# Scalable Subgraph Counting: Methods Behind the Madness
# WWW 2019 Tutorial

C. Seshadhri*
University of California, Santa Cruz, CA
sesh@ucsc.edu

Srikanta Tirthapura†
Iowa State University, Ames, IA
snt@iastate.edu

**Abstract**

This is an auxiliary file containing references to all the papers mentioned in our WWW 2019 tutorial. For each paper, we give a short 1-2 sentence description of the relevance to the field of subgraph counting. We have a separate section for each portion of our tutorial, and papers are listed in chronological order within each section.

## 1 Introduction

We list (most) papers cited in the introductory slides of the tutorial. These are primarily papers on applications of subgraph counting, and a few papers on subgraph counting on attributed graphs. The latter list is by no means comprehensive and merely lists a few papers.

A few papers mentioned in the introduction are talked about in detail, during later sections of tutorial (with regards to algorithmic techniques). We do not mention these papers in this section, and instead cite them in the later relevant section.

- [HL70]: Early social science paper about triad census in social networks. Focus on directed triangles.
- [She71]: Tech report style survey, on use of directed triads. The applications are for international politics.
- [WF94]: Sociology textbook that has early reference to graph transitivity.
- [WS98]: Classic network science paper that defines the clustering coefficients.
- [CL99]: Matrix multiplication algorithm using path sampling. One of the earliest applications of sampling paths in a graph, for solving matrix problems.
- [MSI+02]: Paper defining network motifs. More focus on directed patterns, but uses these to distinguish real-world graphs from random graphs.
- [Bur04]: Sociology paper on "structural holes". These are higher-degree vertices that do not participate in triangles, suggesting the vertex is an information broker.
- [Prz07]: The paper defining the graphlet degree distribution. Fix an orbit. For any vertex $v$, define the graphlet degree to the number of times $v$ appears in that orbit. This can be normalized as a distribution, to compare graphs.
- [SVP+09]: Paper defining the graphlet kernel. This is a distance metric based on the vector of small subgraph counts.
- [SCW+10]: Paper comparing various models, based on properties such as clustering coefficients. One of the older papers empirically showing that Preferential-Attachment has poor clustering.
- [ST10]: Paper doing analysis of Personal Message/Friend/Enemy networks in an MMORPG (Pardus) using clustering coefficients.

*Supported by NSF TRIPODS grant CCF-1740850 and NSF CCF-1813165
†Supported by NSF grants 1527541 and 1725702

- [BHLP11]: Result on using 4-cycle for edge weights, to improving community detection based on modularity
- [SKP12]: Paper using degree-wise clustering coefficients for fitting graph model.
- [SKK+12]: Another paper on study of triadic closure in MMORPG (AION) network.
- [IFMN12]: Paper using triadic closure for graph model.
- [UBK13]: Paper using 4-vertex subgraph counts for understanding social network structure.
- [IMF+14]: Paper on attributed graph models, that uses clustering coefficients to check model fit. The model accounts for clustering in the graph generation process.
- [Tso15]: Paper on dense subgraph discovery using clique counts as vertex weights.
- [BKPS15]: Nearest neighbor search algorithms from sampling 4-cycles. A successor to [CL99].
- [ZG13]: A distributed algorithm for implementation wedge sampling for solving nearest neighbor problems.
- [WFC14]: Paper on counting 4-cycles in bipartite graphs.
- [SSPC15]: Another dense subgraph discovery algorithm using clique counts as vertex and edge weights. Philosophically similar to [Tso15]. but different algorithms and analysis.
- [BGL16]: One of the papers defining motif conductance, gives spectral algorithm that uses local triangle (and other subgraph) counts as edge weight.
- [YYW+16]: Paper on counting temporal subgraphs.
- [TPM17]: Another paper on motif conductance, somewhat of independent discovery of basic definitions from [BGL16]. The algorithms discovered are different.
- [PBL17]: Defines temporal motifs, and gives exact algorithms. Some dynamic programming tricks used, and the focus is on 3-vertex patterns.
- [RKKS17]: Paper using local 4-cycle and 5-cycle counts for weak tie prediction.
- [SSG17]: Another paper on using wedge/path sampling for nearest neighbor search.
- [PBL17]: Paper on counting temporal subgraphs.
- [SST18]: Paper on approximating 4-cycle counts in bipartite graphs
- [YBL18]: Defines higher-order clustering coefficients. The $\ell$-wedge used is an $\ell$-clique, with an adjacent edge. "Closure" means that this is part of an $(\ell+1)$-clique.
- [LBC19]: This paper gives a sampling algorithm for counting temporal subgraphs. The sampling is over the time interval.
- [YBL19]: This paper defines local version of higher-order clustering coefficients. It basically requires local versions of 4, 5-vertex subgraph counting.
- [RAC+19]: Recent paper on counting colored/heterogenous subgraphs. Focus on $k = 3, 4$, uses cutting framework approach to get exact equations for the different orbits.

## 2 Graph orientations

This is a list of papers using graph orientations for subgraph counting. We have mentioned the first few theoretical papers that discovered these ideas first, but all other papers perform empirical studies (or give code).

- [CN85]: The paper that started it all. This is the earliest paper on using orientations for clique counting, and gives the classic $O(m\alpha^{k-2})$ bound. While the algorithm's running time depends on the degeneracy, the algorithm itself only uses the degree ordering. The paper does not explicitly talk of orientations.
- [CE91]: The first paper that uses the core decomposition (though it never uses this specific term) and discusses the problem of triangle counting in terms of orientations.
- [SW05b]: (Also refer to thesis [Sch07]). This gives an independent discovery of the orientation algorithm with degeneracy ordering. This paper considers many variants of the basic triangle counting scheme, and has detailed comparisons. Their fastest algorithm `forward` is basically `TriOrient` with the degree ordering, using a clever implementation.

• [Coh09]: This rediscovers the triangle counting and 4-cycle counting algorithms of Chiba-Nishizeki, but provides Map-Reduce implementations. The interpretation of the 4-cycle counting is much cleaner than that of Chiba-Nishizeki, and is implicitly in terms of orientations.

• [SV11]: This follows the language of Schank-Wagner [SW05b], and one of the results can be thought of as a rediscovery of Cohen's algorithm [Coh09]. There are new ideas to reduce the skew of the computation using sampling methods.

• [AKM13]: Uses degree ordering for MPI implementation, and outperforms Map-Reduce implementations.

• [ELS13]: This paper is on maximal clique enumeration, and uses the degeneracy orientation to get smaller subproblems for clique counting.

• [FFF15]: This paper gives a Map-Reduce algorithm for counting $k$-cliques (for $k < 9$). The basic idea is similar to [ELS13], where degeneracy is used to construct smaller subproblems, which can be parallelized. There are many detailed Map-Reduce experiments, and the final scalable algorithm is obtained by using sampling as well. The sampling uses color coding, and counts monochromatic cliques.

• [JS17]: This paper gives an approximation algorithm to estimate the $k$-clique counting. Like the previous papers, it uses the degeneracy ordering to construct smaller subproblems.

• [PSV17]: This paper gives exact algorithms that count all 5-vertex subgraph counts. Orientations are used to counting cycles and clique-like subgraphs.

• [DBS18]: This paper gives a parallel implementation of the original Chiba-Nishizeki algorithm [CN85] using degree orientations. The aim is to get a multi-threaded algorithm that enumerates $k$-cliques (for $k \leq 10$).

# 3 Subgraph reconstruction

This is the list of papers mentioned on subgraph reconstruction. Following the tutorial, we mention some previous work. These are earlier subgraph counting packages/ideas that were superseded by subgraph reconstruction algorithms. By and large, they are based on subgraph reconstruction techniques.

• [WR06]: The FANMOD paper. This was considered a state of the art package for subgraph counting. It is based on sampling from the brute-force recursion tree that exactly counts subgraphs.

• [MS10]: The RAGE paper, which gives an exact algorithm for counting 4-vertex subgraphs.

• [RS12]: The G-Tries paper. This paper designs a data structure for subgraph search, inspired by data structures for string searching (tries). This builds a more sophisticated version of a recursion tree that searches for all matches of a given subgraph.

• [HD14]: The ORCA paper. This paper gives a method to construct linear equations, where variables are orbit counts and counts of neighborhood intersections (among pairs/triples). The code is specific to 4, 5-vertex subgraph orbits. As of 2019, it is probably the state-of-the-art for this problem.

• [ANRD15]: The PGD paper. This paper uses inclusion/exclusion methods to count 4-vertex subgraphs. One of the key ideas is to get sums of counts of different subgraphs by combinatorial formulas. The code has support for parallelism through an OpenMP implementation.

• [MAM+16]: This paper generalizes the linear equation technique of [HD14]. It gives code that automates the equation generation process, and is able to give these equations for 6-vertex orbits. The result does not give algorithms or code for subgraph counting.

• [PSV17]: The ESCAPE paper. This paper combines orientations with subgraph reconstruction techniques for counting 4, 5-vertex subgraph counts. It is probably the fastest solution for this problem.

# 4 Color Coding

• [AYZ95] is the original paper on color coding, which introduced the technique in the context of *finding* a subgraph isomorphic to a pattern graph (paths or trees) on a small number of vertices.

• [ADH+08] is an application of color coding to *counting* subgraphs (paths, trees, graphs with small tree-width) in the context of applications in bioinformatics.

• [HWZ08] present algorithm engineering ideas for speeding up color coding by increasing the number of colors and efficient storage of color sets.

• [ZWB+12] parallelizes color coding for counting trees to the Hadoop platform.

• [SM15] present methods for parallelizing color coding both using shared memory parallelism (threads) and distributed memory parallelism (MPI).

• [CKM+16] explore bounded treewidth graphs further, and combine with ideas from orientations.

• [BCK+18] is further described later, proves a concentration bound for the estimates of graphlets and also obtains a tradeoff between space and time.

# 5 Edge Sampling

• [BYKS02] is one of the earliest works on algorithms and lower bounds for the problem of counting triangles from a graph stream. This also introduced the models of graph streams, including arbitrary order (which they call "adjacency stream") and adjacency-list order (which they call "incidence stream").

• [BFL+06] used the idea of edge + vertex sampling to improve the space complexity of estimating the number of triangles from a graph stream when compared with [BYKS02].

• [TKMF09, TKM11] present the independent edge sampling algorithm (which we call as Bernoulli edge sparsification) and an analysis.

• [PT12] present the monochromatic edge sampling algorithm in the context of counting triangles in a massive graph, and an analysis of this algorithm using two different methods. This algorithm can also be applied in the streaming model.

• [PTTW13] present the "neighborhood sampling" technique that yields one of the current best tradeoffs between space and accuracy for triangle counting in graphs streams.

• [KP13] present algorithms for computing local clustering coefficients of a graph using application of Bernoulli edge sparsification and monochromatic edge sparsification.

• [BOV13] presents lower bounds on the streaming space complexity of triangle counting, and provides a lower bound that is parameterized by $T$ (a lower bound on the number of triangles), as opposed to an unparameterized lower bound due to [BYKS02].

• [ADNK14] presents the "graph sample and hold" framework for subgraph estimation, which allows non-uniform sampling of edges from a graph stream, with varying probabilities for edges that are adjacent to currently sampled edges, and edges that are not.

• [LK15] present algorithms for *local* triangle counting based on random sampling, using an algorithm similar to wedge sampling [SPK13].

• [MVV16] is on multi-pass algorithms for counting triangles in a graph stream, and so is [CJ17].

• [SERU16] present algorithms for local and global triangle counting based on reservoir sampling of edges. It builds on [LK15, SPK13] and the technical aspect is to handle the dependencies induced by reservoir edge sampling rather than Bernoulli edge sampling.

• [BC17] presents lower and upper bounds for multi-pass streaming estimation of triangle and other subgraph counts. Their lower bounds show that the algorithm due to [PTTW13] is essentially optimal in space.

• [Shi17] presents a method that takes advantage of temporal locality of edges to efficiently count triangles in a stream – works well when triangles are formed by edges that arrive close to each other in stream order.

• [SST18] applied Bernoulli edge sparsification and other methods of substructure sampling to estimate the number of butterflies (4-cycles) in a bipartite graph.

## 6 Substructure sampling

This is the list of papers on substructure sampling. Some of these have been referenced in earlier sections. For completeness, we mention them again, and explain the connection to substructure sampling.

• [SW05a]: This is the first paper that uses the wedge sampling technique for estimating clustering coefficients. Technically, the idea is a special case of sampling from a matrix product, as given in [CL99].

• [SPK13]: The paper that coins the phrase "wedge sampling", and does detailed experiments to show its power in estimating various triangle statistics. This was a (late) rediscovery of the main technique, as given in [SW05a], and the main math overlaps heavily with [SW05a].

• [JSP15]: This paper generalizes the technique to 3-path sampling, for estimating 4-vertex subgraph counts.

• [TT17]: This paper improves on the vanilla wedge sampling convergence rate by sampling from a biased distribution. The idea is to (implicitly) use degree orientations to cut down the space of wedges being sampled from. While this leads to a bias, this can be corrected using standard methods.

• [JS17]: This paper implicitly uses a substructure sampling method to estimate the $k$-clique count (for $k \leq 10$). The substructure sampled is a dense subgraph, and the method to sample these is significantly more involved than previous substructure sampling results. This paper uses the orientation technique to achieve better performance.

• [BCK+18]: This paper combines color coding with spanning tree sampling to get estimates of subgraphs beyond 5 vertices. It is a combinations of techniques, and the specific purpose was to go beyond 5-vertex subgraph counting and beat MCMC methods for this purpose. The results show that color coding is quite effective, but can be space intensive.

• [WZZ+18]: This is the MOSS-5 paper. This result extends the path sampling methods to tree sampling for estimating 5-vertex subgraph counts. Just as [JSP15], the sampling method gives massive improvements over exact counting algorithms.

## References


[ADH+08] Noga Alon, Phuong Dao, Iman Hajirasouliha, Fereydoun Hormozdiari, and Süleyman Cenk Sahinalp. Biomolecular network motif counting and discovery by color coding. In *Intelligent Systems for Molecular Biology (ISMB)*, pages 241–249, 2008. 4

[ADNK14] N. K. Ahmed, N. Duffield, J. Neville, and R. Kompella. Graph sample and hold: A framework for big graph analytics. In *Knowledge Data and Discovery (KDD)*, 2014. 4

[AKM13] S. M. Arifuzzaman, M. Khan, and M. Marathe. Patric: A parallel algorithm for counting triangles and computing clustering coefficients in massive networks. In *Conference on Information and Knowledge Management (CIKM)*, pages 529–538, 2013. 3

[ANRD15] Nesreen K. Ahmed, Jennifer Neville, Ryan A. Rossi, and Nick Duffield. Efficient graphlet counting for large networks. In *International Conference on Data Mining (ICDM)*, 2015. 3

[AYZ95] Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *J. ACM*, 42(4):844–856, 1995. 3

[BC17] Suman K. Bera and Amit Chakrabarti. Towards tighter space bounds for counting triangles and other substructures in graph streams. In *Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 11:1–11:14, 2017. 4

[BCK+18] Marco Bressan, Flavio Chierichetti, Ravi Kumar, Stefano Leucci, and Alessandro Panconesi. Motif counting beyond five nodes. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(4):48:1–48:25, 2018. 4, 5

[BFL+06] Luciana S. Buriol, Gereon Frahling, Stefano Leonardi, Alberto Marchetti-Spaccamela, and Christian Sohler. Counting triangles in data streams. In *Principles of Database Systems (PODS)*, pages 253–262, 2006. 4

[BGL16] A. Benson, D. F. Gleich, and J. Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016. 2

[BHLP11] Jonathan W. Berry, Bruce Hendrickson, Randall A. LaViolette, and Cynthia A. Phillips. Tolerating the community detection resolution limit with edge weighting. *Phys. Rev. E*, 83:056119, May 2011. 2

[BKPS15] Grey Ballard, Tamara G. Kolda, Ali Pinar, and C. Seshadhri. Diamond sampling for approximate maximum all-pairs dot-product (MAD) search. In *International Conference on Data Mining (ICDM)*, pages 11–20, 2015. 2

[BOV13] Vladimir Braverman, Rafail Ostrovsky, and Dan Vilenchik. How hard is counting triangles in the streaming model? In *International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 244–254, 2013. 4

[Bur04] R. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399, 2004. 1

[BYKS02] Ziv Bar-Yossef, Ravi Kumar, and D. Sivakumar. Reductions in streaming algorithms, with an application to counting triangles in graphs. In *Symposium on Discrete Algorithms (SODA)*, pages 623–632, 2002. 4

[CE91] Marek Chrobak and David Eppstein. Planar orientations with low out-degree and compaction of adjacency matrices. *Theor. Comput. Sci.*, 86(2):243–266, 1991. 2

[CJ17] Graham Cormode and Hossein Jowhari. A second look at counting triangles in graph streams (corrected). *Theor. Comput. Sci.*, 683:22–30, 2017. 4

[CKM+16] Venkatesan T. Chakaravarthy, Michael Kapralov, Prakash Murali, Fabrizio Petrini, Xinyu Que, Yogish Sabharwal, and Baruch Schieber. Subgraph counting: Color coding beyond trees. In *International Parallel and Distributed Processing Symposium (IPDPS)*, pages 2–11, 2016. 4

[CL99] Edith Cohen and David D. Lewis. Approximating matrix multiplication for pattern recognition tasks. *J. Algorithms*, 30(2):211–252, 1999. 1, 2, 5

[CN85] Norishige Chiba and Takao Nishizeki. Arboricity and subgraph listing algorithms. *SIAM J. Comput.*, 14:210–223, 1985. 2, 3

[Coh09] Jonathan Cohen. Graph twiddling in a MapReduce world. *Computing in Science & Engineering*, 11:29–41, 2009. 3

[DBS18] Maximilien Danisch, Oana Denisa Balalau, and Mauro Sozio. Listing k-cliques in sparse real-world graphs. In *Conference on World Wide Web (WWW)*, pages 589–598, 2018. 3

[ELS13] David Eppstein, Maarten Löffler, and Darren Strash. Listing all maximal cliques in large sparse real-world graphs. *ACM Journal of Experimental Algorithmics*, 18, 2013. 3

[FFF15] Irene Finocchi, Marco Finocchi, and Emanuele G. Fusco. Clique counting in mapreduce: Algorithms and experiments. *ACM Journal of Experimental Algorithmics*, 20, 2015. 3

[HD14] Tomaz Hocevar and Janez Demsar. A combinatorial approach to graphlet counting. *Bioinformatics*, 2014. 3

[HL70] P. Holland and S. Leinhardt. A method for detecting structure in sociometric data. *American Journal of Sociology*, 76:492–513, 1970. 1

[HWZ08] Falk Hüffner, Sebastian Wernicke, and Thomas Zichner. Algorithm engineering for color-coding with applications to signaling pathway detection. *Algorithmica*, 52(2):114–132, 2008. 4

[IFMN12] Joseph J. Pfeiffer III, Timothy La Fond, Sebastián Moreno, and Jennifer Neville. Fast generation of large scale social networks while incorporating transitive closures. In *International Conference on Privacy, Security, Risk and Trust (PASSAT)*, pages 154–165, 2012. 2

[IMF+14] Joseph J. Pfeiffer III, Sebastián Moreno, Timothy La Fond, Jennifer Neville, and Brian Gallagher. Attributed graph models: modeling network structure with correlated attributes. In *Conference on World Wide Web (WWW)*, pages 831–842, 2014. 2

[JS17] Shweta Jain and C. Seshadhri. A Fast and Provable Method for Estimating Clique Counts Using Turán's Theorem. In *Conference on World Wide Web (WWW)*, pages 441–449, 2017. 3, 5

[JSP15] Madhav Jha, C. Seshadhri, and Ali Pinar. Path sampling: A fast and provable method for estimating 4-vertex subgraph counts. In *Conference on World Wide Web (WWW)*, 2015. 5

[KP13] Konstantin Kutzkov and Rasmus Pagh. On the streaming complexity of computing local clustering coefficients. In *Web Search and Data Mining (WSDM)*, pages 677–686, 2013. 4

[LBC19] Paul Liu, Austin R. Benson, and Moses Charikar. Sampling methods for counting temporal motifs. In *Web Search and Data Mining (WSDM)*, pages 294–302, 2019. 2

[LK15] Yongsub Lim and U. Kang. MASCOT: memory-efficient and accurate sampling for counting local triangles in graph streams. In *Knowledge Data and Discovery (KDD)*, pages 685–694, 2015. 4

[MAM+16] Ine Melckenbeeck, Pieter Audenaert, Tom Michoel, Didier Colle, and Mario Pickavet. An algorithm to automatically generate the combinatorial orbit counting equations. *PLoS One*, 11(1):1–19, 01 2016. 3

[MS10] Dror Marcus and Yuval Shavitt. Efficient counting of network motifs. In *ICDCS Workshops*, pages 92–98, 2010. 3

[MSI+02] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. 1

[MVV16] Andrew McGregor, Sofya Vorotnikova, and Hoa T. Vu. Better algorithms for counting triangles in data streams. In *Principles of Database Systems (PODS)*, pages 401–411, 2016. 4

[PBL17] Ashwin Paranjape, Austin R. Benson, and Jure Leskovec. Motifs in temporal networks. In *Web Search and Data Mining (WSDM)*, pages 601–610, 2017. 2

[Prz07] Natasa Przulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):177–183, 2007. 1

[PSV17] Ali Pinar, C Seshadhri, and Vaidyanathan Vishal. Escape: Efficiently counting all 5-vertex subgraphs. In *Conference on World Wide Web (WWW)*, pages 1431–1440. International World Wide Web Conferences Steering Committee, 2017. 3

[PT12] R. Pagh and C. Tsourakakis. Colorful triangle counting and a mapreduce implementation. *Information Processing Letters*, 112:277–281, 2012. 4

[PTTW13] A. Pavan, K. Tangwongsan, S. Tirthapura, and K.-L. Wu. Counting and sampling triangles from a graph stream. In *International Conference on Very Large Databases (VLDB)*, 2013. 4

[RAC+19] R. Rossi, Nesreen K. Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunyee Koh. Heterogeneous network motifs. Technical Report 1901.10026, arXiv, 2019. 2

[RKKS17] Rahmtin Rotabi, Krishna Kamath, Jon M. Kleinberg, and Aneesh Sharma. Detecting strong ties using network motifs. In *Conference on World Wide Web (WWW)*, pages 983–992, 2017. 2

[RS12] Pedro Ribeiro and Fernando Silva. Querying subgraph sets with g-tries. In *SIGMOD Workshop on Databases and Social Networks*, DBSocial '12, pages 25–30, New York, NY, USA, 2012. ACM. 3

[Sch07] T. Schank. *Algorithmic Aspects of Triangle-Based Network Analysis*. PhD thesis, Universitat Karlsruhe (TH), 2007. 2

[SCW+10] Alessandra Sala, Lili Cao, Christo Wilson, Robert Zablit, Haitao Zheng, and Ben Y. Zhao. Measurement-calibrated graph models for social network experiments. In *Conference on World Wide Web (WWW)*, pages 861–870, 2010. 1

[SERU16] Lorenzo De Stefani, Alessandro Epasto, Matteo Riondato, and Eli Upfal. Trièst: Counting local and global triangles in fully-dynamic streams with fixed memory size. In *Knowledge Data and Discovery (KDD)*, pages 825–834, 2016. 4

[She71] R. Sherwin. Introduction to the graph theory and structural balance approaches to international relations. *World Event/Interaction Survey*, 1971. https://apps.dtic.mil/dtic/tr/fulltext/u2/a080476.pdf. 1

[Shi17] Kijung Shin. WRS: waiting room sampling for accurate triangle counting in real graph streams. In *International Conference on Data Mining (ICDM)*, pages 1087–1092, 2017. 4

[SKK+12] S. Son, A. Kang, H. Kim, T. Kwon, J. Park, and H. Kim. Analysis of context dependence in social interaction networks of a massively multiplayer online role-playing game. *PLoS ONE*, 7(4):e33918, 04 2012. 2

[SKP12] C. Seshadhri, Tamara G. Kolda, and Ali Pinar. Community structure and scale-free collections of Erdös-Rényi graphs. *Physical Review E*, 85(5):056109, May 2012. 2

[SM15] George M. Slota and Kamesh Madduri. Parallel color-coding. *Parallel Computing*, 47:51–69, 2015. 4

[SPK13] C. Seshadhri, Ali Pinar, and Tamara G. Kolda. Fast triangle counting through wedge sampling. In *Proceedings of the SIAM Conference on Data Mining*, 2013. 4, 5

[SSG17] A. Sharma, C. Seshadhri, and A. Goel. When hashes met wedges: A distributed algorithm for finding high similarity vectors. In *World Wide Web (WWW)*, 2017. 2

[SSPC15] Ahmet Erdem Sariyuce, C. Seshadhri, Ali Pinar, and Umit V. Catalyurek. Finding the hierarchy of dense subgraphs using nucleus decompositions. In *Conference on World Wide Web (WWW)*, WWW '15, pages 927–937, New York, NY, USA, 2015. ACM. 2

[SST18] Seyed-Vahid Sanei-Mehri, Ahmet Erdem Sariyüce, and Srikanta Tirthapura. Butterfly counting in bipartite networks. In *Knowledge Data and Discovery (KDD)*, pages 2150–2159, 2018. 2, 4

[ST10] M. Szell and S. Thurner. Measuring social dynamics in a massive multiplayer online game. *Social Networks*, 32:313–329, 2010. 1

[SV11] Siddharth Suri and Sergei Vassilvitskii. Counting triangles and the curse of the last reducer. In *World Wide Web (WWW)*, pages 607–614, 2011. 3

[SVP+09] Nino Shervashidze, S. V. N. Vishwanathan, Tobias Petri, Kurt Mehlhorn, and Karsten M. Borgwardt. Efficient graphlet kernels for large graph comparison. In *Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 488–495, 2009. 1

[SW05a] Thomas Schank and Dorothea Wagner. Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications*, 9:265–275, 2005. 5

[SW05b] Thomas Schank and Dorothea Wagner. Finding, counting and listing all triangles in large graphs, an experimental study. In *Experimental and Efficient Algorithms*, pages 606–609. Springer Berlin / Heidelberg, 2005. 2, 3

[TKM11] C. Tsourakakis, M. N. Kolountzakis, and G. Miller. Triangle sparsifiers. *J. Graph Algorithms and Applications*, 15:703–726, 2011. 4

[TKMF09] Charalampos E. Tsourakakis, U. Kang, Gary L. Miller, and Christos Faloutsos. Doulion: counting triangles in massive graphs with a coin. In *Knowledge Data and Discovery (KDD)*, pages 837–846, 2009. 4

[TPM17] Charalampos E. Tsourakakis, Jakub Pachocki, and Michael Mitzenmacher. Scalable motif-aware graph clustering. In *Conference on World Wide Web (WWW)*, pages 1451–1460, 2017. 2

[Tso15] Charalampos E. Tsourakakis. The k-clique densest subgraph problem. In *Conference on World Wide Web (WWW)*, pages 1122–1132, 2015. 2

[TT17] Duru Türkoglu and Ata Turk. Edge-based wedge sampling to estimate triangle counts in very large graphs. In *International Conference on Data Mining (ICDM)*, pages 455–464, 2017. 5

[UBK13] Johan Ugander, Lars Backstrom, and Jon M. Kleinberg. Subgraph frequencies: mapping the empirical and extremal geography of large graph collections. In *WWW*, pages 1307–1318, 2013. 2

[WF94] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994. 1

[WFC14] Jia Wang, Ada Wai-Chee Fu, and James Cheng. Rectangle counting in large bipartite graphs. In *International Congress on Big Data*, pages 17–24, 2014. 2

[WR06] S. Wernicke and F. Rasche. Fanmod: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006. 3

[WS98] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998. 1

[WZZ+18] Pinghui Wang, Junzhou Zhao, Xiangliang Zhang, Zhenguo Li, Jiefeng Cheng, John C. S. Lui, Don Towsley, Jing Tao, and Xiaohong Guan. MOSS-5: A fast method of approximating counts of 5-node graphlets in large graphs. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 30(1):73–86, 2018. 5

[YBL18] Hao Yin, Austin R. Benson, and Jure Leskovec. Higher-order clustering in networks. *Phys. Rev. E*, 97:052306, 2018. 2

[YBL19] Hao Yin, Austin R. Benson, and Jure Leskovec. The local closure coefficient: A new perspective on network clustering. In *Web Search and Data Mining (WSDM)*, pages 303–311, 2019. 2

[YYW+16] Yi Yang, Da Yan, Huanhuan Wu, James Cheng, Shuigeng Zhou, and John C. S. Lui. Diversified temporal subgraph pattern mining. In *Knowledge Data and Discovery (KDD)*, pages 1965–1974, 2016. 2

[ZG13] Reza Bosagh Zadeh and Ashish Goel. Dimension independent similarity computation. *Journal of Machine Learning Research*, 14(1):1605–1626, 2013. 2

[ZWB+12] Z. Zhao, G. Wang, A. Butt, M. Khan, V. S. Anil Kumar, and M. Marathe. Sahad: Subgraph analysis in massive networks using hadoop. In *Proceedings of International Parallel and Distributed Processing Symposium (IPDPS)*, pages 390–401, 2012. 4