Anchored Densest Subgraph

Yizhou Dai[1] Miao Qiao[1] Lijun Chang[2]

[1]The University of Auckland

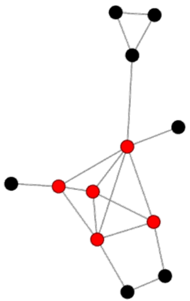
[2]The University of Sydney

June 16, 2022

Densest Subgraph Search

Given an undirected graph $G(V, E)$,

- Density (average-degree) of G: $\rho(G) = \frac{|E|}{|V|}$ $\frac{18}{12}$
- Induced subgraph on $S \subseteq V$ $\frac{9}{5}$
- Densest subgraph $\mathrm{DS}(G) = argmax_{S \subseteq V(G)} \rho(S)$
- Algorithm: Goldberg [Gol84], recent survey [GT15]



Used for detecting

- Communities in social networks
- Biomarkers in bioinformatics and brain networks
- Spam link farms on web graphs

Report only ONE subgraph

Localize densest subgraph search

Given a graph $G(V, E)$ and a reference node set $R \subseteq V$
Report the "densest" subgraph that is "local" to R.

Localize densest subgraph search

Given a graph $G(V, E)$ and a reference node set $R \subseteq V$
Report the "densest" subgraph that is "local" to R.

Existing work under average-degree density

Diversify global densest subgraph search — Report the one that is closest to R

- Find top-k "locally densest" subgraphs [Qin et al. 2015]
  - A subgraph S is locally densest if 1) S is the densest among all its subgraphs and is 2) $\rho(S)$-compact: a notion related to k-core
- Find top-k densest subgraphs [Galbrun et al. 2016]
- All the densest subgraphs [Chang and Qiao, 2020]

Localize densest subgraph search

Given a graph $G(V, E)$ and a reference node set $R \subseteq V$
Report the "densest" subgraph that is "local" to R.

Existing work under average-degree density

Diversify global densest subgraph search — Report the one that is closest to R

- Find top-k "locally densest" subgraphs [Qin et al. 2015]
  - A subgraph S is locally densest if 1) S is the densest among all its subgraphs and is 2) $\rho(S)$-compact: a notion related to k-core
- Find top-k densest subgraphs [Galbrun et al. 2016]
- All the densest subgraphs [Chang and Qiao, 2020]

Global Computation & Possible Degeneration

Localize densest subgraph search

Given a graph $G(V, E)$ and a reference node set $R \subseteq V$
Report the "densest" subgraph that is "local" to R.

An ideal localized densest subgraph search should

- Average-degree density defined bias to R to avoid degeneration
- Have wide real-world applications
- Scalable to billion-scale graphs

~~Global computation~~

Anchored Densest Subgraph

R-subgraph density

Given a graph $G(V, E)$, a reference node set R, the *R-subgraph density* of an arbitrary set S of nodes

$$\rho_R(S) = \frac{2|E(S)| - \sum_{v \in S \text{ and } v \notin R} degree(v)}{|S|}.$$

Anchored Densest Subgraph

R-subgraph density

Given a graph $G(V, E)$, a reference node set R, the *R-subgraph density* of an arbitrary set S of nodes

$$\rho_R(S) = \frac{2|E(S)| - \sum_{v \in S \text{ and } v \notin R} degree(v)}{|S|}.$$

Anchored Densest Subgraph

Given a set $R \subseteq V$ and an optional set $A \subseteq R$, ADS reports the *supergraph* of A that maximizes the R-subgraph density $argmax_{S:\ A \subseteq S \subseteq V} \rho_R(S)$.

Anchored Densest Subgraph

R-subgraph density

Given a graph $G(V, E)$, a reference node set R, the *R-subgraph density* of an arbitrary set S of nodes

$$\rho_R(S) = \frac{2|E(S)| - \sum_{v \in S \text{ and } v \notin R} degree(v)}{|S|}.$$

Anchored Densest Subgraph

Given a set $R \subseteq V$ and an optional set $A \subseteq R$, ADS reports the *supergraph* of A that maximizes the R-subgraph density $argmax_{S:\ A \subseteq S \subseteq V}\, \rho_R(S)$.

Event organization & product recommendation

Locality in node inclusion and node centrality

Localize densest subgraph search

Given a graph $G(V, E)$ and a reference node set $R \subseteq V$
Report the "densest" subgraph that is "local" to R.

Diversify global densest subgraph search

Report the one that is closest to R

An ideal localized densest subgraph search should

- Define the localized density based on R to avoid degeneration
- Have wide real-world applications
- Scalable to billion-scale graphs

~~Global computation~~

Global Algorithm

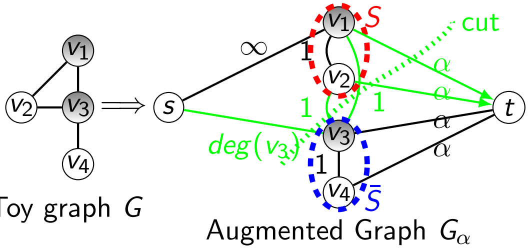


Figure: Augmented graph G_α: $A = \{v_1\}$, $R = \{v_1, v_3\}$ vertices of R are shadowed

Lemma 3.2 (Informal)

The smallest α such that the max-flow of the above network is $\sum_{v \in R} degree(v)$ is the R-subgraph density of the ADS. The nodes reachable with unsaturated edges from the source are the ADS.

Local Algorithm

Binary search on α, for each α value

- B: Initially $\emptyset$
- $\partial(R \cup B)$: the neighbors of $R \cup B$
- $g_{\alpha,B}$: working graph for network flow
- $B \leftarrow B \cup$ the nodes with satuated edges to t
- Terminate when B stops growing.

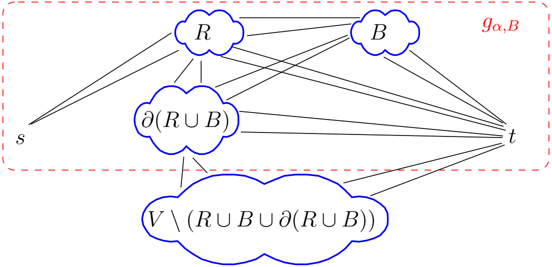


Figure: Illustration of $g_{\alpha,B}$

Local Algorithm

Binary search on α, for each α value

- B: Initially $\emptyset$
- $\partial(R \cup B)$: the neighbors of $R \cup B$
- $g_{\alpha,B}$: working graph for network flow
- $B \leftarrow B \cup$ the nodes with satuated edges to t
- Terminate when B stops growing.

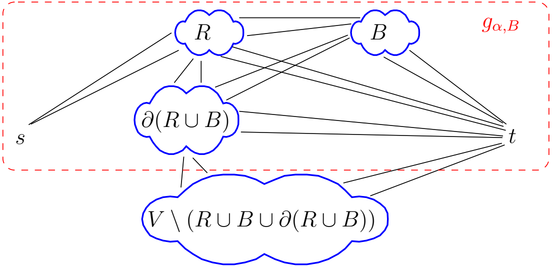
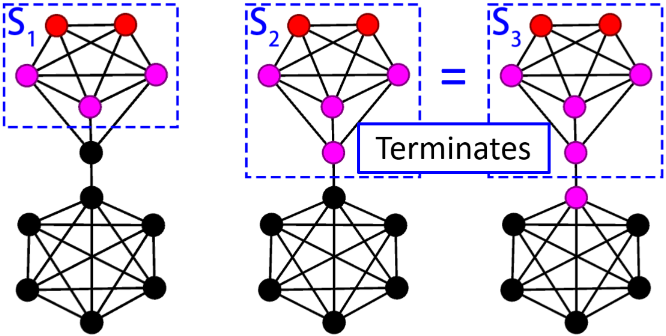


Figure: Illustration of $g_{\alpha,B}$

Lemma 4.7

The number of vertices and edges in g_{α,B^*} is $\mathcal{O}((\sum_{v \in R} degree(v))^2)$.

Local Algorithm



Graph Data

Name	n	m	Density	Type
amazon	334,863	925,872	2.76	Product network
notredame	325,730	1,090,108	3.35	Web graph
digg	279,631	1,548,126	5.54	Social network
citeseer	384,414	1,736,145	4.52	Citation network
livemocha	104,103	2,193,083	21.07	Social network
flickr	105,939	2,316,948	21.87	Image network
hyves	1,402,674	2,777,419	1.98	Social network
youtube	1,134,890	2,987,624	2.63	Social network
google	875,714	4,322,051	4.94	Web graph
trec	1,601,788	6,679,248	4.17	Web graph
flixster	2,523,387	7,918,801	3.14	Social network
dblp	1,653,767	8,159,739	4.93	Citation network
skitter	1,696,416	11,095,299	6.54	Computer network
indian	1,382,868	13,591,473	9.83	Web graph
pokec	1,632,804	22,301,964	13.66	Social network
usaroad	23,947,347	28,854,312	1.20	Road network
livejournal	3,997,962	34,681,189	8.67	Social network
orkut	3,072,441	117,185,083	38.14	Social network
wikipedia	13,593,033	334,591,525	24.61	Web graph
friendster	68,349,466	1,811,849,342	26.51	Social network
uk2007	105,153,952	3,301,876,564	31.40	Web graph

Experiments

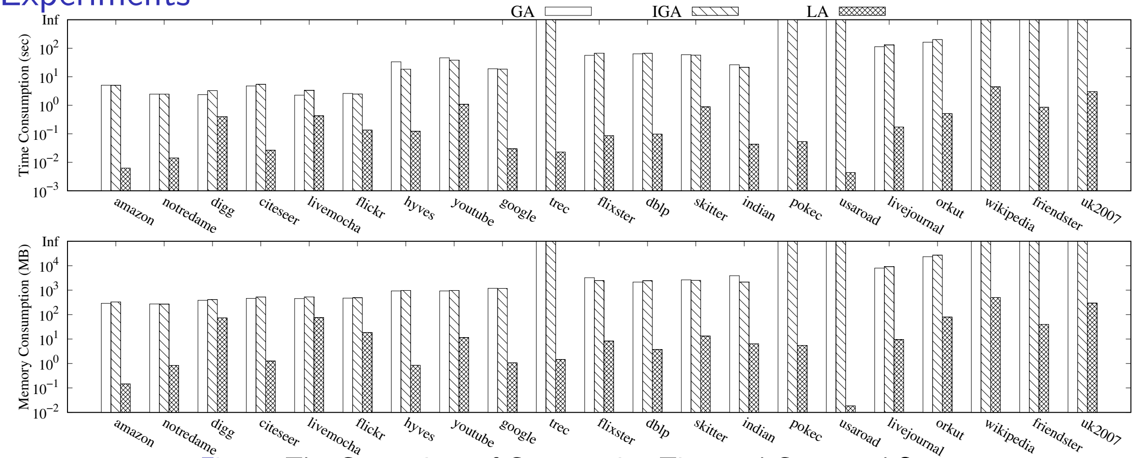


Figure: The Comparison of Computation Time and Consumed Space

LA vs GA: 283× speedup and 1/425 space consumption.

LA: time and memory costs does not increase with the graph size.

Experiments

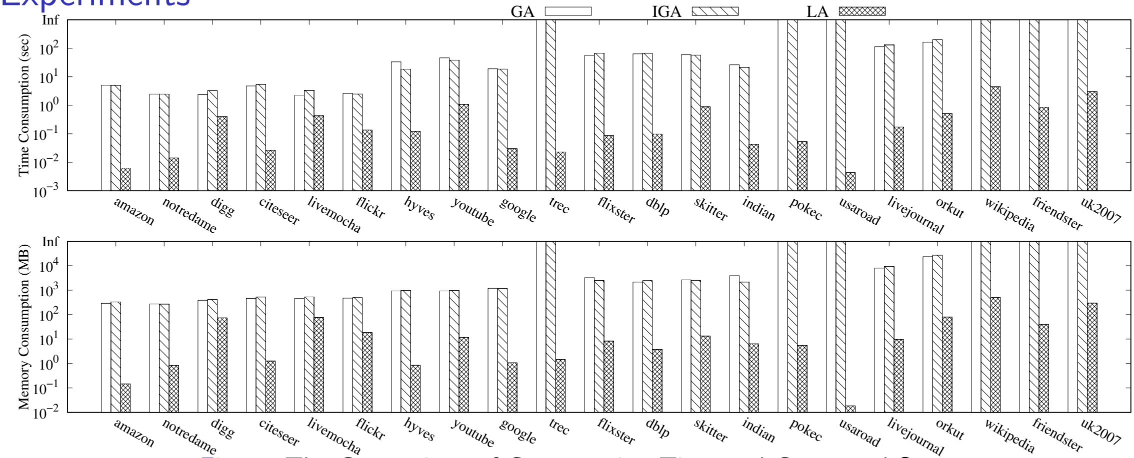


Figure: The Comparison of Computation Time and Consumed Space

LA vs GA: 283× speedup and 1/425 space consumption.

LA: time and memory costs does not increase with the graph size.

Case study, video recording (starting from min 7:01).

Conclusions

- Propose anchored densest subgraph search (ADS) which penalizes non-R nodes proportional to their degree centralities.
- Propose, for ADS, a local (search) algorithm whose complexity is related to R as opposed to the entire graph G.
- Extensive experiments verified the efficiency and effectiveness of the proposed algorithm.
- Provide use cases which apply the techniques to real-life scenarios.

A. V. Goldberg.
Finding a maximum density subgraph.
Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley, 1984.

Aristides Gionis and Charalampos E. Tsourakakis.
Dense subgraph discovery: KDD 2015 tutorial.
In Longbing Cao, Chengqi Zhang, Thorsten Joachims, Geoffrey I. Webb, Dragos D. Margineantu, and Graham Williams, editors, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015*, pages 2313–2314. ACM, 2015.

Lemma 4.6

For any vertex u that is in an anchored densest subgraph, it must satisfy $d_G(u) < \mathrm{vol}(R)$.