A new scalable implementation of the citation exponential random graph model (cERGM) and its application to a large patent citation network

Alex Stivala[1] Alessandro Lomi[1,2]

[1]Università della Svizzera italiana, Lugano, Switzerland
[2]The University of Exeter Business School, Exeter, UK

INSNA Sunbelt XLII, Cairns, Australia (hybrid in-person/online), July 12–16, 2022

Introduction

- The *citation exponential random graph model* (cERGM) was recently described by Schmid et al. (2021),
- where it was used to analyze the United States Supreme Court citation network.
- In this work we will describe a new implementation of the cERGM, in the EstimNetDirected software (Stivala et al., 2020), which uses the "equilibrium expectation" (EE) algorithm (Byshkin et al., 2016, 2018; Borisenko et al., 2020) to allow the analysis of networks far larger than originally possible.
- We will validate the implementation by comparing to the original cERGM results on the US Supreme Court data,
- and describe a new goodness-of-fit (GoF) procedure for cERGM.
- We will use the new cERGM implementation to analyze a citation network of nearly 2 million European patents.

Citation ERGM (cERGM) motivation (Schmid et al., 2021)

The cERGM accounts for specific structural constraints of citation networks:

1. Citation networks are partially acyclic
   - For Supreme Court citations, cases decided in the same term may cite each other.
   - And any case can cite a case in any earlier term; however it cannot cite a case in a future term.
2. New citation arcs can only be created over time (and not deleted).
3. The node set must increase for new citation arcs to be created
   - New arcs (and non-arcs) are created by the addition of new nodes.

cERGM estimation (Schmid et al., 2021)

- The cERGM is estimated by estimating each term separately
- with the network consisting of nodes in the current and all earlier terms
- and constraints so that citations can only be to nodes in current or earlier terms,
- and with ties sent from nodes in earlier terms all fixed.
- In other words, for each term, the model is of citations *from* nodes in the current term, *conditional* on the observed ties from all nodes in earlier terms.
- It is implemented in an R package using statnet.
- It was necessary to use simulated annealing to obtain networks with the same sufficient statistics (but weaker dependence among unfixed ties) as the observed graph (Schmid and Hunter, 2020; Krivitsky et al., 2022) to get starting points with MPLE for the MCMLE estimations to converge.

Data

Data sets

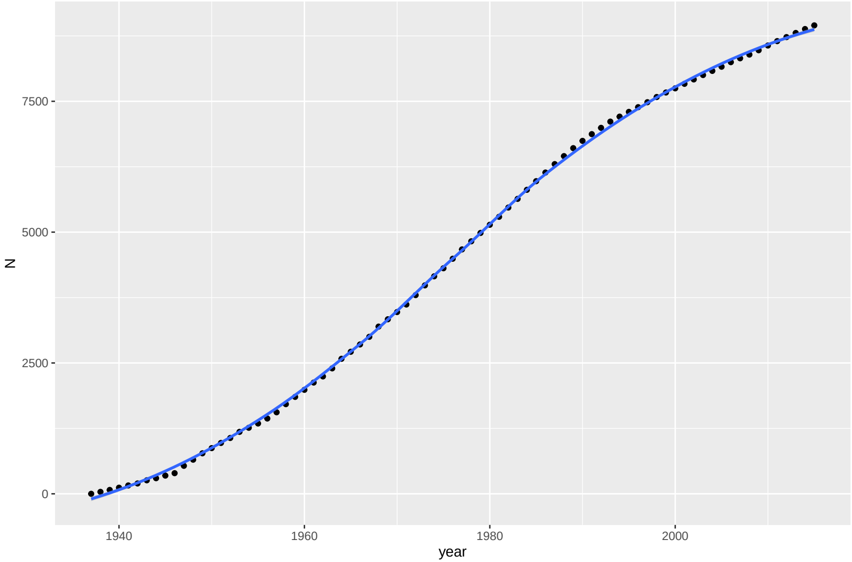
We will use two data sets for demonstration:

1. United States Supreme court citation network
   - As used in the original cERGM paper (Schmid et al., 2021)
   - Data from 1937 – 2015: 9 033 cases
   - Analysing 66 Supreme Court terms, 1950 – 2015
2. European Patent Office (EPO) patent citation network
   - As used in our earlier work (Stivala and Lomi, 2020)
   - 1 933 231 node citation network of patents 1976 – 2008
   - We will analyse 15 time periods from 1984 – 2005
   - (details on following slides)

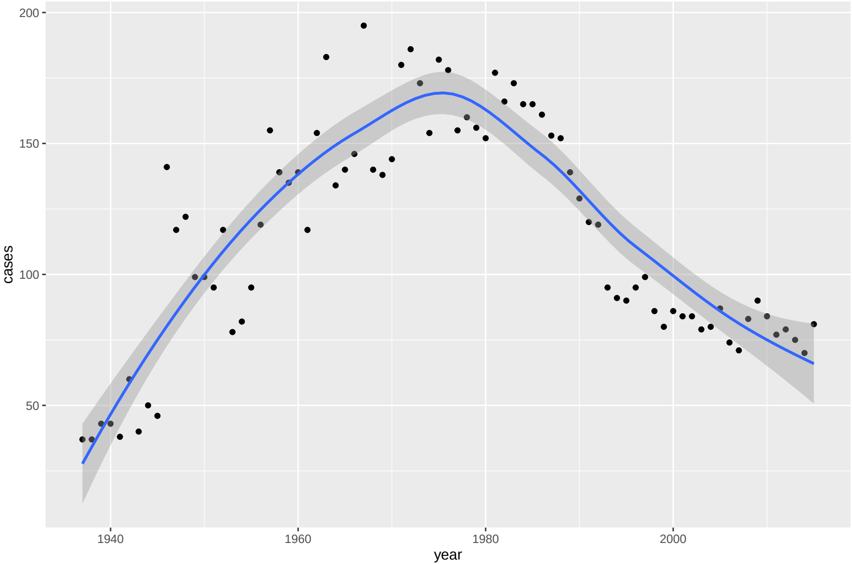
Terms and time periods

- The US Supreme Court sits in “terms” from October to October the following year.
- So the year is a meaningful unit for cERGM analysis: cases in a term can cite other cases in the same term (as well as earlier terms).
- However there is no such meaningful “term” for patent data.
- We instead construct time periods so that there is an approximately equal number of patent applications per time period
  - We construct the time periods so that there are no more than 150 000 applications per time period
  - This limit is chosen as it is larger than the number of applications in 2005, the last year before artifacts due to data set truncation.

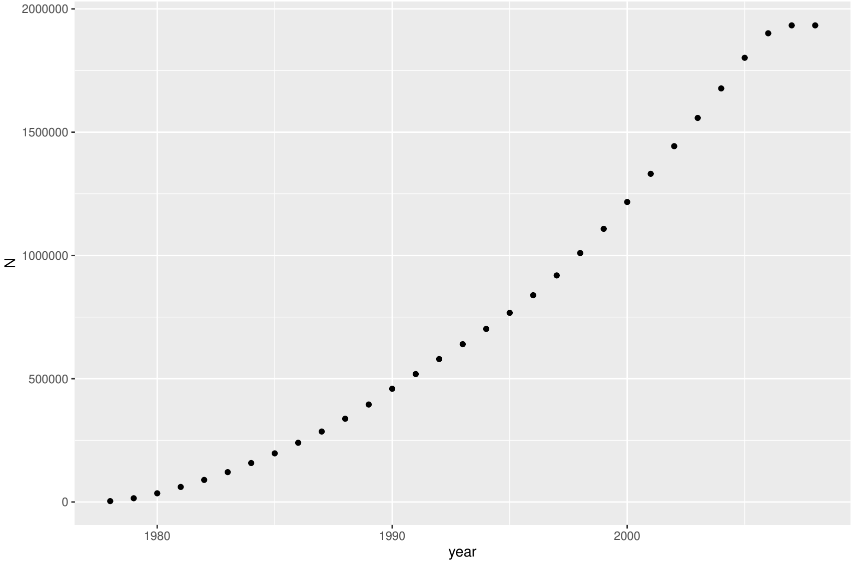
Total number of Supreme Court cases by year



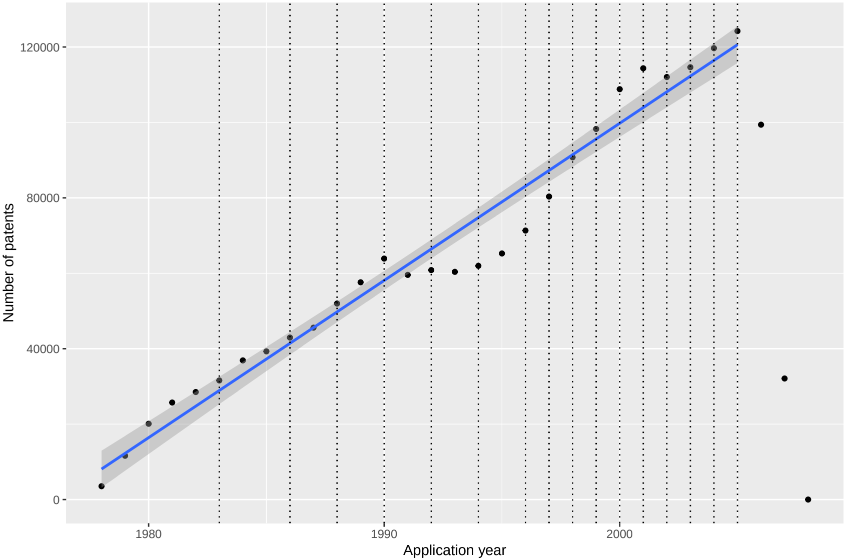
Number of Supreme Court cases per term



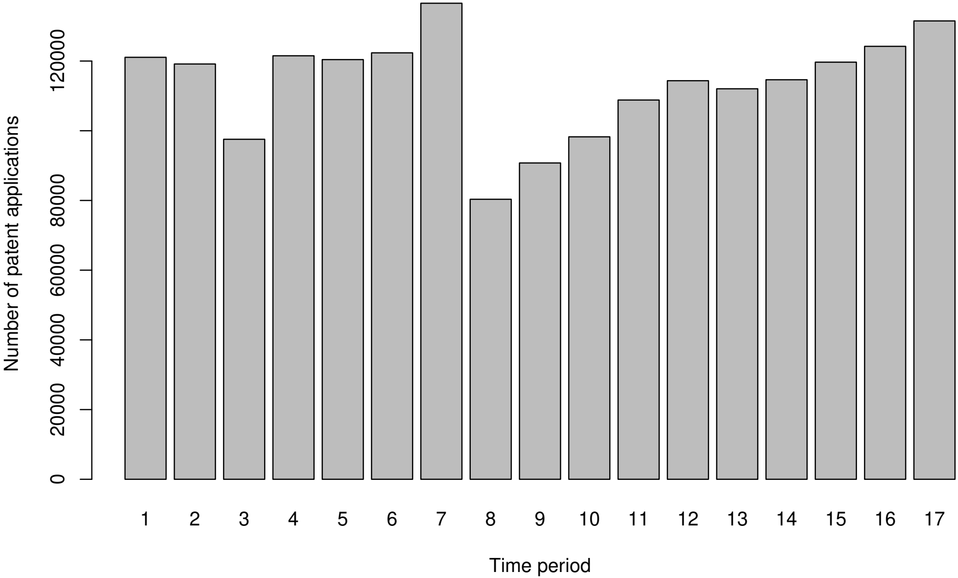
Total number of patents by year



Number of patent applications per year



Patent applications per time period



Validation on Supreme Court data

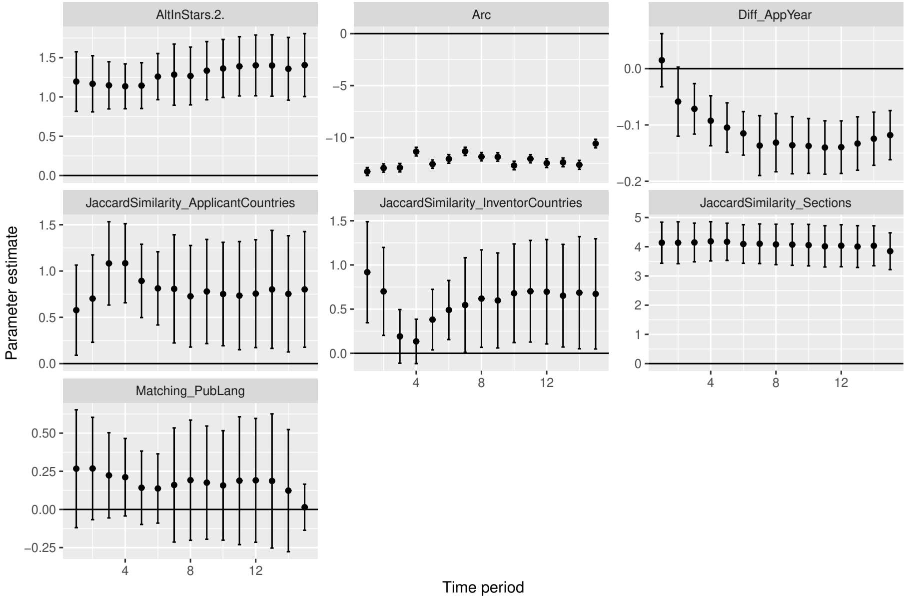
- Comparing to the results from the original cERGM implementation (Schmid et al., 2021), the signs of the parameters are the same and magnitudes are similar to the original (allowing for model differences).
- Simulation-based GoF tests show similar results (we will see these later)
- So we have some confidence that it is working correctly.
- However the estimated confidence intervals are much larger than that of the original statnet cERGM implementation.
  - Resulting in most parameters being not statistically significant, unlike the original.
- The EE algorithm makes less efficient use of the data, and is more suitable for (much) larger data sets,
- Like the patent data will see next
  - Which has $\sim$ 2 million nodes total, with $\sim$ 100 000 per time period
  - Rather than $\sim$ 10 000 nodes total, with $\sim$ 100 per time period.

cERGM results: EPO Patents

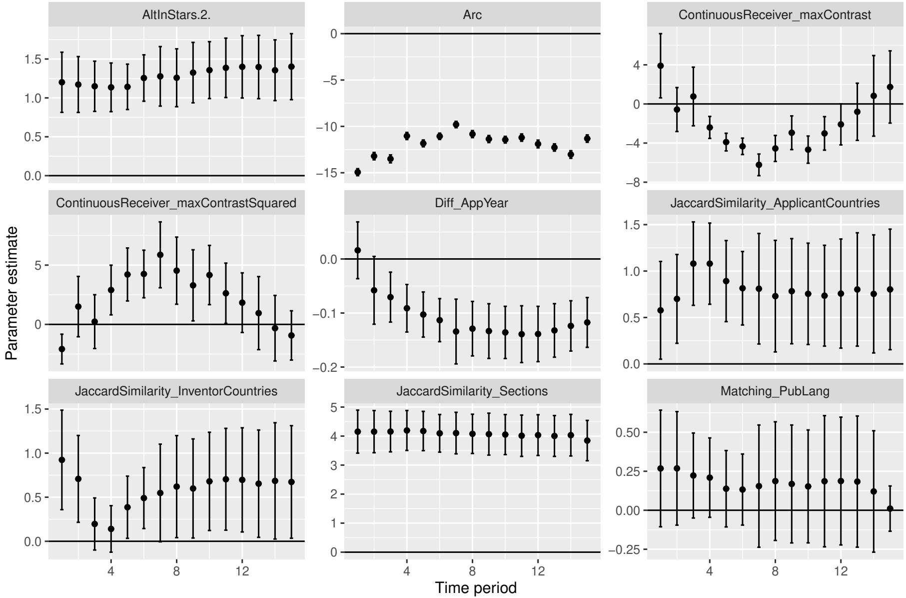
Contrast

- We will use the idea of *categorical contrast* (Hannan et al., 2007; Kovács and Hannan, 2010, 2015).
- The *contrast* of a category captures the idea of sharpness or fuzziness of category boundaries:
  - A category has high contrast (sharp boundaries) if it is seldom assigned low or moderate levels of category membership.
  - A category has lower contrast (fuzzier boundaries) as partial membership is more common.
- A technology class that is seldom assigned together with other classes to a patent has high contrast.
- A technology class that is frequently assigned together with other classes to a patent has low contrast.

cERGM estimates for patent data: simple model



cERGM estimates for patent data: model with max. contrast



Patent cERGM results (1)

Comparing to the ERGM (entire network) and negative binomial results from Stivala and Lomi (2020):

- AltInStars (popularity spread) positive and significant in all time periods, also in the ERGM.
  - Could indicate preferential attachment.
  - The magnitude seems to slightly increase over time.
  - But could this slight increase be an artifact of the tendency for mean in-degree to decrease over time (less time to accumulate citations) together with the tendency to cite more recent (rather than older) patents?
- Jaccard similarity on technology sections and classes is positive and significant in all time periods, also in the ERGM.
  - Patents are more likely to cite those in the same technology section and class.
- Abs. difference in application year is negative and significant in almost all time periods (not the first), also in the ERGM.
  - Patents are more likely to cite recent rather than older patents.
  - Magnitude seems to increase then level out. Not obvious why (and does not happen with the Supreme Court citations).

Patent cERGM results (2)

- Jaccard similarity on both applicant and inventor countries is positive and significant in (almost) all time periods, also in the ERGM.
  - (Geographical knowledge spillover hypothesis): citations are more likely to be geographically localized.
  - But why in time periods 3 and 4 does the effect for inventor become smaller (and not significant) and for applicant slightly larger?
- Matching publication language is always positive, but not significant in any time period.
  - The magnitude seems to slightly decrease over time.
  - In the full ERGM, it is positive and significant in most models (not significant in one).

Patent cERGM results (3)

- Max. contrast receiver effect is negative and significant in most time periods, but positive or not significant near start and end periods, while max. contrast squared is positive and significant in most time periods, and negative or not significant near start and end periods.
  - For the ERGM and negative binomial models, we found that effect for max. contrast is negative, and for max. contrast squared it is positive. There is a quadratic relationship between success and max. contrast, with success decreasing with max. contrast up to a point, but increasing thereafter.
  - But with the cERGM, it seems this relationship is true only in the "middle" of the period studied, reversing at the start and end. (Max. contrast over time looks like a $\cup$, and max. contrast squared, a $\cap$.)
- This is an odd result, perhaps some kind of artifact, but we cannot yet explain it...

cERGM Goodness-of-fit

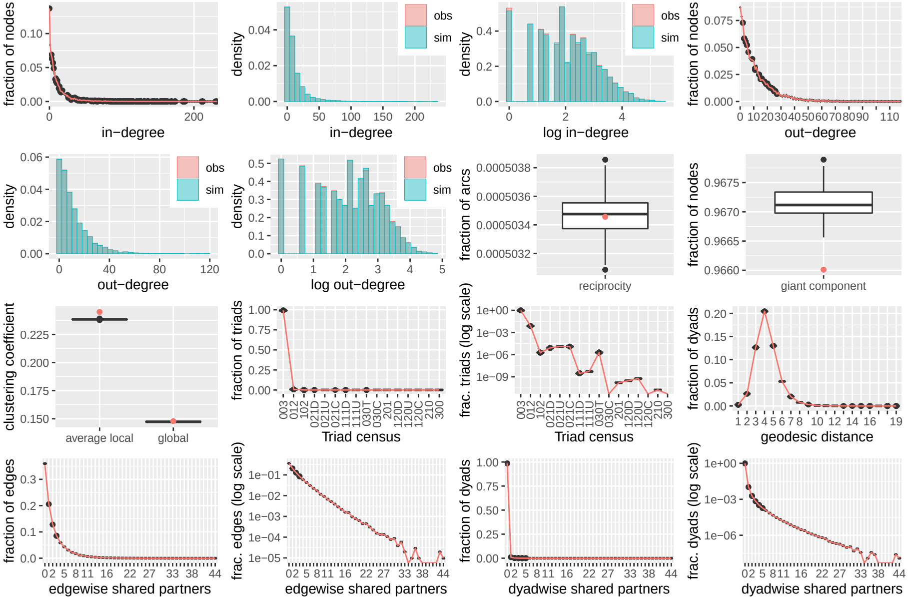
Why can't we use the usual statnet GoF plots?

- Because most of the network is fixed: only ties sent from the current term are modeled.
- So the usual global network statistics are dominated by the contribution of the fixed ties — no matter how badly the model fits the data, we will see a "good" GoF plot.
- Schmid et al. (2021) solve this problem by using only statistics "local" to the ties sent from the current term (the unfixed ties): specifically degree distributions and ESP distributions.
- But this is rather restricted compared to the usual GoF, where we usually include geodesic distance distribution, triad census, and more.
- (Examples on the following slides to demonstrate this.)

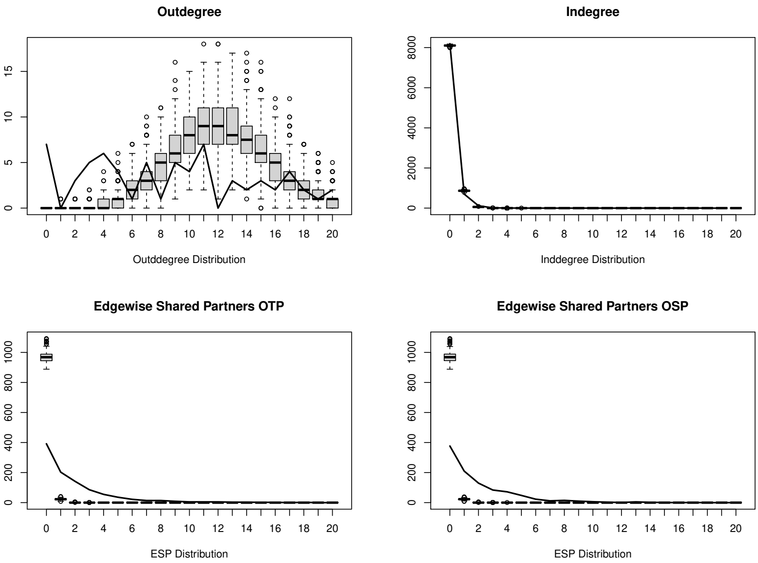
New method for cERGM GoF

- Consider instead (for the observed network and each network simulated from the model),
- the subgraph induced by the union of the nodes in the current term, and all nodes in other (earlier) terms that receive ties from them.
- These subgraphs are likely far smaller than the full graph, and are not dominated by fixed ties (that therefore do not say anything about model fit).
- These networks will not all be the same size, so in the GoF we always plot fractions of nodes (or dyads), not absolute numbers on the y axis.
- We also plot the distribution of the size of the induced subgraphs (and as for all GoF plots, ideally the observed is central in the distribution).
- (Examples on the following slides to demonstrate this.)

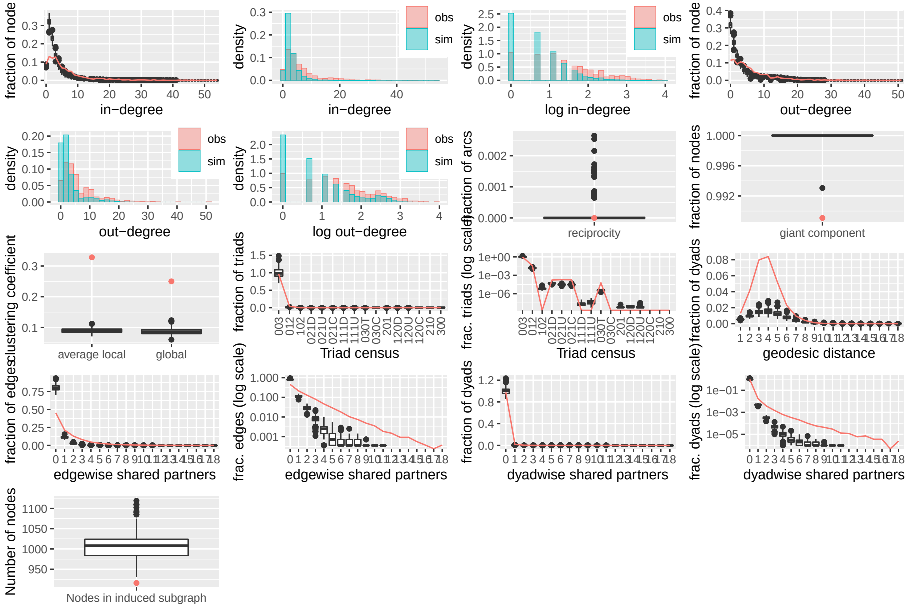
ERGM GoF plot for overly simple Supreme Court model (2015)



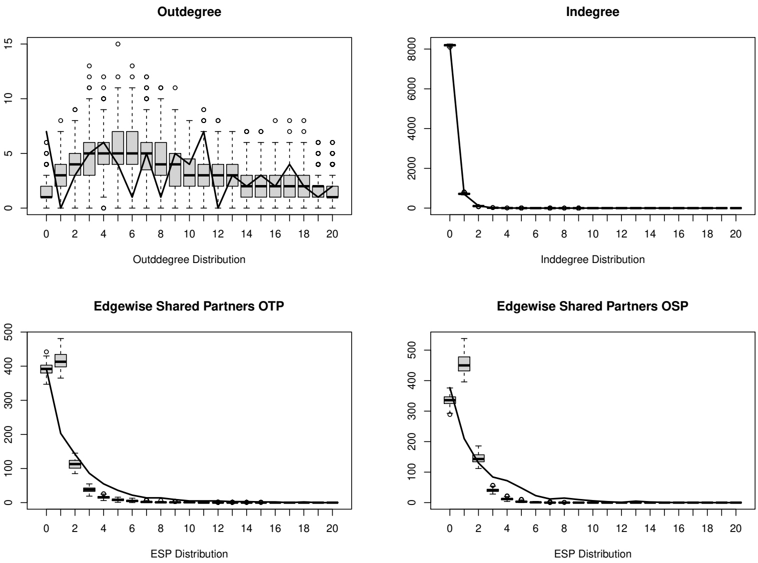
Original cERGM GoF plots for overly simple Supreme Court model (2015)



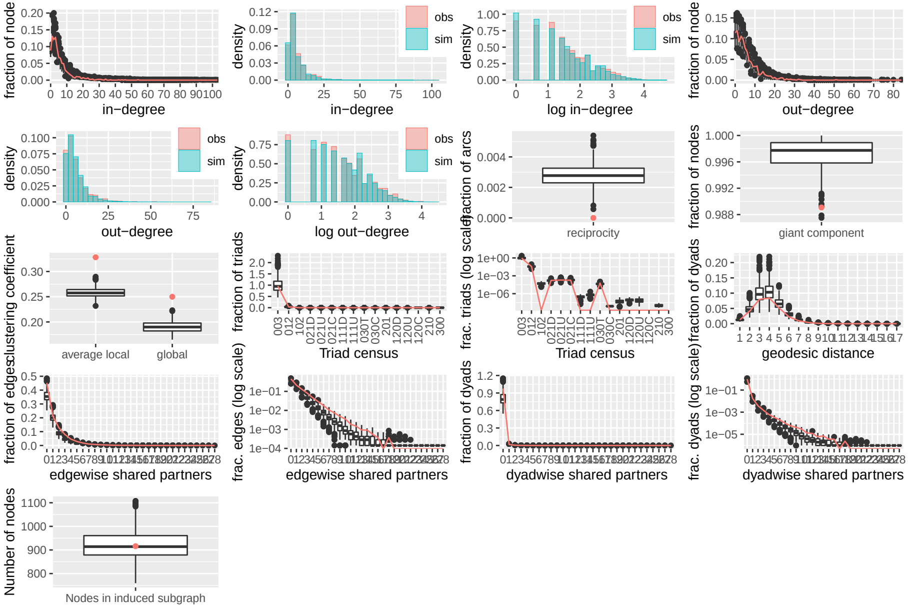
New cERGM GoF plots for overly simple Supreme Court model (2015)



Original cERGM GoF plots for good Supreme Court model (2015)



New cERGM GoF plots for good Supreme Court model (2015)



Conclusion (1)

- The new EE implementation of the cERGM allows estimation of cERGM parameters for networks with over 1 million nodes, and over 100 000 per term (time period)
- with practical time and memory usage (in the patent dataset, 15 minutes elapsed time with 20 cores with 16 GB each, per term).
- However for small data sets, its estimated confidence intervals are much larger than that of the original statnet cERGM implementation.
- So the original statnet cERGM implementation should be used where possible — the MCMCMLE algorithms it uses make more efficient use of the data
  - However even for the Supreme Court network (on the order of 10 000 nodes in total, with less than 200 per term) its resource usage is high ($\sim$ 5 hours and over 64 GB per term, even with ergm 4.1.2)
  - So it is unlikely to be practical for much larger data sets.

Conclusion (2)

- The new simulation-based cERGM goodness-of-fit test allows the use of global network statistics (such as geodesic distribution and clustering coefficients)
- Rather than just local statistics (degree distribution, ESP) for nodes in the last term
- It is practical for cERGM for networks as large as the patent data set studied here — which is not the case for an ERGM for the full network.

Acknowledgments

- This work was funded by Swiss National Science Foundation project numbers 167326 and 200778.
- Thanks to Christian Schmidt for maintaining the cERGM code, and incorporating my changes.
- We used the high performance computing cluster at the Institute of Computational Science, Università della Svizzera italiana, for all data processing and statistical computations.

Software availability

- Original cERGM (Schmid et al., 2021) R package (statnet):
  `https://github.com/schmid86/cERGM/`
- The new large network (EE algorithm) implementation, and R and Python scripts for new cERGM GoF plots:
  `https://github.com/stivalaa/EstimNetDirected`
  - Despite the name, it works with directed and undirected networks.
  - (but since citations are inherently directed, cERGM is for directed networks).

Unpublished work

- This is unpublished work (as of July 2022).
- More details, and references, are in the “hidden bonus slides” after this one.
- I will make these slides available on my website:
- `https://sites.google.com/site/alexdstivala/home/conferences`

Hidden bonus slides

Related work

- For previous applications of ERGM to citation networks, see An and Ding (2018); McLevey et al. (2018); Fanelli and Glänzel (2013); Dubnjakovic (2016); Peng (2015),
- for hyperlink networks, Lusher et al. (2013); Ackland and Shorish (2014),
- and for patent citation networks specifically, Stivala et al. (2019); Chakraborty et al. (2020); Stivala and Lomi (2020).
- The cERGM should also be contrasted to temporal ERGMs of various kinds (Butts, 2022).
- We will compare our new cERGM results on a European patent citation network with some previously presented using other techniques on the same data set (Stivala and Lomi, 2020).

Supreme Court cERGM models (2015)

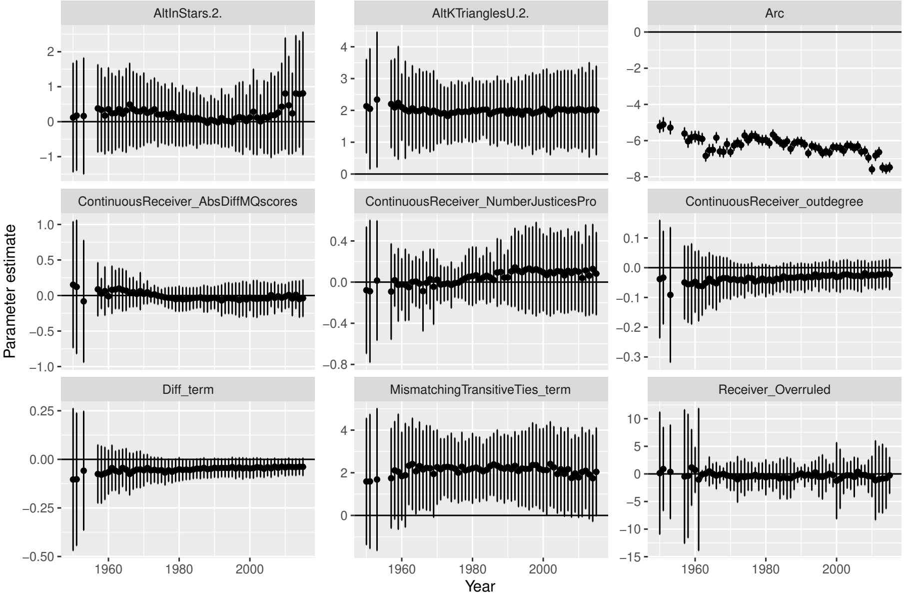
	Model 1	Model 2
Edges	$-6.26\ (0.04)^{***}$	$-6.92\ (0.18)^{***}$
GW indegree ($\alpha = 1$)	$-2.07\ (0.16)^{***}$	$-1.34\ (0.17)^{***}$
Receiver outdegree		$-0.02\ (0.00)^{***}$
Different term transitivity		$2.23\ (0.11)^{***}$
GWESP OSP ($\alpha = 0.15$)		$2.81\ (0.09)^{***}$
Year difference		$-0.04\ (0.00)^{***}$
Receiver Ideological breadth		$-0.05\ (0.02)^{*}$
Receiver Number justices in majority		$0.07\ (0.03)^{**}$
Receiver Overruled		$-0.13\ (0.26)$
AIC	14925.13	10682.63
BIC	14948.14	10786.15
Log Likelihood	-7460.57	-5332.31

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

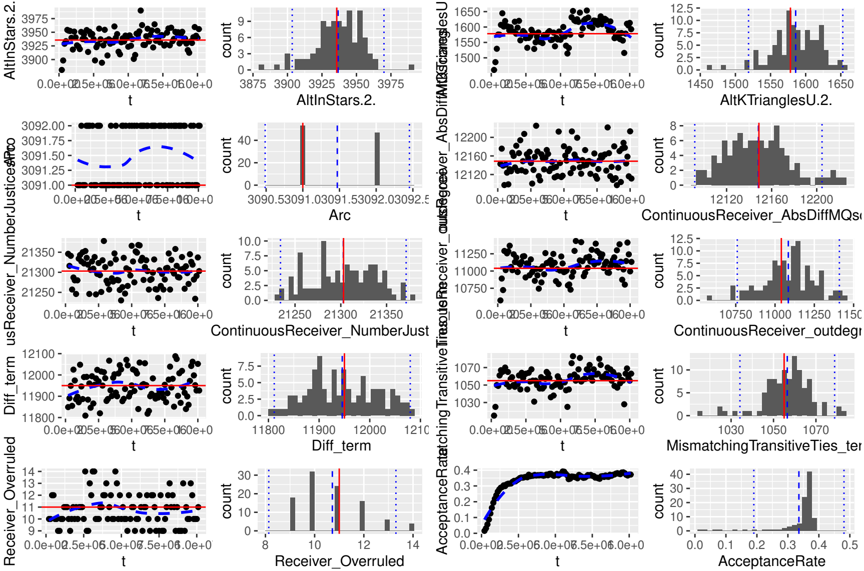
Supreme Court cERGM models (new EE implementation)

Effect	1950	2015
Arc	−5.214 (−5.549,−4.879)	−7.477 (−7.734,−7.219)
AltInStars.2.	0.120 (−1.437,1.676)	0.810 (−0.944,2.564)
ContinuousReceiver outdegree	−0.038 (−0.236,0.159)	−0.023 (−0.074,0.029)
MismatchingTransitiveTies term	1.590 (−1.374,4.555)	2.041 (−0.020,4.102)
AltKTrianglesU.2.	2.131 (0.661,3.602)	1.998 (0.604,3.391)
Diff term	−0.104 (−0.469,0.262)	−0.039 (−0.084,0.007)
ContinuousReceiver AbsDiffMQscores	0.152 (−0.735,1.039)	−0.038 (−0.296,0.220)
ContinuousReceiver NumberJusticesPro	−0.079 (−0.693,0.536)	0.082 (−0.318,0.483)
Receiver Overruled	0.140 (−10.935,11.216)	−0.271 (−3.543,3.002)
ConvergedRuns	20	20
TotalRuns	20	20

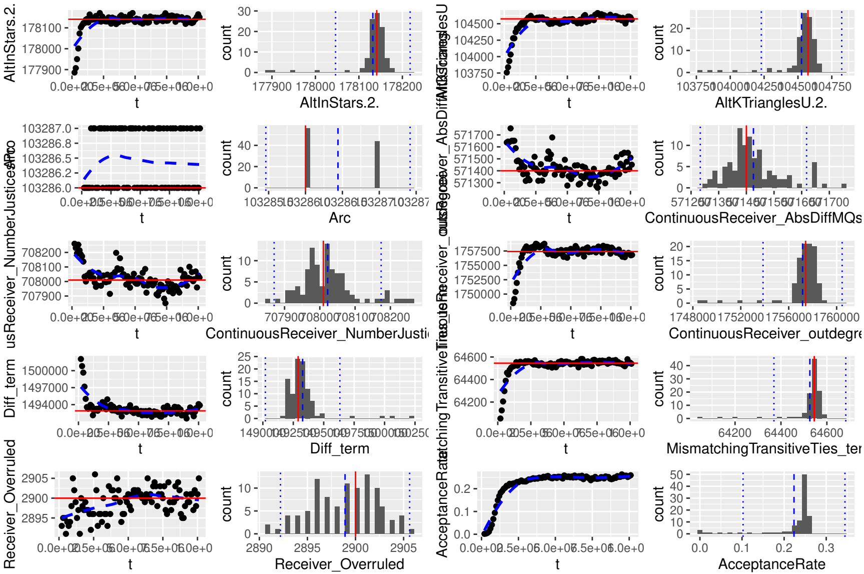
Supreme Court cERGM estimates (new EE implementation)



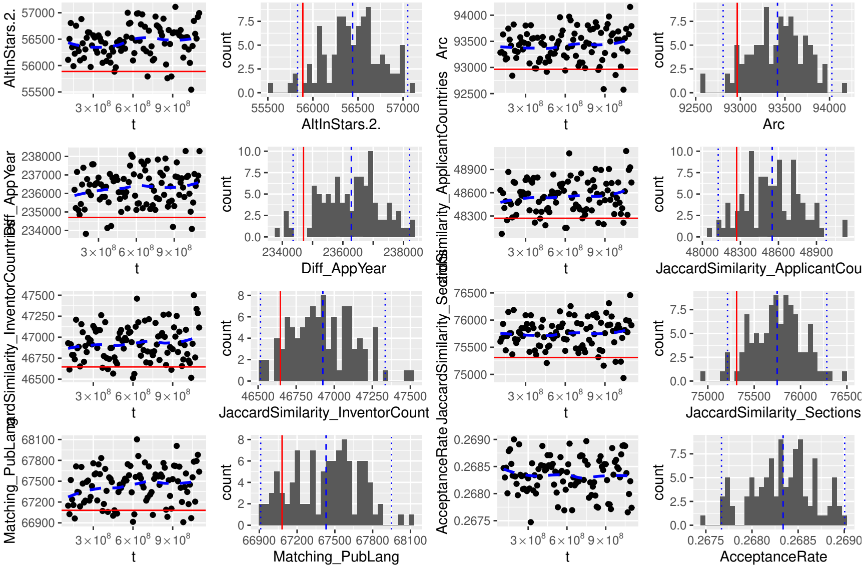
Degeneracy check for Supreme Court model (1950)



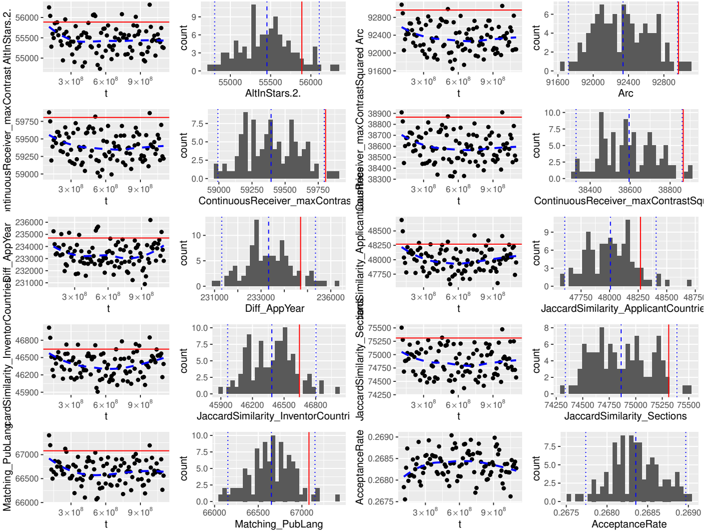
Degeneracy check for Supreme Court model (2015)



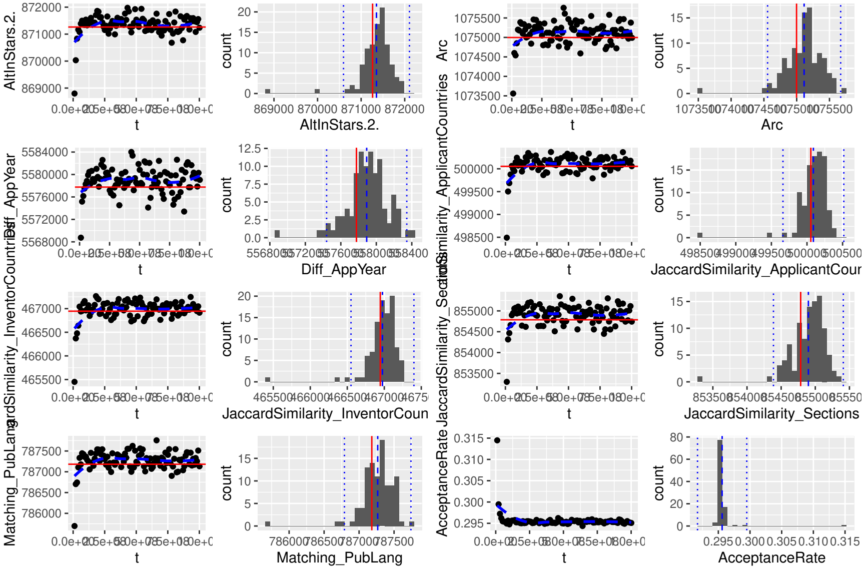
Degeneracy check for simple patent model (time period 1)



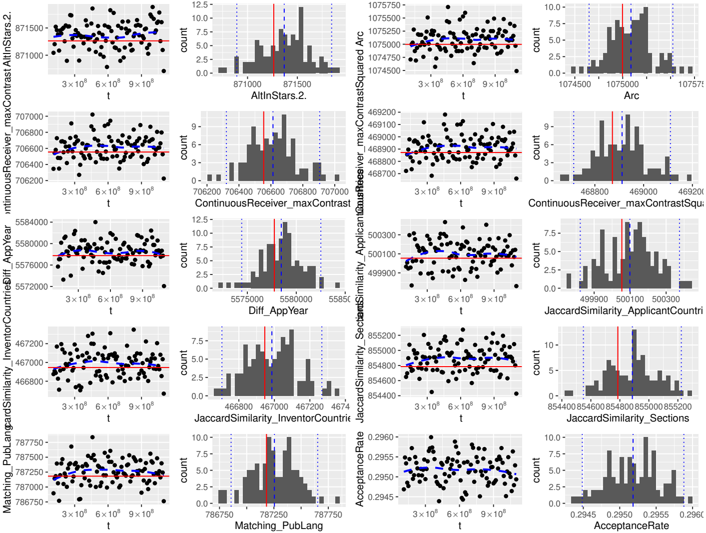
Degeneracy check for max. contrast patent model (time period 1)



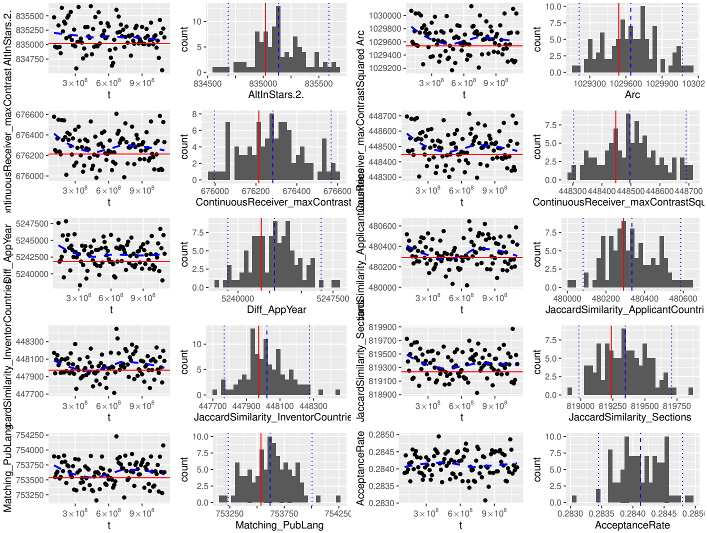
Degeneracy check for simple patent model (time period 15)



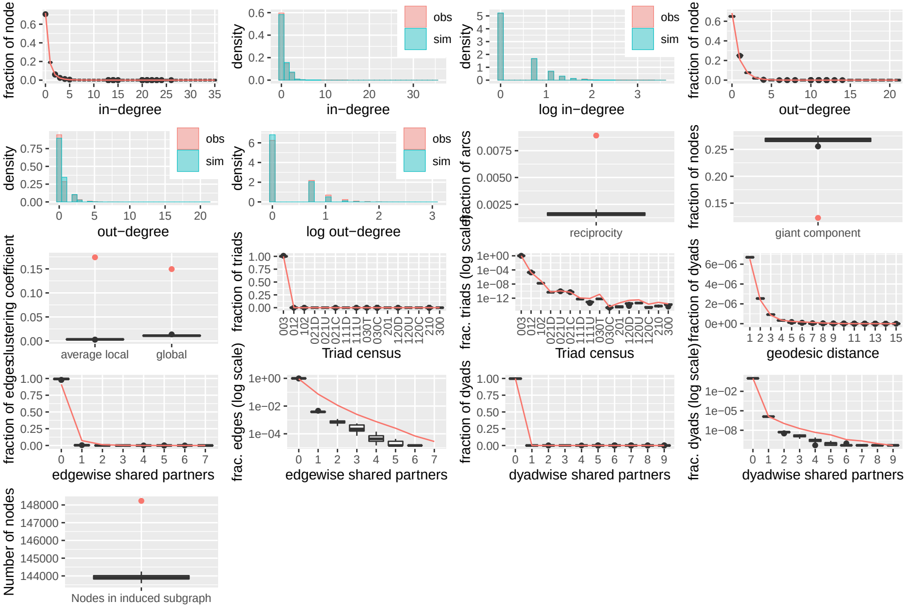
Degeneracy check for max. contrast patent model (time period 15)



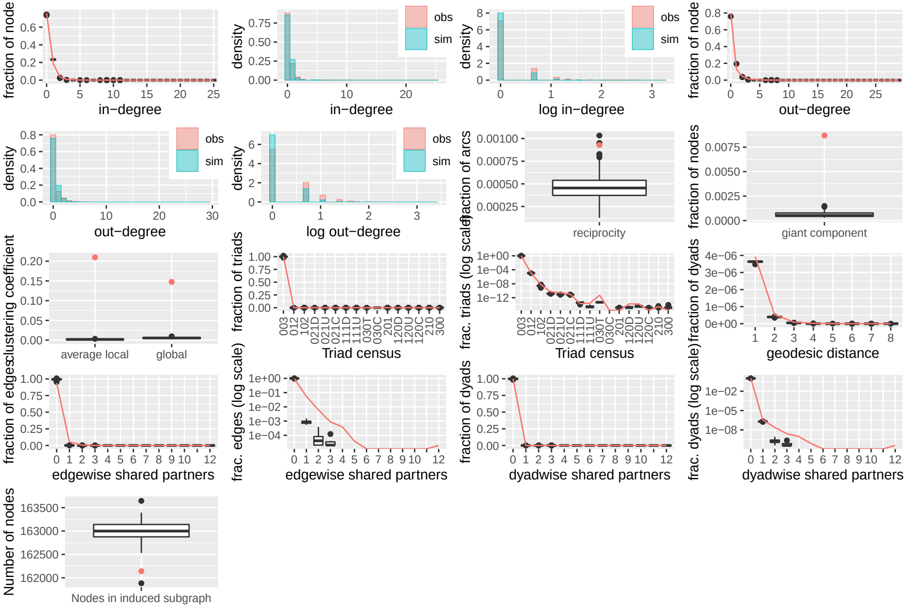
Degeneracy check for max. contrast patent model (time period 14)



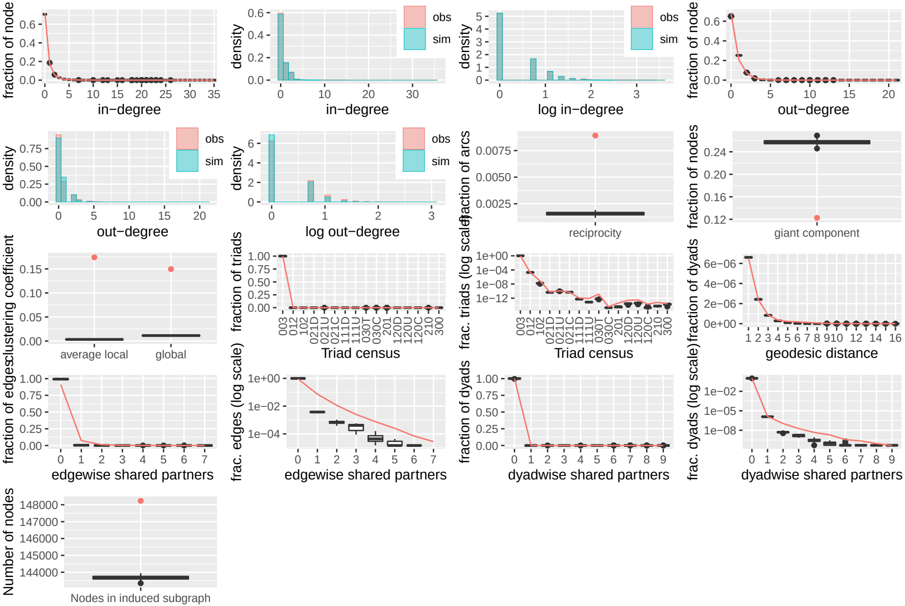
New cERGM GoF plots for simple patent model (time period 1)



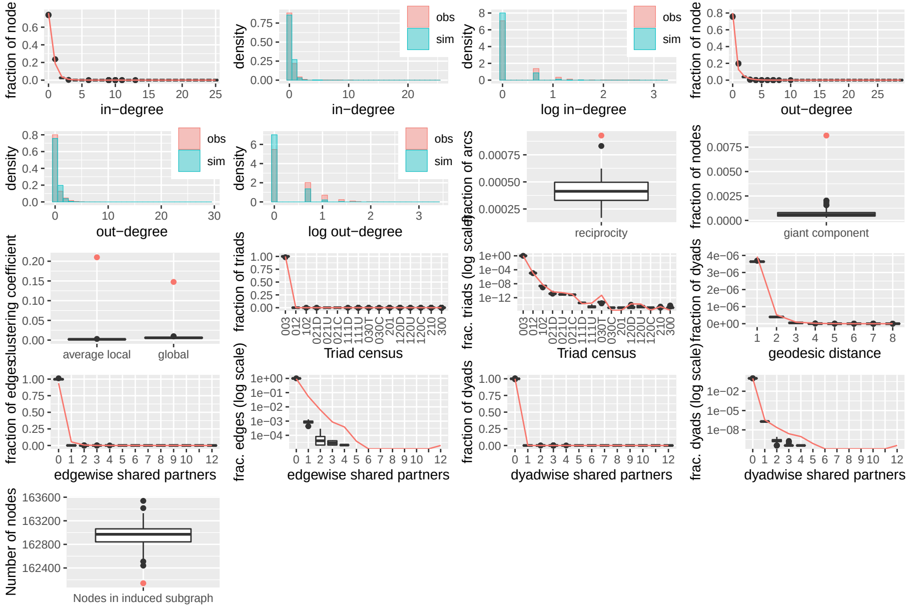
New cERGM GoF plots for simple patent model (time period 15)



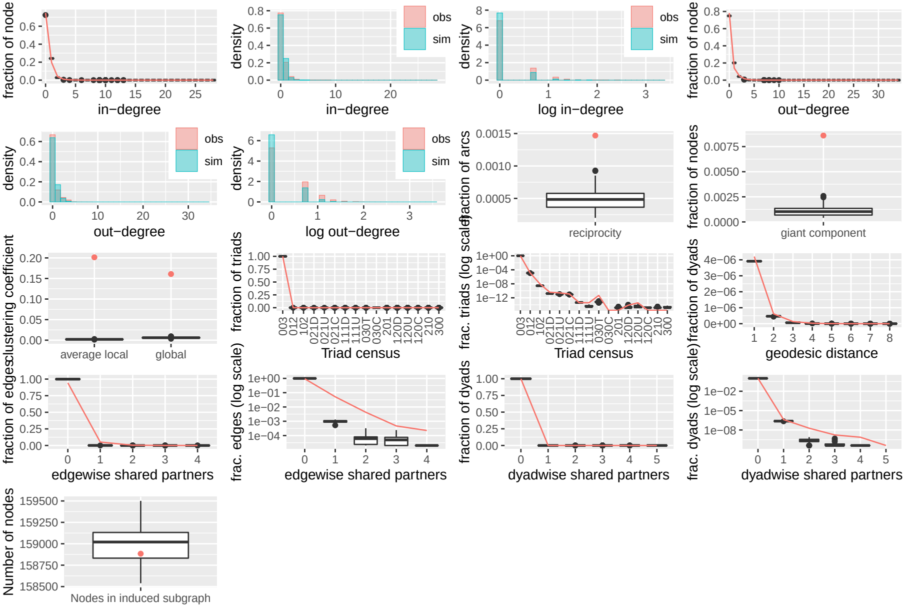
New cERGM GoF plots for max. contrast patent model (time period 1)



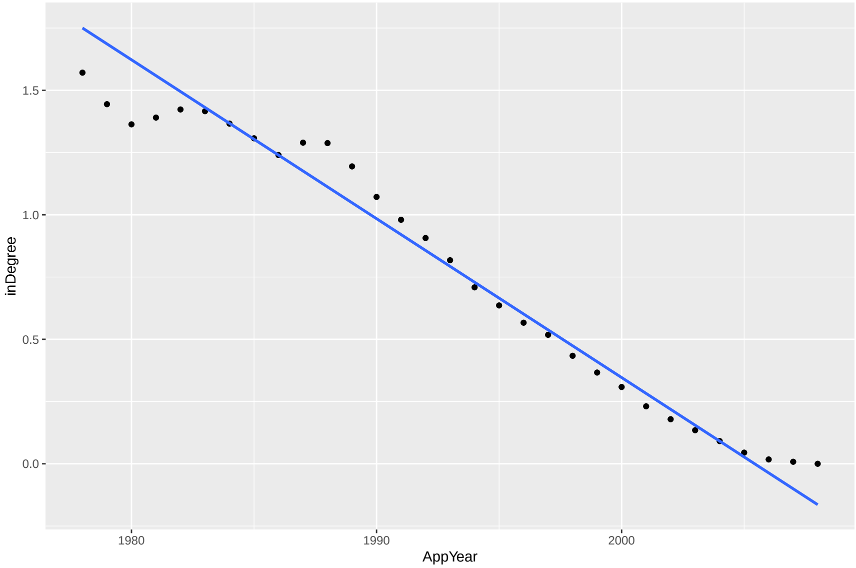
New cERGM GoF plots for max. contrast patent model (time period 15)



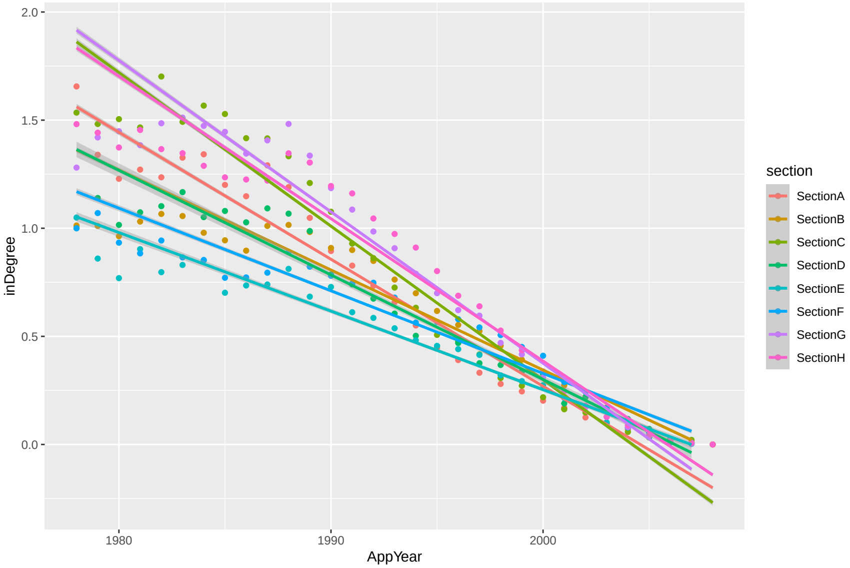
New cERGM GoF plots for max. contrast patent model (time period 14)



EPO patent network mean in-degree by year



EPO patent network mean in-degree by year grouped by technology section



The following slides are from a previous conference presentation (Stivala and Lomi, 2020) using the same EPO patent data set:

- Stivala, A. & Lomi, A. (2020) The network structure of success: Evidence from an empirical study of European patents. *Fifth Annual Australian Social Network Analysis Conference (ASNAC 2020)*, November 25-27, 2020, Perth, Western Australia (Online-only conference due to COVID-19 restrictions).

`https://stivalaa.github.io/AcademicWebsite/slides/patent_contrast_slides.pdf`

Data source

- The patent data is from the Information Retrieval Facility `https://www.ir-facility.org/`
- We used the MAREC (Matrixware Research Collection), of over 19 million patents from 1976 – 2008. `https://www.ir-facility.org/prototypes/marec`
- Specifically we used patents (applications and granted) from the European Patent Office (EPO).
- We extracted bibliographic data for 1 933 231 unique patents from the full text XML data.
- From this a 1 933 231 node citation network is built.
- 149 instances of self-loops are removed.
- Including nodes for patents cited from patents in that data (but for which we have no data other than a unique identifier), a 4 903 886 node citation network is built.
- But this larger network has no attribute data for 61% of the nodes.

Patent technology classifications

- The International Patent Classification (IPC) scheme is hierarchical.
- The highest level is Section (of which there are 8).
- There are then 120 classes and 600 subclasses.
- E.g. Section H is "Electricity" and class H01 is "basic electric elements".
- We will use Section and Class levels.
- Note that the EPO (unlike the USPTO data e.g. from NBER) allows multiple sections and classes to be assigned to a patent.
- Also the EPO assigns classes based on the entire application, not just the "claims" so is determined objectively by the examiner (Gruber et al., 2013).

Summary statistics of the patent data

Statistic	N	Mean	St. Dev.	Min	Max
Forward citations	1933231	0.573	1.448	0	76
App. Year [base 1978]	1933231	18.442	7.297	0	30
Niche width	1928684	0.236	0.282	0.000	0.929
Max. contrast	1928684	0.659	0.064	0.305	0.812
Secondary contrast	817292	0.586	0.071	0.305	0.766
Contrast share	1928684	0.779	0.265	0.087	1.000
Contrast variance	817292	0.006	0.006	0.000	0.086
Num. classes	1933231	1.595	0.841	1	14
Num. subclasses	1933231	1.934	1.190	1	20
Backward citations (subgraph)	1933231	0.573	1.029	0	117
Cited max. contrast	650656	0.666	0.060	0.383	0.812
Cited secondary contrast	374032	0.599	0.070	0.305	0.766
Cited contrast variance	452945	0.004	0.005	0.000	0.086
Cited contrast share	650656	0.680	0.289	0.080	1.000
Class crossing ratio	650511	0.414	0.311	0.000	1.000
Cited niche width	650866	0.325	0.293	0.000	0.923
Num. sections	1933231	1.370	0.579	1	7
Backward citations (all)	1933231	3.251	2.911	1	142

There are 8 technology sections (highest level IPC classification), and at the next level, 123 technology classes. A patent can be assigned multiple classes and multiple sections.

Summary statistics of IPC sections

IPC Section	Description	N
A	Human necessities	405804
B	Performing operations; transporting	497492
C	Chemistry; metallurgy	464874
D	Textiles; paper	54695
E	Fixed constructions	78438
F	Mechanical engineering; lighting; heating ...	227017
G	Physics	477022
H	Electricity	438685
Y	General ...	0

Note that a patent need not be assigned to only a single section; the sections are not mutually exclusive.

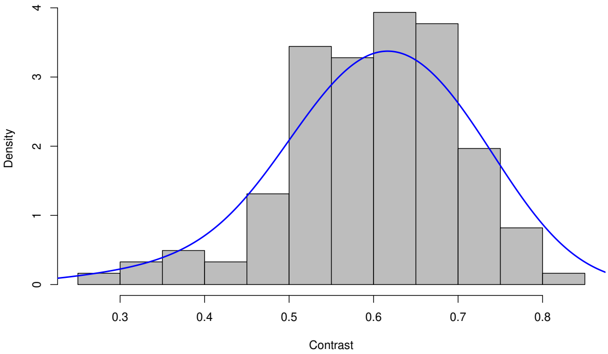
Summary statistics of the patent citation network

Description	N	Components	Giant component	Mean degree	Density
EPO (full)	4903886	746741	3789545	2.30	0.0000002
EPO (subgraph)	1933231	1119794	673306	1.15	0.0000003

Description	Reciprocity	Clustering coefficient	Assortativity coefficient
EPO (full)	0.0005	0.03125	0.08300
EPO (subgraph)	0.0025	0.07862	0.13231

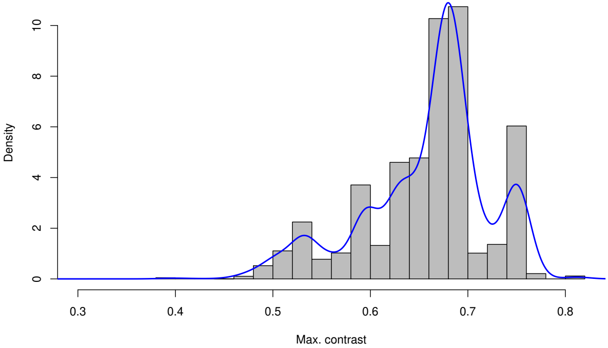
The "full" network is the network containing not only patents in the data set, but also nodes representing patents outside the data set, but which are cited by a patent in the data set. The "subgraph" network is the network induced by only those nodes in the data set itself.

Distribution of contrast values of technology classes



The highest value of contrast (0.812) is for A43 (footwear), and the lowest value (0.250) is for C99 (chemistry; metallurgy).

Distribution of maximum contrast value of patents



Negative binomial models, citations as response variable I

	Model 1	Model 2	Model 3
App. Year [base 1978]	-0.12 (0.00)***	-0.12 (0.00)***	-0.12 (0.00)***
Section A	-0.24 (0.01)***	-0.31 (0.01)***	-0.31 (0.01)***
Section B	0.04 (0.00)***	-0.05 (0.01)***	-0.04 (0.01)***
Section C	0.25 (0.00)***	0.15 (0.01)***	0.14 (0.01)***
Section D	0.07 (0.01)***	-0.00 (0.01)	-0.01 (0.01)
Section E	-0.39 (0.01)***	-0.46 (0.01)***	-0.45 (0.01)***
Section F	-0.07 (0.01)***	-0.15 (0.01)***	-0.14 (0.01)***
Section G	0.19 (0.00)***	0.11 (0.01)***	0.10 (0.01)***
Section H	0.17 (0.01)***	0.09 (0.01)***	0.09 (0.01)***
Pub. Language German	-0.29 (0.00)***	-0.29 (0.00)***	-0.31 (0.00)***
Pub. Language French	-0.31 (0.01)***	-0.31 (0.01)***	-0.32 (0.01)***
Backward citations (all)	0.17 (0.00)***	0.17 (0.00)***	0.17 (0.00)***
Max. contrast	-2.36 (0.44)***	-2.70 (0.44)***	-2.68 (0.44)***
Max. contrast^2	3.45 (0.34)***	3.65 (0.34)***	3.61 (0.35)***
Niche width		0.22 (0.01)***	0.23 (0.01)***
Appplicant Switzerland			-0.05 (0.02)**
Inventor Switzerland			-0.07 (0.03)**
Appplicant Switzerland $\times$ Inventor Switzerland			0.27 (0.03)***
Cited max. contrast			
Cited max. contrast^2			
Cited niche width			
AIC	3331171.47	3330604.41	3248519.42
BIC	3331371.01	3330816.43	3248768.46
Log Likelihood	-1665569.73	-1665285.20	-1624239.71
Deviance	1181391.34	1181445.64	1157693.65
Num. obs.	1927639	1927639	1889616

Negative binomial models, citations as response variable II

	Model 4	Model 5	Model 6
App. Year [base 1978]	-0.11 (0.00)***	-0.11 (0.00)***	-0.11 (0.00)***
Section A	-0.14 (0.01)***	-0.14 (0.01)***	-0.14 (0.01)***
Section B	-0.00 (0.01)	-0.01 (0.01)	-0.00 (0.01)
Section C	0.11 (0.01)***	0.10 (0.01)***	0.10 (0.01)***
Section D	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Section E	-0.35 (0.01)***	-0.35 (0.01)***	-0.35 (0.01)***
Section F	-0.04 (0.01)***	-0.05 (0.01)***	-0.04 (0.01)***
Section G	0.06 (0.01)***	0.06 (0.01)***	0.05 (0.01)***
Section H	-0.03 (0.01)***	-0.03 (0.01)***	-0.03 (0.01)***
Pub. Language German	-0.34 (0.01)***	-0.34 (0.01)***	-0.36 (0.01)***
Pub. Language French	-0.34 (0.01)***	-0.34 (0.01)***	-0.34 (0.01)***
Backward citations (all)	0.04 (0.00)***	0.04 (0.00)***	0.04 (0.00)***
Max. contrast	-2.48 (0.72)***	-2.63 (0.71)***	-2.63 (0.72)***
Max. contrast^2	2.88 (0.57)***	3.21 (0.57)***	3.19 (0.58)***
Niche width	0.23 (0.01)***	0.18 (0.01)***	0.18 (0.01)***
Appplicant Switzerland			-0.07 (0.02)**
Inventor Switzerland			-0.05 (0.03)
Appplicant Switzerland $\times$ Inventor Switzerland			0.23 (0.04)***
Cited max. contrast	0.01 (0.75)	-0.02 (0.75)	0.01 (0.76)
Cited max. contrast^2	0.78 (0.59)	0.55 (0.59)	0.53 (0.60)
Cited niche width		0.11 (0.01)***	0.11 (0.01)***
AIC	1615185.10	1615025.57	1579868.81
BIC	1615401.42	1615253.28	1580130.28
Log Likelihood	-807573.55	-807492.79	-789911.40
Deviance	548718.44	548738.71	538346.76
Num. obs.	650434	650434	639387

ERGM results, 1 933 231 node network I

Effect	Model 1	Model 2	Model 3
Arc	−13.638 (−13.896, −13.380)	−13.932 (−14.224, −13.639)	−13.417 (−13.703, −13.131)
Isolates	−0.182 (−0.253, −0.111)	0.046 (−0.009,0.101)	0.087 (0.023,0.151)
Sink	−0.763 (−0.848, −0.679)	−0.486 (−0.541, −0.430)	−0.490 (−0.559, −0.421)
Source	−0.225 (−0.290, −0.159)	−0.223 (−0.269, −0.176)	−0.222 (−0.285, −0.160)
Popularity spread (AinS)	0.784 (0.697,0.870)	0.757 (0.684,0.831)	0.775 (0.685,0.865)
Activity spread (AoutS)	1.238 (1.096,1.381)	0.841 (0.744,0.937)	0.847 (0.728,0.966)
Two-path (A2P-T)	−0.003 (−0.016,0.010)	−0.023 (−0.041, −0.005)	−0.029 (−0.046, −0.012)
Shared popularity (A2P-D)	−0.213 (−0.246, −0.180)	−0.119 (−0.146, −0.091)	−0.120 (−0.149, −0.092)
Shared activity (A2P-U)	0.074 (0.055,0.092)	0.062 (0.047,0.078)	0.057 (0.038,0.076)
Sender App. Year [base 1978]	0.454 (0.442,0.465)	0.417 (0.402,0.432)	0.449 (0.431,0.466)
Receiver App. Year [base 1978]	−0.523 (−0.540, −0.505)	−0.505 (−0.525, −0.486)	−0.532 (−0.554, −0.509)
DiffSign App. Year	1.872 (1.741,2.003)	2.032 (1.916,2.148)	2.050 (1.937,2.164)
AbsDiff App. Year	−0.625 (−0.650, −0.599)	−0.600 (−0.624, −0.576)	−0.629 (−0.659, −0.600)
Jaccard similarity Applicant countries	0.756 (0.582,0.931)	0.808 (0.646,0.970)	0.786 (0.615,0.957)
Jaccard similarity Inventor countries	0.586 (0.432,0.739)	0.573 (0.443,0.702)	0.551 (0.399,0.704)
Jaccard similarity Sections	3.837 (3.518,4.156)	1.501 (1.360,1.643)	1.402 (1.269,1.535)

ERGM results, 1 933 231 node network II

Matching Pub. Language	0.102 (0.050, 0.154)	0.044 (0.004, 0.083)	−0.025 (−0.061, 0.011)
Sender Max. contrast	−1.409 (−1.596, −1.221)	−0.975 (−1.383, −0.567)	−3.547 (−3.849, −3.245)
Sender Max. contrast^2	−0.788 (−0.946, −0.630)	−1.375 (−1.762, −0.988)	0.668 (0.490, 0.847)
Receiver Max. contrast	−6.515 (−6.802, −6.229)	−5.204 (−5.433, −4.975)	−8.099 (−8.373, −7.825)
Receiver Max. contrast^2	5.169 (4.917, 5.420)	3.303 (3.108, 3.497)	5.067 (4.788, 5.346)
Jaccard similarity Classes	—	4.563 (4.308, 4.817)	5.802 (5.523, 6.080)
DiffSign Max. contrast	0.008 (−0.001, 0.018)	—	—
AbsDiff Max. contrast	−15.999 (−17.996, −14.002)	—	—
Sender Niche width	—	—	1.487 (1.424, 1.551)
Receiver Niche width	—	—	1.978 (1.798, 2.159)
Sender Secondary contrast	—	—	—
Sender Secondary contrast^2	—	—	—
Receiver Secondary contrast	—	—	—
Receiver Secondary contrast^2	—	—	—
Converged runs	20	20	20
Total runs	20	20	20

ERGM results, 1 933 231 node network III

Effect	Model 4
Arc	−13.241 (−13.577, −12.906)
Isolates	0.063 (−0.003,0.130)
Sink	−0.483 (−0.573, −0.393)
Source	−0.252 (−0.324, −0.179)
Popularity spread (AinS)	0.799 (0.710,0.888)
Activity spread (AoutS)	0.834 (0.721,0.947)
Two-path (A2P-T)	−0.022 (−0.041, −0.003)
Shared popularity (A2P-D)	−0.107 (−0.136, −0.077)
Shared activity (A2P-U)	0.058 (0.038,0.078)
Sender App. Year [base 1978]	0.433 (0.416,0.449)
Receiver App. Year [base 1978]	−0.514 (−0.535, −0.492)
DiffSign App. Year	2.046 (1.904,2.189)
AbsDiff App. Year	−0.609 (−0.639, −0.579)
Jaccard similarity Applicant countries	0.764 (0.597,0.931)
Jaccard similarity Inventor countries	0.540 (0.382,0.699)
Jaccard similarity Sections	1.392 (1.259,1.525)

ERGM results, 1 933 231 node network IV

Matching Pub. Language	-0.016 $(-0.051, 0.020)$
Sender Max. contrast	-2.529 $(-2.965, -2.093)$
Sender Max. contrast^2	-1.325 $(-1.736, -0.914)$
Receiver Max. contrast	-6.258 $(-6.603, -5.914)$
Receiver Max. contrast^2	2.104 $(1.910, 2.299)$
Jaccard similarity Classes	5.907 $(5.647, 6.167)$
DiffSign Max. contrast	—
AbsDiff Max. contrast	—
Sender Niche width	1.253 $(1.108, 1.399)$
Receiver Niche width	1.726 $(1.539, 1.914)$
Sender Secondary contrast	-4.322 $(-4.497, -4.147)$
Sender Secondary contrast^2	7.709 $(7.216, 8.203)$
Receiver Secondary contrast	-4.578 $(-4.798, -4.359)$
Receiver Secondary contrast^2	8.102 $(7.661, 8.544)$
Converged runs	20
Total runs	20

Methods I

- Power law and log-normal distributions were fitted using the methods of Clauset et al. (2009) implemented in the poweRlaw package (Gillespie, 2015).
- Negative binomial regression models were estimated using the MASS (Venables and Ripley, 2002) and formatted with the texreg (Leifeld, 2013) packages in R (R Core Team, 2016). Robust standard errors (Hinkley, 1977; MacKinnon and White, 1985) were estimated with the sandwich (Zeileis, 2004, 2006) and lmtest (Zeileis and Hothorn, 2002) packages in R. Residual diagnostics from the DHARMa R package (Hartig, 2019).
- ERGM models were estimated with EstimNetDirected (Byshkin et al., 2018; Borisenko et al., 2020; Stivala et al., 2020).
- The ERGM DiffSign parameter to control for citation temporal direction was introduced by Graham et al. (2018); McLevey et al. (2018) and also used in Stivala et al. (2019).

Methods II

- In the full 4.9 million node network, only 1.9 million nodes represent patents in the data set. The remaining 3 million nodes (61% of the nodes) represent patents cited by one of those in the data set, but for which we have no data.
- An ERGM model with NA for all values on those 3 million nodes does not converge (unlike the 3.7 million node NBER patent citation network where only 27% of the nodes have no data in Stivala et al. (2019)).
- So conditional estimation based on snowball sampling structure (Pattison et al., 2013; Stivala et al., 2016) was used. The 1.9 million nodes (39%) with data are treated as wave 0 (seeds) and the remaining 3 million nodes treated as wave 1, and estimation is conditional on this structure.

References I

R. Ackland and J. Shorish. Political homophily on the web. In *Analyzing Social Media Data and Web Networks*, chapter 1, pages 25–46. Palgrave Macmillan, Houndmills, UK, 2014.

W. An and Y. Ding. The landscape of causal inference: perspective from citation network analysis. *Am Stat*, 72(3):265–277, 2018.

A. Borisenko, M. Byshkin, and A. Lomi. A simple algorithm for scalable Monte Carlo inference. *arXiv preprint arXiv:1901.00533v4*, 2020.

C. T. Butts. Continuous time graph processes with known ERGM equilibria: Contextual review, extensions, and synthesis. *arXiv preprint arXiv:2203.06948*, 2022.

M. Byshkin, A. Stivala, A. Mira, R. Krause, G. Robins, and A. Lomi. Auxiliary parameter MCMC for exponential random graph models. *J Stat Phys*, 165(4):740–754, 2016.

M. Byshkin, A. Stivala, A. Mira, G. Robins, and A. Lomi. Fast maximum likelihood estimation via equilibrium expectation for large network data. *Sci Rep*, 8:11509, 2018.

M. Chakraborty, M. Byshkin, and F. Crestani. Patent citation network analysis: A perspective from descriptive statistics and ERGMs. *PLoS ONE*, 15 (12):e0241797, 2020.

A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM Rev*, 51(4):661–703, 2009.

A. Dubnjakovic. An evaluation of exponential random graph modeling and its use in library and information science studies. *Libr Inf Sci Res*, 38(3): 259–264, 2016.

D. Fanelli and W. Glänzel. Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE*, 8(6):e66938, 2013.

C. S. Gillespie. Fitting heavy tailed distributions: The poweRlaw package. *J Stat Softw*, 64(2), 2015.

A. Graham, P. Browne, J. Barrett, and J. McLevey. Modelling directed acyclic graphs in exponential random graph models. Talk presented at INSNA Sunbelt XXXVIII, Utrecht, The Netherlands, June 2018.

M. Gruber, D. Harhoff, and K. Hoisl. Knowledge recombination across technological boundaries: Scientists vs. engineers. *Manage Sci*, 59(4): 837–851, 2013.

M. T. Hannan, L. Pólos, and G. R. Carroll. *Logics of organization theory: Audiences, codes, and ecologies*. Princeton University Press, Princeton, NJ, 2007.

References II

F. Hartig. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*, 2019. URL `https://CRAN.R-project.org/package=DHARMa`. R package version 0.2.6.

D. V. Hinkley. Jackknifing in unbalanced situations. *Technometrics*, 19(3):285–292, 1977.

B. Kovács and M. T. Hannan. The consequences of category spanning depend on contrast. *Res Sociol Org*, 31:175–201, 2010.

B. Kovács and M. T. Hannan. Conceptual spaces and the consequences of category spanning. *Sociol Sci*, 2:252–286, 2015.

P. N. Krivitsky, D. R. Hunter, M. Morris, and C. Klumb. ergm 4: New features. *arXiv preprint arXiv:2106.04997v2*, 2022.

P. Leifeld. texreg: Conversion of statistical model output in R to LaTeX and HTML tables. *J Stat Softw*, 55(8):1–24, 2013. URL `http://www.jstatsoft.org/v55/i08/`.

D. Lusher, J. Koskinen, and G. Robins, editors. *Exponential Random Graph Models for Social Networks*. Structural Analysis in the Social Sciences. Cambridge University Press, New York, 2013.

J. G. MacKinnon and H. White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J Econometrics*, 29(3):305–325, 1985.

J. McLevey, A. V. Graham, R. McIlroy-Young, P. Browne, and K. S. Plaisance. Interdisciplinarity and insularity in the diffusion of knowledge: an analysis of disciplinary boundaries between philosophy of science and the sciences. *Scientometrics*, 117(1):331–349, 2018.

P. E. Pattison, G. L. Robins, T. A. B. Snijders, and P. Wang. Conditional estimation of exponential random graph models from snowball sampling designs. *J Math Psych*, 57(6):284–296, 2013.

T.-Q. Peng. Assortative mixing, preferential attachment, and triadic closure: A longitudinal study of tie-generative mechanisms in journal citation networks. *J Informetr*, 9(2):250–262, 2015.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

C. S. Schmid and D. R. Hunter. Improving ERGM starting values using simulated annealing. *arXiv preprint arXiv:2009.01202*, 2020.

C. S. Schmid, T. H. Y. Chen, and B. A. Desmarais. Generative dynamics of supreme court citations: Analysis with a new statistical network model. *Polit Anal*, 2021. doi: 10.1017/pan.2021.20.

References III

A. Stivala and A. Lomi. The network structure of success: Evidence from an empirical study of European patents. Talk presented at the Fifth Annual Australian Social Network Analysis Conference (ASNAC 2020), November 2020. URL https://sites.google.com/site/alexdstivala/home/conferences.

A. Stivala, A. Palangkaraya, D. Lusher, G. Robins, and A. Lomi. ERGM parameter estimation of very large directed networks: implementation, example, and application to the geography of knowledge spillovers. Talk presented at INSNA Sunbelt XXXIX, Montréal, Canada, June 2019. URL https://sites.google.com/site/alexdstivala/home/conferences.

A. Stivala, G. Robins, and A. Lomi. Exponential random graph model parameter estimation for very large directed networks. *PLoS ONE*, 15(1): e0227804, 2020.

A. D. Stivala, J. H. Koskinen, D. Rolls, P. Wang, and G. L. Robins. Snowball sampling for estimating exponential random graph models for large networks. *Soc Netw*, 47:167–188, 2016.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. URL http://www.stats.ox.ac.uk/pub/MASS4.

A. Zeileis. Econometric computing with HC and HAC covariance matrix estimators. *J Stat Softw*, 11(10):1–17, 2004. ISSN 1548-7660. doi: 10.18637/jss.v011.i10. URL https://www.jstatsoft.org/v011/i10.

A. Zeileis. Object-oriented computation of sandwich estimators. *J Stat Softw*, 16(9):1–16, 2006. ISSN 1548-7660. doi: 10.18637/jss.v016.i09. URL https://www.jstatsoft.org/v016/i09.

A. Zeileis and T. Hothorn. Diagnostic checking in regression relationships. *R News*, 2(3):7–10, 2002.