# Graph Cities: Their Buildings, Waves, and Fragments 

James Abello ${ }^{1}$ Daniel Nakhimovich Chengguizi Han Mridul Aanjaneya<br>Department of Computer Science, Rutgers University, USA<br>${ }^{1}$ DIMACS, Rutgers University, USA

## 1 APPENDIX

For the sake of completeness, we summarize below the main technical results that support the development of Graph Cities.

```
Alg. 1: Meta-DAG Computation
    Input: \(F_{k}=\left(V_{k}, E_{k}\right)\), a fixed point of peel value k .
    Input: \(S=\left\{S_{0}, S_{1}, S_{2}, \ldots, S_{n}\right\}\) where each \(S_{j} \subset V_{k}\) is a fragment seed set.
    Output: \(G_{D A G}=(V, E, W)\) the meta dag of \(F_{k}\).
    function meta-dag \(\left(F_{k}\right)\) :
        smap \(\leftarrow\) []
        for \(S_{x}\) in \(S\) do
            for \(v\) in \(S_{x}\) do
                \(\operatorname{smap}[v]=x\)
            end
        end
        \(D S \leftarrow\) disjointSet()
        for \((u, v)\) in \(E_{k}\) do
            if \(\operatorname{smap}(u)=\operatorname{smap}(v)\) then
            DS.union \((u, v)\)
            end
        end
        \(V \leftarrow \emptyset\)
        cmap \(\leftarrow\) []
        for \(c\) in DS.sets() do
            \(V \leftarrow V \cup\{c\}\)
            for \(v\) in \(c\) do
                cmap \([v]=c\)
            end
        end
        \(E \leftarrow \emptyset\)
        \(W \leftarrow \emptyset\)
        for \((u, v)\) in \(E_{k}\) do
            \(E \leftarrow E \cup\{(\operatorname{cmap}[u], \operatorname{cmap}[v])\}\)
            \(W[(\operatorname{cmap}[u], \operatorname{cmap}[v])] \leftarrow W[(\operatorname{cmap}[u], \operatorname{cmap}[v])]+1\)
        end
        \(G_{D A G}=(V, E, W)\)
        return \(G_{D A G}\)
    end
```

Proposition 1. The edges of any Fixed Point $F_{k}$ of degree peeling $k$ can be partitioned into an ordered collection of edge fragments whose generating vertex sets partition the vertexes of the $F_{k}$.

Proof Outline 1. Consider the following ordered Wave Identification process, Alg. 1 from [1].

From the Wave computation algorithm described above, it is apparent that the algorithm generates a sequence of disjoint edge fragments each of them with its own disjoint generating set. The basis of induction is clear since the source set is just the set of vertices of degree $k$ in $F_{k}$. The induction hypothesis follows from the fact that the ordered deletion of edge fragments in a wave (except the source fragment) is precisely what is captured by the definition of a Wave. The end fragment is reached when all its neighboring vertices have degree in the wave complement that is greater than or equal to $k$. Resetting the source set to the remaining vertices of degree $k$ in the remaining edges of the fixed

[^0]```
Alg. 2: Floor Visual Parameters
    Input: \(W_{0}=\left(V_{0}, E_{0}\right)\), the wave for this floor.
    Input: \(W_{1}=\left(V_{1}, E_{1}\right)\), the next consecutive wave
    Input: \(F=\left\{F_{1}, F_{2}, \ldots, F_{k}, \ldots, F_{t}\right\}\), the set of fixed points decomposing
        the entire dataset, \(W_{0}, W_{1} \subset F_{k}\)
    Input: \(S=\left\{S_{0}, S_{1}, S_{2}, \ldots, S_{n}\right\}\) where each \(S_{j} \subset V_{0}\) is a fragment seed set.
    Output: \(\left(R_{b}, H, R_{t}, C, L\right)\), the radius of bottom disk, height between
                disks, radius of top disk, color, and light intensity.
    function floor \(\left(F_{k}\right)\) :
        \(R_{b} \leftarrow \log _{2}\left(\left|S_{0}\right|\right)\)
        \(R_{t} \leftarrow \log _{2}\left(|V|-\left|S_{0}\right|\right)\)
        if \(W_{1} \neq \emptyset\) then
            \(E_{h} \leftarrow\left\{(u, v) \in E_{0}: v \in V_{1}\right\}\)
        else
            \(E_{h} \leftarrow\left\{(u, v) \in E_{0}: u \in S_{i}, v \in S_{j} \forall S_{i}, S_{j} \in S\right\}\)
        end
        \(H \leftarrow \log _{2}\left(3 * E_{h} /\left(\pi *\left(R_{t}^{2}+R_{t} R_{b}+R_{b}^{2}\right)\right)\right)\)
        \(C=2 * E_{h} /\left(V_{0} * V_{1}\right)\)
        \(L=\left|\left\{v \in V_{0}: v \in F_{i}, \forall F_{i} \in F \backslash\left\{F_{k}\right\}\right\}\right|\)
        return \(\left(R_{b}, H, R_{t}, C, L\right)\)
    end
```

```
Alg. 3: Wave Decomposition \((O(m))\)
    Input: \(F_{k}=(V, E)\), a fixed point of peel value k .
    Output: \(M=\left\{W_{1}, W_{2}, W_{3}, \ldots, W_{m}\right\}\) where each \(W_{i} \subset M\) are waves.
    Output: \(S=\left\{S_{0}, S_{1}, S_{2}, \ldots, S_{n}\right\}\) where each \(S_{j} \subset V\).
    function waves \(\left(F_{k}\right)\) :
        \(M \leftarrow \emptyset\)
        \(S \leftarrow \emptyset\)
        \(i \leftarrow 1\)
        \(j \leftarrow 0\)
        while \(E \neq \emptyset\) do
            \(W_{i} \leftarrow \emptyset\)
            \(S_{j} \leftarrow\{v \in V: \operatorname{deg}(v)=k\}\)
            while \(S_{j} \neq \emptyset\) do
                \(S \leftarrow S \cup\left\{S_{j}\right\}\)
                \(E \leftarrow E \backslash \mathrm{frag}\left(S_{j}\right)\)
                \(W_{i} \leftarrow W_{i} \cup \mathrm{frag}\left(S_{j}\right)\)
                \(S_{j+1} \leftarrow\left\{v \in \partial S_{j} \mid \operatorname{deg}(v)<k\right\}\)
                \(j \leftarrow j+1\)
            end
            \(M \leftarrow M \cup\left\{W_{i}\right\}\)
            \(i \leftarrow i+1\)
        end
    end
```

point (if any) lets the process starts another wave computation. In summary, a fixed point $F_{k}$ of a graph $G$ minus the edges of one of its waves $W$ leaves either the empty set or a fixed point of the same peel value $k$ in the subgraph remaining after deleting the edges in $W$.

Corollary 1. The subset of edges of any fixed point $F_{k}$ of degree peeling $k$ with end points in different sets of S derived from Alg. 1 can be directed to obtain a Directed Acyclic Graph (DAG) where an edge $(u, v)$ is directed according to the fragment generation process described in the proof of proposition 1.

Proof Outline 2. Direct an edge from a vertex $u$ to a vertex $v$ if $u \in S_{i}, v \in S_{j}$ and $i<j$.

Corollary 2. A fixed point $F_{k}$ of peel value $k$ has a spanning Meta-DAG whose longest path corresponds to the height of a building representing the fixed point.

Proof Outline 3. The subset of edges $(u, v)$ in the DAG defined in corollary 1 where the fragments containing $u$ and $v$ are consecutive is a spanning DAG of the fixed point. It is spanning because the vertex sets of the fragments are connected and they form a partition of the vertex sets of $F_{k}$. Assuming that a fixed point has more than one edge fragment, the longest path from the source set of $F_{k}$ (i.e at the bottom floor of the building) to the ending fragment set of $F_{k}$ (i.e. at the top floor of the building) corresponds precisely to the height of the building because the inter floor distance is a function of the frustum volume between the two floors which in turn encodes the number of edges running between the two floors.

### 1.1 Interpreting a Graph City

"Graph Cities" provide visual representations of the overall macrostructure of graphs with few billion edges, i.e., GigaGraphs. These novel representations are derived by mapping each connected equivalence class, of a special edge partition, into a "city building". Each such edge equivalence class is an edge maximal connected subgraph with a fixed peel value $k$. Figure ?? illustrates the initial edge decomposition derived from the peeling values associated to each vertex of the depicted graph after running plain vanilla vertex core decomposition. Given a partition of the edge set of a graph, a coarsest view of its "graph city" is provided by:
(1) the number of vertices and edges being represented,
(2) the underlying graph degree distribution,
(3) the number of "buildings",
(4) the number of "floors" per "building" and the per "floor" size and density, and
(5) the distribution of inter "floors" "volumes", and "densities".

### 1.2 What do "floors" tell us about a "building" in a Graph City?

An entire "city building" can be fully explored by iteratively removing neighboring vertexes of degree less than or equal to $k$. This "building" exploration partitions the corresponding edge equivalence class into a sequence of $h$ "disjoint edge waves" $\left\{W_{i}, i=0, \ldots, h\right\}$. These "edge waves" are in one to one correspondence with a sequence of $h$ "disjoint seed vertex sets" $\left\{S_{i}, i=0, \ldots, h\right\}$. This vertex partition sequence provides nondecreasing directionality to every edge $(x, y)$ of the graph by directing the edge from the minimum to the maximum of the peel values of $x$ and $y$. When both $x$ and $y$ are in the same seed set, the edge is called horizontal and otherwise it is called a vertical edge. Each seed set $S_{i}$ is characterized by the fact that at some point during the building exploration, all of its vertices become of degree exactly $k$ and their removal leaves all their neighbors with degree strictly less than $k$. In summary, a wave $W_{i}$ is the set of edges traversed in the exploration process when the corresponding seed set $S_{i}$ is identified and it is followed by iterative removal of neighboring vertexes of degree strictly less than $k$. Equivalently, the set of vertices of minimum degree $k$ can be used as originators of message transmissions in its wave with all neighboring vertexes iteratively sending out strictly less than $k$ messages to unvisited neighbors. Each edge wave of peel value $k$ is a localized topology generated by a unique seed set of vertices of minimum degree $k$ and this is conveyed in our representation by associating unique floor with every wave. Namely, the seed set $S_{i}$ of a wave $W_{i}$ is represented by the $i$ th "building floor". Its corresponding edge wave $W_{i}$ is visually represented by the "frustum" consisting of the two corresponding consecutive
building seed sets $S_{i}$ and $S_{i+1}$. The frustum volume between $S_{i}$ and $S_{i+1}$ encodes the number of edges in the wave $W_{i}$.

In summary, the number $h$ of floors in a building (i.e., the number of waves) indicates a fixed point whose full exploration requires the sequential activation of $h$ disjoint seed sets.

It is worth to note that in cases where a "building" is used to represent an edge equivalence class with several connected components then the number of floors in the "building" corresponds to the maximum number of waves in any of its components.

### 1.3 What does a "building" volume represent?

Given two consecutive seed floors representing the sets $S_{i}$ and $S_{i+1}$, the wave $W_{i}$ associated with $S_{i}$ ends at a subset $S_{i}^{\prime}$ of $S_{i+1}$ such that each of its unvisited neighbors has degree bigger that or equal to the peel value of the wave $W_{i}$. The vertical distance $h_{i}$ between $S_{i}$ and $S_{i}^{\prime}$ is the visual vertical distance required so that the volume enclosed by the corresponding frustum is proportional to the number of edges running from $S_{i}$ to $S_{i}^{\prime}$ in the wave $W_{i}$. The sum of all the frustum volumes associated with all the building waves is a lower bound on the size of the entire equivalence class represented by the building. All these "internal" wave edges represent the "building backbone". All the edges running between non-consecutive waves in the building are represented by the volume of the "building enclosure" minus the internal backbone edges. In summary, the visible volume of a building encodes the number of edges of the represented edge equivalence class, i.e., a Fixed Point of degree peeling. A building with no enclosure represents a more localized topology, i.e., is a "tree-like" fixed point with only consecutive edges. The ratio between the overall number of edges in the fixed point and the number of backbone edges could be an interesting and novel fix point measure to explore in the future.

### 1.4 How is the internal detailed structure of a "building" made accessible for user exploration?

The internal structure of a building is represented by a Directed Acyclic Macro Graph obtained by contracting the connected components of each wave seed set . Namely, for each wave $W_{i}$ with a sequence of seed sets $S_{j}, j=0, \ldots, k$, we contract into macrovertices the connected components of the subgraph induced by each $S_{j}$ and direct macro edges from low indexed seed sets to higher numbered ones. These macroedegs are weighted according to the number of original edges interconnecting the two components. The number of internal edges of each of these connected components is encoded by the macrovertex size and the macrovertex is rainbow-colored according to its internal density. The connected components of the last seed set of the wave $W_{i}$ are connected to the connected components of the first seed set of the next wave $W_{i+1}$ (if any). In summary, the internal structure of a building is a DAG representing the connectivity between the connected components of all the seed sets appearing in the building waves. Our interface provides on demand access to this DAG internal structure for user navigation and exploration on a per building basis. Exploration tools include two-way sliders to filter the displayed graph view by a variety of attributes.

## REFERENCES

[1] James Abello and Daniel Nakhimovich. 2020. Graph Waves. In The 3rd International Workshop on Big Data Visual Exploration and Analytics with EDBT/ICDT.


[^0]:    © 2021 Copyright for this paper by its author(s). Published in the Workshop Proceedings of the EDBT/ICDT 2021 Joint Conference (March 23-26, 2021, Nicosia, Cyprus) on CEUR-WS.org. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

