Compression of Preferential Attachment Graphs

Tomasz Łuczak
Adam Mickiewicz University
61-614 Poznań, Poland
Email: tomasz@amu.edu.pl

Abram Magner
University of Michigan
Email: abram10@gmail.com

Wojciech Szpankowski
Purdue University
West Lafayette, IN, USA
Email: spa@cs.purdue.edu

***Abstract*—We study structural properties of preferential attachment graphs (with parameter $m \geq 1$ giving the number of attachment choices that each new vertex makes) which intervene in two complementary algorithmic/statistical/information-theoretic problems involving the information shared between a random graph's labels and its structure: in structural compression, we seek to compactly describe a graph's structure by a bit string, throwing away its label information; in node arrival order recovery, we seek to recover node labels, given only a graph structure. In particular, we study the typical size of the automorphism group, as well as some shape parameters (such as the number of linear extensions and height) of the directed version of the graph, which in turn allows us to estimate the typical number of *admissible labeled representatives* of a given graph structure. Our result on the automorphism group positively settles a conjecture to the effect that, provided that $m \geq 3$, preferential attachment graphs are asymmetric with high probability, and completes the characterization of the number of symmetries for a broad range of parameters of the model (i.e., for all fixed m). These results allow us to give an algorithmically efficient, asymptotically optimal algorithm for compression of unlabeled preferential attachment graphs. To show the optimality of our scheme, we also derive new, precise estimates of the Shannon entropy of both the unlabeled and labeled version of the model. Our results also imply inapproximability results for the problem of node arrival order recovery.**

Index Terms: graph compression, symmetry, preferential attachment, random graphs

I. Introduction

The purpose of this paper is to present mathematical results on structural parameters which are fundamental to statistical and information-theoretic problems involving the information shared between the labels and the structure of a random graph. We first describe two such problems, which are in a sense complementary – compression of graph structures, wherein the goal is to *remove* label information to produce a compact description of a graph structure, and recovery of node arrival order in dynamic networks, wherein the goal is to *recover* label information by examining a graph structure – and then explain the structural parameters involved in their analysis, which form the focus of this work. In a nutshell, we study the question, *how much information about the labels of a random graph is contained in its structure*?

Removing label information – structural compression: Formally, the labeled graph compression problem is as follows: fix a distribution $\mathfrak{G}_n$ on (multi)graphs on n vertices. We would like to exhibit an efficiently computable *source code* [1] $(\mathcal{C}_n, \mathcal{D}_n)$ for $\mathfrak{G}_n$, where $\mathcal{C}_n$ is a function mapping graphs in the support of $\mathfrak{G}_n$ to bit strings, in such a way as to minimize the expected length of the output bit string when the input is a graph distributed according to $\mathfrak{G}_n$, and $\mathcal{D}_n$ inverts $\mathcal{C}_n$ and is efficiently computable. A related problem, and one focus of our paper, seeks to compress graph structures: here, the encoding function $\mathcal{C}_n$ is presented with a multigraph G isomorphic to a sample from $\mathfrak{G}_n$, and $\mathcal{D}_n(\mathcal{C}_n(G))$ is only required to be a labeled multigraph isomorphic to G (that is, only the structural information is preserved). We again insist on a source code with the minimum possible expected code length (which is given by the Shannon entropy of the distribution on unlabeled graphs induced by $\mathfrak{G}_n$; we call this the *structural entropy* of the model).

The structural compression problem is motivated by scenarios in which one only cares to transmit or store information about the isomorphism type of a graph – e.g., its degree sequence, number of occurrences of certain subgraphs, etc. In such scenarios, one *does not* care about labeled graph information, such as the fact that, say, vertex 2 connects to vertex 7. Taking advantage of this fact allows for a more compact description of the relevant information than would result if we naively encoded the labeled graph.

Inferring label information – node arrival order recovery: A complementary problem, *node arrival order recovery* in a dynamic graph, seeks to recover the labels of nodes of a random graph, given its structure. The motivation is as follows: networks in the real world are constructed dynamically, and it is useful to be able to discover node and edge attributes which correlate with time. See [2], [3].

Structural properties: A few structural quantities arise in both of the above problems: as we will see, the structural entropy for a broad class of graph models involves the size of the automorphism group of a sampled graph, as well as the typical number of positive-probability *labeled representatives* of a given structure, and the number of positive-probability

re-labelings (i.e., permutations) of a sampled graph. The same quantities also give lower bounds on the probability of error and the expected number of inversion errors in the node arrival order recovery problem.

We will focus on the analysis of these quantities for *preferential attachment* graphs [4]. Additional structural properties will arise in the analysis of an asymptotically optimal structural compression algorithm which we will give below.

Our contributions: Succinctly, our contributions in this work are threefold: (i) in the setting of preferential attachment graphs, we analyze several structural parameters (explained more precisely below) which arise in both of the motivating problems above and which may be of independent interest; (ii) we use our structural results to precisely determine the entropies of the preferential attachment distributions on both labeled graphs and their structures, giving the fundamental limits of labeled and structural compression; (iii) we give an efficient, asymptotically optimal structural compression algorithm whose analysis relies on our structural results.

The structural properties include the typical size of the automorphism group, as well as some structural characteristics of the *directed* version of the graph (e.g., the number of *admissible labeled representatives* of a given graph structure, which is related to the number of *linear extensions* of the directed version, viewed as a partial order). Our result on the automorphism group (Theorem 1) positively settles a conjecture in [5] to the effect that preferential attachment graphs in which each node makes a sufficiently large number of choices are asymmetric with high probability. This completes the characterization of the number of symmetries for a broad range of parameters of the model.

Regarding structural characteristics of the directed version of the graph (wherein edges are directed from younger nodes to those older nodes that they choose), we analyze a natural partitioning of the vertices into *layers*, which intervenes in the depth-first search process on the directed graph and in the estimation of the number of admissible labeled representatives of the graph (i.e., the number of isomorphic graphs which could have arisen by preferential attachment): in particular, we show that the order of growth of the number of layers is $\Theta(\log n)$ with high probability (see Theorem 3), and almost all vertices occur within the first few layers (Theorem 2). The result on the number of layers is important for our structural compression algorithm (summarized in Theorem 6). We use the above results to provide new, precise estimates of the Shannon entropy of both the labeled and unlabeled models (Theorems 4 and 5).

Full proofs can be found in the journal version [6] of this work. In this conference version, we also present new results on compression algorithms and relevant structural parameters.

Prior work: The general connection between structural compression and the automorphism group of a random graph was pointed out in [7] in the case of unlabeled Erdős-Rényi graphs. The relation between the node arrival order recovery problem, automorphisms, and feasible labeled representatives was pointed out in [2] (but we connect the latter quantity to graph compression in the present work).

There has been significant work on compression of labeled graph and tree models in recent years in both the information theory and computer science communities [8], [9], [10], [7], [11], [12]. In the computer science community, the focus has been on algorithmic complexity, and no attempt seems to have been made to compare with or derive fundamental information- theoretic limits. Work in both communities has largely been restricted to labeled graphs or graphs with strong edge independence assumptions. As we show, additional complications arise when the goal is graph structure compression.

There have been many extensions of the preferential attachment model (as well as models which adopt completely different mechanisms) to provide better fits for certain aspects of real networks: e.g., [13]. It is likely that many of our techniques and results adapt to certain parameter ranges of models extending preferential attachment; we restrict to the plain preferential attachment model (which, in any case, continues to be studied), since the analysis in even this case is quite involved.

Technical challenges: Most of the technical challenge in the entropy derivation and the analysis of the compression algorithm stems from the necessary analysis of the relevant combinatorial quantities. In particular, showing asymmetry for a typical preferential attachment graph requires a detailed understanding of the probability distribution of individual vertex degrees – namely, that there is an initial unique degree region of polynomial size, and that, roughly speaking, all subsequent vertices behave toward this region in a unique way (see Theorem 1). Accomplishing this requires *left* tail bounds on the degree of a vertex at a given time; as vertex degrees, properly normalized, are asymptotically exponentially distributed (and, thus, poorly concentrated), proving these bounds requires techniques beyond the standard Chernoff, Hoeffding, Azuma-style approaches.

Regarding counting admissible relabelings, the conceptual key is tying them (almost) bijectively to linear extensions of the directed version D of the graph and asymptotically counting these objects. Doing this, in turn, requires identification of further structure in D, as described above. Results along the same lines drive the analysis of the asymptotically optimal compression algorithm.

II. Main results

We now introduce the model that we consider and formulate the main results. The *preferential attachment model* $\mathcal{PA}(m; n)$

is a dynamic model of network growth proposed in [4]. For an integer parameter $m \geq 1$ we define the graph $\mathcal{PA}(m;n)$ with vertex set $[n] = \{1, 2, \ldots, n\}$ inductively on n in the following way: the graph $G_1 \sim \mathcal{PA}(m;1)$ is a single node with label 1 with m self-edges (these will be the only self-edges in the graph, and we will only count each such edge once in the degree of vertex 1). Inductively, to obtain a graph $G_{n+1} \sim \mathcal{PA}(m;n+1)$ from G_n, we add vertex $n+1$ and make m random choices (with replacement) $v_1, ..., v_m$ of neighbors in G_n as follows: for each vertex $w \leq n$ (i.e., vertices in G_n), $\mathbb{P}(v_i = w | G_n, v_1, ..., v_{i-1}) = \frac{\deg_n(w)}{2mn}$, where throughout the paper we denote by $\deg_n(w)$ the degree of vertex $w \in [n]$ in the graph G_n (in other words, the degree of w after vertex n has made all of its choices). Our proof techniques adapt to tweaks of the model in which multiple edges are not allowed.

For any graph G, we denote by $S(G)$ its unlabeled version (i.e., the equivalence class of all labeled graphs isomorphic to G). Our structural compression/entropy results will be concerned with the unlabeled preferential attachment model, defined by first generating $G \sim \mathcal{PA}(m;n)$, then taking $S(G)$.

A. Entropy estimates and structural results

Our first concern will be to derive the fundamental lower bound on the expected code length for compression of unlabeled preferential attachment graphs, as described above. As usual, this is given by the Shannon entropy of the distribution on unlabeled graphs induced by $\mathcal{PA}(m;n)$. We are thus interested in $H(S(G))$, where $G \sim \mathcal{PA}(m;n)$.

By the chain rule for conditional entropy, $H(G) = H(S(G)) + H(G|S(G))$. The second term, $H(G|S(G))$, measures our uncertainty about the labeled graph if we are given its structure. We will give a formula for $H(G|S(G))$ in terms of the automorphism group $|\mathrm{Aut}(G)|$ and another quantity, defined as follows: suppose that, after generating G, we relabel G by drawing a permutation π uniformly at random from $\mathbb{S}_n$, the symmetric group on n letters, and computing $\pi(G)$. Then conditioning on $\pi(G)$ yields a probability distribution for possible values of $\pi^{-1} = \sigma$. We can write $H(G|S(G))$ in terms of $H(\sigma|\sigma^{-1}(G)) = H(\sigma|\sigma(G))$ (intuitively, the amount of uncertainty about the value of the random permutation σ upon seeing the result of its application to G) and $\mathbb{E}[\log|\mathrm{Aut}(G)|]$ using the chain rule for entropy, resulting in the following lemma (which is not specific to preferential attachment models).

Lemma 1 (Structural entropy for preferential attachment graphs)**.** *Let $G \sim \mathcal{PA}(m;n)$ for fixed $m \geq 1$, and let σ be a uniformly random permutation from $\mathbb{S}_n$. Then we have*

$$H(G) - H(S(G)) = H(\sigma|\sigma(G)) - \mathbb{E}[\log|\mathrm{Aut}(G)|]. \quad (1)$$

To evaluate $H(S(G))$ and to analyze our compression algorithms, we are thus led to evaluate $\mathbb{E}[\log|\mathrm{Aut}(G)|]$, $H(\sigma|\sigma(G))$, and $H(G)$. The next few results give the structural properties that we need for this. The term $H(\sigma|\sigma(G))$ has multiple interpretations: defining $\Gamma(G)$ to be the set of relabelings of G which produce positive-probability graphs under preferential attachment, we have (at least asymptotically) $H(\sigma|\sigma(G)) = \mathbb{E}[\log|\Gamma(G)|]$. This, in turn, is related to the number of linear extensions of the directed version of G, viewed as a partial order.

Structural results: The proof of Theorem 5 (our expansion of $H(S(G))$) below and the analyses of our algorithms depend on the following structural results.

The next theorem says that with high probability G has no symmetries when $m \geq 3$. As mentioned in the introduction, this essentially completes the analysis of the precise behavior of the number of symmetries of $\mathcal{PA}(m;n)$ for constant m. We will mostly focus on the case $m \geq 3$, since the behaviors for $m = 1, 2$ are qualitatively different (for $m = 1, 2$, there are many symmetries with high probability and with asymptotically positive probability, respectively).

Theorem 1 (Asymmetry for preferential attachment model)**.** *Let $G \sim \mathcal{PA}(m;n)$ for fixed $m \geq 3$. Then, with high probability as $n \to \infty$, $|\mathrm{Aut}(G)| = 1$. More precisely, for $m \geq 3$, $\mathbb{P}(|\mathrm{Aut}(G)| > 1) = O(n^{-\delta})$, for some fixed $\delta > 0$.*

We will also state some results on the *directed* version of G (denoted by $\mathrm{DAG}(G)$). This is the directed multigraph defined on $[n]$, with an edge from w to the older node $v < w$ for each edge between v and w in G. We can partition the vertices of $\mathrm{DAG}(G)$ into *levels* inductively as follows: L_1 consists of the vertices with in-degree 0 (i.e., with total degree m). Inductively, L_j is the set of vertices incident on edges coming from vertices in L_{j-1}. Equivalently, a vertex w is an element of some level $\geq j$ if and only if there exist vertices $v_1 < \cdots < v_j$ such with $v_1 > w$ and the path $v_j v_{j-1} \cdots v_1 w$ exists in G. The *height* of $\mathrm{DAG}(G)$ is then defined to be the number of levels in this partition.

The next result says that almost all of the vertices are concentrated within the first few levels. This will be instrumental in the proof of Theorem 5.

Theorem 2. *For any $\delta = \delta(n) > 0$, there exists $\ell = \ell(\delta)$ for which the number of vertices that are not in the first ℓ layers of $\mathrm{DAG}(G)$ is at most δn, with high probability. In particular, we can take $\ell \geq \frac{15m}{2\delta^4} \log(3/(2\delta^2))$.*

Next, we find the order of growth of the typical height of $\mathrm{DAG}(G)$, which will be useful in the analysis of our structural compression algorithm.

Theorem 3 (Height of $\mathrm{DAG}(G)$)**.** *Consider $G_n \sim \mathcal{PA}(m;n)$ for fixed $m \geq 1$. Then, with probability at least $1 - o(n^{-1})$, the height of $\mathrm{DAG}(G_n)$ is at most $Cm \log n$, for some absolute*

positive constant C.

It is simple to show that with high probability the height is also lower bounded by $\Omega(\log n)$.

Using these results, we will be able to connect $H(\sigma|\sigma(G))$ in (1) to a combinatorial parameter of $\mathrm{DAG}(G)$ (the number of *linear extensions* of $\mathrm{DAG}(G)$, viewed as a partial order), which we will be able to show is estimated by $n\log n + R(n)$, where $C_1 n \le |R(n)| \le C_2 n \log\log n$.

Entropy results: We next evaluate $H(G)$.

Theorem 4 (Entropy of preferential attachment graphs)**.** *Consider $G \sim \mathcal{PA}(m;n)$ for fixed $m \ge 1$. We have*

$$H(G) = mn\log n + m\left(\log 2m - 1 - \log m! - A\right)n + o(n), \tag{2}$$

where $A = A(m) = \sum_{d=m}^{\infty} \frac{\log d}{(d+1)(d+2)}$.

This is a more precise analysis than the one given in [14], which only recovers the first term and the order of the second.

Using the above results, we finally have the following expression for $H(S(G))$.

Theorem 5 (Structural entropy of preferential attachment graphs)**.** *Consider $G \sim \mathcal{PA}(m;n)$ for fixed $m \ge 3$. We have*

$$H(S(G)) = (m-1)n\log n + R(n), \tag{3}$$

where $R(n)$ satisfies $Cn \le |R(n)| \le O(n\log\log n)$ for some nonzero constant $C = C(m)$.

Compared with the naive encoding method which simply stores a labeled representative of the structure using $mn\log(mn)$ bits, the structural entropy is smaller by $n\log n(1+o(1))$ bits.

B. Optimal compression algorithms

We now give our results on efficient algorithms for compression and decompression of unlabeled/labeled samples from $\mathcal{PA}(m;n)$ which asymptotically achieve the entropies.

First, we give an asymptotically optimal algorithm for compression of unlabeled graphs (see Theorem 6 below): that is, given an arbitrary labeled representative G isomorphic to $G' \sim \mathcal{PA}(m;n)$, we construct a code from which $S(G')$ can be efficiently recovered. The algorithm can be run on general undirected graphs; our optimality guarantee is under the assumption that the input is generated by $\mathcal{PA}(m;n)$.

Structural compression algorithm. We state our algorithm and analyze it in the case where the model is preferential attachment with m self-loops on the oldest vertex. Only simple tweaks are needed to generalize to the case where there are no self-loops (and hence where one cannot necessarily uniquely identify the oldest vertex).

Our algorithm starts with finding a certain orientation of the edges of the input graph G to produce a directed, acyclic graph D. In the case where G is isomorphic to a sample G' from $\mathcal{PA}(m;n)$ (say, $G = \pi(G')$), we have $D = \pi(\mathrm{DAG}(G'))$, and all vertices have out-degree m.

We accomplish this by a *peeling* procedure: at each step, consider the set D_{min} of minimum-degree nodes in the graph. We orient the edges incident on those nodes away from them, and then recurse on the subgraph excluding the nodes in D_{min}. This procedure terminates precisely when there are no remaining vertices. For a general input graph G, which might not have arisen by preferential attachment, there may be edges between vertices in D_{min}. We orient edges from nodes with larger labels to those with smaller ones. In general, this yields a directed, acyclic graph (aside from self-loops).

That this yields the directed graph $D = \pi(\mathrm{DAG}(G'))$ when the input is isomorphic to a preferential attachment graph is spelled out in detail in Lemma 2 of [3]. Hence, we are free to apply our structural results (such as Theorem 3) on $\mathrm{DAG}(G')$. We remark that it is not too hard to generalize our algorithm to tweaks of the model, since the only thing that is required is that the height of the resulting directed graph be at most $O(\log n)$; such an orientation of the edges of G exists with high probability, because of Theorem 3.

With this procedure in hand, the structural compression algorithm works as follows, on input G:

1) Construct the directed version $D = \mathrm{DAG}(G)$ by the procedure just described.
2) Starting from the "bottom" vertex (i.e., the vertex with no out-edges except for self-loops), we will do a depth-first search of D (following edges only from their destinations to their sources). To the jth vertex in this traversal, for $j = 1, ..., n$, we will associate a *backtracking number* B_j, which tells us how many steps to backtrack in the DFS process after visiting the jth node; e.g., when there is at least one in-edge leading to an unvisited node (so that we do not backtrack), $B_j = 0$. Upon visiting vertex w from vertex v in the DFS, we do the following:
   a) Denote by k the maximum out-degree of D (which can be determined in a preprocessing step, and which is equal to m if the input arises from preferential attachment). Using $\lceil \log k \rceil$ bits, encode the out-degree d_w of w (for preferential attachment, $d_w = m$, but we encode it for the sake of generality).
   Encode the names of the $d_w - 1$ vertex choices made by w, excluding one choice to connect to vertex v. Here, the *name* of a vertex is the binary expansion of its index in the DFS, which we can represent using exactly $\lceil \log n \rceil$ bits. These can be determined in a preprocessing step, by doing an

initial DFS to label the nodes with their names.

b) We need to know what happens after we visit vertex w: do we go forward in the search, or is there nowhere left to go along the current route (i.e., do we need to backtrack)? Suppose w is the jth vertex to be visited. Then we output an encoding of B_j. We need to more precisely examine how we encode these numbers, since it would be suboptimal to simply encode them in $\Theta(\log n)$ bits. Lemma 2 below tells us how to more efficiently perform this encoding.

3) For the purposes of decoding, we store (once, for the entire graph) the sequence of code words for the code used for the backtracking numbers. This can be done in at most $O(n \log\log n)$ extra bits, at the beginning of the code. We also store k (the maximum out-degree), which can be done with at most $O(\log n)$ bits.

Lemma 2. *The backtracking numbers $B_1, ..., B_n$ can be encoded using a total of $O(n \log\log n)$ bits on average.*

Proof. Consider a random variable X whose distribution is given by the empirical distribution of the collection $B = \{B_1, ..., B_n\}$. That is, $P_X(x) = \frac{|\{j \,:\, B_j = x\}|}{n}$ for each x. Note that this empirical distribution is itself a random variable. We will show that $\mathbb{E}[n \cdot H(X)] = O(n \log\log n)$.

Denote by W the event that the number of levels in D is upper bounded by $O(\log n)$. Under conditioning on this event, X can take on at most $O(\log n)$ values, which implies that $H(X) = O(\log\log n)$. Then we have

$$\begin{aligned}\mathbb{E}[H(X)] &\le \mathbb{E}[H(X)|W] + (1 - \mathbb{P}(W))\mathbb{E}[H(X)|\neg W] \\ &\le \mathbb{E}[H(X)|W] + (1 - \mathbb{P}(W)) \log n = O(\log\log n),\end{aligned}$$

where we have used Theorem 3 to upper bound $1 - \mathbb{P}(W)$.

We can thus construct a prefix code (once, for the entire graph) for the observed values of B_i, whose empirical average length is given by $\sum_{x \,:\, \exists j, B_j = x} \ell_x P_X(x) \le H(X) + 1$, where ℓ_x denotes the length of the code word for x. Now, recalling the definition of $P_X(x)$, this implies

$$\begin{aligned}&\mathbb{E}\left[\sum_{x \,:\, \exists j, B_j = x} \ell_x |\{j \;:\; B_j = x\}|\right] \\ &\le n\mathbb{E}[H(X)] + n = O(n \log\log n).\end{aligned}$$

□

We can show that the code for $S(G)$ is uniquely decodable. Furthermore, its expected length is at most $(m-1)n \log n + O(n \log\log n)$, which recovers the first term of the structural entropy and bounds the second. Construction of the Huffman code for the backtracking numbers takes time $O(n \log n)$, and each step of the DFS takes time at most $O(m \log n)$, so the running time is $O(mn \log n)$. □

We have thus proven the following:

Theorem 6 (Structural compression)**.** *The algorithm given above, on input a graph G isomorphic to $G' \sim \mathcal{PA}(m; n)$, runs in time $O(mn \log n)$ and outputs a code of expected length $(m-1)n \log n + O(n \log\log n)$ from which we can recover $S(G)$ in time $O(mn \log n)$. If self-loops are removed from G' and G (so that the first vertex is hard to identify), then the same code length can be achieved in time $O(mn^2 \log n)$.*

From Theorem 5, our algorithm is optimal at least up to the first term of the lower bound, and we explicitly bound the second term. There is a simple optimal labeled compression algorithm via arithmetic coding. We omit the details.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[2] A. Magner, A. Grama, J. Sreedharan, and W. Szpankowski, "Recovery of vertex orderings in dynamic graphs," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1563–1567.

[3] A. Magner, J. K. Sreedharan, A. Grama, and W. Szpankowski, "TIMES: Temporal information maximally extracted from structure," *Proceedings of the 2018 World Wide Web Conference*, pp. 389–398, 2018.

[4] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys*, vol. 74, pp. 47–97, 2002.

[5] A. Magner, S. Janson, G. Kollias, and W. Szpankowski, "On symmetries of uniform and preferential attachment graphs," *Electronic Journal of Combinatorics*, vol. 21, 2014.

[6] T. Łuczak, A. Magner, and W. Szpankowski, "Asymmetry and structural information in preferential attachment graphs," *Random Structures & Algorithms (to appear)*, 2019. [Online]. Available: http://arxiv.org/abs/1607.04102

[7] Y. Choi and W. Szpankowski, "Compression of graphical structures: Fundamental limits, algorithms, and experiments," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 620–638, 2012.

[8] J. Zhang, E.-H. Yang, and J. C. Kieffer, "A universal grammar-based code for lossless compression of binary trees," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1373–1386, 2014.

[9] A. R. Asadi, E. Abbe, and S. Verdú, "Compressing data on graphs with clusters," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1583–1587.

[10] E. Abbe, "Graph compression: the effect of clusters," in *Proc. of the fifty-fourth annual Allerton conference*, 2016.

[11] P. Delgosha and V. Anantharam, "Universal lossless compression of graphical data," in *2017 IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1578–1582.

[12] F. Chierichetti, R. Kumar, S. Lattanzi, M. Mitzenmacher, A. Panconesi, and P. Raghavan, "On compressing social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 219–228. [Online]. Available: http://doi.acm.org/10.1145/1557019.1557049

[13] C. Cooper and A. Frieze, "A general model of web graphs," *Random Structures & Algorithms*, vol. 22, pp. 311–335, 2003.

[14] M. Sauerhoff, "On the entropy of models for the web graph," *In submission*, 2016.