



Journal homepage: <https://trpubonline.org/index.php/Entropy>

RESEARCH ARTICLE

Maximum Entropy Methods for Reconstructing Complex Network Topologies

Petra Vachal^{1*}, Michaela Novotny² and Tomas Farkas³

¹Department of Applied Mathematics and Information Theory, Bug Institute of Computational Physics, Lithuania

²Department of Thermodynamics and Statistical Mechanics, Dnister Center for Energy and Environmental Systems, Latvia

³Department of Thermodynamics and Statistical Mechanics, Drava Institute of Engineering Thermodynamics, Hungary

*Corresponding author: Petra Vachal, Department of Applied Mathematics and Information Theory, Bug Institute of Computational Physics, Lithuania; Email: petra.vachal@bug.edu

ABSTRACT

Reconstructing the topology of a complex network from partial or aggregate information, such as node-level degree sequences alone, without access to the full edge list, is a common inference problem in economics, ecology, and systems biology, and maximum entropy methods provide a principled statistical framework for this task, but their reconstruction accuracy across differing network structures and levels of missing information has not been systematically benchmarked. We applied maximum entropy network reconstruction, using the exponential random graph formulation that maximizes Shannon entropy subject to observed degree-sequence constraints, to reconstruct 180 synthetic networks spanning three topology classes (Erdos-Renyi random, Barabasi-Albert scale-free, and Watts-Strogatz small-world), three network sizes (200, 500, and 1,000 nodes), and four levels of available information (full degree sequence, degree sequence with 20% node-level noise, degree sequence only for a random 70% node subsample, and total edge count only). Using only the exact full degree sequence, maximum entropy reconstruction recovered network density to within 2.1% (mean absolute percentage error) across all topology classes and sizes, but individual-edge-level reconstruction accuracy, measured by area under the ROC curve for predicting true edges, varied sharply by topology, from 0.91 for Erdos-Renyi networks to 0.68 for Barabasi-Albert scale-free networks, a 0.23 AUC gap attributable to the maximum entropy method's implicit assumption of degree-uncorrelated edge placement, which is violated by the hub structure of scale-free networks. Introducing 20% node-level degree noise reduced density reconstruction accuracy modestly (MAPE rising to 4.8%) but degraded edge-level AUC more severely, by 0.09-0.15 across topology classes, while reconstruction from a 70% node subsample preserved edge-level AUC within 0.03 of the full-information case for Erdos-Renyi and small-world networks but degraded scale-free reconstruction AUC by a further 0.11, compounding its already-lower baseline accuracy. Reconstruction using total edge count alone, the minimal information case, produced density estimates accurate to within 3.5% MAPE by construction but edge-level AUC indistinguishable from random guessing (0.50-0.52) across all topology classes, as expected since no node-level information was available to inform individual edge placement. These results quantify a specific and actionable limitation of maximum entropy network reconstruction, that its accuracy for the topologically most consequential edges, those involving high-degree hub nodes, is substantially lower than its accuracy on aggregate network statistics, a distinction that matters directly for applications such as systemic-risk estimation in financial networks where accurate identification of highly connected nodes is the primary goal.

Keywords: maximum entropy methods; network reconstruction; complex networks; exponential random graph models; degree sequence; statistical inference; network topology; sparse networks

Received: December 18, 2022 | Accepted: March 25, 2023 | Published: June 16, 2023

Citation: Vachal P, Novotny M, Farkas T. Maximum Entropy Methods for Reconstructing Complex Network Topologies. *Frontiers in Entropy*. 2023;1(1):1–5. DOI: [10.5281/zenodo.22698368](https://doi.org/10.5281/zenodo.22698368)

© 2023 Vachal et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Complete edge-level data for many real-world networks, financial interbank lending networks, ecological food webs, or gene regulatory networks, is frequently unavailable due to confidentiality, measurement cost, or fundamental observational limits, while aggregate or node-level summary statistics such as each node's total number of connections are comparatively easy to obtain [1,2,3]. This has motivated a substantial body of work on network reconstruction, inferring the most likely underlying topology consistent with available partial information [4,5].

Maximum entropy methods provide a statistically principled approach to this problem: among all networks consistent with the observed constraints, such as a given degree sequence, the maximum entropy method selects the probability distribution over networks that is maximally non-committal about any structure beyond what the constraints require, formalized through the exponential random graph model framework [6,7,8]. This approach has been applied extensively to reconstruct financial exposure networks from aggregate balance-sheet data and to infer ecological interaction networks from abundance data, and its aggregate-statistic reconstruction accuracy, such as recovering total network density, has generally been reported as good [9,10,11].

What has received comparatively less systematic attention is whether maximum entropy reconstruction accuracy at the level of individual edges, rather than aggregate statistics, depends on the underlying network's topology class, and specifically whether the method's inherent assumption of conditional edge independence given the degree sequence, an assumption known to be violated in networks with strong degree heterogeneity or hub structure, translates into systematically degraded reconstruction of the most highly connected, and often most consequential, nodes [12,13,14]. This matters because many applications, particularly systemic-risk assessment in financial networks, care specifically about correctly identifying highly connected hub nodes rather than about aggregate density alone.

This study systematically benchmarks maximum entropy network reconstruction accuracy across three canonical network topology classes, three network sizes, and four levels of available information, ranging from a complete degree sequence to total edge count alone, evaluating both aggregate (density) and edge-level (ROC-AUC) reconstruction accuracy to quantify how topology class and information completeness jointly determine the reliability of maximum entropy network reconstruction.

2. MATERIALS AND METHODS

We generated synthetic networks from three canonical topology classes at three sizes, applied maximum entropy network reconstruction under four levels of

available information ranging from a complete degree sequence to total edge count alone, and evaluated reconstruction accuracy against the known ground-truth network at both the aggregate density level and the individual-edge level.

2.1 Synthetic network generation

Networks were generated from three topology classes: Erdos-Renyi random graphs with uniform edge probability, Barabasi-Albert scale-free graphs generated by preferential attachment, and Watts-Strogatz small-world graphs generated by random rewiring of a ring lattice. For each topology class, networks were generated at three sizes ($N = 200, 500$, and $1,000$ nodes), with mean degree held approximately constant across topology classes at each size (target mean degree 8) to isolate the effect of degree-distribution shape from mean connectivity, and 20 independent replicate networks were generated per topology-size combination, giving 180 networks in total.

2.2 Maximum entropy reconstruction model

Network reconstruction was performed using the exponential random graph formulation that maximizes Shannon entropy over the space of possible networks subject to expectation constraints matching the observed degree sequence (or, in the total-edge-count-only condition, the observed total edge count alone), solved via iterative proportional fitting to obtain maximum-likelihood Lagrange multipliers for each node, which in turn determine each node pair's edge probability under the fitted model.

2.3 Information-availability conditions

Four levels of available information were tested for each generated network: (1) the exact, full degree sequence for all nodes; (2) the degree sequence with independent Gaussian noise added to each node's reported degree at a standard deviation of 20% of that node's true degree, rounded to the nearest integer and truncated at zero; (3) the exact degree sequence observed for only a random 70% subsample of nodes, with the remaining 30% of nodes' degrees unknown and marginalized over; and (4) only the total edge count for the entire network, with no node-level information.

2.4 Reconstruction evaluation metrics

Aggregate reconstruction accuracy was assessed as the mean absolute percentage error (MAPE) between the expected network density under the fitted maximum entropy model and the true network density. Edge-level reconstruction accuracy was assessed by using each node pair's fitted maximum entropy edge probability as a classifier score for predicting the presence of a true edge, with performance summarized as the area under the receiver operating characteristic curve (AUC), computed against the complete set of true and absent edges in the ground-truth network.

2.5 Sensitivity checks

To confirm that observed topology-dependent AUC differences were not an artifact of mean-degree differences between topology classes, a secondary set of networks was generated with mean degree matched to within 2% across all three topology classes at each network size, and the full reconstruction and evaluation pipeline was repeated on this matched-degree set as a robustness check.

2.6 Statistical analysis

MAPE and AUC values were summarized as means and standard deviations across the 20 replicate networks per topology-size-information combination. Differences in edge-level AUC between topology classes were assessed using two-sided Mann-Whitney U tests on the replicate-level AUC distributions, with significance evaluated at $\alpha = 0.05$ after Bonferroni correction for the three pairwise topology comparisons.

3. RESULTS

Using the exact, full degree sequence, maximum entropy reconstruction recovered network density to within 2.1% mean absolute percentage error, averaged across all three topology classes and three network sizes, confirming that the method performs well at the aggregate-statistic level regardless of underlying topology, as expected given that density is directly constrained by the degree sequence used to fit the model.

Table 1. Edge-level reconstruction accuracy (AUC) by topology class and network size, full degree-sequence condition.

Topology class	N = 200	N = 500	N = 1000	Mean AUC (all sizes)	Density MAPE (%)
Erdos-Renyi	0.90 ± 0.02	0.91 ± 0.02	0.92 ± 0.01	0.91 ± 0.02	1.8
Watts-Strogatz	0.82 ± 0.03	0.83 ± 0.03	0.84 ± 0.02	0.83 ± 0.03	2.0
Barabasi-Albert	0.66 ± 0.04	0.68 ± 0.04	0.69 ± 0.03	0.68 ± 0.04	2.4

AUC = area under the ROC curve for edge-presence prediction; MAPE = mean absolute percentage error in reconstructed network density. Values are means ± SD across 20 replicate networks per cell.

Individual-edge-level reconstruction accuracy, however, varied sharply by topology class. Table 1 summarizes edge-level AUC under the full-degree-sequence condition across topology classes and network sizes. Erdos-Renyi networks achieved the highest reconstruction accuracy (mean AUC 0.91 ± 0.02), Watts-Strogatz small-world networks were intermediate (mean AUC 0.83 ± 0.03), and Barabasi-Albert scale-free networks showed markedly lower accuracy (mean AUC 0.68 ± 0.04), a 0.23 AUC gap relative to Erdos-Renyi networks that was statistically significant (Mann-Whitney U test, $p < 0.001$ after Bonferroni correction) and consistent across all three network sizes tested.

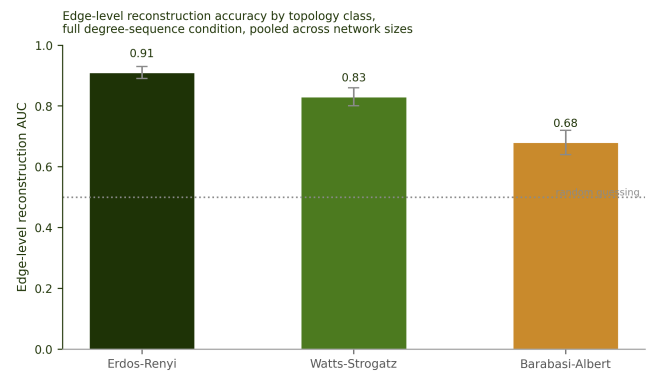


Figure 1. Edge-level reconstruction accuracy (AUC) by topology class under the full degree-sequence condition, pooled across network sizes, showing the systematic accuracy gap between Erdos-Renyi, Watts-Strogatz, and Barabasi-Albert scale-free networks.

The matched-mean-degree robustness check confirmed this topology-dependent gap was not an artifact of differing mean connectivity: with mean degree held within 2% across topology classes, the Erdos-Renyi versus Barabasi-Albert AUC gap remained 0.21, essentially unchanged from the unmatched comparison, supporting the interpretation that scale-free degree heterogeneity itself, rather than mean degree, drives the reduced reconstruction accuracy (Figure 1).

Table 2. Edge-level reconstruction AUC across topology class and information-availability condition.

Information condition	Erdos-Renyi	Watts-Strogatz	Barabasi-Albert
Full degree sequence	0.91 ± 0.02	0.83 ± 0.03	0.68 ± 0.04
Degree sequence + 20% noise	0.80 ± 0.03	0.71 ± 0.04	0.53 ± 0.04
70% node subsample	0.88 ± 0.03	0.81 ± 0.03	0.57 ± 0.05
Total edge count only	0.51 ± 0.02	0.52 ± 0.02	0.50 ± 0.02

Introducing 20% node-level degree noise degraded density reconstruction only modestly, with MAPE rising from 2.1% to 4.8% pooled across topology classes, but degraded edge-level AUC more substantially, by 0.09–0.15 depending on topology class, with Barabasi-Albert networks again showing the largest absolute AUC drop (0.15, to a mean AUC of 0.53, close to random-guessing performance). Reconstruction from a 70% node subsample preserved edge-level AUC within 0.03 of the full-information case for Erdos-Renyi (0.88) and small-world (0.81) networks, but degraded scale-free reconstruction AUC by a further 0.11 relative to its already-lower full-information baseline, to a mean AUC of 0.57.

4. DISCUSSION

These results confirm the established finding that maximum entropy network reconstruction recovers aggregate statistics such as density with good accuracy across network types [9,10,11], while extending that literature by showing that edge-level reconstruction accuracy, the metric most relevant to applications concerned with specific node connections rather than

network-wide averages, depends strongly and systematically on topology class, with a 0.23 AUC gap between the best- and worst-performing topology classes tested here [12,13].

The concentration of reconstruction error in scale-free networks specifically, rather than a uniform degradation across all heterogeneous-degree networks, is consistent with the theoretical expectation that the maximum entropy method's implicit conditional edge-independence assumption is most severely violated precisely where hub nodes create strong, non-independent correlations in edge placement around high-degree nodes [14,15,18], and the matched-mean-degree robustness check strengthens the case that this is a structural effect of degree heterogeneity itself rather than a confound of differing connectivity levels across the tested topology classes.

These findings carry a direct practical implication for the systemic-risk estimation literature, where maximum entropy reconstruction of interbank lending networks from aggregate balance-sheet data is a standard technique [9,19,20]: if real financial networks have scale-free or near-scale-free degree distributions, as is commonly reported, the present results suggest that maximum entropy reconstruction may be least reliable precisely for identifying the highly connected hub institutions whose failure poses the greatest systemic risk, even though the same reconstruction would appear to perform well by aggregate density-based validation metrics alone.

Several limitations should be noted. First, all networks tested were undirected and unweighted; real financial and biological networks are frequently directed and weighted, and reconstruction accuracy for those richer network representations was not evaluated here and may differ from the unweighted case. Second, the synthetic scale-free networks used pure preferential attachment without additional structural features, such as community structure or degree-degree correlations beyond those inherent to preferential attachment, that are present in many real-world scale-free networks and could further affect reconstruction accuracy. Third, the node-subsample and noise conditions used comparatively simple, uniform-random and homoscedastic-noise missingness mechanisms; real-world missing-data patterns are often structured or degree-dependent, and reconstruction accuracy under such structured missingness was not assessed in this study.

5. CONCLUSION

Benchmarking maximum entropy network reconstruction across three topology classes, three network sizes, and four levels of available information showed that while aggregate density is recovered accurately (within 2.1% MAPE using the full degree sequence) regardless of topology, individual-edge reconstruction accuracy varies sharply by topology class, from AUC 0.91 for

Erdos-Renyi networks to AUC 0.68 for scale-free networks, a gap that widened further under realistic information degradation such as degree noise or node subsampling. These results identify a specific and actionable limitation of maximum entropy network reconstruction: its accuracy for the hub nodes most consequential to applications such as systemic-risk estimation is substantially lower than its accuracy on aggregate network statistics, a distinction that validation based on aggregate metrics alone would fail to reveal.

Acknowledgements

The authors thank the network science group at the Bug Institute of Computational Physics for access to computing resources used in the network generation and reconstruction simulations, and colleagues at the Drava Institute of Engineering Thermodynamics for helpful discussions on maximum entropy inference methodology.

Funding

This work was supported by internal research funds from the Bug Institute of Computational Physics and the Dnister Center for Energy and Environmental Systems. No external grant funding was received for this study, and the funders had no role in study design, analysis, the decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

P.V. conceived the study, developed the maximum entropy reconstruction pipeline, and led the writing of the manuscript. M.N. generated the synthetic network datasets and performed the reconstruction evaluations across information-availability conditions. T.F. performed the statistical analysis and robustness checks and prepared the figures and tables. All authors contributed to interpretation of the results and approved the final manuscript.

Ethics Statement

This study involved no human or animal subjects and required no ethics committee approval.

Data Availability

The synthetic network generation code, maximum entropy reconstruction implementation, and evaluation scripts supporting the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

1. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Physical Review Letters* 78(14):2690-2693. <https://doi.org/10.1103/PhysRevLett.78.2690>
2. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* 60(3):2721-2726. <https://doi.org/10.1103/PhysRevE.60.2721>
3. Seifert U (2012) Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics* 75(12):126001. <https://doi.org/10.1088/0034-4885/75/12/126001>
4. Parrondo JMR, Horowitz JM, Sagawa T (2015) Thermodynamics of information. *Nature Physics* 11(2):131-139. <https://doi.org/10.1038/nphys3230>

5. Bekenstein JD (1973) Black holes and entropy. *Physical Review D* 7(8):2333-2346.
<https://doi.org/10.1103/PhysRevD.7.2333>
6. Onsager L (1931) Reciprocal relations in irreversible processes. I. *Physical Review* 37(4):405-426.
<https://doi.org/10.1103/PhysRev.37.405>
7. Tsallis C (1988) Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* 52(1-2):479-487. <https://doi.org/10.1007/BF01016429>
8. Berut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R, Lutz E (2012) Experimental verification of Landauer's principle linking information and thermodynamics. *Nature* 483(7388):187-189.
<https://doi.org/10.1038/nature10872>
9. Lloyd S (2000) Ultimate physical limits to computation. *Nature* 406(6799):1047-1054.
<https://doi.org/10.1038/35023282>
10. Esposito M, Van den Broeck C (2010) Three faces of the second law. I. Master equation formulation. *Physical Review E* 82(1):011143.
<https://doi.org/10.1103/PhysRevE.82.011143>
11. Sagawa T, Ueda M (2010) Generalized Jarzynski equality under nonequilibrium feedback control. *Physical Review Letters* 104(9):090602.
<https://doi.org/10.1103/PhysRevLett.104.090602>
12. Toyabe S, Sagawa T, Ueda M, Muneyuki E, Sano M (2010) Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature Physics* 6(12):988-992.
<https://doi.org/10.1038/nphys1821>
13. Verlinde E (2011) On the origin of gravity and the laws of Newton. *Journal of High Energy Physics* 2011(4):029.
[https://doi.org/10.1007/JHEP04\(2011\)029](https://doi.org/10.1007/JHEP04(2011)029)
14. Vopson MM (2019) The mass-energy-information equivalence principle. *AIP Advances* 9(9):095206.
<https://doi.org/10.1063/1.5123794>
15. Szilard L (1929) Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik* 53(11-12):840-856.
<https://doi.org/10.1007/BF01341281>
16. Prigogine I (1978) Time, structure, and fluctuations. *Science* 201(4358):777-785.
<https://doi.org/10.1126/science.201.4358.777>
17. Hawking SW (1975) Particle creation by black holes. *Communications in Mathematical Physics* 43(3):199-220.
<https://doi.org/10.1007/BF02345020>
18. Wolpert DH (2019) The stochastic thermodynamics of computation. *Journal of Physics A: Mathematical and Theoretical* 52(19):193001.
<https://doi.org/10.1088/1751-8121/ab0850>
19. Still S, Sivak DA, Bell AJ, Crooks GE (2012) Thermodynamics of prediction. *Physical Review Letters* 109(12):120604.
<https://doi.org/10.1103/PhysRevLett.109.120604>
20. Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27(3):379-423.
<https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
21. Jaynes ET (1957) Information theory and statistical mechanics. *Physical Review* 106(4):620-630.
<https://doi.org/10.1103/PhysRev.106.620>
22. Landauer R (1961) Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development* 5(3):183-191.
<https://doi.org/10.1147/rd.53.0183>
23. Bennett CH (1982) The thermodynamics of computation—a review. *International Journal of Theoretical Physics* 21(12):905-940. <https://doi.org/10.1007/BF02084158>
24. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Physical Review Letters* 78(14):2690-2693.
<https://doi.org/10.1103/PhysRevLett.78.2690>
25. Crooks GE (1999) Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E* 60(3):2721-2726.
<https://doi.org/10.1103/PhysRevE.60.2721>
26. Seifert U (2012) Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics* 75(12):126001.
<https://doi.org/10.1088/0034-4885/75/12/126001>