

Praca Magisterska

Wykaz i uwagi do literatury

Paweł Polerowicz

Grudzień 2023

1 Zasadnicze struktury danych

1.1 gSketch[18]

Autorzy: P. Zhao, C. C. Aggarwal, and M. Wang

Tytuł: gSketch: On Query Estimation in Graph Streams

Rok publikacji: 2011

Artykuł ten stanowi jedną z pierwszych prób zastosowania szkiców danych do obsługi prostych zapytań dotyczących grafu, takich jak częstość występowania danej krawędzi w strumieniu lub gęstość wybranego podgrafu. Autorzy proponują wykorzystanie probabilistycznej metody Count Min, opartej o dwuwymiarową tablicę do składowania częstotliwości krawędzi. Odpowiednie komórki są wyznaczane przez funkcje haszujące zastosowane na krawędziach. Algorytm wykorzystuje próbkę testową do podziału zbioru krawędzi na podzbiory, bazując na ich częstotliwości w taki sposób, aby efektywnie wykorzystać dostępną pamięć, a następnie przetwarza właściwy strumień danych. Metoda jest stratna, a praktycznym wyzwaniem jest odpowiednie wyważenie parametrów tak, aby zachować balans między zużyta pamięcią a dokładnością wyników.

1.2 TCM[16]

Autorzy: N. Tang, Q. Chen, and P. Mitra

Tytuł: Graph Stream Summarization. From Big Bang to Big Crunch

Rok publikacji: 2016

Autorzy proponują strukturę TCM do przechowywania informacji o strumieniowanym grafie. Ma ona postać macierzy o boku długości m , gdzie m jest pewną stałą. W jej komórkach składowane są wagi krawędzi. Rząd i kolumna komórki odpowiadającej danej parze wierzchołków są wyznaczane przez wynik funkcji haszującej. Czas obliczania hasza jest stały, a co za tym idzie, złożoność czasowa zapytań i wstawiania nowych krawędzi również. Teoretyczna złożoność pamięciowa jest natomiast stała i wynosi $O(m^2)$. W praktycznych zastosowaniach wybór m zależy jednak często od liczby krawędzi i przejmuje się m rzędu $O(\sqrt{|V|})$. Dokładność rezultatów może być niska ze względu na kolizje haszy i zależy od rozmiaru macierzy. Użyteczność tej struktury w bazowej formie jest dyskusyjna, stanowi ona jednak punkt wyjściowy dla bardziej zaawansowanych rozwiązań.

1.3 gMatrix[11]

Autorzy: A. Khan and C. Aggarwal.

Tytuł: Query-friendly compression of graph streams

Rok publikacji: 2016

Autorzy spoglądają w kierunku gSketch, zauważając potencjał tej metody, ale jednocześnie celnie wskazując jej ograniczenia. Przede wszystkim, wymaga ona próbek testowej dobrze reprezentującej właściwości grafu. Jej skuteczne

wyznaczenie może być trudne w wielkich grafach o dynamicznej strukturze. Dodatkowo, wspierane są głównie zapytania o częstotliwość występowania krawędzi, brakuje natomiast np. tych dotyczących osiągalności lub efektywnego wyznaczenia wierzchołków o szczególnie wysokiej częstotliwości. W odpowiedzi na te wyzwania zaproponowana zostaje 3-wymiarowa struktura gMatrix, gdzie dwa wymiary odpowiadają haszom wierzchołków, a trzeci numerowi użytej funkcji haszującej. Trzeci wymiar w tym przypadku pozwala zwiększyć dokładność zapytań, co odróżnia tę metodę od TCM. Podobnie jak tam, pamięć jest ograniczona przez wybrane rozmiary macierzy, a złożoność czasowa zależy od liczby użytych funkcji haszujących.

1.4 GSS[8, 9]

Autorzy: X. Gou, L. Zou, C. Zhao, and T. Yang

Tytuł: Graph Stream Sketch: Summarizing Graph Streams with High Speed and Accuracy

Rok publikacji: 2019 (2023)

1.5 GS4[2]

Autorzy: N. Ashrafi-Payaman, M. R. Kangavari, S. Hosseini, and A. M. Fander.

Tytuł: GS4: Graph stream summarization based on both the structure and semantics

Rok publikacji: 2020

1.6 Scube[3]

Autorzy: M. Chen, R. Zhou, H. Chen, and H. Jin.

Tytuł: Efficient summarization for skewed graph streams

Rok publikacji: 2022

1.7 Horae[4]

Autorzy: M. Chen, R. Zhou, H. Chen, J. Xiao, H. Jin, and B. Li.

Tytuł: Horae: A graph stream summarization structure for efficient temporal range query

Rok publikacji: 2022

1.8 Auxo[10]

Autorzy: Z. Jiang, H. Chen, and H. Jin

Tytuł: Auxo: A scalable and efficient graph stream summarization structure.

Rok publikacji: 2023

Przedstawiona zostaje struktura Auxo. Jest ona w dużej mierze rozwinięciem idei TCM i GSS. Głównym założeniem nowej struktury jest wprowadzenie wielu macierzy haszy, tworzących drzewo prefiksowe. Podpisy wierzchołków są wbudowywane w strukturę drzewa, co pozwala zredukować zużytą pamięć. Czas aktualizowania i podstawowych zapytań jest logarytmiczny względem liczby krawędzi. Koszt pamięciowy autorzy dość niefortunnie określają jako $O(|E|(1 - \log|E|))$. W rzeczywistości maksymalny rozmiar struktury jest ograniczony i zależy od długości użytych podpisów.

2 Najkrótsze ścieżki

2.1 All Pairs SP for undirected graphs[15]

Autorzy:

Tytuł:

Rok publikacji:

2.2 QbS[17]

Autorzy:

Tytuł:

Rok publikacji:

2.3 SP2[5]

Autorzy:

Tytuł:

Rok publikacji:

2.4 Road networks[1]

Autorzy:

Tytuł:

Rok publikacji:

2.5 $(1 + \epsilon)$ -approximate shortest paths[6]

Autorzy:

Tytuł:

Rok publikacji:

3 Literatura pomocnicza

3.1 Semi-streaming model[7]

Autorzy:

Tytuł:

Rok publikacji:

3.2 Subgraph Search[13]

Autorzy:

Tytuł:

Rok publikacji:

3.3 Graph summarization survey[14]

Autorzy:

Tytuł:

Rok publikacji:

3.4 Efficient Sketching algorithm[12]

Autorzy:

Tytuł:

Rok publikacji:

Literatura

- [1] G. Aggarwal, S. Gollapudi, R. Raghavender, and A. K. Sinop. Sketch-based algorithms for approximate shortest paths in road networks. *Proceedings of the Web Conference 2021*, 2021.
- [2] N. Ashrafi-Payaman, M. R. Kangavari, S. Hosseini, and A. M. Fander. Gs4: Graph stream summarization based on both the structure and semantics. *The Journal of Supercomputing*, 77(3):2713–2733, 2020.
- [3] M. Chen, R. Zhou, H. Chen, and H. Jin. Scube: Efficient summarization for skewed graph streams. *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*, 2022.
- [4] M. Chen, R. Zhou, H. Chen, J. Xiao, H. Jin, and B. Li. Horae: A graph stream summarization structure for efficient temporal range query. *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022.
- [5] B. Dolgorsuren, W. Xu, K. U. Khan, B.-S. Jeong, and Y.-K. Lee. Sp2: Spanner construction for shortest path computation on streaming graph. *Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory*, 2016.
- [6] M. Elkin and C. Trehan. Brief announcement: $(1+\epsilon)$ -approximate shortest paths in dynamic streams. *Proceedings of the 2022 ACM Symposium on Principles of Distributed Computing*, 2022.
- [7] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. *Theoretical Computer Science*, 348(2–3):207–216, 2005.
- [8] X. Gou, L. Zou, C. Zhao, and T. Yang. Fast and accurate graph stream summarization. *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019.
- [9] X. Gou, L. Zou, C. Zhao, and T. Yang. Graph stream sketch: Summarizing graph streams with high speed and accuracy. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5901–5914, 2023.
- [10] Z. Jiang, H. Chen, and H. Jin. Auxo: A scalable and efficient graph stream summarization structure. *Proceedings of the VLDB Endowment*, 16(6):1386–1398, 2023.
- [11] A. Khan and C. Aggarwal. Query-friendly compression of graph streams. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.

- [12] J. Lemiesz. Efficient framework for operating on data sketches. *Proceedings of the VLDB Endowment*, 16(8):1967–1978, 2023.
- [13] Y. Li, L. Zou, M. T. Ozsü, and D. Zhao. Time constrained continuous subgraph search over streaming graphs. *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019.
- [14] Y. Liu, T. Safavi, A. Dighe, and D. Koutra. Graph summarization methods and applications. *ACM Computing Surveys*, 51(3):1–34, 2018.
- [15] L. Roditty and U. Zwick. Dynamic approximate all-pairs shortest paths in undirected graphs. *SIAM Journal on Computing*, 41(3):670–683, 2012.
- [16] N. Tang, Q. Chen, and P. Mitra. Graph stream summarization. *Proceedings of the 2016 International Conference on Management of Data*, 2016.
- [17] Y. Wang, Q. Wang, H. Koehler, and Y. Lin. Query-by-sketch: Scaling shortest path graph queries on very large networks. *Proceedings of the 2021 International Conference on Management of Data*, 2021.
- [18] P. Zhao, C. C. Aggarwal, and M. Wang. gsketch: On query estimation in graph streams. *Proceedings of the VLDB Endowment*, 5(3):193–204, 2011.