Multilevel Algorithm for Network Compression-friendly Ordering

Ilya Safro (Argonne National Laboratory), Boris Temkin (Weizmann Institute of Science)

Network Compression

Finding a suitable compressed representation of large-scale networks has been intensively studied in both practical and theoretical branches of data mining and network analysis [CS09, AM01, CKL+09, BV04, SBBA08]. In particular, the success of applying some of the recently proposed compression schemes [CKL+09, BV04, AD09] strongly depends on the "compression-friendly" arrangement of network nodes. Usually, the goal of these arrangements is to order the nodes such that the endpoints of network links (edges) are located as close as possible. Doing so leads to a more compact representation of links and allows a better performance of compression schemes and network element access operations.

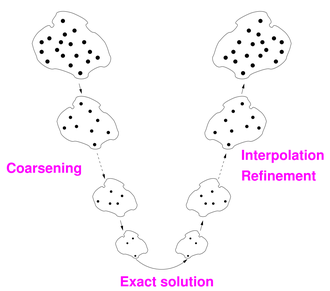
In [CKL+09], Chierichetti et al. propose to use ordering by *the minimum logarithmic arrangement problem* (MLogA), that minimizes the gap encodings of edges stretched between their endpoints. This is achieved by ordering the network nodes and assigning to them unique integer values (ids) such that the endpoints of a link will obtain close values. MLogA seeks a nearly optimal information-theoretical compressed encoding size for all network links as it minimizes the total number of bits to spend for this purpose. The problem is NP-hard. For example, instead of the popular network/matrix compressed row representation, which contains a sorted lists of neighbors per node, the following link gap encoding can be used.

	Compressed row format	Gap format
node number	sorted neighbors	sorted gaps
1	i,j,k,...	i,j-i,k-j,...
2	p,q,r,...	p,q-p,r-q,...

Problem (GMLogA): Given a weighted graph $G = (V, E)$, the goal of the link-weighted generalized minimum logarithmic arrangement problem is to minimize $\sum_{ij\in E} w_{ij} \lg |\pi(i) - \pi(j)|$ over all permutations π. Given node volumes, v, this is equivalent to the continuous version

$$\min_{\pi} \sum_{ij\in E} w_{ij} \lg |x_i - x_j| \quad \text{such that} \quad x_i = \frac{v_i}{2} + \sum_{k,\pi(k)<\pi(i)} v_k \ .$$

Multilevel Algorithm



Coarsening. Algebraic Multigrid-based projections $\mathcal{L}_c = \uparrow_f^c \mathcal{L}_f (\uparrow_f^c)^T$ with interpolation operator $\uparrow_f^c$ based on the *algebraic distances* between vertices.

Algebraic Distance. Extended p-normed algebraic distance between nodes i and j after k iterations $x^{(k+1)} = H_{JOR} x^{(k)}$ on R random initialial $x^{(0,r)}$

$$\rho_{ij}^{(k)} := \left(\sum_{r=1}^{R} |x_i^{(k,r)} - x_j^{(k,r)}|^p \right)^{1/p},$$

where $H_{JOR} = (D/\omega)^{-1} ((1/\omega - 1)D + L + U)$.

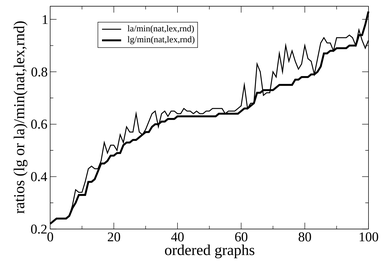
Uncoarsening. Given a (partial) solution, improving a contribution of one node requires minimization of $\sum_{j\in N_i} w_{ij} \lg |x_i - \tilde{x}_j|$. Since for every $j \in N_i$, $x_i = \tilde{x}_j$ implies the best solution, resolve this ambiguity by setting

$$x_i = \tilde{x}_t \iff t = \arg\min_{k\in N_i} \sum_{k\neq j\in N_i} w_{kj} \lg |\tilde{x}_k - \tilde{x}_j|.$$

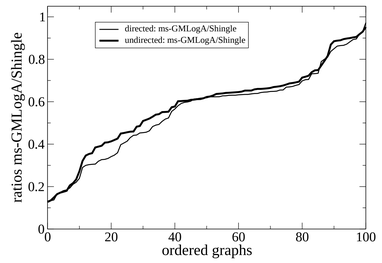
To preserve linear complexity, solved using linear time approximation that seeks nearly minimum sum at the point of maximal density using Gaussian kernel.

Computational Results and Applications

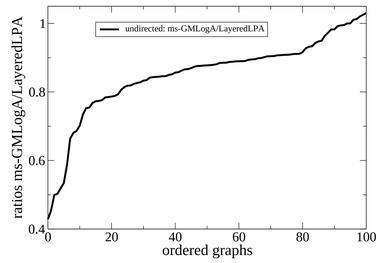
The main unit for empirical comparison is *number of bits per link*. Benchmark: 100 networks from UFL matrix collection, SNAP, and IBM VLSI benchmark.



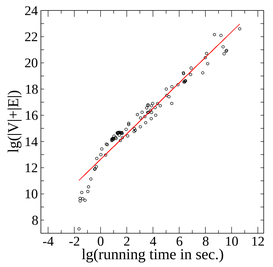
Multilevel vs min(Natural, Lexicographic, Randomized)



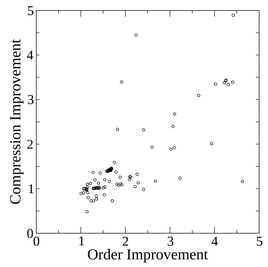
Multilevel vs Shingle (based on Jaccard coefficient)



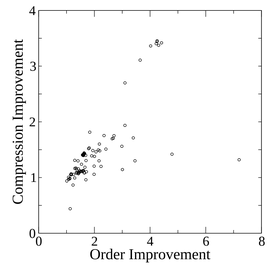
Multilevel vs LayeredLPA (based on Potts model)



Scalability



Further Network Compression (vs natural ordering)



Further Network Compression (vs best(Shingle, Gray)

Highlights

- Algebraic distance is very important, otherwise the interpolation order for AMG has to be at least 2.
- Most beneficial networks are *not* from one collection: VLSI graphs, roads, social networks, web links.
- Spectral ordering produced poor results on most of the networks.
- Strong solver for minimum linear arrangement ordering with GMLogA postprocessing can produce good results.
- GMLogA reordering of network can improve general performance of algorithms with intensive node/link access operations via improving operations with cache. Currently work in solver for optimal response to epidemics and cyber attacks. Also see Peter Lindstrom's poster.

More information in

◇ D. Ron, I. Safro, A. Brandt, "Relaxation-based coarsening and multiscale graph organization", SIAM Multiscale Modeling and Simulations, 2011
◇ I. Safro, B. Temkin, "Multiscale approach for the network compression-friendly ordering", Journal of Discrete Algorithms, 2011

References **[AD09]** Apostolico and Drovandi. Graph compression by BFS. Algorithms, 2009; **[AM01]** Adler and Mitzenmacher. Towards compressing web graphs. Proceedings of the Data Compression Conference, 2001; **[BV04]** Boldi and Vigna. The Webgraph framework I: compression techniques. Proceedings of the 13th WWW, 2004; **[CKL+09]** Chierichetti, Kumar, Lattanzi, Mitzenmacher, Panconesi, and Raghavan. On compressing social networks, Proceedings of ACM SIGKDD 2009; **[CS09]** Choi and Szpankowski. Compression of graphical structures. IEEE Symposium on Information Theory, 2009; **[SBBA08]** Sun, Bollt, and Ben-Avraham. Graph compression-save information by exploiting redundancy. Journal of Statistical Mechanics: Theory and Experiment, 2008.

References

[AD09] Alberto Apostolico and Guido Drovandi. Graph compression by BFS. *Algorithms*, 2(3):1031–1044, 2009.

[AM01] Micah Adler and Michael Mitzenmacher. Towards compressing web graphs. In *DCC '01: Proceedings of the Data Compression Conference*, page 203, Washington, DC, USA, 2001. IEEE Computer Society.

[BV04] P. Boldi and S. Vigna. The Webgraph framework I: compression techniques. In *WWW '04: Proceedings of the 13th International Conference on World Wide Web*, pages 595–602, New York, NY, USA, 2004. ACM.

[CKL+09] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Michael Mitzenmacher, Alessandro Panconesi, and Prabhakar Raghavan. On compressing social networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 219–228, New York, NY, USA, 2009. ACM.

[CS09] Yongwook Choi and Wojciech Szpankowski. Compression of graphical structures. In *ISIT'09: Proceedings of the 2009 IEEE Symposium on Information Theory*, pages 364–368, Piscataway, NJ, USA, 2009. IEEE Press.

[SBBA08] Jie Sun, Erik M Bollt, and Daniel Ben-Avraham. Graph compression-save information by exploiting redundancy. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(06):P06001, 2008.