Bottom-up approaches. Newman [19] proposed a bottom-up approach for hierarchical community detection, where communities are merged, if the modularity of the resulting community can be increased, starting from vertices as seeds. In an attempt to improve the efficiency of this algorithm [19], Clauset et al. [4] proposed several optimization techniques, which is yet unable to scale to networks with millions of vertices. Expansion of every single vertex in a network can be very time consuming and it can result in many redundant communities, therefore in LFM [15] only a randomly selected set of vertices are expanded until the value of modularity is locally maximal. Similarly, Sales-Pardo et al [27] considered a set of local maxima communities according to modularity fitness metric. However, the modularity used in [15], [19] and [27] does not allow single vertices to be a community. Haveman et al. [13] modified the fitness function to allow a single vertex to be a community. Many researchers [7, 17, 21, 29] have attempted to use small subsets of vertices as seeds. For example, in the works EAGLE [29] and COCD [7] it was proposed to start with small cliques as seeds and merge two communities with the maximum similarity into one, where the similarity between two communities is proportional to the number of edges between them. However, cliques are a very strict condition for real-world communities. Lancichinetti et al. [16] presented the algorithm OSLOM, which uses an estimated statistical significance of a community as a fitness metric and locally optimizes the statistical significance of communities. Rosvall et al. [26] proposed a method called Informap, which is a generalization of their flow-based clustering method [25] to uncover hierarchical communities. Ahn et al. [1], proposed a different approach called link clustering, where edges are partitioned via hierarchical clustering.

8 CONCLUSIONS

In this paper, we studied the hierarchical community detection problem. We formally defined the problem of hierarchical community detection, as finding a rooted tree of communities where each community is a subset of its parent in the tree, and the information centrality of communities is no less than that of their parent in the hierarchical tree. We showed that the problem of finding hierarchical communities is NP-hard and devised an efficient and scalable heuristic algorithms for this problem. We further incorporated a fast sparsification method to reduce the network size for finding global cuts. We also proposed a fast randomized algorithm to estimate the value of information centrality in large-scale networks. We finally validate the effectiveness of our proposed algorithms using extensive experiments over five large-scale real-world datasets.

9 ACKNOWLEDGEMENT

This work is supported by the grant of Australian Research Council Discovery Project No. DP120102627.

REFERENCES

[1] Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. 2010. Link communities reveal multiscale complexity in networks. *Nature* 466, 7307 (2010), 761–764.
[2] Lijun Chang, Jeffrey Xu Yu, Lu Qin, Xuemin Lin, Chengfei Liu, and Weifa Liang. 2013. Efficiently computing k-edge connected components via graph decomposition. In *SIGMOD'13*. 205–216.
[3] James Cheng, Yiping Ke, Shumo Chu, and M Tamer Özsu. 2011. Efficient core decomposition in massive networks. In *ICDE*. 51–62.
[4] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. 2004. Finding community structure in very large networks. *Physical review E* 70, 6 (2004), 066111.
[5] Jonathan Cohen. 2008. Trusses: Cohesive subgraphs for social network analysis. *National Security Agency Technical Report* (2008), 16.
[6] Thomas H.. Cormen, Charles Eric Leiserson, Ronald L Rivest, and Clifford Stein. 2001. *Introduction to algorithms*. Vol. 6. MIT press Cambridge.
[7] Nan Du, Bai Wang, and Bin Wu. 2008. Overlapping community structure detection in networks. In *CIKM'08*. 1371–1372.
[8] David Eppstein and Joseph Wang. 2001. Fast approximation of centrality. In *Proc. of SODA'01*. 228–229.
[9] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3 (2010), 75–174.
[10] Santo Fortunato, Vito Latora, and Massimo Marchiori. 2004. Method to find community structures based on information centrality. *Physical review E* 70, 5 (2004), 056104.
[11] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *PNAS'02* 99, 12 (2002), 7821–7826.
[12] Prem K Gopalan and David M Blei. 2013. Efficient discovery of overlapping communities in massive networks. *PNAS'13* 110, 36 (2013), 14534–14539.
[13] Frank Havemann, Michael Heinz, Alexander Struck, and Jochen Gläser. 2011. Identification of overlapping communities and their hierarchy by locally calculating community-changing resolution levels. *Journal of Statistical Mechanics: Theory and Experiment* 2011, 01 (2011), P01023.
[14] Wassily Hoeffding. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association* 58, 301 (1963), 13–30.
[15] Andrea Lancichinetti, Santo Fortunato, and János Kertész. 2009. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* 11, 3 (2009), 033015.
[16] Andrea Lancichinetti, Filippo Radicchi, José J Ramasco, and Santo Fortunato. 2011. Finding statistically significant communities in networks. *PloS one* 6, 4 (2011), e18961.
[17] Conrad Lee, Fergal Reid, Aaron McDaid, and Neil Hurley. 2010. Detecting highly overlapping community structure by greedy clique expansion. *SNA/KDD'10* (2010), 33–42.
[18] Hiroshi Nagamochi and Toshihide Ibaraki. 1992. Computing edge-connectivity in multigraphs and capacitated graphs. *SIAM Journal on Discrete Mathematics* 5, 1 (1992), 54–66.
[19] Mark EJ Newman. 2004. Fast algorithm for detecting community structure in networks. *Physical review E* 69, 6 (2004), 066133.
[20] Mark EJ Newman and Michelle Girvan. 2003. Mixing patterns and community structure in networks. In *Statistical mechanics of complex networks*. Springer, 66–87.
[21] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 7043 (2005), 814–818.
[22] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. 2004. Defining and identifying communities in networks. *PNAS'04* 101, 9 (2004), 2658–2663.
[23] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. 2002. Hierarchical organization of modularity in metabolic networks. *Science* 297, 5586 (2002), 1551–1555.
[24] Mojtaba Rezvani, Weifa Liang, Wenzheng Xu, and Chengfei Liu. 2015. Identifying top-k structural hole spanners in large-scale social networks. In *CIKM'15*. 263–272.
[25] Martin Rosvall and Carl T Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *PNAS'08* 105, 4 (2008), 1118–1123.
[26] Martin Rosvall and Carl T Bergstrom. 2011. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PloS one* 6, 4 (2011), e18209.
[27] Marta Sales-Pardo, Roger Guimera, André A Moreira, and Luís A Nunes Amaral. 2007. Extracting the hierarchical organization of complex systems. *PNAS'07* 104, 39 (2007), 15224–15229.
[28] Satu Elisa Schaeffer. 2007. Graph clustering. *Computer science review* 1, 1 (2007), 27–64.
[29] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. 2009. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications* 388, 8 (2009), 1706–1712.
[30] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world' networks. *nature* 393, 6684 (1998), 440–442.
[31] Joyce Jiyoung Whang, David F Gleich, and Inderjit S Dhillon. 2013. Overlapping community detection using seed set expansion. In *CIKM'13*. 2099–2108.
[32] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. 2013. Overlapping community detection in networks: The state-of-the-art and comparative study. *Comput. Surveys* 45, 4 (2013), 43.
[33] Jaewon Yang and Jure Leskovec. 2013. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *WSDM'13*. ACM, 587–596.
[34] Ying Zhao, George Karypis, and Usama Fayyad. 2005. Hierarchical clustering algorithms for document datasets. *Data mining and knowledge discovery* 10, 2 (2005), 141–168.