Contents lists available at ScienceDirect
Physica A
journal homepage: www.elsevier.com/locate/physa

Comment on 'An optimal algorithm for counting networks motifs' [Physica A 381 (2007) 482–490]

Sebastian Wernicke
Institut für Informatik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany

ARTICLE INFO

Article history:
Received 1 March 2009
Received in revised form 28 February 2010
Available online xxxx

Keywords:
Network motif detection algorithm
FANMOD

ABSTRACT

Network motifs in a given network are small connected subnetworks that occur at significantly higher frequencies than would be expected for a random network. In their 2007 article "An optimal algorithm for counting network motifs", Itzhack, Mogilevski, and Louzoun present an algorithm for detecting network motifs. Based on an experimental comparison with a motif detection software called FANMOD, they claim that their algorithm is "more than a thousand times faster" than any previous motif detection algorithm. We show that this claim is not correct and based on a significant flaw in the experimental setup. Once the experimental data of Itzhack et al. is corrected for this flaw, the implementation of their algorithm actually turns out to be a little *slower* than FANMOD for random Erdős–Rényi graphs. For random scale-free networks, the implementation of Itzhack et al. is faster only by a factor of ∼1.5, not the orders of magnitude claimed by Itzhack et al.

1. Introduction

Network motifs in a given network are small connected subnetworks that occur at significantly higher frequencies than would be expected for a random network with the same degree distribution [1,2]. The task of algorithmically detecting network motifs is computationally expensive and has therefore received considerable attention due to its biological relevance [3–7]. In 2005 and 2006, Wernicke described an algorithm called "ESU" [7,8] for the efficient enumeration of connected fixed-size subnetworks in a given network, a crucial step of network motif detection. This algorithm was implemented in early 2006 by Wernicke and Rasche in a tool called "FANMOD" [6].

In 2007, Itzhack et al. [3] published a "motif count (sic!) algorithm", which we shall refer to as "MCA" for short. Based on an experimental comparison to the FANMOD tool, one of the main claims by Itzhack et al. is that MCA is "more than a thousand times faster" [3] than any previous motif detection algorithm, in particular the algorithm employed by FANMOD. In this comment, we show that this claim is not correct and draw attention to two errors made by Itzhack et al.:

1. Itzhack et al. [3] wrongly claim that FANMOD is not capable of enumerating all fixed-size subnetworks in a given network.
2. Itzhack et al. [3] wrongly claim that MCA is much faster than FANMOD. We show that their assertion is based on a distorted experimental setup. Once this distortion is accounted for, the data of Itzhack et al. [3] indicates that MCA is in fact *slower* than FANMOD on random Erdős–Rényi graphs and only slightly faster on random scale-free graphs.

Finally, Itzhack et al. [3] claim their algorithm to be "novel", whereas in fact it is very similar to the ESU algorithm [7,8] on which FANMOD [6] is based: in short, both algorithms start with a seed node v, then separate the neighbourhood of v into layers based on the minimum distance to v, and finally explore these layers in an iterative fashion before removing v from further consideration.

E-mail address: sebastian.wernicke@uni-jena.de.

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.05.035

2. FANMOD is capable of subnetwork enumeration

Ithak et al. [3] claim that FANMOD [6,9] is only capable of "sampling" subnetworks, i.e., that FANMOD is not able to perform a full enumeration of all connected subnetworks of a given size. This claim is not correct. As outlined in the FANMOD manual [9], the original FANMOD publication [6], and the paper describing the underlying algorithmics [7], FANMOD is fully capable of performing this task. If the user desires, FANMOD is *also* capable of sampling subnetworks in a uniform manner, but this is turned off in the default setting.

Somewhat surprisingly, Itzhack et al. [3] state that the FANMOD software is always sampling "exactly 100 000" subnetworks. This statement is not correct:

1. As outlined above, FANMOD is not confined to sampling.
2. If FANMOD is used in sampling mode, the number of samples can be adjusted by the user and is not restricted to any particular number
3. As outlined in Refs. [7,8], FANMOD cannot sample an *exact number* of motifs, only an *expected fraction*.

The prominent use of the particular number "100 000" appears to stem from a misunderstanding of Itzhack et al. [10, 3]: FANMOD prominently displays the word "sampling" and the number 100 000 together in its graphical user interface. However, this only refers to a *preprocessing* step of the algorithm that enables the display of an accurate progress bar (this is discussed in more detail in the FANMOD manual [9]).

3. FANMOD is often faster than the algorithm of Itzhack et al.

The experimental setup that Itzhack et al. [3] used to compare running times between MCA and FANMOD is based on the wrong assumption that the FANMOD software is only capable of sampling subnetworks. Since their MCA algorithm performs a full enumeration, Itzhack et al. [3] therefore designed their experiments such that FANMOD was always run on *ten times* the number of networks that the MCA algorithm was run on. In other words: Despite both FANMOD and the MCA implementation performing exactly the same task, the experiments of Ithak et al. [3] gave FANMOD ten times the algorithmic load compared to MCA, yet running times were directly compared.[1] It comes as no surprise that the experiments of Itzhack et al. [3] indicated MCA to be faster than FANMOD.

Since we were not able to obtain the data or implementation used for the experimental results in Ref. [3], no direct measurements between MCA and FANMOD could be made to see how running times compare in a "fair" experimental setting. In order to nevertheless gain an understanding of how MCA compares with FANMOD, we used the following procedure in order to adjust the published experimental results of Itzhack et al. [3] so that the different loads are accounted for:

1. FANMOD was run with 1 000 random Erdős–Rény graphs and 1 000 random scale-free networks, each consisting of between 100 and 50 000 nodes.
2. For each of the 2 000 runs, two scenarios were tested, one analogous to the experimental setup of Itzhack et al. (generating 1 000 random comparison networks), another one representing a "fair" experimental setup (generating 100 random comparison networks).
3. The results show that the running time of FANMOD as measured by Itzhack et al. [3] is too high by a factor of ∼9.91.

Fig. 1 shows the adjusted experimental results once the factor of 9.91 is accounted for.

For random Erdős–Rényi graphs, Fig. 1 shows that FANMOD clearly outperforms the implementation of Itzhack et al. [3]. FANMOD is slightly slower for very large scale-free networks, which might be the result of a specific algorithmic strategy employed by Itzhack et al. [3], where nodes with high connectivity are considered early on in the enumeration process and can then be discarded from further consideration (this is advantageous, e.g., considering memory locality on modern computer architectures). As a suggestion for future research, it would be interesting to test this strategy with FANMOD in future implementations.

Finally, it should be noted that the experiments of Itzhack et al. [3] do not provide a precise comparison of the ESU algorithm to MCA, since both implementations utilize different algorithms for grouping the detected subnetworks into isomorphic classes and most of the underlying random networks were not identical (for further details on correcting for these effects, see, e.g., the experimental section of Ref. [7]).

4. Conclusion

As shown above, the algorithm of Itzhack et al. [3] is not generally faster than FANMOD [6,7]; their respective claims in [3] are only due to a misunderstanding of the underlying algorithms, which led to a strongly biased experimental setup.[2]

---

1 Actually, the "correction" factor is not precisely 10, since Itzhack et al. ran their algorithm on 101 networks whereas they ran FANMOD on 1001 networks. Thus, the load is roughly increased by a factor of $1001/101 = 9.91$. As subsequently outlined, this scale factor can also been experimentally confirmed.

2 Note that Itzhack et al. [3] do not distinguish between the motif detection algorithms (called "ESU" and "RAND-ESU") and the respective implementation (called "FANMOD").

Please cite this article in press as: S. Wernicke, Comment on 'An optimal algorithm for counting networks motifs' [Physica A 381 (2007) 482–490], Physica A (2010), doi:10.1016/j.physa.2010.05.035

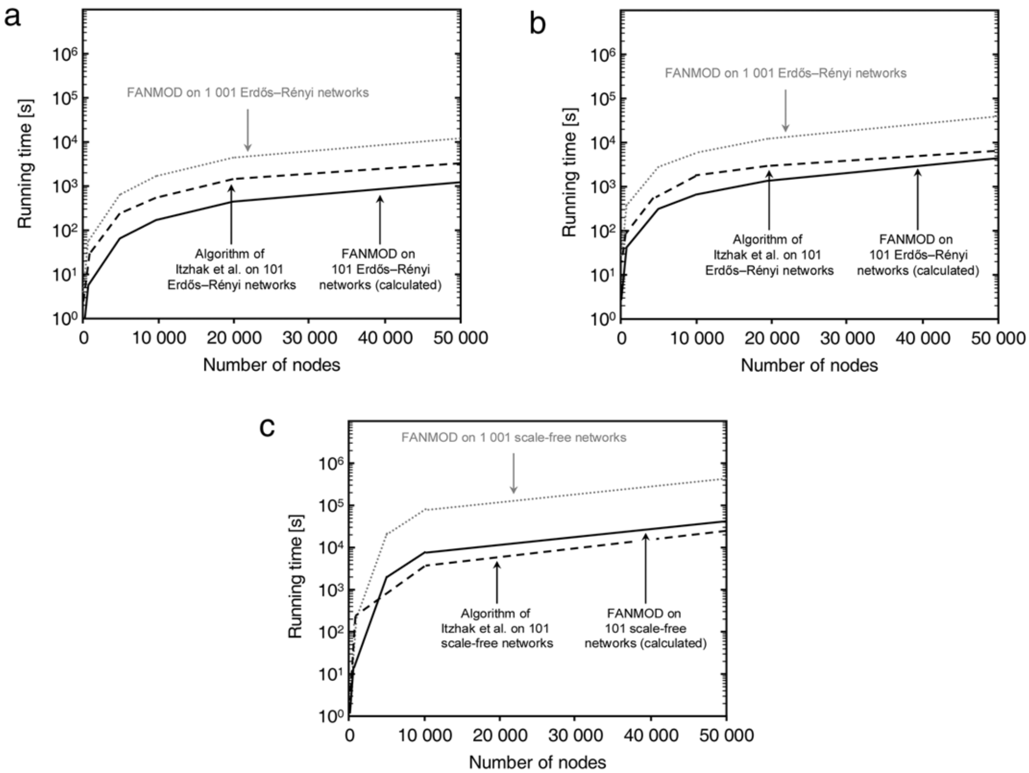

Fig. 1. Correcting the data of Itzhack et al. [3] for flaws in the experimental setup strongly indicates that FANMOD [6] remains one of the fastest motif detection tools. Figure (a) shows the detection speed for size-3 motifs; Figures (b) and (c) show the detection speed for size-4 motifs.

For future work, it would be interesting to see if the algorithmic techniques and algorithm engineering methods employed by Wernicke [7] and Itzhack et al. [3] could be combined in order to arrive at a motif detection algorithm that is indeed much faster than existing implementations.

References

[1] R. Milo, et al., Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827.
[2] R. Milo, et al., Superfamilies of designed and evolved networks, Science 303 (2004) 1538–1542.
[3] R. Itzhack, Y. Mogilevski, Y. Louzoun, An optimal algorithm for counting network motifs, Physica A 381 (2007) 482–490.
[4] N. Kashtan, et al., Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs, Bioinformatics 20 (2004) 1746–1758.
[5] F. Schreiber, H. Schwöbbermeyer, Mavisto: a tool for the exploration of network motifs, Bioinformatics 21 (2005) 3572–3574.
[6] S. Wernicke, F. Rasche, FANMOD: a tool for fast network motif detection, Bioinformatics 22 (9) (2006) 1152–1153.
[7] S. Wernicke, Efficient detection of network motifs, IEEE/ACM Transactions on Computational Biology and Bioinformatics 3 (4) (2006) 347–359.
[8] S. Wernicke, A faster algorithm for detecting network motifs, in: Proceedings of the 5th Workshop on Algorithms in Bioinformatics, WABI'05, in: Lecture Notes in Bioinformatics, vol. 3692, 2005, pp. 165–177.
[9] S. Wernicke, F. Rasche, FANMOD manual, 2006, Available online at http://theinf1.informatik.uni-jena.de/motifs/fanmod-manual.pdf.
[10] R. Itzhack, Personal correspondence on October 28th, 2008.

Please cite this article in press as: S. Wernicke, Comment on 'An optimal algorithm for counting networks motifs' [Physica A 381 (2007) 482–490], Physica A (2010), doi:10.1016/j.physa.2010.05.035