KURTOSIS: IS IT AN APPROPRIATE MEASURE TO COMPARE THE EXTENT OF FAT-TAILEDNESS OF THE DEGREE DISTRIBUTION FOR ANY TWO REAL-WORLD NETWORKS?

Natarajan Meghanathan

Jackson State University, MS, USA

ABSTRACT

"Kurtosis" has long been considered an appropriate measure to quantify the extent of fat-tailedness of the degree distribution of a complex real-world network. However, the Kurtosis values for more than one real-world network have not been studied in conjunction with other statistical measures that also capture the extent of variation in node degree. In this paper, we determine the Kurtosis values for a suite of 48 real-world networks along with measures such as SPR(K), Max(K)-Min(K), Max(K)-Avg(K), SD(K)/Avg(K), wherein SPR(K), Max(K), Min(K), Avg(K) and SD(K) represent the spectral radius ratio for node degree, maximum node degree, minimum node degree, average and standard deviation of node degree respectively. Contrary to the conceived notion in the literature, we observe that real-world networks whose degree distribution is Poisson in nature (characterized by lower values of SPR(K), Max(K)-Min(K), Max(K)-Avg(K), SD(K)/Avg(K)) could have Kurtosis values that are larger than that of real-world networks whose degree distribution is scale-free in nature (characterized by larger values of SPR(K), Max(K)-Min(K), Max(K)-Avg(K), SD(K)/Avg(K)). When evaluated for any two real-world networks among all the 48 real-world networks, the Kendall's concordance-based correlation coefficients between Kurtosis and each of SPR, Max(K)-Min(K), Max(K)-Avg(K) and SD(K)/Avg(K) are 0.40, 0.26, 0.34 and 0.50 respectively. Thus, we seriously question the appropriateness of using Kurtosis to compare the extent of fat-tailedness of the degree distribution of the vertices for any two real-world networks.

KEYWORDS

Fat-tailedness, Degree Distribution, Kurtosis, Real-World Networks, Kendall's Concordance-based Correlation Coefficient

1. INTRODUCTION

Complex network analysis is about analyzing complex real-world networks from a graph theoretic perspective [1]. Several measures from Statistics are also used to infer the distribution of the node-level metrics [2]. One such metric and distribution that are of interest in this paper is the degree centrality metric and the fat-tailedness of its distribution. The degree of a vertex is the number of neighbors for the vertex. A degree distribution is considered to be fat-tailed if the maximum degree for a vertex is much different from the minimum or the average degree for the vertex (correspondingly, the standard deviation of node degree is also comparable or even larger than that of the average node degree) [3]. Poisson degree distributions (characteristic of random networks [4]) are not fat-tailed; whereas, power-law degree distributions (characteristic of scale-free networks [5]) are fat-tailed. Real-world networks typically exhibit power-law degree

Natarajan Meghanathan et al. (Eds) : ACSIT, SIPM, FCST, CoNeCo, CMIT - 2017
pp. 89– 99, 2017. © CS & IT-CSCP 2017 DOI : 10.5121/csit.2017.70807

distribution [3]; however the extent of fat-tailedness of the distribution differs among the networks.

Until now, the Kurtosis measure has been perceived to be the most appropriate measure that could be used to quantify the extent of fat-tailedness of the degree distribution of the vertices in a real-world network [2]. But, there is no formal work that determined the Kurtosis of a suite of real-world networks of diverse degree distributions and analyzed whether the Kurtosis of a network with smaller variation in node degree (i.e., less fat-tailed) is more likely to be larger than the Kurtosis of a network with a relatively larger variation in node degree (i.e., more fat-tailed). This forms the motivation for our research in this paper. We measure the Kurtosis of the degree distributions for a suite of 48 real-world networks in conjunction with several other relevant metrics that also capture the extent of variation in node degree. Let SPR(K), Max(K), Min(K), Avg(K) and SD(K) represent the spectral radius ratio for node degree, maximum node degree, minimum node degree, average and standard deviation of node degree respectively. The metrics that are explored in this research along with Kurtosis for node degree are: SPR(K), Max(K)-Min(K), Max(K)-Avg(K) and SD(K)/Avg(K). The spectral radius ratio for node degree (SPR(K)) [6] is defined as the ratio of the principal eigenvalue of the adjacency matrix of the network graph to that of the average node degree. According to literature [7], Min(K) ≤ Avg(K) ≤ Principal Eigenvalue(K) ≤ Max(K). The smaller the difference between Max(K) and Min(K) for a network, the lower the value for SPR(K) = Principal Eigenvalue(K) / Avg(K). SPR(K) values start from 1.0 and this is the value expected for a truly random network.

We seek to explore whether or not a real-world network A with larger Kurtosis for node degree than a real-world network B also incurs larger values for one of these above metrics that also capture the extent of variation in node degree. We measure the Kendall's concordance-based correlation coefficient [8] for Kurtosis with each of the above four metrics for the suite of 48 real-world networks. We say two networks A and B are concordant with respect to any two metrics (say, X and Y) if X(A) < X(B) and Y(A) < Y(B) or X(A) > X(B) and Y(A) > Y(B) or X(A) = X(B) and Y(A) = Y(B). Surprisingly, we observe the Kendall's concordance-based correlation coefficient for Kurtosis with each of SPR(K), Max(K)-Min(K), Max(K)-Avg(K) and SD(K)/Avg(K) to be low: 0.40, 0.26, 0.34 and 0.50 respectively; thus, seriously raising the question of using Kurtosis to compare the extent of fat-tailedness of the degree distribution of real-world networks when it has lower correlation with metrics that also capture the extent of variation in node degree.

The rest of the paper is organized as follows: Section 2 illustrates the computation of Kurtosis for a real-world network with an example graph. Section 3 illustrates the computation of the Kendall's concordance-based correlation coefficient (Kurtosis vs. Spectral radius ratio for node degree) for a subset of 8 real-world networks from the suite of 48 real-world networks studied in this research. Section 4 first provides a brief overview of the 48 real-world networks and then presents the values for Kurtosis and the other metrics (stated above) that capture the extent of variation in node degree. Section 4 also discusses the correlation for Kurtosis with each of these metrics. Section 5 concludes the paper. Throughout the paper, the terms 'node' and 'vertex', 'link' and 'edge', 'network' and 'graph' are used interchangeably. They mean the same.

2. KURTOSIS: FORMULATION AND ILLUSTRATION

Kurtosis has been traditionally used to quantify the extent of fat-tailedness of the distribution of a random variable. In the context of complex network analysis, Kurtosis has been used to quantify the extent of fat-tailedness of the degree distribution of the vertices in a real-world network. However, there is no formal work that has evaluated its appropriateness for comparing two real-world networks on the basis of the fat-tailedness of the degree distribution of the vertices in conjunction with other statistical and spectral metrics that also capture the extent of variation in node degree. In this section, we first present the formulation to compute the Kurtosis of the degree distribution of the vertices and then illustrate the computation with an example graph.

Let K be the set of all degree values for the vertices in a graph. Let P(k) indicate the probability of finding a vertex with degree k, where $k \in K$. The average, standard deviation and kurtosis for node degree are computed as follows:

$$\text{Avg}(K) = \sum_{k \in K} P(k) * (k) \quad \text{.......... (1)}$$

$$\text{SD}(K) = \sqrt{\sum_{k \in K} P(k) * (k - Avg(K))^2} \quad \text{.......... (2)}$$

$$\text{Kurtosis}(K) = \frac{\sum_{k \in K} P(k) * (k - Avg(K))^4}{SD(K)^4} \quad \text{.......... (3)}$$

0	1
0	2
0	3
0	4
0	5
0	7
0	8
1	3
1	5
1	6
1	9
3	4
4	6
4	7
5	9
7	8

Edge List

0 → 1, 2, 3, 4, 5, 7, 8
1 → 0, 3, 5, 6, 9
2 → 0
3 → 0, 1, 4
4 → 0, 3, 6, 7
5 → 0, 1, 9
6 → 1, 4
7 → 0, 4, 8
8 → 0, 7
9 → 1, 5

ID	Degree
0	7
1	5
2	1
3	3
4	4
5	3
6	2
7	3
8	2
9	2

Degree	#Nodes	P(K)
1	1	1/10 = 0.1
2	3	3/10 = 0.3
3	3	3/10 = 0.3
4	1	1/10 = 0.1
5	1	1/10 = 0.1
7	1	1/10 = 0.1

Avg. Degree = $\sum_{k} k * P(k)$

Avg. Degree, Avg(K)
= (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) + (5)(0.1) + (7)(0.1)
= 3.2

Degree (k)	P(k)	$(k\text{-Avg}(K))^2$	$(k\text{-Avg}(K))^4$	$P(k)^*(k\text{-Avg}(K))^2$	$P(k)^*(k\text{-Avg}(K))^4$
1	0.1	4.84	23.43	0.484	2.343
2	0.3	1.44	2.07	0.432	0.621
3	0.3	0.04	0.0016	0.012	0.00048
4	0.1	0.64	0.4096	0.064	0.04096
5	0.1	3.24	10.498	0.324	1.0498
7	0.1	14.44	208.51	1.444	20.851

$$\textbf{SD(K)} = \sqrt{\sum_{k} P(k) * (k - Avg(K))^2} = \sqrt{2.76} = \underline{\mathbf{1.66}}$$

$$\textbf{Kurtosis(K)} = \frac{\sum_{k} P(k) * (k - Avg(K))^4}{SD^4} = \frac{24.91}{(1.66)^4} = \underline{\mathbf{3.27}}$$

Figure 1. Example Illustration to Compute the Average, Standard Deviation and Kurtosis of the Degree Distribution of the Vertices in a Graph

Figure 1 presents an illustration of the computation of the Avg(K), SD(K) and Kurtosis(K) for node degree for an example undirected graph of 10 vertices (whose list of edges is given). We first compute the node degree (the number of neighbors for a vertex) and determine the number of vertices that are of a particular degree. The probability of finding a vertex with a certain degree is simply the fraction of the total number of vertices with the particular degree. Once we have the k vs. P(k) values for a graph, we can compute the above three statistical metrics using formulations (1), (2) and (3).

3. KENDALL'S CONCORDANCE-BASED CORRELATION

The Kendall's concordance-based correlation measure could be used to evaluate the relative ranking of two networks with respect to any two network-level metrics; in our case, Kurtosis vs. any statistical or spectral metric. In this section, we illustrate the computation of the Kendall's concordance-based correlation coefficient for a set of 8 real-world networks (taken from the suite of 48 real-world networks analyzed in Section 4) with respect to Kurtosis and Spectral radius ratio for node degree. Figure 2 illustrates the calculations. We count the number of concordat pairs of networks and the number of discordant pairs of networks and calculate the Kendall's correlation coefficient as the ratio of the sum of the number of concordant pairs and discordant pairs to that of the difference of the number of concordant pairs and discordant pairs.

Net. #	Net. Name	Kurtosis (K)	SPR(K)
1	ADJ	15.41	1.73
8	DON	2.25	1.40
15	FON	5.89	1.01
23	KCN	6.30	1.47
25	LMN	8.89	1.82
33	PBN	4.35	1.42
36	SJN	6.34	1.29
44	APN	12.77	3.22

Net. Pairs	Kurtosis (K) Values	SPR(K) Values	C/D
1, 8	15.41, 2.25	1.73, 1.40	C
1, 15	15.41, 5.89	1.73, 1.01	C
1, 23	15.41, 6.30	1.73, 1.47	C
1, 25	15.41, 8.89	1.73, 1.82	D
1, 33	15.41, 4.35	1.73, 1.42	C
1, 36	15.41, 6.34	1.73, 1.29	C
1, 44	15.41, 12.77	1.73, 3.22	D
8, 15	2.25, 5.89	1.40, 1.01	D
8, 23	2.25, 6.30	1.40, 1.47	C
8, 25	2.25, 8.89	1.40, 1.82	C
8, 33	2.25, 4.35	1.40, 1.42	C
8, 36	2.25, 6.34	1.40, 1.29	D
8, 44	2.25, 12.77	1.40, 3.22	C

Net. Pairs	Kurtosis Values	SPR(K) Values	C/D
15, 23	5.89, 6.30	1.01, 1.47	C
15, 25	5.89, 8.89	1.01, 1.82	C
15, 33	5.89, 4.35	1.01, 1.42	D
15, 36	5.89, 6.34	1.01, 1.29	C
15, 44	5.89, 12.77	1.01, 3.22	C
23, 25	6.30, 8.89	1.47, 1.82	C
23, 33	6.30, 4.35	1.47, 1.42	C
23, 36	6.30, 6.34	1.47, 1.29	D
23, 44	6.30, 12.77	1.47, 3.22	C
25, 33	8.89, 4.35	1.82, 1.42	C
25, 36	8.89, 6.34	1.82, 1.29	C
25, 44	8.89, 12.77	1.82, 3.22	C
33, 36	4.35, 6.34	1.42, 1.29	D
33, 44	4.35, 12.77	1.42, 3.22	C
36, 44	6.34, 12.77	1.29, 3.22	C

#. Concordant Pairs = 21
#. Discordant Pairs = 7

$$\text{Kendall's Concordance based Correlation Coefficient} = \frac{\#\ \text{Concordant Pairs} - \#\ \text{Discordant Pairs}}{\#\ \text{Concordant Pairs} + \#\ \text{Discordant Pairs}}$$

$$\text{Kendall's Concordance based Correlation Coefficient} = \frac{21-7}{21+7} = \frac{14}{28} = 0.50$$

Figure 2. Example Illustration to Compute the Kendall's Concordance-based Correlation Coefficient between Kurtosis and Spectral Radius Ratio for Node Degree for a Subset of the Real-World Networks

A pair of networks X and Y are said to be concordant with respect to Kurtosis(K) and SPR(K) if either one of the following are true:

(i) $\text{Kurtosis}_X(K) < \text{Kurtosis}_Y(K)$ and $\text{SPR}_X(K) < \text{SPR}_Y(K)$ or
(ii) $\text{Kurtosis}_X(K) > \text{Kurtosis}_Y(K)$ and $\text{SPR}_X(K) > \text{SPR}_Y(K)$ or
(iii) $\text{Kurtosis}_X(K) = \text{Kurtosis}_Y(K)$ and $\text{SPR}_X(K) = \text{SPR}_Y(K)$

A pair of networks X and Y are said to be discordant with respect to Kurtosis(K) and SPR(K) if either one of the following are true:

(i) $\text{Kurtosis}_X(K) > \text{Kurtosis}_Y(K)$ and $\text{SPR}_X(K) \leq \text{SPR}_Y(K)$ or
(ii) $\text{Kurtosis}_X(K) < \text{Kurtosis}_Y(K)$ and $\text{SPR}_X(K) \leq \text{SPR}_Y(K)$

For the set of 8 real-world networks considered in Figure 2 and their Kurtosis(K) and SPR(K) values, we observe 21 concordant pairs of networks and 7 discordant pairs of networks; this leads to the Kendall's concordance-based correlation coefficient of (21-7) / (21+7) = 0.50.

4. REAL-WORLD NETWORKS AND THEIR CORRELATION ANALYSIS

In this section, we first introduce the 48 real-world networks analyzed in this paper. Table 1 lists the three character code acronym, name and the network type as well as the number of nodes and edges. The networks considered cover a broad range of categories (as listed below along with the number of networks in each category): Acquaintance network (12), Friendship network (9), Co-appearance network (6), Employment network (4), Citation network (3), Literature network (3), Collaboration network (2), Political network (2), Biological network (2), Game network (2), Geographical Network, Transportation network and Trade network (1 each). A brief description about each category of networks is as follows: An *acquaintance network* is a kind of social network in which the participant nodes slightly (not closely) know each other, as observed typically during an observation period. A *friendship network* is a kind of social network in which the participant nodes closely know each other and the relationship is not captured over an observation period. A *co-appearance network* is a network typically extracted from novels/books in such a way that two characters or words (modeled as nodes) are connected if they appear alongside each other. An *employment network* is a network in which the interaction/relationship between people is primarily due to their employment requirements and not due to any personal liking. A *citation network* is a network in which two papers (nodes) are connected if one paper cites the other paper as reference. A *collaboration network* is a network of researchers/authors who are listed as co-authors in at least one publication. A *biological network* is a network that models the interactions between genes, proteins, animals of a species, etc. A *political network* is a network of entities (typically politicians) involved in politics. A *game network* is a network of teams or players playing for different teams and their associations. A *literature network* is a network of books/papers/terminologies/authors (other than collaboration, citation or co-authorship) involved in a particular area of literature. A *transportation network* is a network of entities (like airports and their flight connections) involved in public transportation. A *trade network* is a network of countries/people involved in certain trade. The reader is referred to [9] for a more detailed description of the individual real-world networks.

Table 1. Real-World Networks used in the Correlation Analysis

#	Net.	Net. Description	Ref.	Network Type	#nodes	#edges
1	ADJ	Word Adjacency Network	[10]	Co-appearance Net.	112	425
2	AKN	Anna Karnenina Network	[11]	Co-appearance Net.	140	494
3	JBN	Jazz Band Network	[12]	Employment Net.	198	2742
4	CEN	C. Elegans Neural Network	[13]	Biological Net.	297	2148
5	CLN	Centrality Literature Net.	[14]	Citation Net.	118	613
6	CGD	Citation Graph Drawing Net	[15]	Citation Net.	259	640
7	CFN	Copperfield Network	[11]	Co-appearance Net.	89	407
8	DON	Dolphin Network	[16]	Acquaintance Net.	62	159
9	DRN	Drug Network	[17]	Acquaintance Net.	212	284
10	DLN	Dutch Literature 1976 Net.	[18]	Literature Net.	37	81
11	ERD	Erdos Collaboration Net.	[19]	Collaboration Net.	433	1314
12	FMH	Faux Mesa High School Net	[20]	Friendship Net.	147	202
13	FHT	Friendship in Hi-Tech Firm	[21]	Friendship Net.	33	91
14	FTC	Flying Teams Cade Net.	[22]	Employment Net.	48	170
15	FON	US Football Network	[23]	Game Net.	115	613
16	CDF	College Dorm Fraternity Net	[24]	Acquaintance Net.	58	967
17	GD96	Graph Drawing 1996 Net	[19]	Citation Net.	180	228
18	MUN	Marvel Universe Network	[25]	Co-appearance Net.	167	301

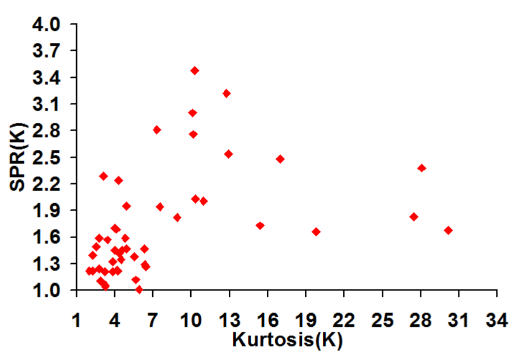
19	GLN	Graph Glossary Network	[19]	Literature Net.	67	118
20	HTN	Hypertext 2009 Network	[26]	Acquaintance Net.	115	2164
21	HCN	Huckleberry Coappear. Net.	[11]	Co-appearance Net.	76	302
22	ISP	Infectious Socio-Patterns Net	[26]	Acquaintance Net.	309	1924
23	KCN	Karate Club Network	[27]	Acquaintance Net.	34	78
24	KFP	Korea Family Planning Net.	[28]	Acquaintance Net.	37	85
25	LMN	Les Miserables Network	[11]	Co-appearance Net.	77	254
26	MDN	Macaque Dominance Net.	[29]	Biological Net.	62	1167
27	MTB	Madrid Train Bombing Net.	[30]	Acquaintance Net.	64	295
28	MCE	Manufact. Comp. Empl. Net.	[31]	Employment Net.	77	1549
29	MSJ	Soc. Net. Journal Co-authors	[32]	Co-author Net.	475	625
30	AFB	Author Facebook Network	-	Friendship Net.	171	940
31	MPN	Mexican Political Elite Net.	[33]	Political Net.	35	117
32	MMN	ModMath Network	[19]	Friendship Net.	30	61
33	PBN	US Politics Books Network	[34]	Literature Net.	105	441
34	PSN	Primary School Contact Net.	[35]	Acquaintance Net.	238	5539
35	PFN	Prison Friendship Network	[36]	Friendship Net.	67	142
36	SJN	San Juan Sur Family Net.	[37]	Acquaintance Net.	75	155
37	SDI	Scotland Corp. Interlock Net	[38]	Employment Net.	230	359
38	SPR	Senator Press Release Net.	[39]	Political Net.	92	477
39	SWC	Soccer World Cup 1998 Net	[19]	Game Net.	35	118
40	SSM	Sawmill Strike Comm. Net.	[40]	Acquaintance Net.	24	38
41	TEN	Taro Exchange Network	[41]	Acquaintance Net.	22	39
42	TWF	Teenage Female Friend Net.	[42]	Friendship Net.	47	77
43	UKF	UK Faculty Friendship Net.	[43]	Friendship Net.	83	578
44	APN	US Airports 1997 Network	[19]	Transportation Net.	332	2126
45	USS	US States Network	[44]	Geographical Net.	49	107
46	RHF	Residence Hall Friend Net.	[45]	Friendship Net.	217	1839
47	WSB	Windsurfers Beach Network	[46]	Friendship Net.	43	336
48	WTN	World Trade Metal Network	[47]	Trade Net.	80	875

Table 2 lists the values for SPR(*K*), Avg(*K*), SD(*K*), Min(*K*), Max(*K*) and Kurtosis(*K*) obtained for these 48 real-world networks. Figure 3 plots the distribution of Kurtosis(*K*) vs. each of the following: SPR(*K*), SD(*K*)/Avg(*K*), Max(*K*) - Min(*K*) and Max(*K*) - Avg(*K*). We also mention the values for the Kendall's correlation coefficient obtained for Kurtosis(*K*) vs. each of these metrics. We observe all the four correlation coefficient values to be less than or equal to 0.50; the largest being 0.50 for Kurtosis vs. SD(*K*)/Avg(*K*) ratio and the lowest being 0.26 for Kurtosis(*K*) vs. Max(*K*) - Min(*K*), an appreciable measure of the extent of variation in node degree and fat-tailedness nature of the degree distribution.

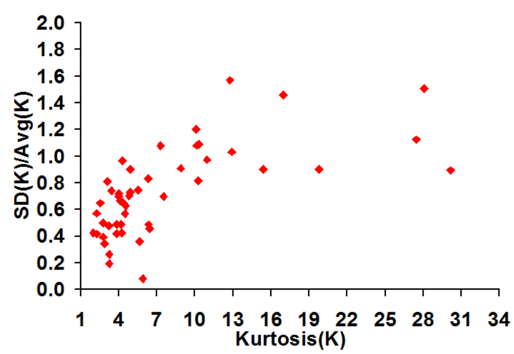
Table 2. SPR, Avg, SD, Min, Max and Kurtosis Values for the Degree Distribution of the Real-World Networks

#	Net.	SPR(*K*)	Avg(*K*)	SD(*K*)	Min(*K*)	Max(*K*)	Kurtosis (*K*)
1	ADJ	1.73	7.59	6.85	1	49	15.41
2	AKN	2.48	7.06	10.43	1	71	16.97
3	JBN	1.45	27.70	17.41	1	100	4.54
4	CEN	1.68	14.47	12.94	1	134	30.18
5	CLN	2.03	10.39	10.35	0	66	10.30
6	CGD	2.24	4.94	3.98	0	20	4.27
7	CFN	1.83	9.15	10.49	1	82	27.46
8	DON	1.40	5.13	2.93	1	12	2.25
9	DRN	2.76	2.68	2.06	0	15	10.12
10	DLN	1.49	4.38	2.96	1	12	2.52
11	ERD	3.00	6.07	6.69	0	41	10.11

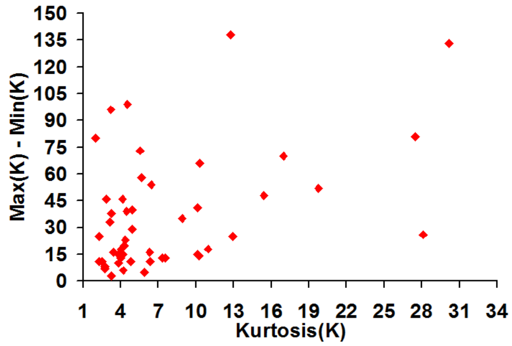
12	FMH	2.81	2.75	2.12	0	13	7.29
13	FHT	1.57	5.52	3.74	0	16	3.41
14	FTC	1.21	7.08	2.97	1	16	3.82
15	FON	1.01	10.66	0.88	7	12	5.89
16	CDF	1.11	33.35	11.43	6	52	2.87
17	GD96	2.38	2.53	3.82	1	27	28.07
18	MUN	2.54	3.61	3.76	1	26	12.92
19	GLN	2.01	3.52	3.19	0	18	10.96
20	HTN	1.21	37.64	18.30	1	97	3.21
21	HCN	1.66	7.95	7.34	1	53	19.77
22	ISP	1.69	12.45	8.33	1	47	4.14
23	KCN	1.47	4.59	3.82	1	17	6.30
24	KFP	1.70	4.59	3.11	0	13	3.99
25	LMN	1.82	6.60	6.00	1	36	8.89
26	MDN	1.04	37.65	7.40	17	55	3.24
27	MTB	1.95	9.22	6.27	0	29	4.91
28	MCE	1.12	40.23	12.53	18	76	5.64
29	MSJ	3.48	2.63	2.15	1	15	10.25
30	AFB	2.29	10.99	8.16	0	33	3.11
31	MPN	1.23	6.69	3.27	2	17	4.18
32	MMN	1.59	4.07	2.26	0	11	4.81
33	PBN	1.42	8.40	5.45	2	25	4.35
34	PSN	1.22	46.55	19.85	8	88	2.00
35	PFN	1.32	4.24	2.07	1	11	3.83
36	SJN	1.29	4.13	2.02	1	12	6.34
37	SDI	1.94	3.12	2.04	0	13	7.53
38	SPR	1.47	10.37	7.55	1	41	4.91
39	SWC	1.45	6.74	4.71	1	19	4.02
40	SSM	1.22	3.17	1.34	1	7	4.20
41	TEN	1.06	3.55	0.94	3	6	3.24
42	TWF	1.59	3.28	1.55	0	7	2.75
43	UKF	1.35	13.93	8.11	2	41	4.48
44	APN	3.22	12.81	20.10	1	139	12.77
45	USS	1.24	4.37	1.72	1	9	2.75
46	RHF	1.27	16.95	7.76	2	56	6.42
47	WSB	1.22	15.63	6.53	6	31	2.26
48	WTN	1.38	21.88	16.33	4	77	5.54



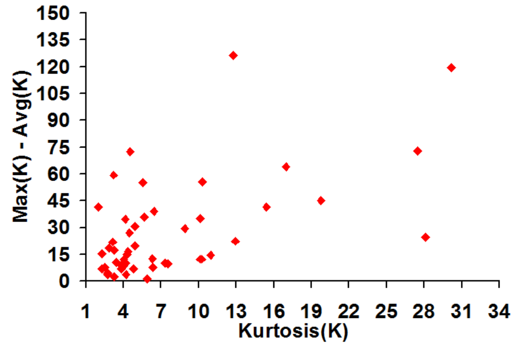
Kurtosis(K) vs. SPR(K)
Kendall's Correlation Coefficient = 0.40



Kurtosis(K) vs. SD(K)/Avg(K)
Kendall's Correlation Coefficient = 0.50



Kurtosis(K) vs. Max(K) - Min(K)
Kendall's Correlation Coefficient = 0.26



Kurtosis(K) vs. Max(K) - Avg(K)
Kendall's Correlation Coefficient = 0.35

Figure 3. Distribution of the Kurtosis(K) Values vs. {SPR(K), SD(K)/Avg(K), Max(K)-Min(K) and Max(K)-Avg(K)} Values and the Kendall's Correlation Coefficient for the 48 Real-World Networks

5. RELATED WORK

In the context of complex network analysis, Kurtosis has been typically used to capture the extent of fat-tailedness of degree distribution of the vertices and make an initial educated guess on the type (i.e., Poisson random networks or Power-law scale-free networks) of degree distribution for an underlying network graph [3]. A real-world network with Kurtosis for the degree distribution greater than 3 is typically considered to be fat-tailed [48]. Kurtosis has also been used to analyze the possibility of existence of outlier(s) in a data set [49]. In the context of complex network analysis, a larger Kurtosis for the degree distribution of a network could imply that the network has one or more nodes with degree(s) that is extremely different from the rest of the nodes in the network [3]. But, the existence of few such outlier nodes is not sufficient to classify a network as a fat-tailed network. We would need the degree distribution to exhibit non-zero probability values for degree values spanning a broader range and exhibit a decreasing trend as the degree values approach the extreme value.

Instead of Kurtosis, several other approaches have also been attempted in the literature to capture the extent of variation in node degree (inclusive of fat-tailedness). For example, graph traversal algorithms like Breadth First Search (BFS) [50] have been used in the literature to analyze the fat-tailed nature of real-world networks. The BFS algorithm could be used to determine the diameter of a network. The idea proposed in [50is to calculate the diameter (D_0) of the unperturbed network (with all nodes in the network) and calculate the diameter (D_i) of the network due to the removal of node i. The $\Delta_i = D_i - D_0/D_0$ value for each node is then calculated. A distribution of probability(Δ_i) vs. the Δ_i values (for $\Delta_i > 0$) is plotted and if it appears to mimic a power-law distribution, then the network is considered to be fat-free.

6. CONCLUSIONS

The high-level contribution of this paper is to illustrate that the Kurtosis measure may not be appropriate to compare any two real-world networks with respect to the extent of fat-tailedness. The Kurtosis of a network with a lower variation in node degree (less fat-tailed) could be larger than the Kurtosis of a network with a relatively larger variation in node degree (relatively more fat-tailed). We measure the Kendall's concordance-based correlation coefficient for Kurtosis with four different statistical/spectral measures that effectively capture the variation in node degree. We observe the correlation coefficients to be no more than 0.50. A possible solution to comprehensively measure and compare the fat-tailedness of degree distributions of a suite of real-world networks is to compute the normalized scores of Kurtosis and every other metric (that are also a measure of the extent of variation in node degree) and use a weighted score of all the metrics as a measure of the fat-tailedness of the degree distribution.

REFERENCES

[1] M. E. J. Newman, Networks: An Introduction. 1st Edition, Oxford University Press, Oxford, UK, May 2010.

[2] E. D. Kolaczyk, Statistical Analysis of Network Data, Springer, Berlin, Germany, March 2009.

[3] A. L. Barabasi, Network Science, 1st Edition, Cambridge University Press, Cambridge, UK.

[4] P. Erdos and A. Renyi, "On Random Graphs I," Publicationes Mathematicae, vol. 6, pp. 290-297, 1959.

[5] A. L. Barabasi and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, no. 5439, pp. 509-512, October 1999.

[6] N. Meghanathan, "Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs," Proceedings of the 3rd International Conference on Digital Contents and Applications, pp. 30-33, Hainan, China, December 2014.

[7] B. G. Horne, "Lower Bounds for the Spectral Radius of a Matrix," Linear Algebra and its Applications, September 1997, vol. 263, pp. 261-273.

[8] G. Strang, G. Linear Algebra and its Applications, 4th Edition, Brooks Cole, Pacific Grove, CA, USA, 2006.

[9] N. Meghanathan, "A Computationally-Lightweight and Localized Centrality Metric in lieu of Betweenness Centrality for Complex Network Analysis," Springer Vietnam Journal of Computer Science, vol. 4, no. 1, pp. 23-38, February 2017.

[10] M. E. J. Newman, "Finding Community Structure in Networks using the Eigenvectors of Matrices," Physical Review E, vol. 74, no. 3, 036104, September 2006.

[11] D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing, 1st Edition, Addison-Wesley, Reading, MA, December 1993.

[12] P. Geiser and L. Danon, "Community Structure in Jazz," Advances in Complex Systems, vo. 6, no. 4, pp. 563-573, July 2003.

[13] J. G. White, E. Southgate, J. N. Thomson and S. Brenner, "The Structure of the Nervous System of the Nematode Caenorhabditis Elegans," Philosophical Transactions B, vol. 314, no. 1165, pp. 1-340, November 1986.

[14] N. P. Hummon, P. Doreian and L. C. Freeman, "Analyzing the Structure of the Centrality-Productivity Literature Created between 1948 and 1979," Science Communication, vol. 11, no. 4, pp. 459-480, May 1990. DOI: 10.1177/107554709001100405.

[15] T. Biedl and B. J. Franz, "Graph-Drawing Contest Report," Proceedings of the 9th International Symposium on Graph Drawing, pp. 513-521, September 2001.

[16] 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, vol. 54, no. 3, pp. 396-405, September 2003.

[17] J.-S. Lee, "Generating Networks of Illegal Drug Users Using Large Samples of Partial Ego-Network Data," Intelligence and Security Informatics, Lecture Nodes in Computer Science, vol. 3073, pp. 390-402, 2004.

[18] W. de Nooy, "A Literary Playground: Literary Criticism and Balance Theory," Poetics, vol. 26, no. 5-6, pp. 385-404, August 1999.

[19] V. Batagelj and A. Mrvar, Pajek Datasets, http://vlado.fmf.uni-lj.si/pub/networks/data/, 2006.

[20] M. D. Resnick, P. S. Bearman, R. W. Blum, K. E. Bauman, K. M. Harris, J. Jones, J. Tabor, T. Beuhring, R. E. Sieving, M. Shew, M. Ireland, L. H. Bearinger and J. R. Udry, "Protecting Adolescents from Harm. Findings from the National Longitudinal Study on Adolescent Health," Journal of the American Medical Association, vol. 278, no. 10, pp. 823-832, September 1997.

[21] D. Krackhardt, "The Ties that Torture: Simmelian Tie Analysis in Organizations," Research in the Sociology of Organizations, vol. 16, pp. 183-210, 1999.

[22] J. L. Moreno, The Sociometry Reader, The Free Press, pp. 534-547, Glencoe, IL, USA, 1960.

[23] M. Girvan and M. E. J. Newman, "Community Structure in Social and Biological Networks," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 12, pp. 7821-7826, June 2002.

[24] H. R. Bernard, P. D. Killworth and L. Sailer, "Informant Accuracy in Social Network Data IV: A Comparison of Clique-level Structure in Behavioral and Cognitive Network Data," Social Networks, vol. 2, no. 3, pp. 191-218, 1980.

[25] P. M. Gleiser, "How to become a Superhero," Journal of Statistical Mechanics: Theory and Experiments, P09020, September 2007.

[26] L. Isella, J. Stehle, A. Barrat, C. Cattuto, J. F. Pinton and W. Van den Broeck, "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks," Journal of Theoretical Biology, vol. 271, no. 1, pp. 166-180, February 2011. DOI: 10.1016/j.jtbi.2010.11.033.

[27] W. W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups," Journal of Anthropological Research, vol. 33, no. 4, pp. 452-473, 1977.

[28] E. M. Rogers and D. L. Kincaid, Communication Networks: Toward a New Paradigm for Research, Free Press, June 1980.

[29] Y. Takahata, "Diachronic Changes in the Dominance Relations of Adult Female Japanese Monkeys of the Arashiyama B Group," The Monkeys of Arashiyama, pp. 124-139, Albany: State University of New York Press, 1991.

[30] B. Hayes, "Connecting the Dots," American Scientist, vol. 94, no. 5, pp. 400-404, 2006.

[31] R. L. Cross, A. Parker and R. Cross, The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations, Harvard Business Review Press, 1st Edition, June 2004.

[32] C. McCarty and L. Freeman, http://moreno.ss.uci.edu/data.html, 2008.

[33] J. Gil-Mendieta and S. Schmidt, "The Political Network in Mexico," Social Networks, vol. 18, no. 4, pp. 355-381, October 1996.

[34] V. Krebs, "Proxy Networks: Analyzing One Network to Reveal Another," Bulletin de Méthodologie Sociologique, vol. 79, pp. 61-40, July 2003.

[35] V. Gemmetto, A. Barrat and C. Cattuto, "Mitigation of Infectious Disease at School: Targeted Class Closure vs. School Closure," BMC Infectious Diseases, vol. 14, no. 695, pp. 1-10, December 2014.

[36] D. MacRae, "Direct Factor Analysis of Sociometric Data," Sociometry, vol. 23, no. 4, pp. 360-371, December 1960.

[37] C. P. Loomis, J. O. Morales, R. A. Clifford and O. E. Leonard, Turrialba Social Systems and the Introduction of Change, pp. 45-78, The Free Press, Glencoe, IL, USA, 1953.

[38] J. P. Scott, The Anatomy of Scottish Capital: Scottish Companies and Scottish Capital, 1900-1979, Croom Helm, 1st Edition, 1980.

[39] J. Grimmer, "A Bayesian Hierarchical Topic Mode for Political Texts: Measuring Expressed Agendas in Senate Press Releases," Political Analysis, vol. 18, no. 1, pp. 1-35, January 2010.

[40] J. H. Michael, "Labor Dispute Reconciliation in a Forest Products Manufacturing Facility," Forest Products Journal, vol. 47, no. 11-12, pp. 41-45, October 1997.

[41] E. Schwimmer, Exchange in the Social Structure of the Orokaiva: Traditional and Emergent Ideologies in the Northern District of Papua, C Hurst and Co-Publishers Ltd., December 1973.

[42] M. Pearson and L. Michell, "Smoke Rings: Social Network Analysis of Friendship Groups, Smoking and Drug-taking," Drugs: Education, Prevention and Policy, vol. 7, no. 1, pp. 21-37, 2000.

[43] T. Nepusz, A. Petroczi, L. Negyessy and F. Bazso, "Fuzzy Communities and the Concept of Bridgeness in Complex Networks," Physical Review E, vol. 77, no. 1, 016107, January 2008.

[44] N. Meghanathan, " Complex Network Analysis of the Contiguous United States Graph," Computer and Information Science, vol. 10, no. 1, pp. 54-76, February 2017.

[45] L. C. Freeman, C. M. Webster and D. M. Kirke, "Exploring Social Structure using Dynamic Three-Dimensional Color Images," Social Networks, vol. 20, no. 2, pp. 109-118, April 1998.

[46] L. C. Freeman, S. C. Freeman and A. G. Michaelson, "How Humans See Social Groups: A Test of the Sailer-Gaulin Models," Journal of Quantitative Anthropology, vol. 1, pp. 229-238, 1989.

[47] D. A. Smith and D. R. White, "Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980," Social Forces, vol. 70, no. 4, pp. 857-893, June 1992.

[48] K. P. Balanda and H. L. MacGillivray, "Kurtosis: A Critical Review," The American Statistician, vol. 42, no. 2, pp. 111-119, 1988.

[49] J. H. Livesey, "Kurtosis Provides a Good Omnibus Test for Outliers in Small Samples," Clinical Biochemistry, vol. 40, no. 13-14, pp. 1032-1036, September 2007.

[50] K. Goh, E. Oh, C-M. Ghim, B. Kahng and D. Kim, "Classes of the Shortest Pathway Structures in Scale Free Networks," Complex Networks, Lecture Notes in Physics, vol. 650, pp. 105-125, August 2004.