Analysis of Core and Truss Decompositions on Real-World Networks

Penghang Liu
penghang@buffalo.com
University at Buffalo
Buffalo, New York

Ahmet Erdem Sarıyüce
erdem@buffalo.edu
University at Buffalo
Buffalo, New York

ABSTRACT

Finding the dense region in a graph is a crucial problem in network analysis. Core decomposition and truss decomposition address this problem from two different perspectives. The core decomposition is a vertex-driven approach that gives each vertex a core number based on the degree, while the truss decomposition is an edge-driven approach that gives each edge a truss number based on the triangles count. Some previous works explored the common patterns in real-world networks through a vertex-driven approach. Our ongoing research aims to explore the pervasive patterns and anomalies in real-world networks from both the vertex and edge perspective. We introduce an analysis of truss decomposition and its relation to core decomposition in various types of real-world networks. We first investigate the characteristics of the core and truss degeneracy of real-world networks as well as random graphs and check how the clique counts relate to those properties. Then we analyze the interplay between core and truss decomposition by checking the truss numbers (and triangle counts) of edges with respect to the core numbers (and degrees) of their endpoints.

KEYWORDS

network analysis, dense subgraph discovery, k-core, k-truss

ACM Reference Format:
Penghang Liu and Ahmet Erdem Sarıyüce . 2019. Analysis of Core and Truss Decompositions on Real-World Networks. In *Proceedings of ACM Conference (MLG'19)*. ACM, Anchorage, AK, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Network is an effective and intuitive model to address many real-world problems. The structural and statistical characteristics of networks provide great insight to real-world applications in many fields. Vertices and edges are the two fundamental components of networks, yet most of the research on network analysis aims to explore the networks by interpreting the characteristics of vertices whereas the patterns of edges are rarely addressed. In fact, the analysis on edges reveals valuable information of networks. For instance, in a road network, we can estimate the cost of detour due to the road construction, by evaluating the length of circle containing the edge, which cannot be addressed by the vertex-driven analysis.

Detecting the dense structures in various granularities and finding the hierarchical relations among them is a fundamental problem in graph mining. For instance, in a citation network, the hierarchical relations of dense parts in various granularities can reveal how new research areas are initiated or which research subjects became popular in time [15]. k-core [11, 16] and k-truss decompositions [4, 14, 21, 22] are effective ways to find many dense regions in a graph and construct a hierarchy among them. k-core is based on the vertices and their degrees and assigns core numbers to the vertices, whereas k-truss relies on the edges and their triangle counts and yields truss numbers on the edges. Core and truss numbers can be considered as density pointers on vertices and edges that indicate the cohesiveness around a given vertex/edge.

Our ongoing research aims to understand the characteristics of core and truss decompositions on real-world networks and random graphs. Here we explore various real-world networks and investigate the interplay between core and truss numbers.

Outline. We first give some background on the core and truss decompositions in Section 2. Then we introduce the real-world datasets in Section 3. In Section 4, we present an analysis of core and truss degeneracy of real-world networks and explore their relations with the number of k-cliques for $2 \leq k \leq 10$. Next, we investigate the interplay between core and truss numbers inSection 5 by looking at the truss numbers of edges and the core numbers of their endpoints. We give the related work in Section 6 and conclude our work in Section 7.

2 BACKGROUND

Our study explores the real-world networks which can be represented as an undirected unweighted graph $G = (V, E)$, where V is the set of vertices and E is the set of edges. Here we remind the core and truss decompositions.

2.1 Core decomposition

The k-core subgraph was introduced by Seidman [16] for social networks analysis, and also by Matula and Beck [11] for clustering and graph coloring. **The k-core is a connected, maximal subgraph such that every node in the subgraph has degree at least k within the subgraph**, and the core number of a node is the highest value of k such that the node belongs to a k-core.

Core decomposition is the process of finding the core numbers of nodes, which are used to locate all the k-core subgraphs. Batagelj and Zaversnik introduced an efficient iterative peeling algorithm that uses a bucket data structure to find the core numbers of nodes

MLG'19, August 2019, Anchorage, AK, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

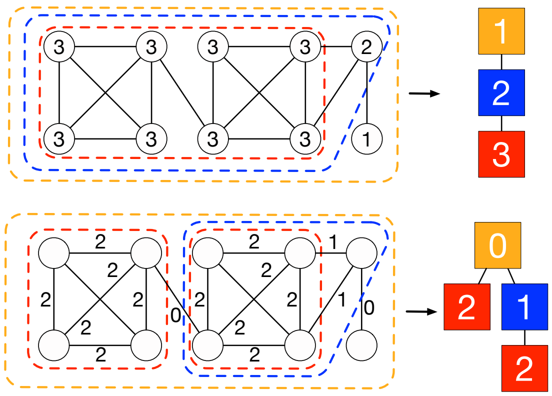


Figure 1: Examples for core (top) and truss (bottom) decompositions. On the left, core numbers are shown for each node and red, blue, and orange regions show the 3-, 2-, and 1-cores. They form a hierarchy by containment as denoted; 1-core contains 2-core and 2-core subsumes 3-core. For the same graph, trusses and truss numbers of edges are presented on the right. Entire graph is a 0-truss and the five nodes on right form a 1-truss. There are two 2-trusses and one of them is a subset of the 1-truss, as denoted by the tree hierarchy.

in $O(|E|)$ time [2]. Starting from the node with the minimum degree, the peeling algorithm assigns the degree of a node as its core number and remove it from the graph — thus decrementing the neighbor degrees if larger. This process continues until the graph is empty. k-core subgraph (for any k) is found by a traversal that visits all reachable nodes whose core numbers are at least k. The nested structure of k-cores reveals a hierarchy by containment. Figure 1 (top) presents a toy graph, core numbers of the nodes, all the k-core subgraphs, and their hierarchy.

2.2 Truss Decomposition

Higher-order structures, also known as motifs or graphlets, have been used to locate dense regions that cannot be detected otherwise with edge-centric methods [3, 20]. Finding the frequency and distribution of triangles and other small motifs in real-world networks is a simple yet effective approach used in data analysis [1, 7, 12, 13]. The truss decomposition is inspired by the k-core and considers the edges and the triangles they participate in [4, 14, 21, 22]. **k-truss is a connected, maximal subgraph such that every edge in the subgraph participates in at least k triangles within the subgraph**, and the truss number of an edge is the highest k such that the edge is part of a k-truss. Similar to the core decomposition, finding truss numbers has two phases; 1) Counting the triangles that each edge participates in, 2) Peeling those counts by choosing the edge with minimum count, assigning that as truss number, and decrement triangle counts of neighbors. This requires $O(\sum_{v \in V} d(v)^2)$ time. k-trusses also exhibit a hierarchy by containment. Figure 1 (bottom) presents the truss numbers on edges, k-trusses, and their hierarchy on the same toy graph.

3 DATASETS

In order to explore patterns in various types of real-world networks, our datasets cover networks from five different categories: social networks, router networks, citation networks, collaboration networks, and web graphs. The datasets are obtained from SNAP [9], DBLP [10], and Konect [8] and the statistics are summarized in Table 1 with details.

Social Networks. Catster is a network contains friendships between users of the website catster.com. Dogster is a friendship network between users of the website dogster.com. Email is a email network between employee of Enron Corporation. Flickr is a social network of Flickr users and their friendship connections. Hamster is a friendship networks between users of the website hamsterster.com. LiverJournal is a social network of LiveJournal. Orkut is a social network of Orkut users and their connections. YouTube is a friendship network between YouTube users. We also use the Facebook 100 dataset [19] that contains the Facebook friendship networks of 100 colleges (details are omitted for brevity).

Router Networks. As-733 is the router network contains 733 daily instances of an autonomous system. Caida is a network of autonomous systems from the CAIDA project. Gnutella is a network of Gnutella hosts. Oregon-2 is a network of autonomous system inferred from Oregon route-views. Skitter is a network of autonomous system on the Internet connected to each other, from the Skitter project.

Citation Networks. CiteSeer is a citation network extracted from the CiteSeer digital library. Cora is a citation network of scientific papers. DBLP is a citation network of scientific papers from DBLP computer science bibliography. HepTh is a network of publications in the arXiv's High Energy Physics – Theory (hep-th) section. Patent is a citation network of patents registered with the United States Patent and Trademark Office.

Collaboration Networks. DBLP-dbs is the co-authorship network for the authors of top database conference papers (VLDB, SIGMOD, and ICDE) in last ten years DBLP-dm and DBLP-pp are similar where the former is for top data mining conferences (SIGKDD, WWW, WSDM, ICDM, and SDM) and the latter is for the top parallel processing conferences (IPDPS, HPDC, SC, and ICS). All are obtained from the DBLP computer science bibliography.

Web Networks. BerkStan is a hyperlink network of the websites of the Universities in Berkley and Stanford. Blogs is a network contains front-page hyperlinks between blogs in the context of the 2004 US election. Google is a hyperlink network released in 2002 by Google as a part of the Google Programming Contest. NotreDame is a network of hyperlinks between the web pages from the website of the University of Notre Dame. Stanford is a network of the websites of the Stanford University.

Others. Drug is a network of interactions between drugs, which are approved by the U.S. Food and Drug Administration. PGP is a interaction network of users of the Pretty Good Privacy (PGP) algorithm. PowerGrid is a network contains information about the power grid of the Western States of the United States of America.

To validate if the patterns are exclusive for real-world networks, we generate two sets of random graphs based on Erdős-Rényi model (E-R) and configuration model. For the E-R model, a graph is generated randomly giving the number of vertices n and the number of edges m. For the configuration model, a graph is generated randomly based on the number of vertices n and the degree sequence of these vertices. For each model we generated 10 corresponding random graphs for each of the real-world networks. We took the average clique count and degeneracy of these 10 graphs for further validation. The core and truss degeneracy information of these random graphs are summarized in the Table 1. The core degeneracy of real-world networks ranges from 5 to 568, and the truss degeneracy ranges from 2 to 350. The degeneracy of social networks are commonly higher than other types of networks.

4 DEGENERACY AND CLIQUE COUNTS

The maximum core number of a vertex in the graph is defined as the *(core) degeneracy* [5]. It is a measure of sparseness and also quantifies the resilience of the graph with respect to cohesiveness. An equivalent definition in the context of truss decomposition is the *truss degeneracy*, which is the maximum truss number of an edge in the graph. Core and truss degeneracy concepts are closely related to the abundance of cliques in a given graph; e.g., a maximum k-clique in a graph implies a lower bound of $k-1$ on core degeneracy and of $k-2$ on truss degeneracy.

4.1 Analysis on Real-World Graphs

Here we first explore the characteristics of core and truss degeneracies of real-world networks with respect to number of the cliques (up to 10). We use 130 networks that includes 30 networks from various domains (given in Table 1) and the Facebook100 dataset [19] which has Facebook friendship networks of 100 colleges with the number of nodes between 770-41K and number of edges in the range of 16K-1.6M. We use the fast sampling algorithms proposed by Jain and Seshadhri [6] to find the clique counts. We measure the Spearman's correlation coefficient [18] between the core degeneracy numbers and the the number of k-cliques for $2 \le k \le 10$. Table 2 presents the results. We have the following observations:

Core degeneracy is best correlated with the number of 5-cliques. We observe that the correlation coefficient (ρ) between core degeneracy and number of 5-cliques is 0.895 (left part of Table 2), and highest among all. Number of edges ($|E|$) yields the lowest correlation, 0.703, and there is a significant difference between $|E|$

Category	Name	$\|V\|$	$\|E\|$	Core degeneracy			Truss degeneracy		
				Exact	E-R	Conf.	Exact	E-R	Conf.
Citation	CiteSeer	384K	1.74M	**15**	6	11	**11**	1	2
	Cora	23.2K	89.2K	**13**	5	7	**9**	1	2
	DBLP	12.6K	49.6K	**12**	4	11	**7**	1	4
	HepTh	27.7K	352K	**37**	18	24	**28**	1	8
	Patent	3.78M	16.5M	**64**	6	8	**34**	1	1
Collaboration	DBLP_dbs	8.10K	23.0K	**35**	5	5	**34**	1	1
	DBLP_dm	16.4K	33.9K	**24**	4	4	**23**	1	1
	DBLP_pp	8.41K	22.9K	**44**	3	6	**43**	1	1
	Erdos	5.10K	7.52K	**7**	2	8	**6**	1	2
Router	As-733	6.47K	12.6K	**12**	3	12	**8**	1	9
	Caida	26.5K	53.4K	**22**	3	25	**14**	1	18
	Gnutella	62.6K	148K	**6**	3	6	**2**	1	1
	Oregon-2	10.9K	31.2K	**31**	4	21	**23**	1	13
	Skitter	1.70M	11.1M	**111**	9	146	**66**	1	104
Social	Catster	150K	5.45M	**419**	57	279	**205**	1	132
	Dogster	427K	8.54M	**248**	29	256	**91**	1	145
	Email	36.7K	184K	**43**	7	35	**20**	1	15
	Flickr	1.72M	15.6M	**568**	21	271	**276**	1	54
	Hamster	1.86K	12.5K	**20**	9	16	**7**	1	5
	LiveJournal	4.00M	34.7M	**360**	20	28	**350**	1	3
	Orkut	3.07M	117M	**253**	60	69	**76**	1	22
	YouTube	1.13M	2.99M	**51**	4	65	**17**	1	33
Web	BerkStan	685K	6.65M	**201**	15	162	**199**	1	92
	Blogs	1.22K	16.7K	**36**	23	29	**23**	2	8
	Google	876K	4.32M	**44**	8	43	**42**	1	25
	NotreDame	326K	1.09M	**155**	6	40	**153**	1	20
	Stanford	282K	1.99M	**71**	11	82	**60**	1	60
Other	Drug	1.51K	48.5K	**65**	50	55	**36**	3	14
	PGP	10.7K	24.3K	**31**	3	7	**25**	1	2
	PowerGrid	4.94K	6.59K	**5**	2	2	**4**	1	1

Table 1: Statistics of the networks (FB100 dataset is not shown due to space). Last six columns show the core degeneracy and truss degeneracy numbers. In each group, 'Exact' shows the core and truss degeneracy numbers. 'E-R' and 'Conf.' give the same numbers in the random graphs generated for each network with respect to Erdos-Renyi (same $|V|$ and $|E|$) and configuration models (same degree sequence).

	Core degeneracy		Truss degeneracy	
	ρ	slope	ρ	slope
Edges	0.703	0.278	0.775	0.280
Triangles	0.845	0.277	0.863	0.274
4-Cliques	0.893	0.213	0.919	0.217
5-Cliques	**0.895**	**0.173**	0.946	0.179
6-Cliques	0.887	0.141	0.961	0.149
7-Cliques	0.865	0.117	0.970	0.130
8-Cliques	0.848	0.099	0.976	0.112
9-Cliques	0.833	0.083	0.980	0.098
10-Cliques	0.818	0.070	**0.984**	**0.085**

Table 2: Correlation between the core/truss degeneracy and the number of k-cliques for $2 \leq k \leq 10$ for all networks in Table 1. Spearman's correlation coefficient (ρ) and the slope of the regression line in log-log comparison (see Fig. 2 and Fig. 3) are given for each comparison.

and larger clique counts ($\geq$ 3) in terms of their correlation with the core degeneracy. As the clique size increases, the correlation decreases (Fig. 2c and Fig. 2d).In a previous work by Shin et al. [17], it is observed that the number of triangles has a strong correlation with the core degeneracy. Our extensive experimental evaluation on a larger (and more diverse) dataset suggests that the number of 5-cliques (and 4-cliques) is a stronger indicator (with $\rho = 0.895$ and $\rho = 0.893$) for the core degeneracy than the number of triangles ($\rho = 0.845$). Note that the maximum k-cores in our networks are not cliques (except 3 collaboration networks). Another observation we make is on the slope of relationships between clique counts and core degeneracy (Fig. 2). Shin et al. [17] propose that there is a 3-to-1 power law (log-log line slope) between the triangle count and core degeneracy, which we also observe as shown in Fig. 2a. In our evaluation, we also observe that the core degeneracy seems to have around 6-to-1 log-log line ratio with the number of 5-cliques, which is presented in Fig. 2b.

Truss degeneracy is best correlated with the number of 10-cliques. Our evaluation suggests that the truss degeneracy has a stronger correlation with the number of k-cliques ($2 \leq k \leq 10$) than the core degeneracy, as shown in Table 1. In addition, the correlation gets stronger with the larger k values, reaching up to 0.984 for 10-cliques. In general, we observe that the maximum truss subgraphs are denser than the maximum cores and consequently larger truss degeneracy implies the existence of more cliques of large size. We also observe that truss degeneracy numbers are always smaller than the core degeneracies, 36% less on average, and the difference is usually more than one meaning that the maximum trusses and cores are different subgraphs. Regarding the slope of log-log regression lines (Fig. 3), we observe ratios 4-to-1 with respect to the number of triangles (Fig. 3a) to 12-to-1 with respect to the number of 10-cliques (Fig. 3b).

4.2 Analysis on Erdős-Renyi and configuration models

Next we investigate the core and truss degeneracy characteristics of two random graph models: Erdős-Renyi (E-R) and configuration model. For each network in our dataset, we generate 10 E-R graphs that has the same number of nodes and edges, and 10 random networks that has the same degree distribution. We find the core/truss degeneracy numbers in each random network. We observe the following:

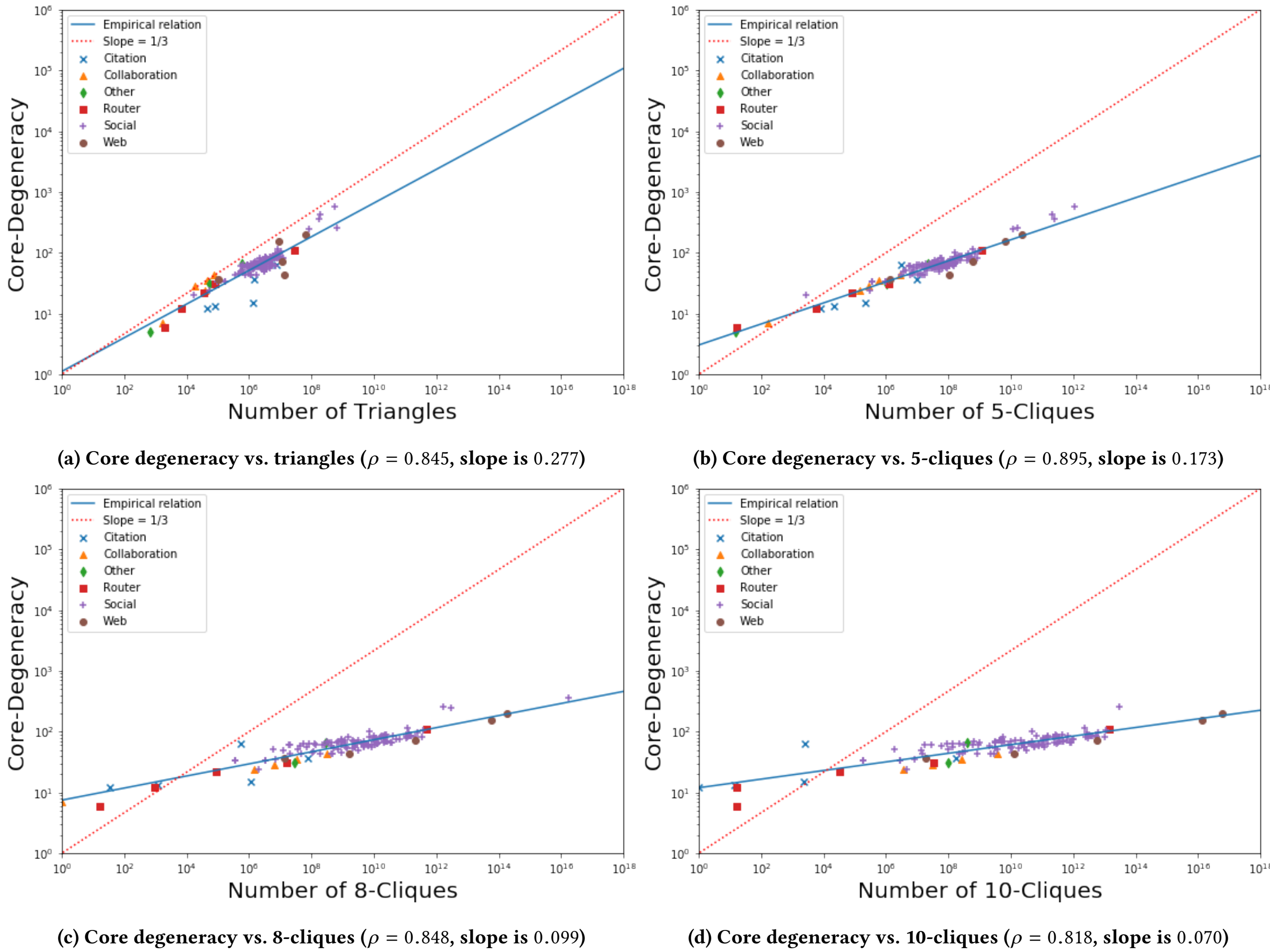
Core degeneracy numbers can be approximated to some extent by both models. Configuration model, in particular, can yield similar core degeneracy results for several networks, such as As-733, Gnutella, Dogster, and Google. For 17 (of 30) networks, random graphs that follow the same degree sequence yields less than 20% error margin for the core degeneracy. For some networks such as LiveJournal and Orkut, however, configuration model fails to provide good estimates: 28 vs. 360 and 69 vs. 253.

Truss degeneracy is a more distinctive characteristic for real-world networks as it cannot be simulated by neither random graph model. E-R model fails to provide any meaningful estimate for the truss degeneracy. Configuration model provides good approximations for only a few networks and fails on the most.

5 EDGE DISTRIBUTION

From a local perspective, real-world networks contain interesting characteristics bound to the edges. To reveal these characteristics, we explore the interplay between core and truss numbers by checking the edge distribution of real-world networks with respect to the two endpoints. The information lies in the edges are represented by the average truss number or triangle count, while the information lies in vertices are represented by the core number or degree. We proposed four representations of edge distribution regarding the four different combinations of edge information and vertex information. In Fig. 4, each edge is represented as a point in the scatter plot, with coordinates representing the degree or core number of the two endpoints. Without loss of generality, the smaller degree or core number of the two is denoted on the x-axis and the larger is denoted on the y-axis. The color of the point gives the truss number or triangle count information of the edge. For multiple edges with the same core/degree information on its vertices, we aggregate the edge charateristics by taking the average of truss numbers or triangle counts. Among the four charts, the upper left chart shows the truss number and core number relations, the upper right denotes the truss number and degree relations, the bottom left presents the triangle count and core number relations, and the bottom right gives the triangle count and degree relations. We have the following observations:

The truss number of an edge infers the core numbers of its two endpoints, and the triangle count of an edge infers the degree of the two endpoints. We observe that the truss number of edge is consistent with the core numbers of the its two endpoints. As shown in the upper left charts, edges with large truss numbers (deep red) are connecting vertices with larger core numbers (located on upper right), and vice versa. The consistency stays true for the case of triangle count and degree, as shown in the bottom right charts. For most of the real-world networks, the truss number can also infer the degree of the two endpoints, as well as the triangle count can infer the core numbers. It indicates that edges between vertices with large degrees or core numbers are likely to have large truss numbers and triangle counts. The deviation from this pattern occurs particularly in the citation networks. As shown in



(a) Core degeneracy vs. triangles (ρ = 0.845, slope is 0.277)

(b) Core degeneracy vs. 5-cliques (ρ = 0.895, slope is 0.173)

(c) Core degeneracy vs. 8-cliques (ρ = 0.848, slope is 0.099)

(d) Core degeneracy vs. 10-cliques (ρ = 0.818, slope is 0.070)

Figure 2: The correlations between core degeneracy and the number of k-cliques for k = 3, 5, 8, and 10. Each point in the charts represents a real-world network and is shape coded for the categories. Regression lines for the data points are shown in blue.

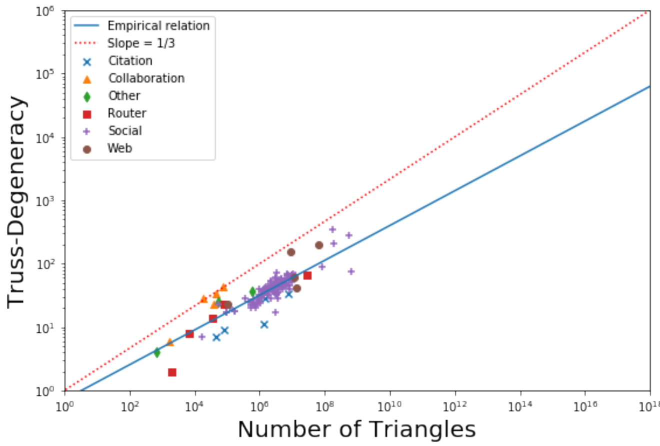
the distribution of Patent network (Figure 4d), there is an obvious mismatch in dense-color region between the distributions of core number and degree. It indicates that the edges with highest truss number are connecting to two vertices with large core number but relatively low degree. Similar deviations have been discovered in HepTh network and CiteSeer network, as shown in Figure 4e and 4f.

Core Decomposition leads to a concentration of the edge distribution. We identified a common pattern in the edge distribution, showing as an concentration from the degree to core number (from right charts to left charts). This pattern indicates that the non-trivial edges can be better inferred from the core numbers rather than degrees. However, the concentration process breaks especially for web graphs. As shown in Figure 4b, in the BerkStan web graph, there is an obvious break in transition from degree to core number. We can infer the existence of three clusters having the highest truss number but different degrees (max degree of 2,000; max degree of 7,000; and max degree of 20,000). There is another cluster with highest truss number but low degree (around 100), which might be the small cluster connecting the three big clusters. In the NotreDame web graphs (Figure 4c), we distinguished three clusters having the highest truss number but different degrees (max degree of 1,000; max degree of 3,000; and max degree of 10,000).There is another cluster with highest truss number but low degree (around 100), which is a small cluster connecting the three big cluster.

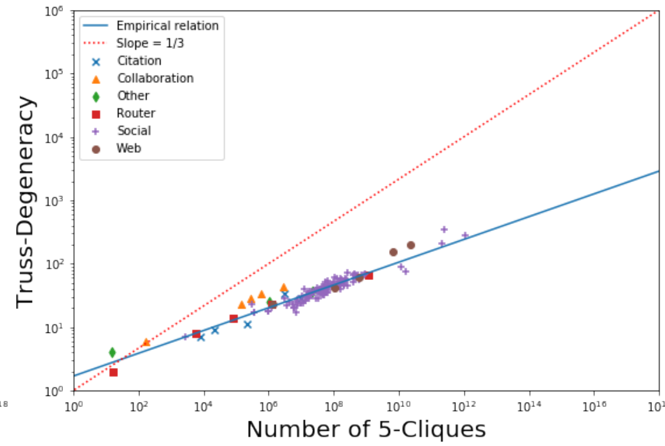
6 RELATED WORK

Previous works on real-world network analysis mostly focus on patterns related to the vertices. Our ongoing work describe the real-world network from an edge perspective, which uncover the interesting information lies in the connections.

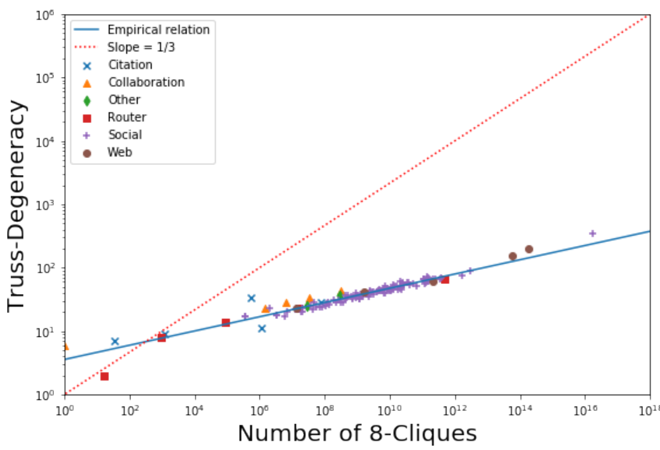
For clique counting, we used the clique number estimation algorithm called Turan-shadow, proposed by Jain and Seshadhri [6]. Their algorithm has better accuracy and efficiency for large size



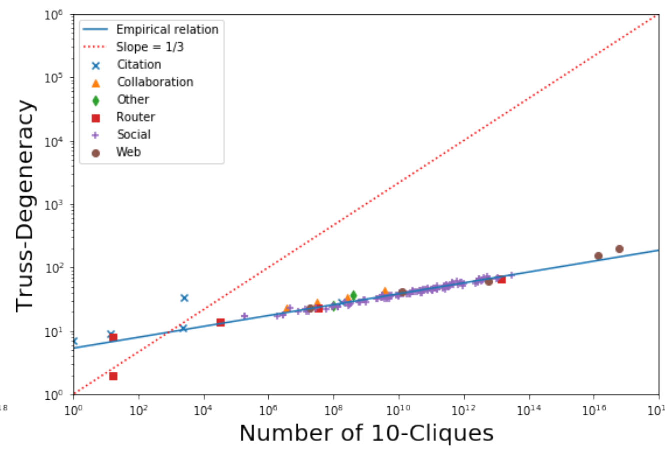
(a) Truss degeneracy vs. triangles (ρ = 0.863, slope is 0.274)



(b) Truss degeneracy vs. 5-cliques (ρ = 0.946, slope is 0.179)



(c) Truss degeneracy vs. 8-cliques (ρ = 0.976, slope is 0.112)



(d) Truss degeneracy vs. 10-cliques (ρ = 0.984, slope is 0.085)

Figure 3: The correlations between truss degeneracy and the number of k-cliques for k = 3, 5, 8, and 10. Each point in the charts represents a real-world network and is shape coded for the categories. Regression lines for the data points are shown in blue.

clique counting than other sampling algorithms. In our work, we computed the clique counts based on this estimating algorithm.

One of the most related research is done by Shin et al [17]. Their study on the degeneracy pattern is limited to the correlation between core-degeneracy and triangles count. Our study explore the correlation between both core and truss degeneracy with cliques count from 2-cliques to 10-cliques, and we compared the significance of these correlations. Since their research was mostly focus on the vertices, they considered the truss number as a supplementary information for vertex analysis. We utilized the truss number for analysis on edge itself. In addition, we designed the analysis on both vertices and edges, using both core numbers and truss numbers.

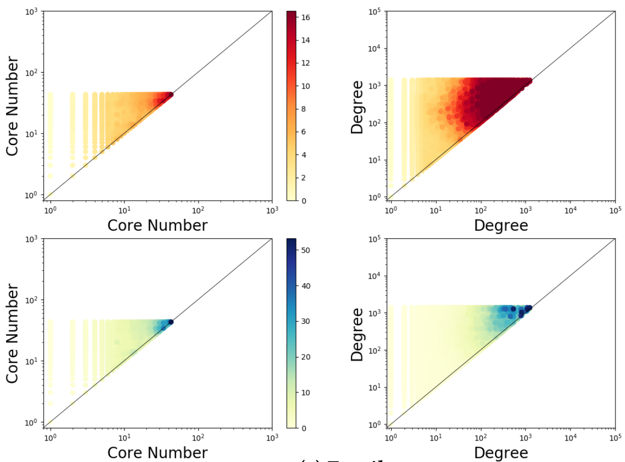
7 CONCLUSION

In this paper, we studied the the pervasive patterns and interesting anomalies of both core and truss decompositions on a large variety of real-world network. We first analyzed the networks from a global perspectives, and validated our findings on two random graph models. The global pattern discovered by our approach reveals the nature of real-world networks regarding the graph degeneracy and cliques number. We also investigated the local interplay between core and truss decompositions regarding the truss numbers (and triangle counts) of edges and the core numbers (and degrees) of their endpoints. Interestingly, we observed very distinct patterns for different type of networks.

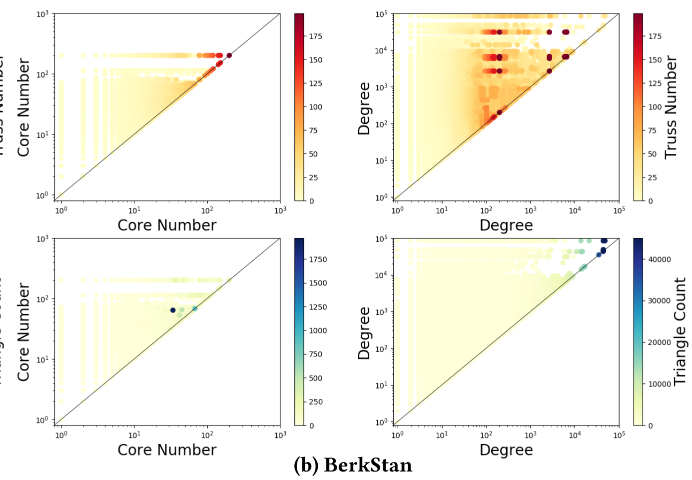
In our future work, we intend to discover other random graph models to approximate the degeneracy characteristics of real-world networks. In addition, we plan to explain our findings on the local interplay by investigating the hierarchical structures of core and truss decompositions.

REFERENCES

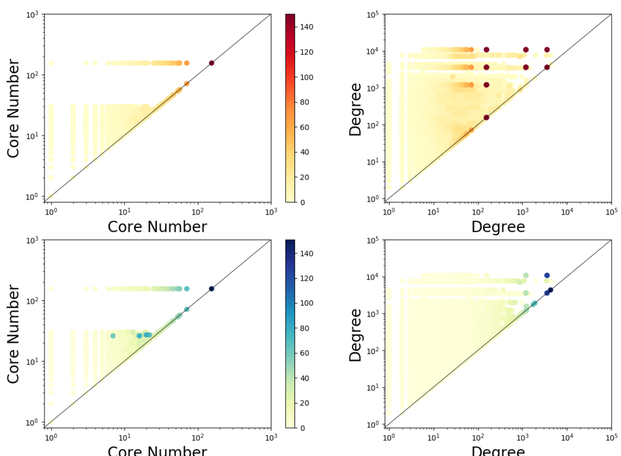
[1] Nesreen K. Ahmed, Jennifer Neville, Ryan A. Rossi, and Nick G. Duffield. 2015. Efficient Graphlet Counting for Large Networks. In *2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015.* 1–10. http://dx.doi.org/10.1109/ICDM.2015.141



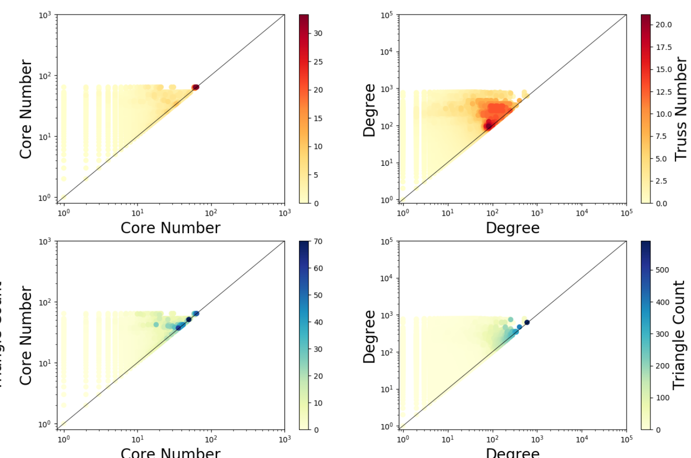
(a) Email



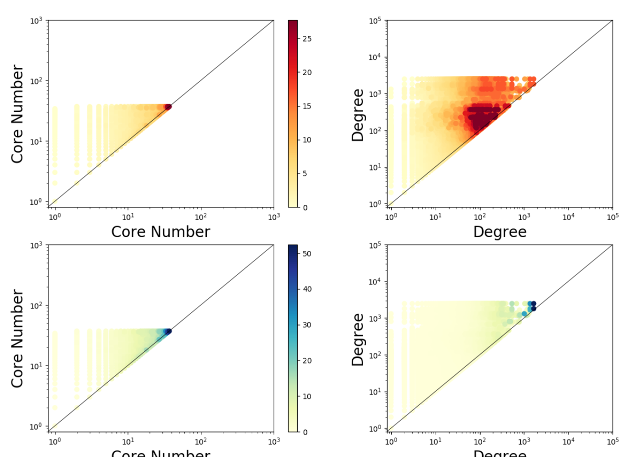
(b) BerkStan



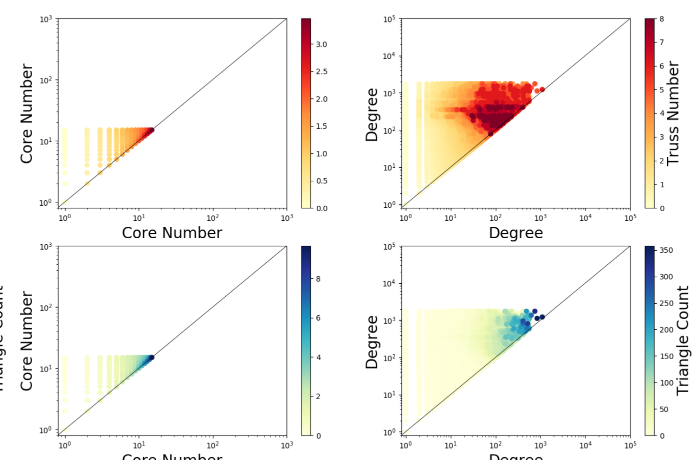
(c) NotreDame



(d) Patent



(e) HepTh



(f) Citeseer

Figure 4: The Edge Distribution. Each edge is represented as a scatter point, with respect to the two vertices that edge connected to. Without lost of generality, the vertex with smaller core number or degree is denoted by the X coordinate, while the large one is denoted by the Y coordinate. The coordinates of two charts on the left represent the core number of the two vertices, while the coordinates of two charts on the right represent the degree. The color represents the average truss number in the upper two charts, and the triangles count in the bottom two charts.

[2] V. Batagelj and M. Zaversnik. 2003. *An O(m) Algorithm for Cores Decomposition of Networks*. Technical Report cs/0310049. Arxiv.

[3] Austin R. Benson, David F. Gleich, and Jure Leskovec. 2016. Higher-order organization of complex networks. *Science* 353, 6295 (2016), 163–166. https://doi.org/10.1126/science.aad9029

[4] J. Cohen. 2008. Trusses: Cohesive subgraphs for social network analysis. National Security Agency Technical Report (2008).

[5] P. Erdős and A. Hajnal. 1966. On chromatic number of graphs and set-systems. *Acta Mathematica Hungarica* 17 (1966), 61–99.

[6] Shweta Jain and C Seshadhri. 2017. A Fast and Provable Method for Estimating Clique Counts Using Turán's Theorem. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 441–449.

[7] Madhav Jha, C. Seshadhri, and Ali Pinar. 2015. Path Sampling: A Fast and Provable Method for Estimating 4-Vertex Subgraph Counts. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. 495–505. https://doi.org/10.1145/2736277.2741101

[8] Jérôme Kunegis. 2013. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1343–1350.

[9] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets.

[10] Michael Ley. 2002. The DBLP computer science bibliography: Evolution, research issues, perspectives. In *International symposium on string processing and information retrieval*. Springer, 1–10.

[11] D. Matula and L. Beck. 1983. Smallest-last ordering and clustering and graph coloring algorithms. *JACM* 30, 3 (1983), 417–427.

[12] Ali Pinar, C. Seshadhri, and V. Vishal. 2016. ESCAPE: Efficiently Counting All 5-Vertex Subgraphs. *CoRR* abs/1610.09411 (2016). http://arxiv.org/abs/1610.09411

[13] Ryan A. Rossi, Rong Zhou, and Nesreen K. Ahmed. 2017. Estimation of Graphlet Statistics. *CoRR* abs/1701.01772 (2017). http://arxiv.org/abs/1701.01772

[14] K. Saito and T. Yamada. 2006. Extracting Communities from Complex Networks by the k-dense Method. In *ICDMW*.

[15] A. E. Sarıyüce, C. Seshadhri, A. Pınar, and Ü. V. Çatalyürek. to appear. Nucleus Decompositions for Identifying Hierarchy of Dense Subgraphs. *ACM Transactions on Web (TWEB)* (to appear).

[16] S. B. Seidman. 1983. Network structure and minimum degree. *Social Networks* 5, 3 (1983), 269–287.

[17] Kijung Shin, Tina Eliassi-Rad, and Christos Faloutsos. 2018. Patterns and anomalies in k-cores of real-world graphs with applications. *Knowledge and Information Systems* 54, 3 (2018), 677–710.

[18] Charles Spearman. 1904. The proof and measurement of association between two things. *American journal of Psychology* 15, 1 (1904), 72–101.

[19] Amanda L. Traud, Peter J. Mucha, and Mason A. Porter. 2012. Social structure of Facebook networks. *Physica A: Statistical Mechanics and its Applications* 391, 16 (2012), 4165 – 4180.

[20] C. Tsourakakis. 2015. The K-clique Densest Subgraph Problem. In *Proc. of the 24th International Conf. on World Wide Web (WWW '15)*. 1122–1132. http://dl.acm.org/citation.cfm?id=2736277.2741098

[21] Anurag Verma and Sergiy Butenko. 2012. Network clustering via clique relaxations: A community based approach. In *DIMACS Graph Part. and Clustering*. 129–140.

[22] Y. Zhang and S. Parthasarathy. 2012. Extracting Analyzing and Visualizing Triangle K-Core Motifs Within Networks. In *ICDE*. 1049–1060.