodesic cycle size (top), and distribution of geodesic cycle sizes (bottom) for the high school friendship network. In the top plot, the dashed red line is the value in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series or ERGM as labelled on the  $x$ -axis. In the bottom plots, the points shown as red diamonds joined by the red line are the values in the observed network, with the box plots showing the values in 100 networks simulated from the  $dk$ -series 2.5 $k$  distribution (left) and from the ERGM (right). In both ERGM plots the ERGM is Model 2 in Table B7.



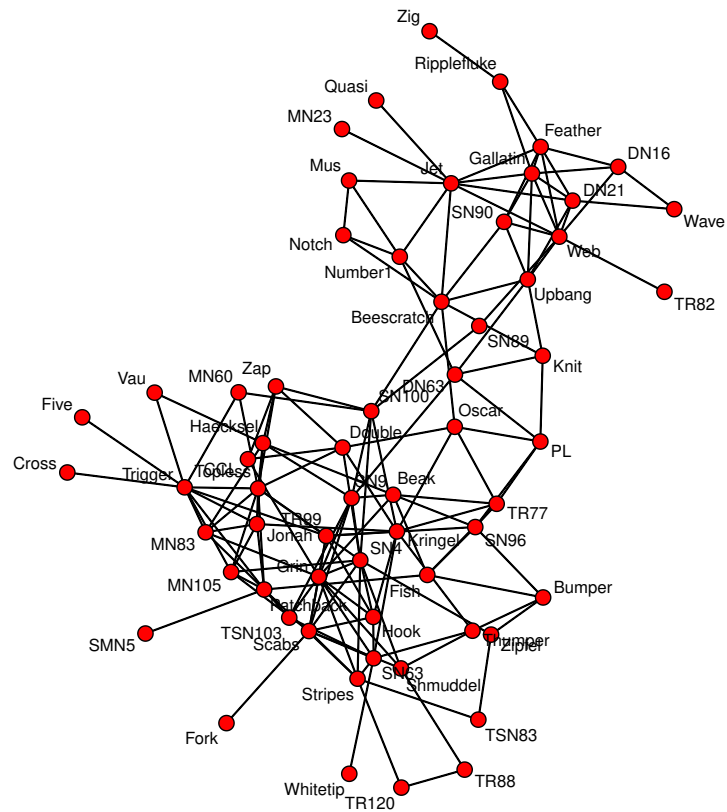


Figure A2: Dolphin social network. Visualization created using the network R package (Butts, 2008, 2015).



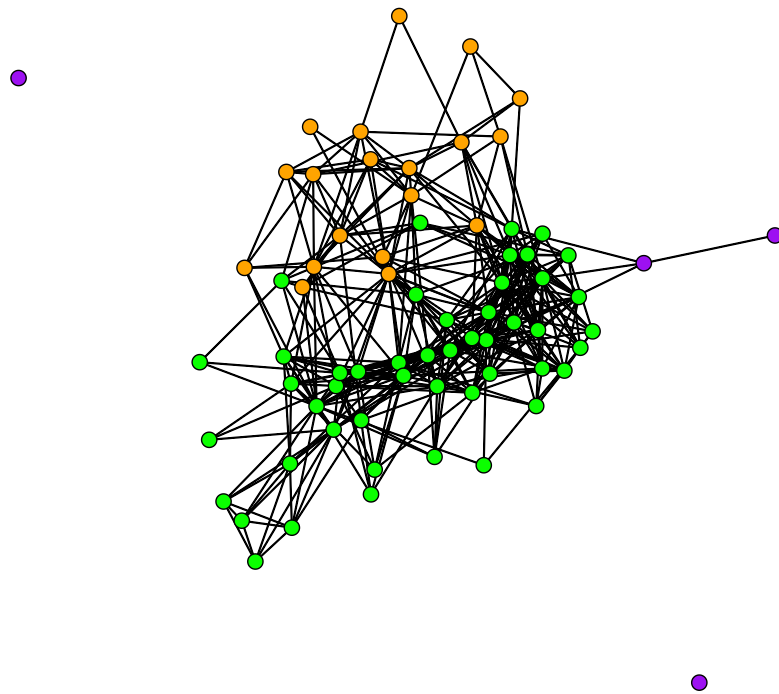


Figure A3: Lazega law firm friendship network. Nodes are colored according to the office the person works at. Visualization created using the network R package (Butts, 2008, 2015).

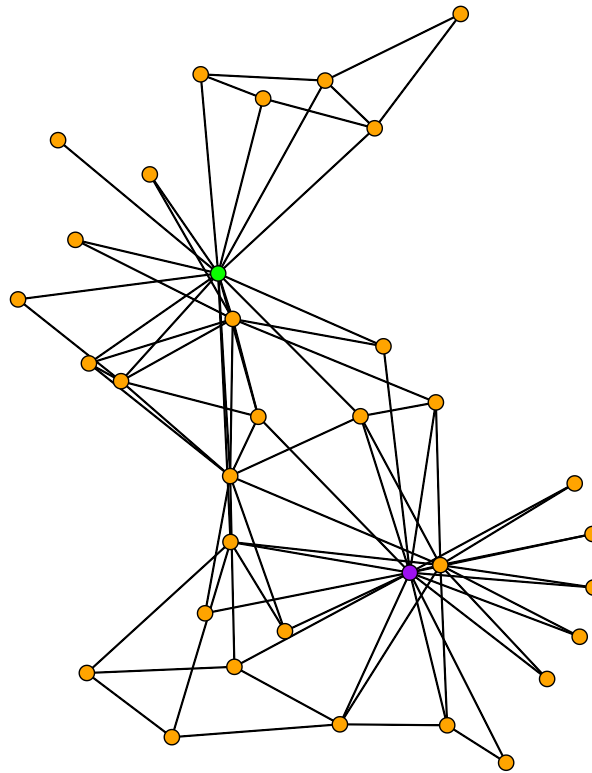


Figure A4: Zachary karate club network. Nodes are colored according to role (Instructor [green, Mr. Hi], Member [orange], or President [purple, John A.]). Visualization created using the network R package (Butts, 2008, 2015).

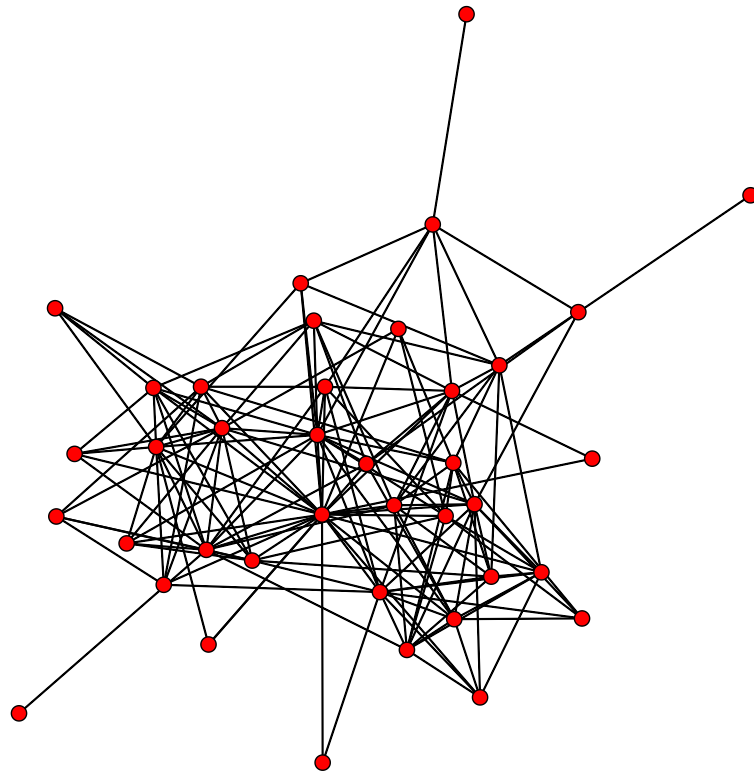


Figure A5: Kapferer tailor shop network. Visualization created using the network R package (Butts, 2008, 2015).

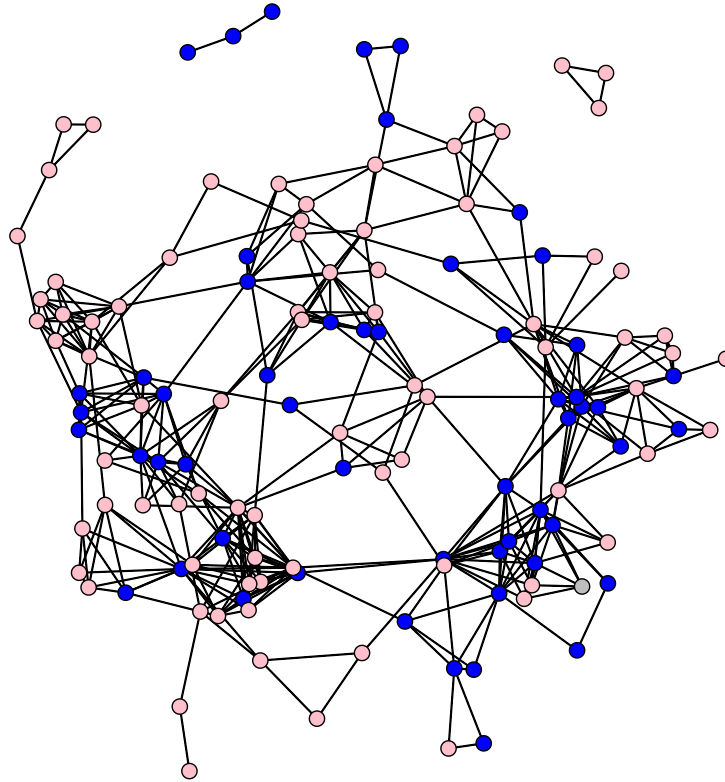
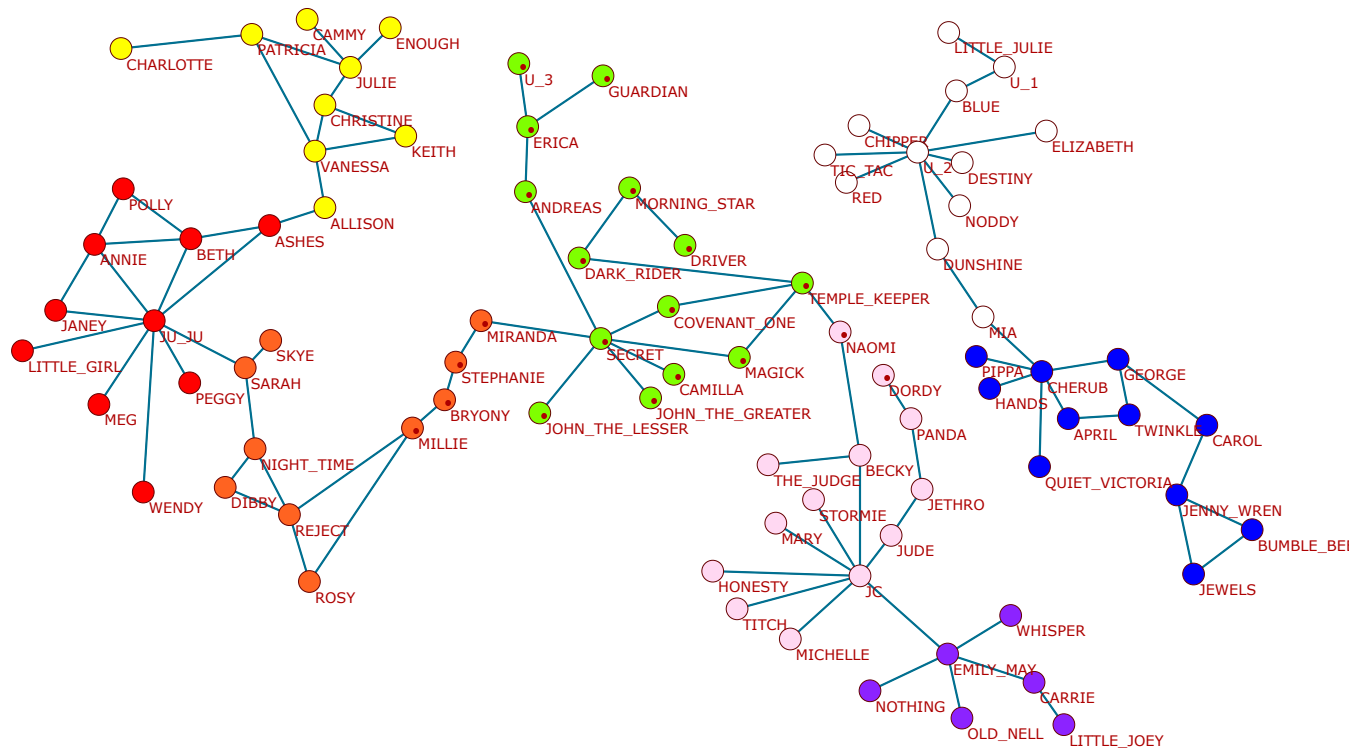


Figure A6: High school friendship network. Male students are colored blue, and female pink (there is one unknown colored gray). Visualization created using the network R package (Butts, 2008, 2015).



Pajek

Figure A7: Patricia's 1992 network. Nodes marked with a dot are included inside the "Sphere of the Blue Flame" in Patricia's original diagram. Coloring is according to network communities found with the Louvain algorithm (Blondel et al., 2008). It is apparent that the Sphere of the Blue Flame is largely included in, but does not correspond exactly with, a network community using this method, with Miranda, Stephanie, Bryony and Millie being part of another community according to the algorithm. In addition Naomi and Dordy are in the Sphere, but not in the corresponding community according to the Louvain algorithm. Dordy in particular would never be in the same community as the rest of the Sphere according to any network community algorithm, unless it also includes JC and the surrounding nodes, as Dordy is at a geodesic distance of six (via JC) from the nearest other member of the Sphere. The inclusion of Dordy in the Sphere appears to be due to spatial rather than network logic (according to the layout of Patricia's original drawing). See Martin (2017) for more discussion of this point. Community detection and visualization were done with Pajek (Batagelj and Mrvar, 2004; Mrvar and Batagelj, 2016).

Table B1: Parameters for undirected networks.

Effect	Description
Edges	Baseline density
Degree $k$	Nodes of degree $k$ . A positive parameter indicates over-representation of nodes of degree $k$ . “Degree $k - l$ ” indicates nodes of degree between the values of $k$ and $l$ inclusive.
GWDEGREE	Geometrically weighted degree distribution. A positive parameter indicates anti-preferential attachment (Hunter, 2007). See also Levy (2016).
GWDSP	Geometrically weighted dyadwise shared partner.
GWESP	Geometrically weighted edgewise shared partner. A positive parameter indicates transitivity (closure).
Homophily $c$	Homophily on categorical attribute $c$ . A positive parameter value indicates an edge preferentially forming between nodes with the same value of the categorical attribute. This may be shown instead as “ $a - b$ ” for differential homophily: homophily specifically between values $a$ and $b$ of a categorical attribute, rather than uniform homophily on an attribute.
Heterophily $c$	Heterophily on continuous attribute $c$ . This is based on the absolute value of the difference of the attribute values of two nodes.
$a$ Activity	Activity on binary attribute $a$ . A positive parameter value indicates that nodes with the binary attribute are more likely to have an incident edge.
$a$ Interaction	interaction on binary attribute $a$ . A positive parameter value indicates that two nodes with the binary attribute are more likely to have an edge between them.

## Appendix B ERGM models for other networks

Table B2: ERGM models of the Grey's Anatomy sexual network.

Effect	Model 1	Model 2	Model 3
Edges	-1.442 (0.241) <sup>***</sup>	-0.287 (0.564)	-0.844 (0.636)
Homophily Sex	-3.133 (0.718) <sup>***</sup>	-3.428 (0.741) <sup>***</sup>	-3.542 (0.732) <sup>***</sup>
Degree 1	2.026 (0.500) <sup>***</sup>	3.533 (1.058) <sup>***</sup>	3.393 (1.000) <sup>***</sup>
Degree 2		1.828 (0.911) <sup>*</sup>	1.743 (0.858) <sup>*</sup>
Degree 3		0.988 (0.805)	0.983 (0.769)
Heterophily Birth year		-0.132 (0.030) <sup>***</sup>	-0.142 (0.032) <sup>***</sup>
Attending – Attending		1.172 (0.508) <sup>*</sup>	1.085 (0.533) <sup>*</sup>
Attending – Chief		1.137 (0.682)	1.004 (0.699)
Attending – Non-Staff		-0.714 (0.642)	-0.834 (0.648)
Attending – Nurse		0.109 (0.988)	-0.058 (1.215)
Attending – Other		0.490 (0.789)	0.345 (0.874)
Attending – Resident		1.041 (0.502) <sup>*</sup>	1.004 (0.502) <sup>*</sup>
Chief – Non-Staff		-0.156 (1.183)	-0.517 (1.316)
Chief – Resident		0.438 (1.080)	0.427 (1.096)
Intern – Intern		5.003 (1.904) <sup>**</sup>	4.611 (1.753) <sup>**</sup>
Non-Staff – Non-Staff		-1.289 (1.312)	-1.431 (1.266)
Non-Staff – Resident		0.395 (0.593)	0.398 (0.604)
Nurse – Resident		1.147 (0.847)	1.554 (0.861)
Homophily Black			2.326 (0.754) <sup>**</sup>
Homophily White			0.856 (0.392) <sup>*</sup>
AIC	302.10	278.40	271.30
BIC	316.70	365.80	368.40

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. Models based on those described in Lind (2012); Leavitt and Clark (2014); Weissman (2019). \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

Table B3: ERGM models of the dolphin social network.

Effect	Model 1	Model 2
Edges	-0.821 (0.642)	-1.553 (3.122)
GWDEGREE	-2.148 (0.647) <sup>***</sup>	-0.522 (3.260)
GWDSP ( $\alpha = 0.7$ )	-0.305 (0.067) <sup>***</sup>	
GWESP ( $\alpha = 0.1$ )	0.984 (0.151) <sup>***</sup>	
GWDSP		-0.250 (0.368)
GWDSP $\alpha$		0.834 (0.669)
GWESP		0.630 (0.421)
GWESP $\alpha$		1.082 (0.324) <sup>***</sup>
AIC	1014.00	1014.00
BIC	1036.00	1047.00

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b) using the “Stepping” algorithm (Hummel et al., 2012). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.5. \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

Table B4: ERGM models of the Lazega law firm friendship network.

Effect	Model 1
Edges	-5.256 (0.317)***
GWDEGREE	1.290 (0.874)
GWESP	0.597 (0.072)***
GWESP $\alpha$	1.398 (0.030)***
Homophily GENDER	0.535 (0.103)***
Homophily LAW_SCHOOL	0.137 (0.130)
Homophily OFFICE	0.767 (0.111)***
Homophily PRACTICE	0.485 (0.105)***
Homophily STATUS	0.759 (0.104)***
Heterophily AGE	-0.019 (0.009)*
Heterophily SENIORITY	-0.019 (0.009)*
AIC	1697.00
BIC	1761.00

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.1. \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .

Table B5: ERGM models of the Zachary karate club network.

Effect	Model 1	Model 2
Edges	-3.830 (0.405)***	-2.095 (0.491)***
GWDEGREE	5.566 (3.376)	0.988 (1.228)
GWESP ( $\alpha = 0.5$ )	1.102 (0.211)***	0.358 (0.230)
Instructor		2.345 (0.527)***
President		2.369 (0.543)***
Faction abs. diff. 1		-0.246 (0.316)
Faction abs. diff. 2		-1.542 (0.492)**
Faction abs. diff. 3		-2.179 (0.605)***
Faction abs. diff. 4		-2.672 (0.626)***
AIC	419.40	346.80
BIC	432.40	385.80

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.2. The faction identifier is coded in the Zachary karate club data from the `ergm.count` package (Krivitsky, 2016) as -2 (strongly Mr. Hi's), -1 (weakly Mr. Hi's), 0 (neutral), +1 (weakly John's), and +2 (strongly John's), and so the absolute difference in faction id ranges from 0 (same faction) to 4 (strongly Mr Hi's and strongly John's). \*\*\*  $p < 0.0001$ , \*\*  $p < 0.001$ , \*  $p < 0.05$ .



Table B6: ERGM models, reproducing those in Hummel et al. (2012), of the Kapferer tailor shop network.

Effect	Model 1	Model 2
Edges	−3.082 (0.567) <sup>***</sup>	−2.997 (0.523) <sup>***</sup>
GWDEGREE	0.360 (0.935)	
GWDSP ( $\alpha = 0.25$ )	−0.129 (0.051) <sup>*</sup>	−0.130 (0.052) <sup>*</sup>
GWESP ( $\alpha = 0.25$ )	1.491 (0.343) <sup>***</sup>	1.436 (0.286) <sup>***</sup>
AIC	732.70	732.20
BIC	751.10	746.00

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b) using the “Stepping” algorithm (Hummel et al., 2012). Estimated standard errors in parentheses. The GWDEGREE decay parameter is fixed at 0.25. <sup>\*\*\*</sup>  $p < 0.0001$ , <sup>\*\*</sup>  $p < 0.001$ , <sup>\*</sup>  $p < 0.05$ .

Table B7: ERGM models of the high school friendship network.

Effect	Model 1	Model 2	Model 3
Edges	−7.000 (0.570) <sup>***</sup>	−8.542 (0.471) <sup>***</sup>	−8.574 (0.457) <sup>***</sup>
GWDEGREE	2.447 (0.385) <sup>***</sup>	3.062 (0.266) <sup>***</sup>	3.037 (0.259) <sup>***</sup>
GWDEGREE decay	1.563 (0.113) <sup>***</sup>		
GWDSP ( $\alpha = 0.5$ )	−0.014 (0.031)	0.049 (0.023) <sup>*</sup>	0.047 (0.022) <sup>*</sup>
GWESP	1.210 (0.079) <sup>***</sup>		
GWESP $\alpha$	1.157 (0.029) <sup>***</sup>		
GWESP ( $\alpha = 1.2$ )		1.344 (0.069) <sup>***</sup>	1.334 (0.067) <sup>***</sup>
Homophily Class	1.057 (0.086) <sup>***</sup>	1.090 (0.084) <sup>***</sup>	1.081 (0.080) <sup>***</sup>
Homophily Sex			0.171 (0.086) <sup>*</sup>
AIC	2410.00	1937.00	1975.00
BIC	2459.00	1972.00	2018.00

Note: Estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b). Estimated standard errors in parentheses. In models 2 and 3, the GWDEGREE decay parameter is fixed at 1.7, the GWESP  $\alpha$  parameter is fixed at 1.2, and the GWDSP  $\alpha$  parameter is fixed at 0.5. Models without homophily on class included do not converge (there is a very noticeable community structure in the network, caused by the classes; a parameter needs to be included in the ERGM to account for this). <sup>\*\*\*</sup>  $p < 0.0001$ , <sup>\*\*</sup>  $p < 0.001$ , <sup>\*</sup>  $p < 0.05$ .

## References

- Amati, V., Lomi, A., and Mira, A. (2018). Social network modeling. *Annual Review of Statistics and its Application*, 5:343–369.
- Batagelj, V. and Mrvar, A. (2004). Pajek — analysis and visualization of large networks. In *Graph Drawing Software*, Mathematics and Visualization, pages 77–103. Springer.
- Batten, L. M. (1983). Geodesic subgraphs. *Journal of Graph Theory*, 7(2):159–163.
- Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110(1):44–91.
- Benjamini, I., Hoppen, C., Ofek, E., Prałat, P., and Wormald, N. (2011). Geodesics and almost geodesic cycles in random regular graphs. *Journal of Graph Theory*, 66(2):115–136.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.
- Borisenko, A., Byshkin, M., and Lomi, A. (2020). A simple algorithm for scalable Monte Carlo inference. Preprint. Available from: arXiv:1901.00533v4. Cited 18 February 2020.
- Butts, C. T. (2008). network: a package for managing relational data in R. *Journal of Statistical Software*, 24(2).
- Butts, C. T. (2015). *network: Classes for Relational Data*. The Statnet Project (<http://statnet.org>). R package version 1.13.0.
- Byshkin, M., Stivala, A., Mira, A., Krause, R., Robins, G., and Lomi, A. (2016). Auxiliary parameter MCMC for exponential random graph models. *Journal of Statistical Physics*, 165(4):740–754.
- Byshkin, M., Stivala, A., Mira, A., Robins, G., and Lomi, A. (2018). Fast maximum likelihood estimation via equilibrium expectation for large network data. *Scientific Reports*, 8:11509.
- Caimo, A. and Friel, N. (2011). Bayesian inference for exponential random graph models. *Social Networks*, 33:41–55.
- Colomer-de Simón, P. and Boguñá, M. (2014). Double percolation phase transition in clustered complex networks. *Physical Review X*, 4(4):041020.
- Colomer-de Simón, P., Serrano, M. A., Beiró, M. G., Alvarez-Hamelin, J. I., and Boguñá, M. (2013). Deciphering the global organization of clustering in real complex networks. *Scientific Reports*, 3:2517.
- Corander, J., Dahmström, K., and Dahmström, P. (1998). Maximum likelihood estimation for Markov graphs. Technical Report 8, Stockholm University, Department of Statistics.
- Corander, J., Dahmström, K., and Dahmström, P. (2002). Maximum likelihood estimation for exponential random graph models. In Hagberg, J., editor, *Contributions to social network analysis, information theory, and other topics in statistics; a Festschrift in honour of Ove Frank*, pages 1–17. Department of Statistics, University of Stockholm.
- Csárdi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- David, A., Kemp, R., Smith, L., and Fahy, T. (1996). Split minds: Multiple personality and schizophrenia. In Halligan, P. W. and Marshall, J. C., editors, *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, chapter 7, pages 122–146. Psychology Press, Hove, UK.
- Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–60.
- Farber, M. (1987). Bridged graphs and geodesic convexity. *Discrete Mathematics*, 66(3):249–257.
- Farber, M. and Jamison, R. E. (1986). Convexity in graphs and hypergraphs. *SIAM Journal on Algebraic Discrete Methods*, 7(3):433–444.
- Freud, S. (1895/1950). Extracts from the Fliess papers. In Strachey, J., editor, *The standard edition of the complete works of Sigmund Freud*, volume 1, pages 173–280. Hogarth Press and the Institute of Psychoanalysis, London.

- Garey, M. R. and Johnson, D. S. (1979). *Computers and intractability: a guide to the theory of NP-completeness*. Freeman, San Francisco.
- Gashler, M. (2011). Waffles: A machine learning toolkit. *Journal of Machine Learning Research*, 12(Jul):2383–2387.
- Gashler, M. and Martinez, T. (2012). Robust manifold learning with CycleCut. *Connection Science*, 24(1):57–69.
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., Bender-deMoll, S., and Morris, M. (2016a). *statnet: Software Tools for the Statistical Analysis of Network Data*. The Statnet Project (<http://www.statnet.org>). R package version 2016.4.
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., and Morris, M. (2016b). *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (<http://www.statnet.org>). R package version 3.6.0.
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Morris, and Martina (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1):1548.
- Hellmuth, M., Leydold, J., and Stadler, P. F. (2014). Convex cycle bases. *Ars Mathematica Contemporanea*, 7(1):123–140.
- Hummel, R. M., Hunter, D. R., and Handcock, M. S. (2012). Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4):920–939.
- Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, 29(2):216–230.
- Hunter, D. R. and Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29.
- Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012). Computational statistical methods for social network models. *Journal of Computational and Graphical Statistics*, 21(4):856–882.
- Kapferer, B. (1972). *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*. Manchester University Press.
- Kihlstrom, J. F. (2005). Dissociative disorders. *Annual Review of Clinical Psychology*, 1:227–253.
- Krivitsky, P. N. (2016). *ergm.count: Fit, Simulate and Diagnose Exponential-Family Models for Networks with Count Edges*. The Statnet Project (<http://www.statnet.org>). R package version 3.2.2.
- Lazega, E. (2001). *The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership*. Oxford University Press.
- Lazega, E. and Pattison, P. E. (1999). Multiplexity, generalized exchange and cooperation in organizations: a case study. *Social Networks*, 21(1):67–90.
- Leavitt, A. and Clark, J. (2014). SNA in R workshop. <https://github.com/alexleavitt/SNAinRworkshop>. Accessed 12 July 2019.
- Levy, M. (2016). gwdegree: Improving interpretation of geometrically-weighted degree estimates in exponential random graph models. *Journal of Open Source Software*, 1(3):36.
- Li, Y. and Shi, L. (2018). Geodesic cycles in random graphs. *Discrete Mathematics*, 341(5):1275–1281.
- Lind, B. (2012). Lessons on exponential random graph modeling from Grey’s Anatomy hook-ups. <https://badhessian.org/2012/09/lessons-on-exponential-random-graph-modeling-from-greys-anatomy-hook-ups/>. Accessed 12 July 2019.
- Lokshtanov, D. (2009). Finding the longest isometric cycle in a graph. *Discrete Applied Mathematics*, 157(12):2670–2674.

- Łuczak, T. (1991). Holes in random graphs. *Discrete Applied Mathematics*, 30(2-3):213–219.
- Łuczak, T. (1993). The size of the largest hole in a random graph. *Discrete Mathematics*, 112(1-3):151–163.
- Lusher, D., Koskinen, J., and Robins, G., editors (2013). *Exponential random graph models for social networks*. Structural Analysis in the Social Sciences. Cambridge University Press, New York.
- Lusseau, D., Schneider, K., Boisseau, O. J., Haase, P., Slooten, E., and Dawson, S. M. (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405.
- Mahadevan, P., Krioukov, D., Fall, K., and Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *Computer Communication Review*, 36(4):135–146.
- Martin, J. L. (2017). The structure of node and edge generation in a delusional social network. *Journal of Social Structure*, 18(1):1–21.
- Mastrandrea, R., Fournet, J., and Barrat, A. (2015). Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, 10(9):e0136497.
- Mrvar, A. and Batagelj, V. (2016). Analysis and visualization of large networks with program package Pajek. *Complex Adaptive Systems Modeling*, 4(1):1–6.
- Negami, S. and Xu, G.-H. (1986). Locally geodesic cycles in 2-self-centered graphs. *Discrete Mathematics*, 58(3):263–268.
- Orsini, C., Dankulov, M. M., Colomer-de Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A., Bassler, K. E., Toroczkai, Z., Boguná, M., Caldarelli, G., et al. (2015). Quantifying randomness in real networks. *Nature Communications*, 6:8627.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph ( $p^*$ ) models for social networks. *Social Networks*, 29(2):173–191.
- Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007b). Recent developments in exponential random graph ( $p^*$ ) models for social networks. *Social Networks*, 29(2):192–215.
- Rolls, D. A., Wang, P., McBryde, E., Pattison, P., and Robins, G. (2015). A simulation study comparing epidemic dynamics on exponential random graph and edge-triangle configuration type contact network models. *PLoS ONE*, 10(11):e0142181.
- Roman, S. (1989). *An Introduction to Discrete Mathematics*. Harcourt Brace Jovanovich, Orlando, FL, 2nd edition.
- Schweinberger, M., Krivitsky, P. N., Butts, C. T., and Stewart, J. (2019). Exponential-family models of random graphs: Inference in finite-, super-, and infinite-population scenarios. Preprint. Available from: arXiv:1707.04800v4. Cited 18 February 2020. To appear in *Statistical Science*.
- Snijders, T. A. B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., and Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153.
- Stivala, A., Robins, G., and Lomi, A. (2020). Exponential random graph model parameter estimation for very large directed networks. *PLoS ONE*, 15(1):e0227804.
- Takács, L. (1988). On the limit distribution of the number of cycles in a random graph. *Journal of Applied Probability*, 25(A):359–376.
- Uno, T. and Satoh, H. (2014). An efficient algorithm for enumerating chordless cycles and chordless paths. In *International Conference on Discovery Science*, pages 313–324. Springer.
- Wang, P., Robins, G., and Pattison, P. (2009). *PNet: program for the simulation and estimation of exponential random graph ( $p^*$ ) models*. Department of Psychology, The University of Melbourne.

- Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105(2):493–527.
- Weissman, G. (2019). Grey’s Anatomy network of sexual relations. <https://gweissman.github.io/post/grey-s-anatomy-network-of-sexual-relations/>. Accessed 12 July 2019.
- Weisstein, E. W. (2020). Graph cycle. From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/GraphCycle.html>. Accessed 12 February 2020.
- White, H. C., Boorman, S. A., and Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology*, 81(4):730–780.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- Zachary, W. W. (1977). An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473.