Network motifs

1. RELATED WORK

We review algorithms and techniques for finding significantly overrepresented subgraphs in a (large) network. We can distinguish two main sub-areas:

- Network motif discovery usually refers to the discovery of subgraphs that are overrepresented with respect to network randomizations, with p-value higher than a certain threshold. Recent algorithms for network motifs are discussed in [20, 14, 12].
- Frequent subgraph mining refers to the discovery of subgraphs that occur more than a specified threshold. A survey on frequent subgraph mining can be found in [6].

1.1 Network motif discovery

Network motifs have been extensively studied by Alon et al [10, 8], with special focus on biological applications [1].

Some of the most popular algorithms are:

- mFinder [10, 8] (introduce network motifs, brute force and edge sampling)
- Moda [11]
- ESU (FANMOD) [19] (avoid duplication without simmetry breaking, node sampling)
- Grochow [4] (introduces simmetry breaking)
- Kavosh [7]
- NeMoFinder [3] (maximal, not necessarily induced motifs)
- gTrie [13]

Some of the above algorithms (mfinder, ESU, Grochow, GTrie) can be speed-up by sampling.

1.2 Frequent subgraph mining

Frequent subgraph mining is widely discussed in [6]. Some of the most known algorithms are gSpan, FSG, FFSM, Gaston and SUBDUE. Sampling has been recently introduced by Zou et al [21].

We are interested in occurrence-based counting (as opposed to transaction-based counting). This is usually applied to single large graphs. One of the main isues is to find a suport measure that has the downward closure property, since the number of disinct occurrences does not have this property. Vanetik et al [17, 5, 18] provided some theoretical conditions for the downward closure property to hold and defined a measure (hard to compute) based on the concept of occurrences-overlap graph. Calders et al [2] built upon it and defined a polynomial support measure.

Similarly, Karypis et al [9] and Shreiber at al [15, 16] (MaVisto) defined the concept of "number of edge-disjoint subgraphs" for counting, and proposed an a-priori-like algorithm.

2. REFERENCES

[1] U. Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6), 2007.

[2] T. Calders, J. Ramon, and D. Van Dyck. Anti-monotonic overlap-graph support measures. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 73–82. IEEE, 2008.

[3] J. Chen, W. Hsu, M. L. Lee, and S.-K. Ng. Nemofinder: Dissecting genome-wide protein-protein interactions with meso-scale network motifs. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 106–115. ACM, 2006.

[4] J. A. Grochow and M. Kellis. Network motif discovery using subgraph enumeration and symmetry-breaking. In *Research in Computational Molecular Biology*, pages 92–106. Springer, 2007.

[5] E. Gudes, S. E. Shimony, and N. Vanetik. Discovering frequent graph patterns using disjoint paths. *Knowledge and Data Engineering, IEEE Transactions on*, 18(11):1441–1456, 2006.

[6] C. Jiang, F. Coenen, and M. Zito. A survey of frequent subgraph mining algorithms. *Knowledge Engineering Review*, 28(1):75–105, 2013.

[7] Z. R. M. Kashani, H. Ahrabian, E. Elahi, A. Nowzari-Dalini, E. S. Ansari, S. Asadi, S. Mohammadi, F. Schreiber, and A. Masoudi-Nejad. Kavosh: a new algorithm for finding network motifs. *BMC bioinformatics*, 10(1):318, 2009.

[8] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11), 2004.

[9] M. Kuramochi and G. Karypis. Finding frequent patterns in a large sparse graph*. *Data mining and knowledge discovery*, 11(3):243–271, 2005.

[10] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple

building blocks of complex networks. *Science Signalling*, 298(5594):824, 2002.

[11] S. Omidi, F. Schreiber, and A. Masoudi-Nejad. MODA: an efficient algorithm for network motif discovery in biological networks. *Genes & genetic systems*, 84(5):385–395, 2009.

[12] P. Ribeiro. *Efficient and Scalable Algorithms for Network Motifs Discovery*. PhD thesis, PhD thesis, University of Porto, 2011.

[13] P. Ribeiro and F. Silva. G-Tries: a data structure for storing and finding subgraphs. *Data Mining and Knowledge Discovery*, 28(2):337–377, 2014.

[14] P. Ribeiro, F. Silva, and M. Kaiser. Strategies for network motifs discovery. In *e-Science, 2009. e-Science'09. Fifth IEEE International Conference on*, pages 80–87. IEEE, 2009.

[15] F. Schreiber and H. Schwobbermeyer. Towards motif detection in networks: Frequency concepts and flexible search. *Proc. Intl. Wsh. Network Tools and Applications in Biology (NETTAB'04)*, pages 91–102, 2004.

[16] F. Schreiber and H. Schwöbbermeyer. Frequency concepts and pattern detection for the analysis of motifs in networks. In *Transactions on computational systems biology III*, pages 89–104. Springer, 2005.

[17] N. Vanetik, E. Gudes, and S. E. Shimony. Computing frequent graph patterns from semistructured data. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 458–465. IEEE, 2002.

[18] N. Vanetik, S. E. Shimony, and E. Gudes. Support measures for graph data*. *Data Mining and Knowledge Discovery*, 13(2):243–260, 2006.

[19] S. Wernicke. Efficient detection of network motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 3(4):347–359, 2006.

[20] E. Wong, B. Baur, S. Quader, and C.-H. Huang. Biological network motif detection: principles and practice. *Briefings in bioinformatics*, 13(2):202–215, 2012.

[21] R. Zou and L. B. Holder. Frequent subgraph mining on a single large graph using sampling techniques. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, pages 171–178. ACM, 2010.