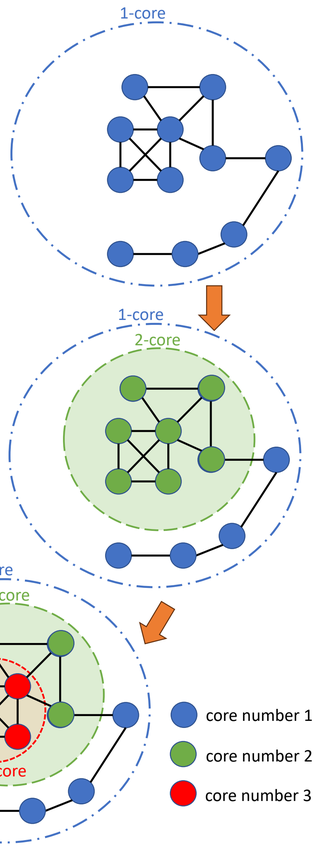
Privacy and Security in Distributed Graph Analytics: Secure Distributed k-core Decomposition

Bin Guo (binguo@trentu.ca)
Department of Computing & Information System, Trent University, Peterborough, ON, Canada

Motivation

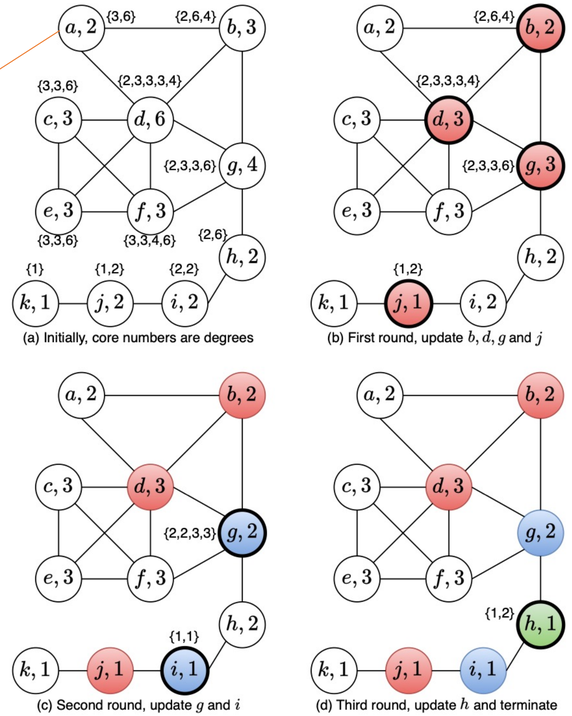
- Graphs are important data structures used in many applications:
  - Social Networks: Facebook, Twitter
  - Knowledge Networks: Dbpedia
  - Financial networks
- Data graphs can be distributed:
  - Decentralized Social Networks where each user is a client
  - Different organizations contribute subgraphs for analysis with keeping privacy
- Large distributed data graphs require algorithms for data analytics:
  - Strongly Connected Components
  - Minimum Spanning Forest
  - ***k*-Core**



- ***k*-Core Decomposition** [2] is to Find the largest subgraph, in which each node has at least k neighbours
- The core number is the largest value of k
- It is to find the dense part in a graph

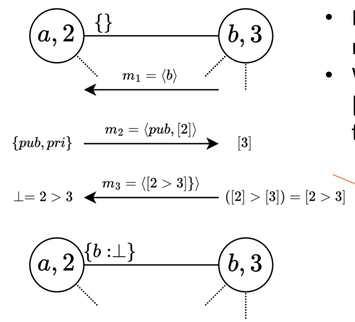
Distributed k-Core Decomposition

- a is ID of vertices
- 2 is core number
- {3, 6} is neighbours' core number
- a has core number as 2 such that it has two neighbours with core numbers at least 2



- **Locality** [3]: for a vertex u, u has at least k neighbours whose core numbers are k Then, k is the core number of u

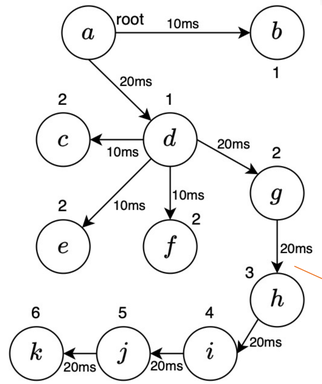
Contribution 1: Secure Core Number Comparison



- Vertex a compares core number with b
- No leaking the values and only get the result of true or false
- We use Homomorphic Encryption (HE) [3] together with asymmetric encryption for **Secure Integer Compare operation**

- m_1, m_2, m_3 are messages
- pub: public key; pri: private key
- [2]: encrypted value 2
- [2] > [3]: compare directly on encrypted values
- a only receive result as false without know b's core number

Contribution 2 and 3: Decentralized Termination Detection & Releasing Core Number with Distributed BFS Tree

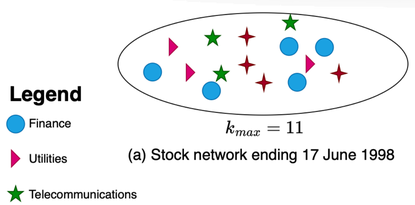


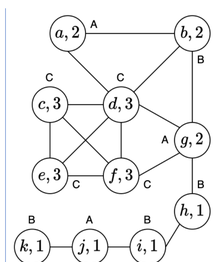
- A latency **BFS tree** is built with root vertex a
- The longest latency T is round time between a and k, $T = 240$ ms
- **Live vertices** send heartbeats forwarded to all the other vertices in BFS tree
- Each vertex is **live** if receive heartbeat within T, otherwise the termination detected

- The numbers beside edges are latency
- Each vertex has children and one parent
- The root does not have parent

- The core number distribution is released by accumulation in **BFS tree**
- By choosing a label and a core number like 'A1', starting from **leaves**, the counter will be accumulated added and send to parent repeatedly to root
- We use HE with asymmetric encryption for **Secure Integer Add operation,** only the root a obtain the core number distribution, e.g. |A1| = 1

Application of k-Core Distribution in Economy



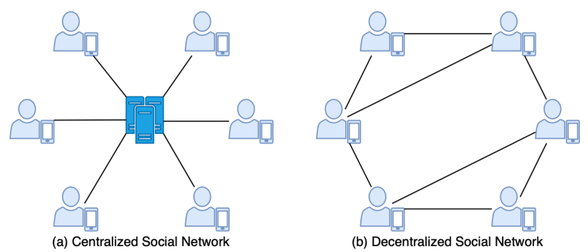


Label & Core Number	#Vertices
A1	1
A2	2
B1	3
B2	1
C3	4

- In Stock Networks, the max core is dominated by **Finance** in 2003 [1], so that **Finance** has huge effects to economy

- Only the **distribution of core number** can be released as results
- Not leak each vertex's core number

Application of Decentralized Social Networks



- **Each user is a client**, and the information is stored locally
- There not exists a single centralized server to store the global information
- Each client only know the directedly connected neighbours
- Users are not willing to share private information, e.g. core numbers and connection to other users

References

- [1] Burleson-Lesser, Kate, et al. "K-core robustness in ecological and financial networks." Scientific reports 2020.
- [2] V. Batagelj and M. Zaversnik, "An o(m) algorithm for cores decomposition of networks," *CoRR*, vol. cs.DS/0310049, 2003. [Online]. Available: http://arxiv.org/abs/cs/0310049
- [3] A. Montresor, F. De Pellegrini, and D. Miorandi, "Distributed k-core decomposition," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 2, pp. 288–300, 2013.
- [4] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys (Csur)*, vol. 51, no. 4, pp. 1–35, 2018.
- [5] **Bin Guo**, Emil Sekerinski, and Lingyang Chu. "Federated k-Core Decomposition: A Secure Distributed Approach." *arXiv preprint arXiv:2410.02544* (2024).