CCS/ITALY 2023
Naples, 9th – 11th October 2023
Across Complexity and Risk

universidade de aveiro
theoria poiesis praxis

Statistics of remote regions of networks

João Gama Oliveira[a], S. N. Dorogovtsev[a,b], J. F. F. Mendes[a]
jgo@ua.pt sdorogov@ua.pt jfmendes@ua.pt
[a]Departamento de Física da Universidade de Aveiro & I3N,
Campus Universitário de Santiago, Aveiro, 3810-193, Portugal
[b]Ioffe Physico-Technical Institute, St Petersburg, 194021, Russia

Introduction

Abstract

We delve into the statistical properties of regions within complex networks that are distant from vertices with high centralities, such as hubs or highly connected clusters [1]. These remote regions play a pivotal role in shaping the asymptotic behaviours of various spreading processes and the features of associated spectra. We investigate the probability distribution $P_{\geq m}(s)$ of the number s of vertices located at distance m or beyond from a randomly chosen vertex in an undirected network. Earlier, this distribution and its large m asymptotics $1/s^2$ were obtained theoretically for undirected uncorrelated networks [2]. Employing numerical simulations and analysing empirical data, we explore a wide range of real undirected networks and their models, including trees and loopy networks, and reveal that the inverse square law is valid even for networks with strong correlations. We observe this law in the networks demonstrating the small-world effect and containing vertices with degree 1. We find the specific classes of networks for which this law is not valid. Such networks include the finite-dimensional networks and the networks embedded in finite-dimensional spaces. We notice that long chains of nodes in networks reduce the range of m for which the inverse square law can be spotted. Interestingly, we detect such long chains in the remote regions of the undirected projection of a large Web domain.

Introduction (cont.)

Highlights

– For a wide array of nets, their remote regions obey a compelling pattern.

– This pattern emerges in the distributions of the number of nodes located at long distances from a randomly chosen node.

– These distributions follow the inverse square law, s^{-2}, remarkably well.

– Trees and loopy nets, including those marked by strong correlations, are observed to follow this quasi-universal scaling.

Methods

Some basic definitions

– *Intervertex distance*, ℓ_{ij} :

Length (in edges) of the shortest path between a pair of vertices

– *Nearest neighbors* :

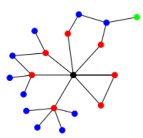
A pair of vertices separated by a distance m are m^{th} nearest neighbors

(1^{st} neighbors are distanced by one edge, 2^{nd} neighbors by two, ...)

– m^{th} *neighbors' shell of vertex* i :

The set of vertices which are m^{th} neighbors from a vertex i

The figure illustrates an example of a small, simple graph with 1^{st} neighbors (red), 2^{nd} neighbors (blue) and a 3^{rd} neighbor (green) from a given vertex (black).



– $P_m(s)$: The probability that, by choosing, uniformly, at random, a vertex i in a graph, then the number of its m^{th} nearest neighbors is s

In particular, if $m = 1$, then $P_1(s)$ is the degree distribution of the graph (where *degree* is the number of edges connected to a given vertex).

Methods (cont.)

– $P_{\geq m}(s)$: The probability that number of nodes located at distance m or beyond from a randomly chosen node is s.

– Cumulative $P_{\geq m}(s)$: $P^{(cum)}_{\geq m}(s) = \sum_{u \geq s} P_{\geq m}(u)$

– *Average distance* between pairs of vertices in a connected component of size n : $\bar{\ell} = \frac{2}{n(n-1)} \sum_{i,j>i} \ell_{ij}$

– *Diameter*, $\ell_{\max}$: The largest of all distances ℓ_{ij}

– *Remote distance* d : A distance in the interval $\bar{\ell} \lesssim d \lesssim \ell_{max}$

Notes

– For various networks, we measured the distribution $P_{\geq m}(s)$ through numerical computation of the number s of vertices located at distance m or beyond from each (and every) vertex in that specific realization of the net.

– For computations, we used C programming language and, when network size demanded it, MPI routines in the University of Aveiro's HPC cluster.

– Unless otherwise stated, dashed lines in the figures have slope -1.

Results

Inverse square law in synthetic networks

Network	(Fig.)	N	k_{max}	ℓ	ℓ_{max}
Erdős-Rényi	(1)	10^6	20	8.76	17
Recursive random 2.2 tree	(2a)	10^6	78,743	6.24	24
Recursive BA tree	(2b)	10^5	514	12.45	32
Recursive random tree	(2c)	10^5	16	20.54	51
Uniform random tree	(3)	10^4	8	106.2	285
Recursive 2.2 mixed	(4a)	10^6	119,422	4.173	13
Recursive 2.2 to two	(4b)	10^6	178,791	3.435	8
Recursive BA mixed	(4c)	10^6	2,241	7.66	18
Recursive BA to two	(4d)	10^6	2,647	6.73	12
Recursive random mixed	(4e)	10^6	32	11.24	23
Recursive random to two	(4f)	2×10^7	43	10.89	16

Table 1: Basic structural characteristics of the synthetic networks considered.

Results (cont.)

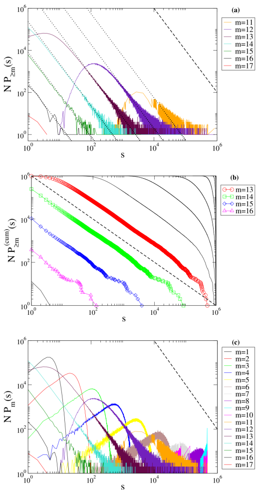


Fig. 1: Erdős–Rényi random graph, size 10^6 vertices, with average vertex degree $\langle q \rangle = 5$.

(a) Distribution $P_{\geq m}(s)$. Dotted lines show the theoretical asymptotics provided in [2]. Dashed line has slope -2.

(b) Cumulative distribution $P^{(cum)}_{\geq m}(s)$.

(c) $P_m(s)$ plot. Dashed line has slope -2.

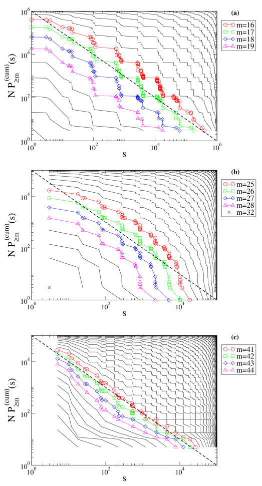


Fig. 2: Random growing trees.

(a) Scale-free recursive tree with γ=2.2 (linear preferential attachment)

(b) Scale-free BA tree (with γ=3) (proportional preference)

(c) Random recursive tree (with γ=∞) (attachment to random vertices)

Results (cont.)

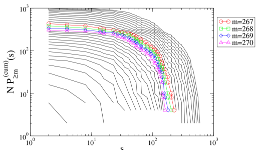


Fig. 3: Connected uniform random tree. Only highest 30 values of m are plotted.

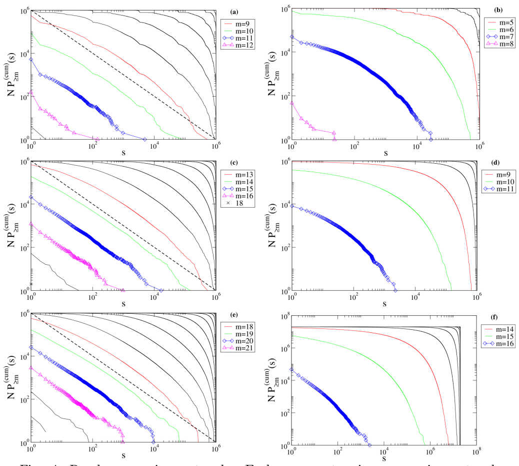


Fig. 4: Random growing networks. Each new vertex in a recursive network attaches, with equal probability, to one or two existing vertices, (a,c,e), or to two existing vertices, (b,d,f), selected by the same attachment rules as for the recursive trees in Fig. 2(a,b,c), respectively.

Results (cont.)

The statistics of remote regions in real-world networks

Network	(Fig.)	N	k_{max}	ℓ	ℓ_{max}
FP5	(5a)	25,287	2,783	3.14	8
CiteSeer	(5b)	365,154	1,739	6.470	34
YouTube	(5c)	1,134,890	28,754	5.279	24
Facebook	(5d)	63,392	1,098	4.322	15
Routers CAIDA	(6a)	192,244	1,071	6.98	26
AS CAIDA	(6b)	26,475	2628	3.876	17
US power grid	(7a)	4,941	19	18.99	46
Road newtork PA	(7b)	1,087,562	9	308.0	794
Google web	(8a)	15,763	11,401	2.517	7
Web Stanford	(9)	255,265	38,625	6.815	164

Table 2: Basic structural characteristics of the empirical networks considered.

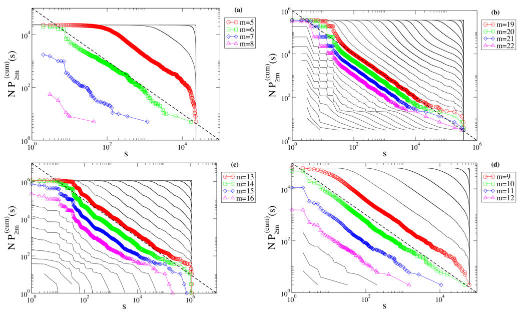


Fig. 5: Four large collaboration social networks. (a) FP5 net [3,4]. (b) CiteSeer net [5]. (c) YouTube friends net [6]. (d) Facebook [7].

Results (cont.)

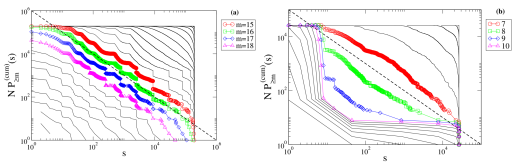


Fig. 6: Two internet networks. (a) CAIDA map of routers [8]. (b) CAIDA map of Autonomous Systems [9].

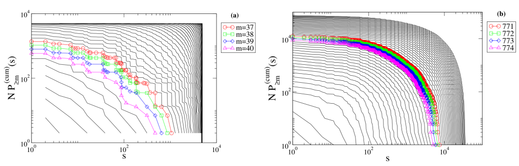


Fig. 7: Two networks embedded in geometric space. (a) US power grid net [10]. (b) Road network PA [11].

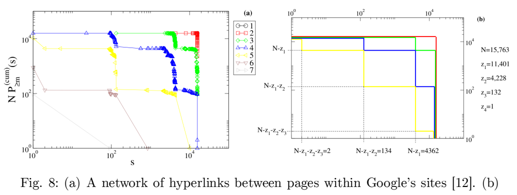


Fig. 8: (a) A network of hyperlinks between pages within Google's sites [12]. (b) Model tree-like network mimicking the Google net: it has the hub with the same numbers of the first-, second-, third-, and fourth-nearest neighbours, $z_1 = 11{,}401$, $z_2 = 4228$, $z_3 = 132$, and $z_4 = 1$, as the Google net.

Results (cont.)

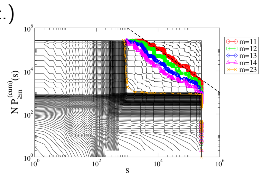


Fig. 9: Stanford Web domain [11]. These empirical cumulative distributions have the $1/s$ asymptotics for m within the range between 10 and 15, but for larger m, the cumulative distributions become step-like.

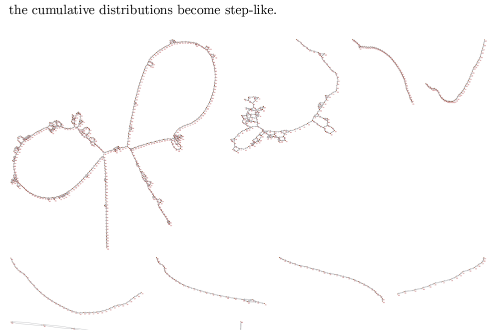


Fig. 10: Visualization, using Pajek [13], of the Stanford Web clusters formed by the vertices at a distance m = 25 or beyond from the largest hub in the undirected projection of the network, and edges between them. Notably, only 3 of these chains have one of their ends free, and the remaining 7 chains are parts of long cycles. Loosely speaking, the Web in this remote region is one-dimensional.

Conclusions & Perspectives

– We observed the s^{-2} scaling of the distribution $P_{\geq m}(s)$ of the number s of vertices located at distance m or beyond from a randomly chosen vertex in a large set of synthetic and real nets – small worlds – with surprisingly diverse architectures. Such nets include trees and loopy nets, with strong and weak correlations, one-partite projections of bipartite nets (FP5 net), undirected projections of directed nets (Stanford Web), collaboration, social, Internet and Web nets. This inverse square law is *not* observed in nets having no vertices of degree 1 and in finite-dimensional nets (power grids, roads).

– The structure of connections between vertices within the remote regions of networks differs dramatically from the main part of the network, Fig. 10.

– One should emphasize that the theoretical results of Ref. [2] for uncorrelated networks still do not offer a compelling explanation for the consistent observation of the inverse square law across such a wide spectrum of networks. The explanation of this law is a challenge for future work.

– Other challenging directions for future work are the exploration of remote regions of directed networks and examining the role of the chain structures observed in this work in network processes.

References

[1] J.G. Oliveira, S.N. Dorogovtsev, J.F.F. Mendes, Statistics of remote regions of networks, Chaos, Solitons & Fractals, accepted.
[2] S. N. Dorogovtsev, J. F. F. Mendes, A. N. Samukhin, Metric structure of random networks, Nucl. Phys. B 653 (2003) 307.
[3] J. A. Almendral, J. G. Oliveira, L. López, J. F. F. Mendes, M. A. Sanjuán, The network of scientific collaborations within the European Framework Programme, Physica A: Statistical Mechanics and its Applications 384 (2007) 675–683.
[4] J. A. Almendral, J. G. Oliveira, L. López, M. A. F. Sanjuán, J. F. F. Mendes, The interplay of universities and industry through the FP5 network, New Journal of Physics 9 (2007) 183 – 183.
[5] J. Kunegis, Konect: the koblenz network collection, Proceedings of the 22nd International Conference on World Wide Web (2013). URL: https://api.semanticscholar.org/CorpusID:15005100.
[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and Analysis of Online Social Networks, in: Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC'07), San Diego, CA, 2007.
[7] B. Viswanath, A. Mislove, M. Cha, K. P. Gummadi, On the evolution of user interaction in facebook, in: Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09), 2009.
[8] CAIDA Skitter Router-Level Topology and Degree Distribution, from April 21 to May 8, 2003, https://www.caida.org/catalog/datasets/router-adjacencies/, 2013.
[9] The dataset contains 122 CAIDA AS graphs, from January 2004 to November 2007, http://www.caida.org/data/active/as-relationships/, 2013.
[10] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[11] J. Leskovec, K. J. Lang, A. Dasgupta, M. W. Mahoney, Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters, Internet Math. 6 (2009) 29–123.
[12] G. Palla, I. J. Farkas, P. Pollner, I. Derényi, T. Vicsek, Directed network modules, New J. Phys. 9 (2007) 186.
[13] V. Batagelj, Mrvar: Pajek - program for large network analysis, Connections 21 (1998) 47–57.

Acknowledgments This work was supported by project i3N, UIDB/50025/2020, UIDP/50025/2020 and LA/P/0037/2020, financed by national funds through the FCT and the MEC of Portugal.