Graph Cities [ANCA21]

Supervised by James Abello, Computer Science Department/DIMACS, Rutgers, abello@dimacs.rutgers.edu

Abstract

"*Graph Cities*" are 3D representations of maximal edge graph partitions. Each connected equivalence class corresponds to a "*building*" that is formed by stacking graph "*Edge Fragments*". The number of such graph edge fragments determines the height of the building.
The overall number of buildings is the number of equivalence classes in the edge partition.

A poly-log bucketization of the size distribution of the graph edge equivalence classes is used to generate a 2D position for each bucket. For the buckets containing more than one equivalence class, we also generate a visual "bush" representation**.**

The Delaunay triangulation of the building locations determines the "*street network*" of the *Graph City*. The weight of a connection between two buildings on this street network is proportional to the intersection of the subgraph vertex sets represented by the two buildings. To handle equivalence classes (i.e., buildings) consisting of a large number of fragments, we use the notion of "*Graph Waves*" from [AN20]. *Graph Waves* are intervals of graph edge fragments with a "well-defined" beginning and end fragment. For computational purposes, the beginning and end fragments should satisfy a computationally "easy to verify" property. We illustrate *Graph Cities* obtained from the maximal edge partitions defined by the iterative edge core decomposition introduced in [AQ14]. The graphs used include the Friendster social network (1.8 billion edges), a co-occurrence keywords network derived from the internet movie database (115 million edges), and a patents citation network (16.5 million edges).

For graphs with up to 2 billion edges, all the main elements of their corresponding Graph Cities (including buildings and street networks) are built in minutes (excluding I/O time) and storage proportional to the number of edges and vertices of a graph. **Our ultimate goal is to obtain humanly-interpretable hierarchical descriptions of any graph that are accessible via a Unified Web Interface for Graph Analytics, without being constrained by the graph size.**

In this project, REU participants will build "*Graph Cities*" from data collections containing at least a **billion related** entity pairs. Each created "*City*" will come equipped with a "*Story*" plus an accessible catalogue of its most prominent "*locations"*, "*entities"*, "*landmarks", and "Subgraph Motifs".* Each discovered motif will be judged according to its "*provenance*" and a profile that may include measures of "*density*", "*diversity*" , "*surprise*", and "*interestingness*".

Contact; James Abello. MSCS Director, abelloj@cs.rutgers.edu, Computer Science Department, Rutgers University

Pre-Requisites

Programming Languages: JavaScript and two languages from {Python, Java, C/C++}.
Algorithms/Optimization: A class on Algorithms and a second class in Optimization or Graph Theory.
Reading List: Become familiar with at least 7 papers from the list below before the beginning of the REU.

References

[ANCA21 J. Abello, D. Nakhimovich, L. Chang, M. Aanjaneya, "Graph Cities: Their Buildings, Waves and Fragments', submitted for publication.

[ABW98] ABELLO, JAMES, BUCHSBAUM, ADAM L, and WESTBROOK, JEFFERY R. "*A Functional Approach to External Graph Algorithms*". European Symposium on Algorithms. Springer. 1998, 332–343 4.

[AH17] ANGEL, ED and HAINES, ERIC. "*An Interactive Introduction to WEBGL and Three.JS*". ACM SIGGRAPH 2017 Courses. SIGGRAPH '17. Los Angeles, California, 2017 8, 9.

[AKM17] ARLEO, ALESSIO, KWON, OH-HYUN, and MA, KWAN-LIU, "*GraphRay: Distributed Pathfinder Network Scaling*". 2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV). IEEE. 2017, 74–83 3.

[AMS19] ABELLO, JAMES, MAWHIRTER, DANIEL, and SUN, KEVIN. "*Taming a Graph Hairball: Local Exploration in a Global Context*". Business and Consumer Analytics: New Ideas. Springer, 2019, 467–490 4.

[AN20] ABELLO, JAMES and NAKHIMOVICH, DANIEL. "*Graph Waves*". The 3rd International Workshop on Big Data Visual Exploration and Analytics with EDBT/ICDT. 2020 1–5, 7–9, 11.

[AQ14] ABELLO, JAMES and QUEYROI, FRANÇOIS. "*Network decomposition into fixed points of degree peeling*". Social Network Analysis and Mining 4.1 (2014), 1–14 1–5.

[AVK06] ABELLO, JAMES, VAN HAM, FRANK, and KRISHNAN, NEERAJ. "*Ask-Graphview: a large scale graph visualization system*". IEEE Transactions on Visualization and Computer Graphics 12.5 (2006), 669–676 3, 9.

[BDH14] BEREND, DANIEL, DOLEV, SHLOMI, and HANEMANN, ARIEL. "*Graph Degree Sequence Solely Determines the Expected Hopfield Network Pattern Stability"*. Neural computation 27 (Nov. 2014), 1–9. DOI: 10.1162/NECO_a_00685 2.

[BGG*07] BAUR, MICHAEL, GAERTLER, MARCO, GÖRKE, ROBERT, et al. "*Generating graphs with predefined k-core structure*". Proceedings of the European Conference of Complex Systems. Citeseer. 2007 3.

[BGHK20] BEN-ELIEZER, OMRI, GISHBOLINER, LIOR, HEFETZ, DAN, and KRIVELEVICH, MICHAEL. "*Very fast construction of bounded degree spanning graphs via the semi-random graph process*". Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM. 2020, 718–737 9.

[FK18] FRANKL, PETER and KUPAVSKII, ANDREY. "T*he Erdös Matching Conjecture and concentration inequalities*". arXiv preprint arXiv:1806.08855 (2018) 9.

[FKW*17] FANG, DEZHI, KEEZER, MATTHEW, WILLIAMS, JACOB, et al. "*Carina: interactive million-node graph visualization using web browser technologies*". International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee. 2017, 775–776 3.

[JRDX12] JIN, RUOMING, RUAN, NING, DEY, SAIKAT, and XU, JEFFREY YU. "*SCARAB: scaling reachability computation on large graphs"*. Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. 2012, 169–180 3.

[LK14] LESKOVEC, JURE and KREVL, ANDREJ. "*SNAP Datasets, Stanford Large Network Dataset Collection*", http://snap.stanford. edu/data. June 2014 7, 8.

[MSM*16] MI, PENG, SUN, MAOYUAN, MASIANE, MOETI, et al. *"Interactive graph layout of a million nodes"*. Informatics. Vol. 3. Multidisciplinary Digital Publishing Institute. 2016, 23 3.

[NHEM17] NGUYEN, QUAN HOANG, HONG, SEOK-HEE, EADES, PETER, and MEIDIANA, AMYRA. "*Proxy graph: visual quality metrics of big graph sampling*". IEEE transactions on visualization and computer graphics 23.6 (2017), 1600–1611 3.

[PHE*18] PIENTA, ROBERT, HOHMAN, FRED, ENDERT, ALEX, et al. "*VIGOR: interactive visual exploration of graph query results*". IEEE Transactions on Visualization and Computer Graphics 24.1 (2018), 215–225 3.

[RRAS08] ROYER, LOIC, REIMANN, MATTHIAS, ANDREOPOULOS, BILL, and SCHROEDER, MICHAEL. "*Unraveling protein networks*".