The hollow ring of randomness: Large worlds in small data

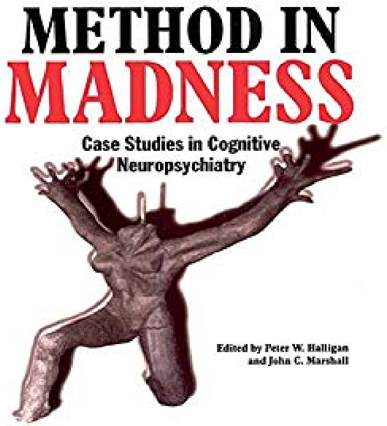
Alex Stivala
Università della Svizzera italiana, Lugano, Switzerland

ASNAC 2019, Adelaide, South Australia, November 28–29, 2019

Introduction

- Martin (2017) analyzed social networks of personalities from a patient with Multiple Personality Disorder (MPD) [now known as dissociative identity disorder (DID)].
- using the "dk-series" model (Orsini et al., 2015), a sequence of nested network distributions of increasing complexity.
- One network contains a large "hollow ring" — a cycle with no shortcuts so the shortest path is along the cycle.
- The other two networks however have much smaller largest such cycles — smaller than expected under the dk-series model.

Data source



A. David, R. Kemp, L. Smith, and T. Fahy. Split minds: Multiple personality and schizophrenia. In P. W. Halligan and J. C. Marshall, editors, *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, chapter 7, pages 122-146. Psychology Press, Hove, UK, 1996

"Patricia"

- "Patricia" is a patient diagnosed with Multiple Personality Disorder described by David et al. (1996).
- She drew maps of her alternate personalities, with the alters connected by lines.
- As noted by Martin (2017) the exact nature of what these edges represent is obscure, but according to David et al. (1996):

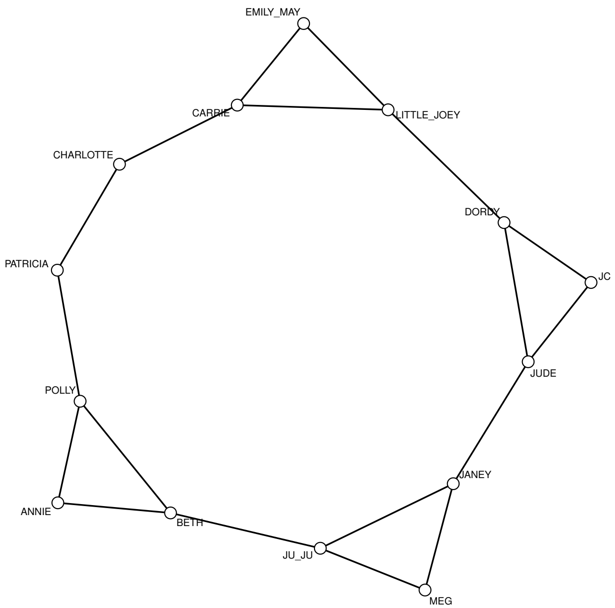
> *Patricia illustrated the multiplication and branching of her personality with diagrams ...*

(David et al., 1996, p. 139)

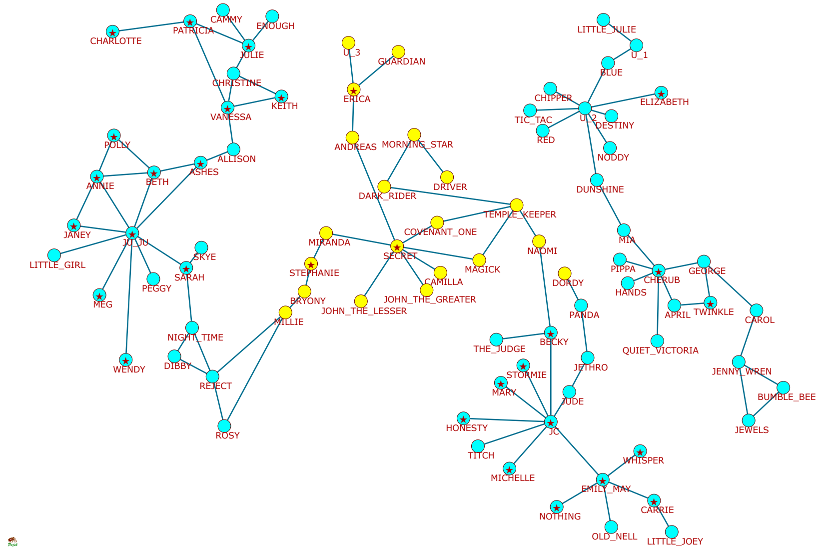
> *It appears that certain personalities occupied a central position in her mind and these in turn have given rise to subsidiary personalities (as indicated by the connecting lines).*

(David et al., 1996, p. 136).

Patricia's 1990 network

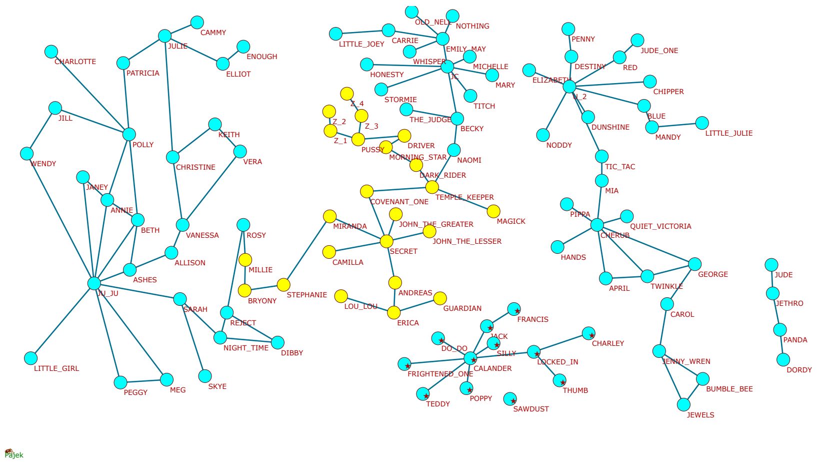


Patricia's 1992 network



Yellow nodes are in the "Sphere of the Blue Flame" and nodes marked with a star are marked as Christian on Patricia's original diagram.

Patricia's 1993 network



Yellow nodes are in the "Sphere of the Blue Flame" and nodes marked with a star are marked as "(Behind)" on Patricia's original diagram.

What does it mean?

> *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)

> *The illusion of unity of consciousness is not exposed in MPD but is repeated over and over again. One illusory consciousness gives way (or joins) another. Each is a coherent autonomous homonculus. 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)

Freudian interpretation

- 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, 1895/1950, pp. 211–212)

Martin's interpretation

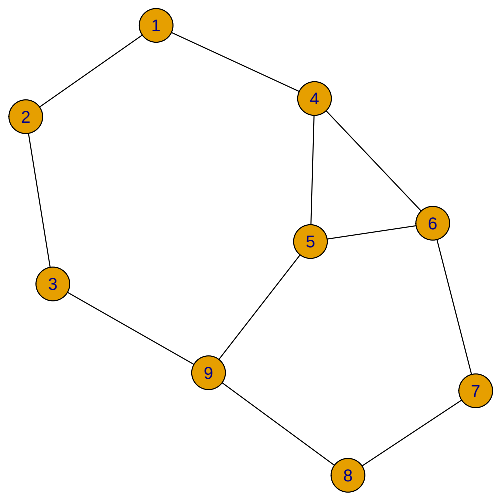
> *In many ways, this returns us to core intuitions of early network analysts: that we are interested in some sort of intertwining of lives, but we might be unable to specify a single content to the nature of the ties involved. ... Patricia's maps presumably are indicating a more fundamental connection,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

(Martin, 2017, p. 5)

An alternative interpretation

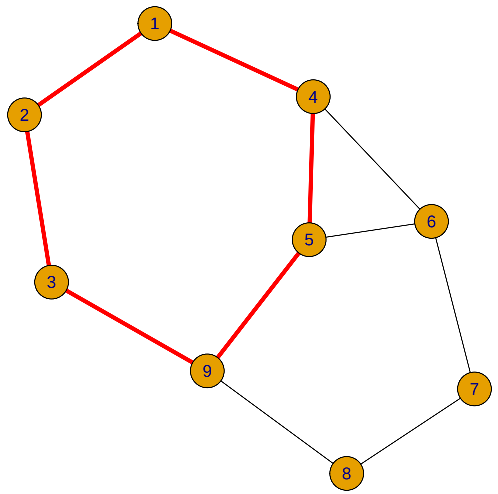
- But is not "a single mind capable of sustaining dozens of other minds, lives, and relationships" exactly (part of) what a novelist (or playwright or screenwriter, or other fiction writer), must have have, to some degree?
- The characters ("coherent autonomous homonculi") in a work of fiction are individual and coherent, but they perhaps have a common hallmark, from the author.
- So (leaving aside any considerations as to how "real" MPD alters are) it makes sense to 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.

Teach Yourself Graph Theory in 60 Seconds!

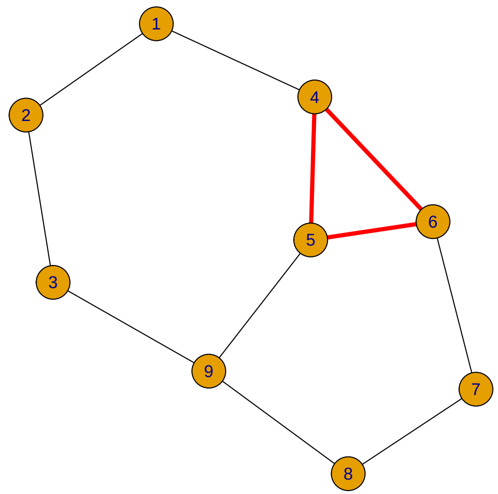


This graph has 7 cycles (one is Hamiltonian). 4 are **chordless** and 3 of those are also **geodesic** (and, in addition, **convex**).

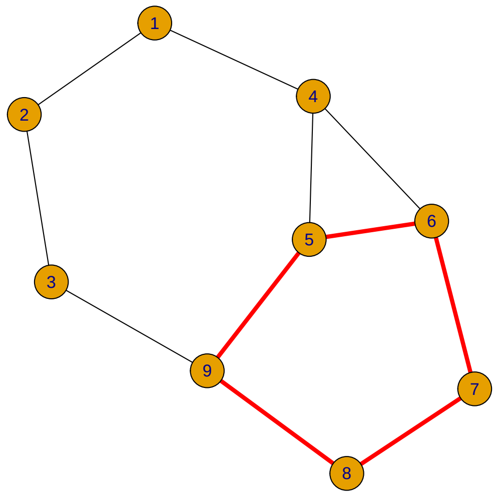
The cycle in red is geodesic (and so also chordless)



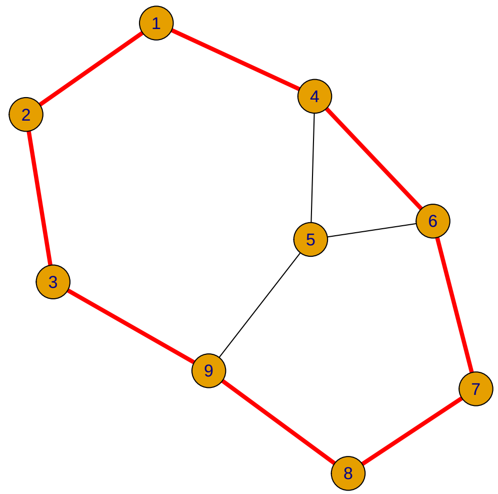
The cycle in red is geodesic (and so also chordless)



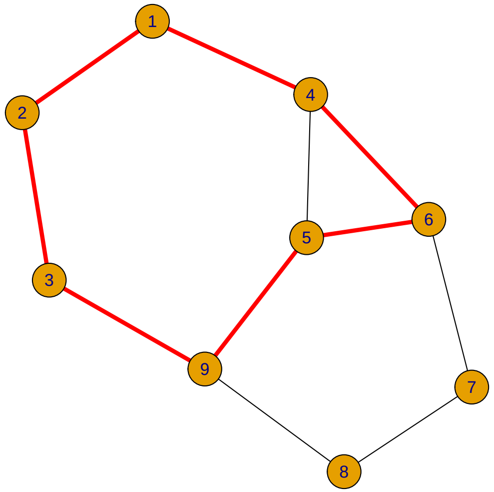
The cycle in red is geodesic (and so also chordless)



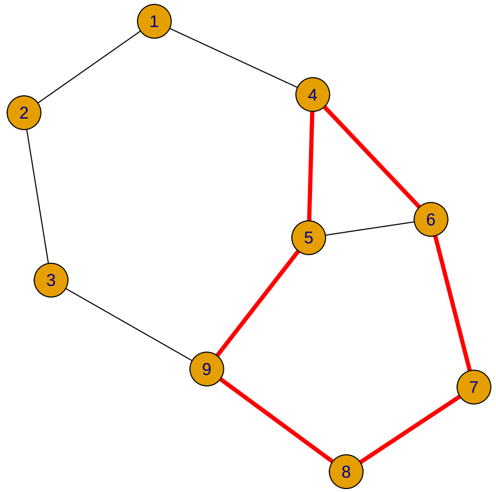
The cycle in red is chordless but not geodesic



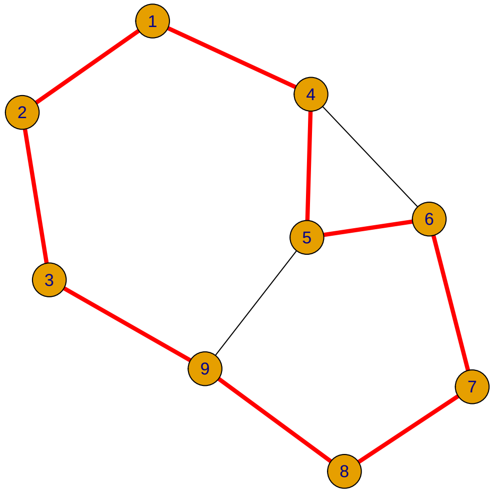
The cycle in red is not chordless (and so not geodesic)



The cycle in red is not chordless (and so not geodesic)



The cycle in red is not chordless (and so not geodesic)



This cycle is Hamiltonian.

Ring / Cycle

- Martin (2017) describes a “hollow ring” as “a cycle ... no pair of [nodes of] which have a distance in the graph lower than that in the cycle. (In other words, there are no “short cuts” between nodes in the ring).” [p. 16]
- But in fact we can see that this is just the same as a “geodesic cycle”.
- I prefer the term “geodesic cycle” which makes sense mathematically given its definition, without confusing it with other properties such as atomicity, and dates back to Negami and Xu (1986).

High school dating network (Bearman et al., 2004)

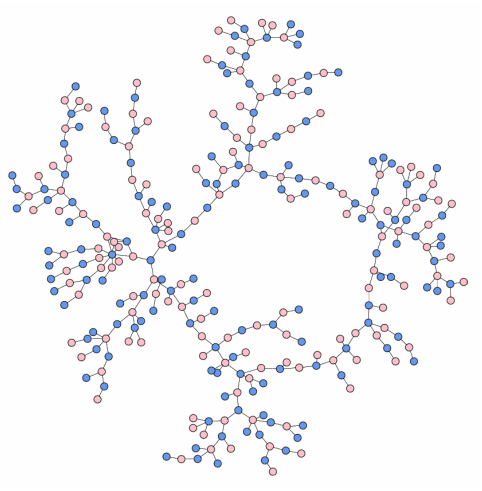
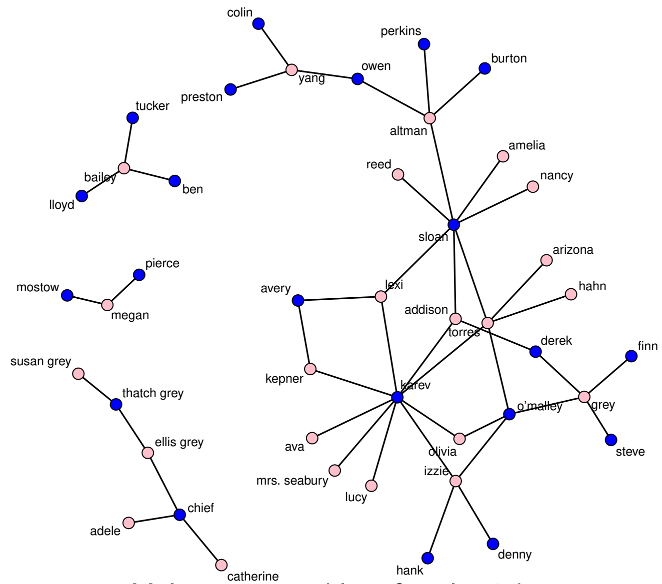


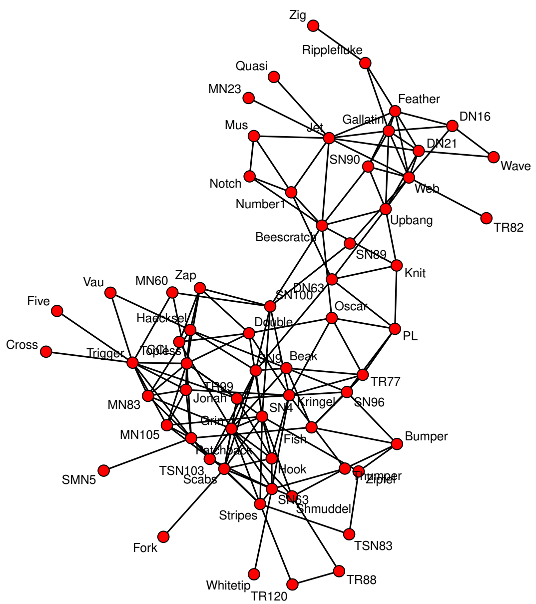
Figure by Mark Newman downloaded from `http://www-personal.umich.edu/~mejn/networks/addhealth.gif`

Grey's Anatomy sexual relations network

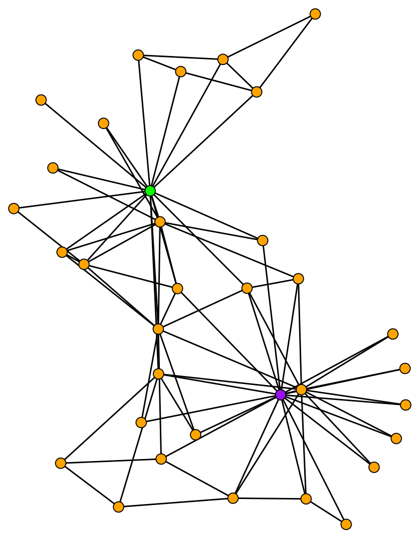


Male actors are blue, female pink.

Dolphin social network

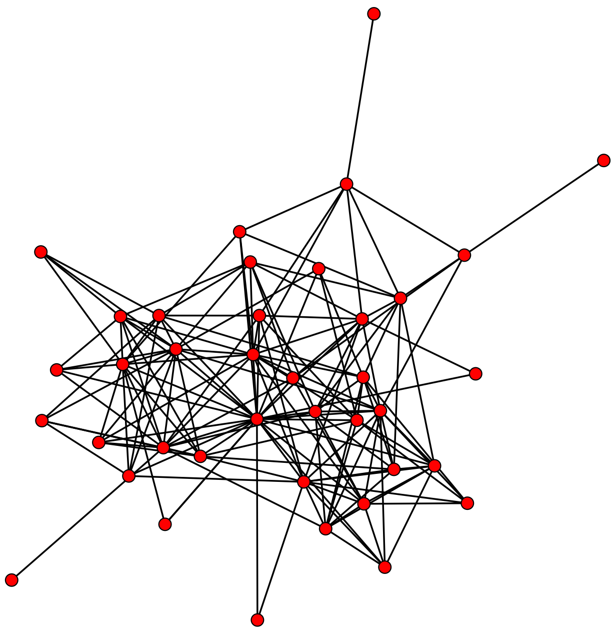


Zachary karate club network

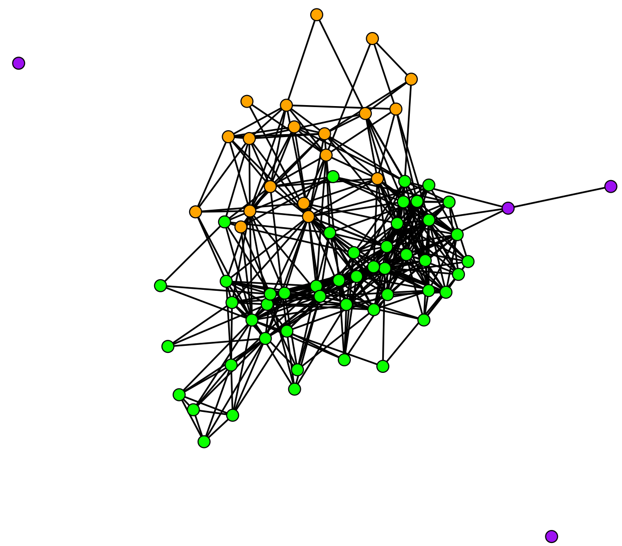


The Instructor (Mr. Hi) is green and the President (John A.) is purple.

Kapferer tailor shop network

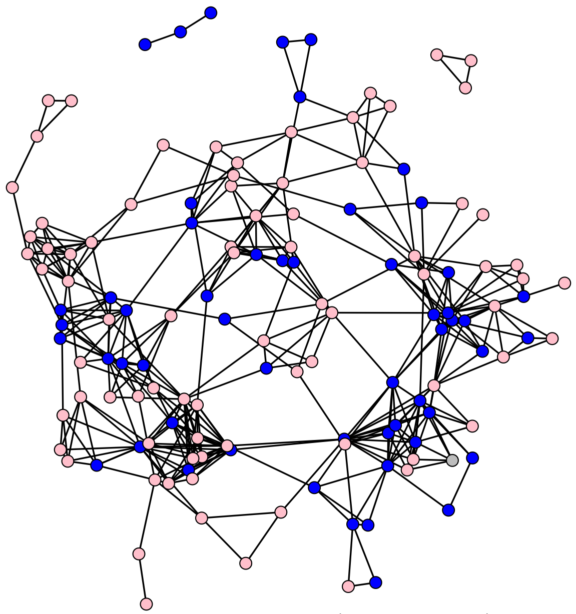


Lazega law firm friendship network



Nodes are colored according to the office the person works at.

High school friendship network

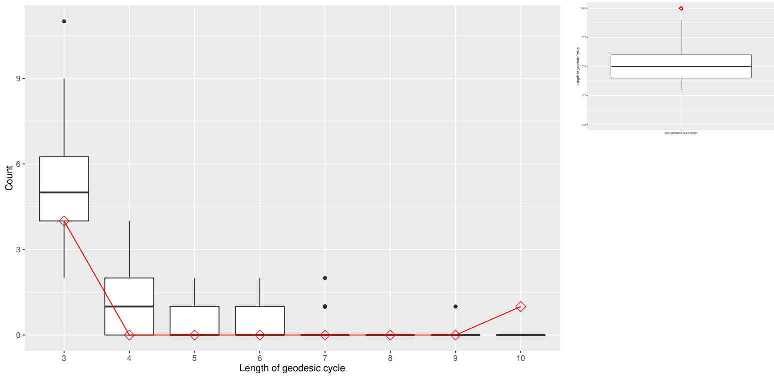


Male students are blue, female pink (one unknown is gray).

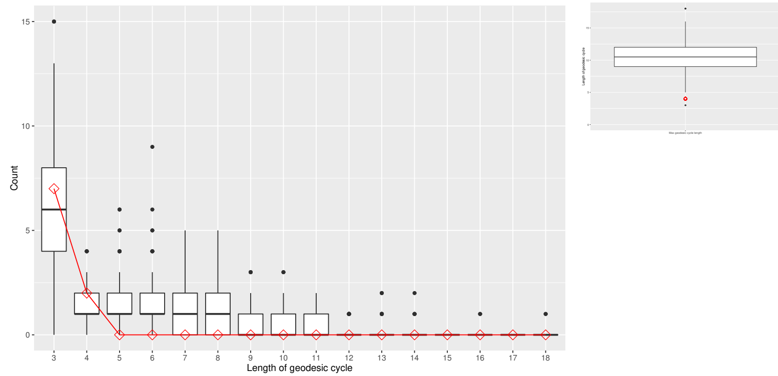
Network descriptive statistics

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

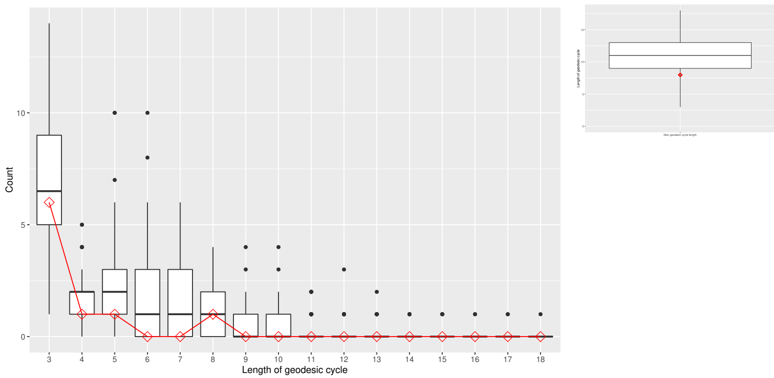
Patricia 1990 geodesic cycle length distribution



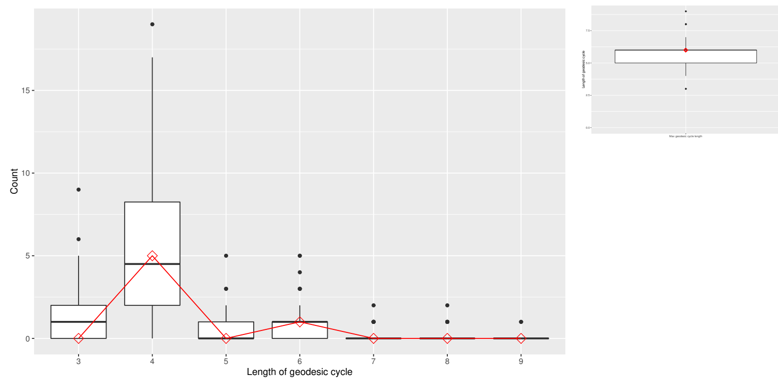
Patricia 1992 geodesic cycle length distribution



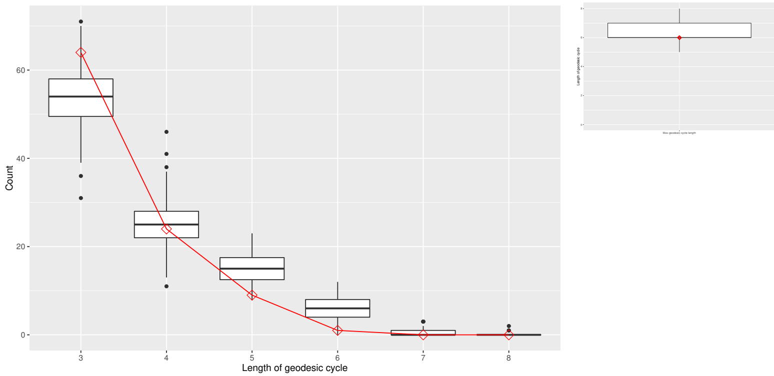
Patricia 1993 geodesic cycle length distribution



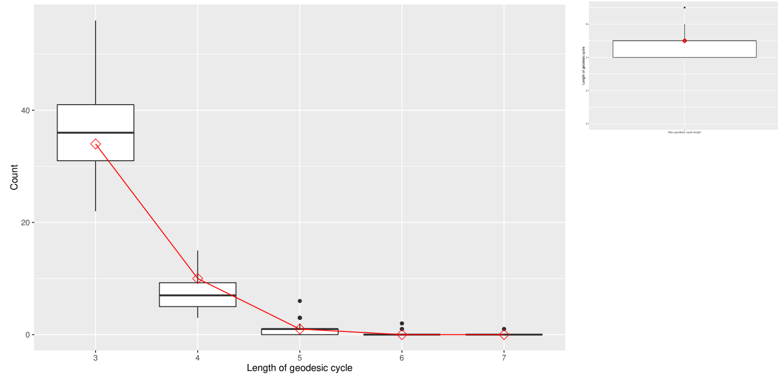
Grey's Anatomy sexual relations network geodesic cycle length distribution



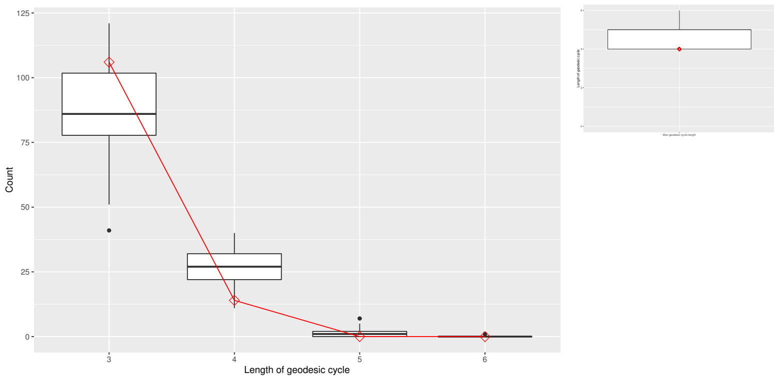
Dolphin social network geodesic cycle length distribution



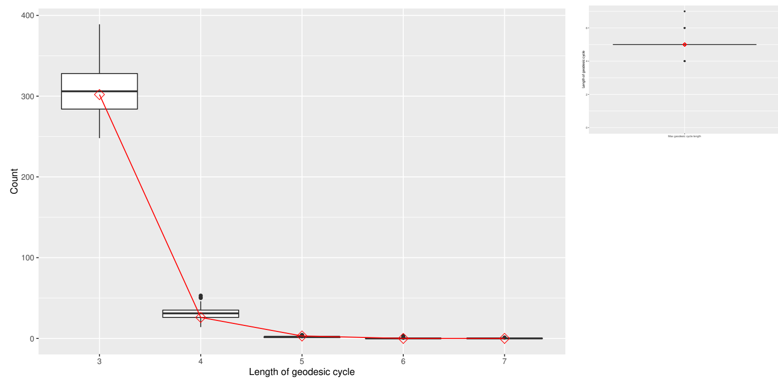
Zachary karate club network geodesic cycle length distribution



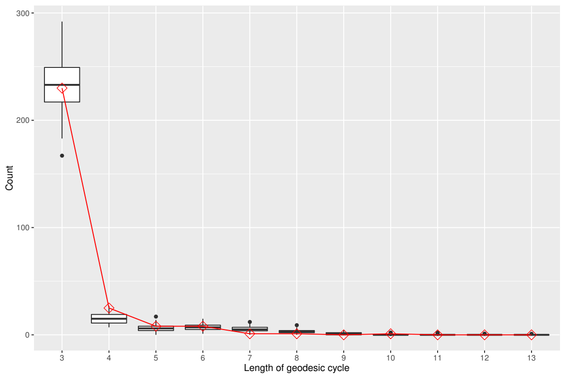
Kapferer tailor shop geodesic cycle length distribution

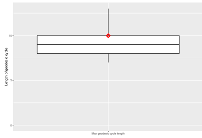


Lazega law firm friendship geodesic cycle length distribution



High school friendship geodesic cycle length distribution





Results

- Using a different model (ERGM rather than dk-series), and considering not just the largest geodesic cycle size but the distribution of geodesic cycle sizes,
- I confirm the results of Martin (2017) that Patricia's 1990 network has significantly larger geodesic cycle than expected, but that the largest geodesic cycle in the 1992 and 1993 networks is significantly smaller than expected.
- Further, in the latter two networks, in general geodesic cycles of size 4 or more are under-represented.
- But in a selection of other (fictional and real) social networks, the observed cycle length distribution is a good match to that expected from an ERGM model.
- So it appears there is something "special" about Patricia's networks in this respect.

Discussion

- But why is this? Perhaps because Patricia's networks are the product of a a single mind (multiple personalities aside), and hence reflect her (spatial) idea of what a network "should" look like.
- While the other networks are either empirical, or (the sole fictional network, Grey's Anatomy) the product of multiple creators (over an extended period).
- That ERGMs, explicitly on their local (social circuit) assumption, reproduce well the (non-local) geodesic cycle size distribution is an encouraging confirmation that local interactions really do produce these observed global structures.

Unpublished work

- As of November 2019 this is unpublished work.
- However these slides will be available from my website after the talk (ASNAC 2019, 28–29 November, Adelaide, South Australia).
- Details including methods, graph theoretic definitions, ERGM models, and full references are in the "hidden bonus slides" after this one.
- `https://sites.google.com/site/alexdstivala/home/conferences`

Hidden bonus slides

Introduction (2)

- Martin (2017) concludes from this that the logic of these networks is spatial, the long path lengths indicate a "large world" and 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, he concludes with the hypothesis that the root schema for social networks is local and spatial.
- I will re-examine these networks using ERGMs, make precise the "hollow ring" definition, and compare the empirical distribution of such cycle sizes with those expected under ERGM models of the networks.
- I will repeat this with several other empirical (human and animal social networks and fictional character networks) networks and compare the results with those for the MPD networks.

Data sources

"Patricia" David et al. (1996); Martin (2017). I manually coded the networks from Patricia's hand drawings as reproduced in David et al. (1996).

Grey's Anatomy Lind (2012); Weissman (2019); Leavitt and Clark (2014). ERGM models also based on those described in the citations.

Dolphins Lusseau et al. (2003), downloaded from `http://www-personal.umich.edu/~mejn/netdata/`.

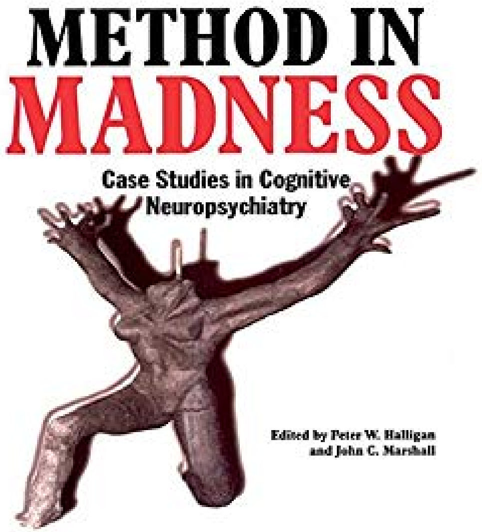
Lazega law firm Lazega and Pattison (1999); Lazega (2001); Snijders et al. (2006) downloaded from `http://moreno.ss.uci.edu/data.html`.

Zachary karate club Zachary (1977) via statnet.

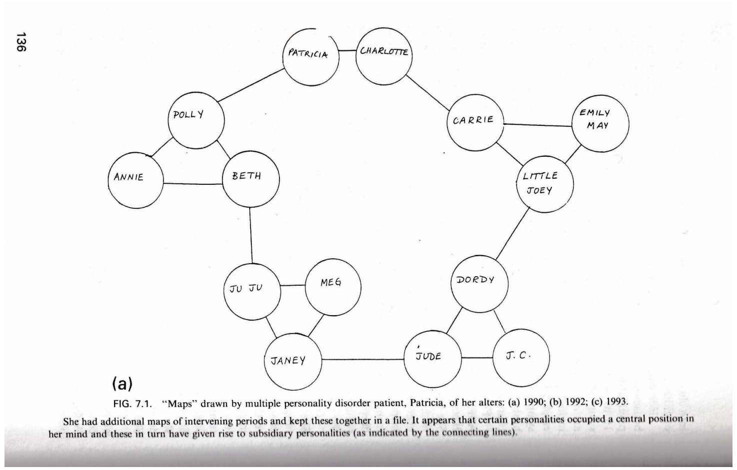
Kapferer tailor shop Kapferer (1972) via statnet. ERGM models are replications of those in Hummel et al. (2012).

High school friendship Mastrandrea et al. (2015) downloaded from `http://www.sociopatterns.org/datasets/high-school-contact-and-friendship-networks/`.

Halligan & Marhsall (Eds.) 1996

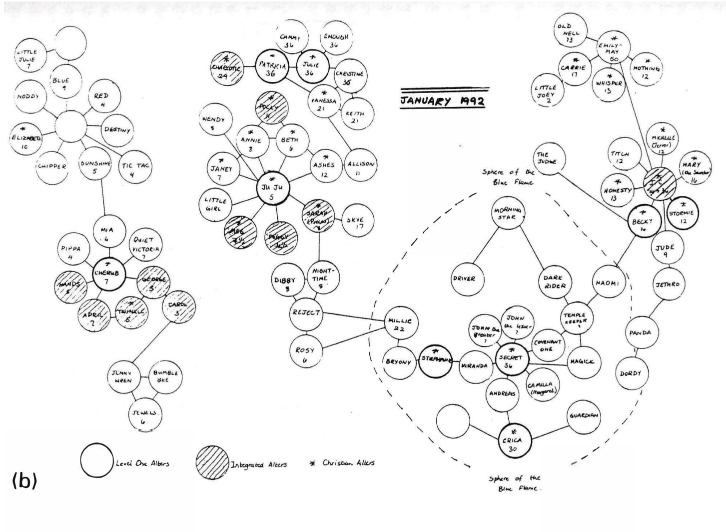


“Map” drawn by Patricia of her alters, 1990



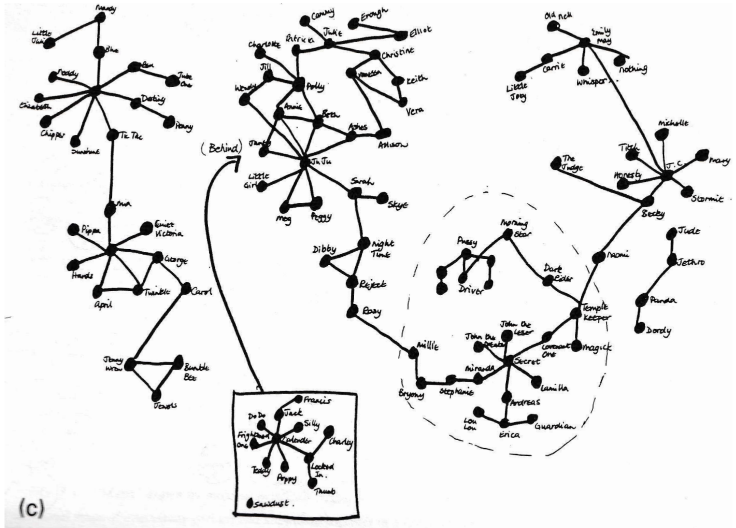
David et al. (1996, Fig. 7.1(a))

“Map” drawn by Patricia of her alters, 1992



David et al. (1996, Fig. 7.1(b))

“Map” drawn by Patricia of her alters, 1993



David et al. (1996, Fig. 7.1(c))

“Small world” statistics

Network	N_g	L_g	L_r	L_l	C^*_g	C^*_r	C^*_l	ω	SWI
Patricia 1990	14	2.96	2.66	3.10	0.47619	0.09690	0.37262	-0.377	0.464
Patricia 1992	62	8.99	4.83	8.46	0.12769	0.02325	0.16736	-0.226	—
Patricia 1993	66	9.73	5.27	9.93	0.09411	0.02000	0.14853	-0.092	0.025
Grey’s Anatomy	31	3.58	3.40	3.75	0.00000	0.04984	0.12373	0.949	—
Dolphins	62	3.36	2.71	4.22	0.25896	0.09749	0.53410	0.321	0.211
Zachary karate club	34	2.41	2.25	2.30	0.57064	0.35880	0.56298	-0.079	—
Kapferer tailor shop	39	2.04	1.99	2.03	0.45803	0.32829	0.42702	-0.099	—
Law firm friendship	69	2.19	2.01	2.14	0.49778	0.26779	0.53389	-0.011	—
High school friendship	128	4.02	2.83	5.90	0.54006	0.06319	0.64029	-0.140	0.505

- Small world coefficient ω (Telesford et al., 2011) ranges from -1 to 1, with values close to -1 indicating the graph has lattice characteristics, values close to 1 indicating random graph characteristics, and values near 0 indicating small world characteristics.
- Small world index (SWI) (Neal, 2017) ranges from 0, when the network displays neither of the small-world characteristics (relatively small shortest path lengths and relatively high clustering coefficient), to 1 when it has both characteristics.

Some graph theory definitions (always undirected graphs here)

- 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 non-empty trail in which the first and last vertices are repeated.
- A **cycle** (or **simple circuit**) is a circuit where only the first and last vertices are repeated.
- An **induced subgraph** of a graph is a subgraph formed by a subset of vertices of the graph and all edges connecting vertices in that subset.
- An **induced path** in G is a path that is an induced subgraph of G, i.e. 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.

Chordless cycles

- 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. a cycle with no chords.
- Also known as an **induced cycle** as (just as for an induced path) in a chordless cycle in 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.
- Also sometimes called a **hole**.
- The terms from the this and the previous slide are well-known, e.g. you can find them in Wikipedia, Wolfram MathWorld, computer science and discrete mathematics textbooks.

Geodesic or isometric cycles

- 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 Negami and Xu (1986).
- Note that although finding the largest cycle and largest chordless cycle in a graph are both $\mathcal{NP}$-hard problems (Garey and Johnson, 1979), finding the largest geodesic (isometric) cycle is in $\mathcal{P}$ (Lokshtanov, 2009).

Convex cycles I

- A cycle C of a graph G is **(geodetically) convex** if for any pair of distinct vertices $u, v \in V(C)$,

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

- 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) [not cited by Hellmuth et al. (2014)].

Convex cycles II

- 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 iff 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).

Atomic cycles

- 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.
- "Atomic cycle" with its definition above by Gashler and Martinez (2012) is referred to in the Wikipedia entry for "Induced path" but neither that entry nor Gashler and Martinez (2012) refer to (geodetic) convexity or "convex cycle", "isometric cycle", "geodesic cycle", or any of the literature on those concepts.

Chordless and geodesic cycles

- Every geodesic cycle is chordless.
- But not every chordless cycle is geodesic.
- A chordless cycle is an induced cycle. Indeed a cycle is chordless if and only if it is an induced cycle.

What's the frequency, Kenneth?

In this work:

- Cycles and chordless cycles are counted using the CYPATH program (Uno and Satoh, 2014).
- Geodesic cycles (atomic cycles) are counted using the algorithm of Gashler and Martinez (2012) implemented in the Waffles machine learning toolkit (Gashler, 2011).

Expected number of cycles in a random graph

- The expected number of cycles of length k ($3 \leq k \leq n$) in an Erdős-Rényi random graph $G(n, p)$ is

$$\mathrm{E}[\gamma_n(p, k)] = \binom{n}{k} \frac{(k-1)!}{2} p^k$$

  (Erdős and Rényi, 1960; Takács, 1988)
- On the probability (Łuczak, 1991) and size of the largest chordless cycle in a random graph see Łuczak (1993) [see also citations therein].
- Geodesic cycles in random graphs were studied in Benjamini et al. (2011). For the length of longest geodesic (isometric) cycles in random graphs see Li and Shi (2018) and for finding the longest such cycle in a graph, Lokshtanov (2009) [the former does not cite the latter].

ERGM introduction

- 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, etc.
- 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.
- I.e. it tells us about the process occurring significantly more or less than by chance, *given all the other effects in the model occurring simultaneously*.

ERGM details

- All ERGMs here were estimated with the statnet software (Handcock et al., 2008; Hunter et al., 2008; Handcock et al., 2016a,b), and convergence and goodness-of-fit found to be acceptable using the statnet MCMC diagnostics and gof functions.
- Unless otherwise noted, the statnet default MCMLE algorithm was used for estimation.
- For each network the "best" model (using AIC, BIC, and goodness-of-fit plots) was selected, and 100 networks simulated from that model using statnet.
- Geodesic cycle length distributions were computed from these simulated networks.

Patricia 1990 ERGM

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: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWESP α parameter is fixed at 0.

Patricia 1992 ERGM

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: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWDEGREE decay parameter is fixed at 0.5 and the GWESP α parameter is fixed at 0.

Patricia 1993 ERGM

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: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWDEGREE decay parameter is fixed at 0.5 and the GWESP α parameter is fixed at 0.

Grey's Anatomy sexual contacts ERGM

Effect	Model 1	Model 2	Model 3
Edges	−1.442 (0.241)***	−0.287 (0.564)	−0.844 (0.636)
Homophily Sex	−3.133 (0.718)***	−3.428 (0.741)***	−3.542 (0.732)***
Degree 1	2.026 (0.500)***	3.533 (1.058)***	3.393 (1.000)***
Degree 2		1.828 (0.911)*	1.743 (0.858)*
Degree 3		0.988 (0.805)	0.983 (0.769)
Heterophily Birth year		−0.132 (0.030)***	−0.142 (0.032)***
Attending – Attending		1.172 (0.508)*	1.085 (0.533)*
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)*	1.004 (0.502)*
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)**	4.611 (1.753)**
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)**
Homophily White			0.856 (0.392)*
AIC	302.10	278.40	271.30
BIC	316.70	365.80	368.40

Note: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

Dolphin social network ERGM

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

Note: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWDEGREE decay parameter is fixed at 0.5.
The “Stepping” algorithm (Hummel et al., 2012) was used for estimation.

Lazega law firm friendship network ERGM

Effect	Model 1
Edges	$-5.256\ (0.317)^{***}$
GWDEGREE	$1.290\ (0.874)$
GWESP	$0.597\ (0.072)^{***}$
GWESP α	$1.398\ (0.030)^{***}$
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: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

Zachary karate club ERGM

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: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWDEGREE decay parameter is fixed at 0.2.
Faction categorical attribute is coded from -2 (strongly Mr. Hi's) to +2 (strongly John's).

Kapferer tailor shop ERGM

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

Note: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWDEGREE decay parameter is fixed at 0.25.
The "Stepping" algorithm (Hummel et al., 2012) was used for estimation.

High school friendship ERGM

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

Note: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$. par

In models 2 and 3 the GWDEGREE decay parameter is fixed at 1.7.

Sexual contact networks and 4-cycles I

- Bearman et al. (2004) propose a normative proscription against 4 cycles ("don't date your old partner's current partner's old partner"), based on low number of 4-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 GoF that small numbers of 4-cycles are generated.
- The (much smaller) fictional Grey's Anatomy sexual contact network contains seven 4-cycles (of which all 7 are chordless and 5 are geodesic) — violating this proposed proscription could be because it makes compelling entertainment (Lind, 2012).
- Could not find a converged ERGM with a 4-cycle term for this network to explicitly test for over- or under-representation (neither could Lind (2012)).

Sexual contact networks and 4-cycles II

- But note the model fits geodesic cycles of length 4 (and also cycles and chordless cycles of length 4) well, similar to the case for 4-cycles in Rolls et al. (2015) for the Bearman et al. (2004) network.

Future work I

- I wanted to compare to other fictional networks, with a single author, specifically to test the hypothesis that such cases would also produce anomalous geodesic cycle length distributions like Patricia's.
- However fitting ERGMs to these (the ones I have been able to acquire anyway) is proving difficult.
- There is usually a main character, who is linked to most of the other characters, which causes problems for ERGM.
- Note that this is not the case in Patricia's networks: "Patricia" is not the star of her own story in relation to her alters.
- This can be handled by fixing the ties for that character, but I have still not been able to get good ERGMs for the data I have available (Les Misérables, David Copperfield, Anna Karenina, Huckleberry Finn).

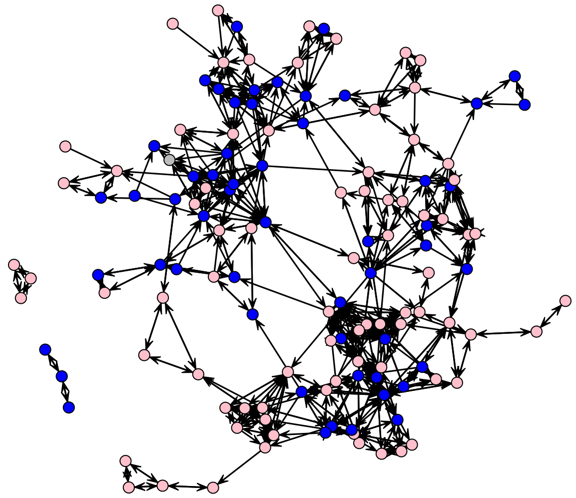
Future work II

- This difficulty with ERGMs (having to only use data for which a model fit can be obtained) is one reason Martin (2017) gives for using dk-series instead (although I was able to fit all 3 Patricia networks well with ERGM, and more besides).
- ERGMs also have the advantage of allowing node attributes, while dk-series is purely structural.
- Despite "little evidence of structural implications of these characteristics" (Martin, 2017, p. 8) I found statistically significant effects of some of them using ERGM (although the resulting simulated networks are indeed not too different wrt the geodesic cycle size distributions).

Future work III

- It would also be interesting to test the geodesic cycle length distribution of the Add Health high school romantic network (Bearman et al., 2004). 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 (recoded manually from the figure in Bearman et al. (2004)).

High school friendship network (original, directed)



High school friendship directed ERGM

Effect	Model 1	Model 2
Edges	$-5.253\ (0.240)^{***}$	$-5.295\ (0.247)^{***}$
GW out-degree	$1.612\ (0.751)^{*}$	$1.516\ (0.756)^{*}$
GW in-degree	$3.058\ (1.209)^{*}$	$2.915\ (1.132)^{*}$
Reciprocity	$4.888\ (0.332)^{***}$	$5.055\ (0.372)^{***}$
GWDSP ($\alpha = 1.4$)	$-0.121\ (0.017)^{***}$	$-0.130\ (0.017)^{***}$
GWESP ($\alpha = 0.7$)	$1.156\ (0.064)^{***}$	$1.148\ (0.063)^{***}$
Homophily Class	$1.378\ (0.164)^{***}$	$1.350\ (0.166)^{***}$
Reciprocity Class	$-1.722\ (0.360)^{***}$	$-1.671\ (0.364)^{***}$
Homophily Sex		$0.246\ (0.168)$
Reciprocity Sex		$-0.336\ (0.364)$
AIC	2571.00	2564.00
BIC	2633.00	2642.00

Note: *** $p < 0.0001$, ** $p < 0.001$, * $p < 0.05$.

The GWIDEGREE and GWODEGREE decay parameters are fixed at 0.2 The MCMC burnin parameter is set to 1×10^5 and the MCMC interval parameter to 16384.

References I

L. M. Batten. Geodesic subgraphs. *Journal of Graph Theory*, 7(2):159–163, 1983.

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

I. Benjamini, C. Hoppen, E. Ofek, P. Prałat, and N. Wormald. Geodesics and almost geodesic cycles in random regular graphs. *Journal of Graph Theory*, 66(2):115–136, 2011.

A. David, R. Kemp, L. Smith, and T. Fahy. Split minds: Multiple personality and schizophrenia. In P. W. Halligan and J. C. Marshall, editors, *Method in Madness: Case Studies in Cognitive Neuropsychiatry*, chapter 7, pages 122–146. Psychology Press, Hove, UK, 1996.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–60, 1960.

M. Farber. Bridged graphs and geodesic convexity. *Discrete Mathematics*, 66(3):249–257, 1987.

M. Farber and R. E. Jamison. Convexity in graphs and hypergraphs. *SIAM Journal on Algebraic Discrete Methods*, 7(3):433–444, 1986.

S. Freud. Extracts from the Fliess papers. In J. Strachey, editor, *The standard edition of the complete works of Sigmund Freud*, volume 1, pages 173–280. Hogarth Press and the Institute of Psychoanalysis, London, 1895/1950.

M. R. Garey and D. S. Johnson. *Computers and intractability: a guide to the theory of NP-completeness*. Freeman, San Francisco, 1979.

M. Gashler. Waffles: A machine learning toolkit. *Journal of Machine Learning Research*, 12(Jul):2383–2387, 2011.

M. Gashler and T. Martinez. Robust manifold learning with CycleCut. *Connection Science*, 24(1):57–69, 2012.

M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, Morris, and Martina. statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1): 1548, 2008. URL `http://www.jstatsoft.org/v24/i01`.

M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, P. N. Krivitsky, S. Bender-deMoll, and M. Morris. *statnet: Software Tools for the Statistical Analysis of Network Data*. The Statnet Project (`http://www.statnet.org`), 2016a. URL `CRAN.R-project.org/package=statnet`. R package version 2016.4.

References II

M. S. Handcock, D. R. Hunter, C. T. Butts, S. M. Goodreau, P. N. Krivitsky, and M. Morris. *ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks*. The Statnet Project (http://www.statnet.org), 2016b. URL http://CRAN.R-project.org/package=ergm. R package version 3.6.0.

M. Hellmuth, J. Leydold, and P. F. Stadler. Convex cycle bases. *Ars Mathematica Contemporanea*, 7(1):123–140, 2014.

R. M. Hummel, D. R. Hunter, and M. S. Handcock. Improving simulation-based algorithms for fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4):920–939, 2012.

D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3):1–29, 2008. URL https://www.jstatsoft.org/v024/i03.

B. Kapferer. *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*. Manchester University Press, 1972.

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

E. Lazega and P. E. Pattison. Multiplexity, generalized exchange and cooperation in organizations: a case study. *Social Networks*, 21(1):67–90, 1999.

A. Leavitt and J. Clark. SNA in R workshop. https://github.com/alexleavitt/SNAinRworkshop, March 2014. Accessed 12 July 2019.

Y. Li and L. Shi. Geodesic cycles in random graphs. *Discrete Mathematics*, 341(5):1275–1281, 2018.

B. Lind. 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/, September 2012. Accessed 12 July 2019.

D. Lokshtanov. Finding the longest isometric cycle in a graph. *Discrete Applied Mathematics*, 157(12):2670–2674, 2009.

T. Łuczak. Holes in random graphs. *Discrete Applied Mathematics*, 30(2-3):213–219, 1991.

T. Łuczak. The size of the largest hole in a random graph. *Discrete Mathematics*, 112(1-3):151–163, 1993.

References III

D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

J. L. Martin. The structure of node and edge generation in a delusional social network. *Journal of Social Structure*, 18(1):1–21, 2017. doi: 10.21307/joss-2018-005.

R. Mastrandrea, J. Fournet, and A. Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE*, 10(9):e0136497, 2015.

Z. P. Neal. How small is it? comparing indices of small worldliness. *Network Science*, 5(1):30–44, 2017.

S. Negami and G.-H. Xu. Locally geodesic cycles in 2-self-centered graphs. *Discrete Mathematics*, 58(3):263–268, 1986.

C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguná, G. Caldarelli, et al. Quantifying randomness in real networks. *Nature Communications*, 6:8627, 2015.

D. A. Rolls, P. Wang, E. McBryde, P. Pattison, and G. Robins. A simulation study comparing epidemic dynamics on exponential random graph and edge-triangle configuration type contact network models. *PLoS ONE*, 10 (11):e0142181, 2015.

T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.

L. Takács. On the limit distribution of the number of cycles in a random graph. *Journal of Applied Probability*, 25 (A):359–376, 1988.

Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette, and P. J. Laurienti. The ubiquity of small-world networks. *Brain Connectivity*, 1(5):367–375, 2011.

T. Uno and H. Satoh. An efficient algorithm for enumerating chordless cycles and chordless paths. In *International Conference on Discovery Science*, pages 313–324. Springer, 2014.

G. Weissman. Grey's Anatomy network of sexual relations. https://gweissman.github.io/post/grey-s-anatomy-network-of-sexual-relations/, April 2019. Accsessed 12 July 2019.

W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 33(4):452–473, 1977.