



Biological networks: the microscope of the twenty-first century?

Frank Emmert-Streib^{1,2*} and Matthias Dehmer^{3,4}

¹ Computational Medicine and Statistical Learning Laboratory, Department of Signal Processing, Tampere University of Technology, Tampere, Finland, ² Institute of Biosciences and Medical Technology, Tampere, Finland, ³ Department of Computer Science, Universität der Bundeswehr München, Germany, ⁴ Department of Mechatronics and Biomedical Computer Science, UMIT, Hall in Tyrol, Austria

Keywords: biological networks, computational biology, network biology, translational research, systems medicine

The development of science is accompanied by a number of seminal inventions that contributed significantly to make science what it is today. In biology and the biomedical sciences, the introduction of the light microscope in the seventeenth century by Antony van Leeuwenhoek and others can certainly be seen as such a contribution because it literally opened a door to a new universe that cannot be seen by the naked eye, the molecular world of biology (Wollman et al., 2015). Within 100 years after its introduction, new phenomena in biology and medicine have been discovered, e.g., different types of cells such as red blood cells or the existence of single-cell organisms like bacteria. Furthermore, performing tissue analysis enabled histopathology and new concepts of diseases based on cells rather than whole organs.

Extending the principle idea of the microscope, a variety of imaging technologies have been introduced since then allowing to visualize for instance, the 3D-structure of a protein (crystallography), broken bones (x-ray), the structure of atoms (electron microscope), or the neuronal activity of the brain (MRI). Despite the technical differences between all of these different imaging technologies they share the common principle idea to enhance the capabilities of our eyes by making them more sensitive to the scale of the microscopic and atomic world. Hence, this allows it for everyone to easily grasp the capabilities of any of the above imaging methods and to understand the resulting visualizations as a direct reflection of reality.

A recent publication by Cohen (2004) attracted wide attention, because the author put the provocative argument forward that *mathematics is biology's next microscope*. In our paper, we provide a succinct example for this. In our opinion *biological networks* are a prime example for a mathematical approach that has indeed much in common with a microscope, but differs in one important point making them even more potent for biology and medicine. Namely, biological networks are mathematical models, as we will argue below.

The history of networks, or graphs in general, started at about the same time as that of the microscope with Euler studying the Königsberger bridges problem (Seven Bridges of Königsberg; Euler, 1736) while the term “graph” was coined by König much later in the 1930s (König, 1936). The first wave of general interest outside mathematics was sparked by the introduction of *random networks* in the 1950s (Solomonoff and Rapoport, 1951; Erdős and Rényi, 1959) followed by the second wave in the mid 1990s discovering that many different types of “real” networks have structural properties quite different from random, which can be better described by scale-free and small-world networks (Watts and Strogatz, 1998; Barabási and Albert, 1999). One highlight for the application of random networks in biology can be found in the seminal work of Kauffman (1969) introducing models for gene regulations, whereas scale-free and small-world networks helped shaping the emergence of systems biology, network biology, and network medicine (Barabási et al., 2011). Fueled by big data from genome-scale screening experiments in combination with powerful statistical inference methods this enabled finally the emergence of gene regulatory networks for mammalian organisms including Human (Basso et al., 2005; Emmert-Streib et al., 2014). Note

OPEN ACCESS

Edited by:

Yasset Perez-Riverol,
EMBL European Bioinformatics
Institute, UK

Reviewed by:

Ali Masoudi-Nejad,
University of Tehran, Iran
Antonio Marco,
University of Essex, UK

*Correspondence:

Frank Emmert-Streib
v@bio-complexity.com

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 03 August 2015

Accepted: 23 September 2015

Published: 13 October 2015

Citation:

Emmert-Streib F and Dehmer M
(2015) Biological networks: the
microscope of the twenty-first
century? *Front. Genet.* 6:307.
doi: 10.3389/fgene.2015.00307

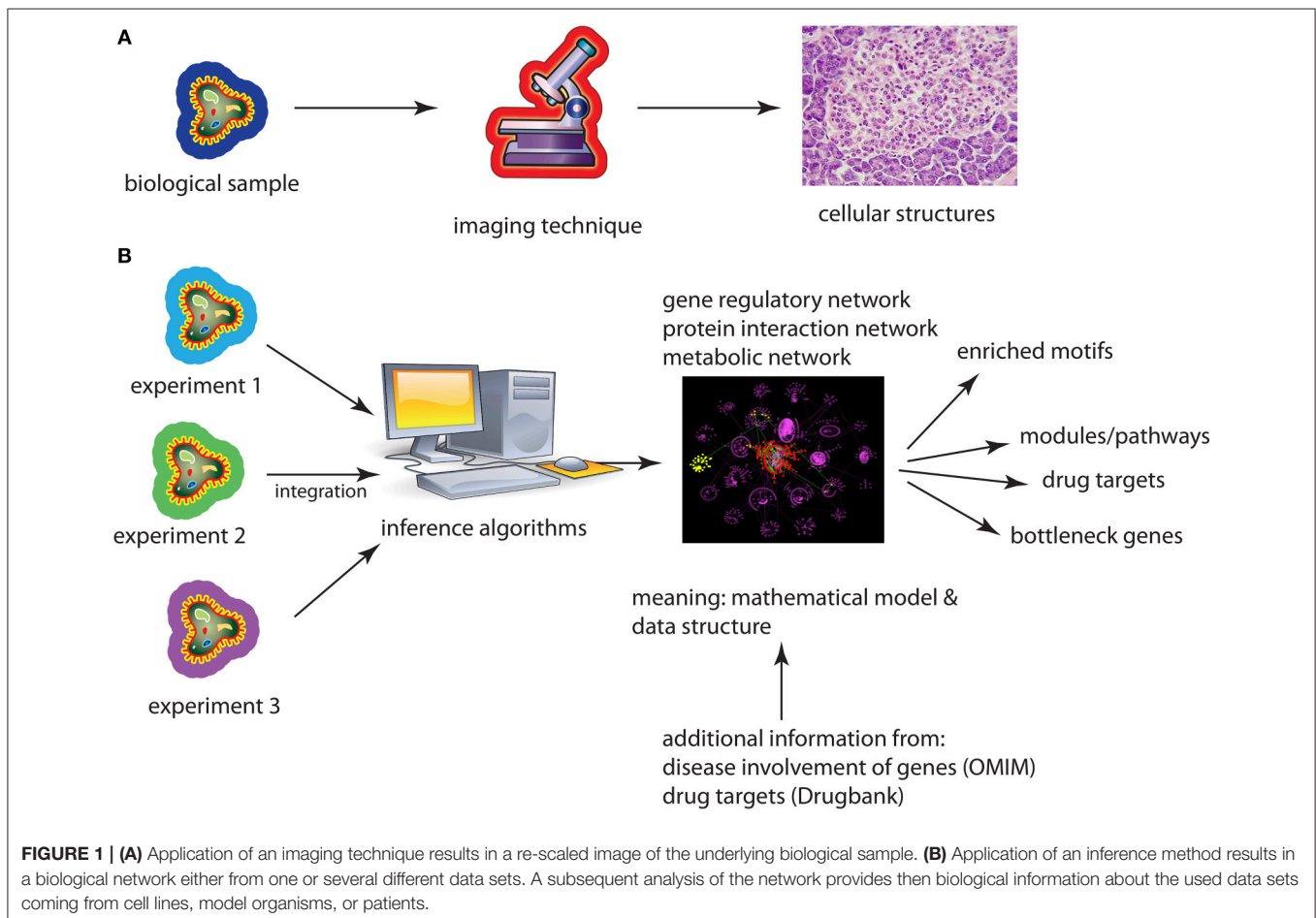
that graph-theoretical methods such as graph comparison to analyze biological networks have been discussed in Emmert-Streib and Dehmer (2011).

A similarity between microscopes and networks is that both allow the visualization of an underlying phenomenon so it can be explored by means of the eyes. However, a fundamental difference is that microscopes visualize *geometric objects* whereas networks show *topological objects*, see **Figure 1**. The difference is that neither the nodes/vertices of a graph nor their edges/links are associated with 2- or 3-dimensional coordinates that specify their Euclidean position, e.g., on a canvas or screen uniquely, but their positions are not defined by the definition of the graph at all. In other words, a graph is arbitrarily deformable with respect to the positions of its constituting components.

The implications that result from this fundamental difference are enormous and we want to highlight just three examples. First, networks are not merely a magnification of a microscopic world in a different scale, like a model car. Instead, networks are mathematical models that cannot be observed with a microscope at all, even with an hypothetically large magnification. For instance, a transcriptional regulatory network (TRN) is a directed network that represents the regulations of transcription factors on the transcription of genes on a genomic-scale. While it is true that one specific transcription factor binds at a particular

time and space to the promotor region of a gene to initiate its transcription, not all transcription regulation events happen at this particular time or space. Also, no matter how long you observe such events, they never happen all at once. As such the TRN does not represent what can be observed but it corresponds to averaged observations over time and space making it a mathematical model of transcription regulation. Second, since all types of biological networks are *models*, which are per se abstract, one needs actually to derive the information captured in them by application of abstract (mathematical/statistical) methods. Examples for such an information extraction analysis step are the identification of modules/pathways, identification of bottleneck genes or the detection of evolutionary conserved motifs (Milo et al., 2002; Yu et al., 2007). Third, considering networks as data structures (Emmert-Streib and Dehmer, 2011) allows a convenient integration of high-throughput data from different experiments, e.g., by the application of Bayesian methods. This also allows the integration across heterogeneous data types, e.g., from gene expression and proteomics, resulting in hierarchical, multi-scale models of molecular interactions.

We think that the long standing visualization tradition in biology and medicine by means of microscopes causes also one of the biggest misconceptions of biological networks. Namely, that their purpose is their visualization to “see” what information



the networks contain. However, the fact that one can visualize biological networks does not establish this as their main purpose. Instead, the main purpose of any mathematical model is to be used for extracting information from it, and a visualization is just one side of this, e.g., by enabling an exploratory analysis (Tukey, 1977).

We are of the opinion that, so far, biological networks are not utilized up to their full potential and there are at least three reasons for this. First, when biological networks are inferred from data it is not clear where to deposit them so they can be re-used by other groups. In contrast, when generating gene expression or next-generation sequencing data there are a number of well known databases that are open to anyone to deposit the data, e.g., GEO or Ensembl. These databases enforce also the documentation of the data in a standardized way that ensures an easy re-use of the data with a minimal (ideally no) interaction with the data generating groups. At the moment, it is unclear where to store inferred biological networks, in what format and what additional information needs to be provided because, presently, there are no official standards. However, the R packages *igraph* and *graph* (Csardi and Nepusz, 2006; Gentleman et al., 2010) should be notably mentioned for providing low level graph operations. In addition, depositing networks into databases would be also an important step toward reproducible results in this area.

Second, due to the lack of databases from which biological networks can be acquired, necessarily, an analysis starts with the inference of the networks. Unfortunately, the application of causal inference algorithms to experimental or observational data (Emmert-Streib et al., 2012) is far from being trivial leading frequently to low quality networks, even when starting with high quality data. This makes the resulting downstream analysis at least problematic or even controversial. Due to the intricacy of this analysis step, we suggest a collaborative approach by teaming up with a computational biology or biostatistics group that is specialized in such an analysis in order to ensure the highest

quality of the inferred networks and all derived subsequent results.

As a third issue, we think that complementing biological networks with additional information, e.g., from associations between genes and disorders (OMIM) or drug targets (Drugbank) would be very beneficial for studying problems beyond biology, e.g., in translational medicine (Dudley and Karczewski, 2013). Notable example studies can be found in Ideker and Sharan (2008); Kotlyar et al. (2012). This connection would allow to bring a systems approach down to the patient side, e.g., for investigating diagnostic or prognostic questions. Furthermore, in this way, naturally, a door to pharmacogenomics is opened placing the drug targets into their molecular interaction networks. In this way, new insights into the side effects of drugs may be gained.

Returning back to our initial analogy, biological network research is surely still at the stage of the early microscope and it maybe too early to state without uncertainty that biological networks will indeed acquire the status of a microscope. However, implementing some of the above measures will quickly allow biological networks to further blossom in biological, medical, and pharmacogenomic research to drive these fields to a level unachievable without the usage of biological networks.

AUTHOR CONTRIBUTIONS

FE and MD conceived the study, wrote the paper, and approved the final version.

FUNDING

MD thanks the Austrian Science Funds for supporting this work (project P26142). MD gratefully acknowledges financial support from the German Federal Ministry of Education and Research (BMBF) (project RiKoV, Grant No. 13N12304).

REFERENCES

- Barabási, A. L., and Albert, R. (1999). Emergence of scaling in random networks. *Science* 206, 509–512.
- Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918
- Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human b cells. *Nat. Genet.* 37, 382–390. doi: 10.1038/ng1532
- Cohen, J. E. (2004). Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PLoS Biol* 2:e439. doi: 10.1371/journal.pbio.0020439
- Csardi, G., and Nepusz, T. (2006). The *igraph* software package for complex network research. *InterJ. Complex Syst.* 1695.
- Dudley, J., and Karczewski, K. (2013). *Exploring Personal Genomics*. Oxford: Oxford University Press.
- Emmert-Streib, F., de Matos Simoes, R., Mullan, P., Haibe-Kains, B., and Dehmer, M. (2014). The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. *Front. Genet.* 5:15. doi: 10.3389/fgene.2014.00015
- Emmert-Streib, F., and Dehmer, M. (2011). Networks for systems biology: conceptual connection of data and function. *IET Syst. Biol.* 5, 185. doi: 10.1049/iet-syb.2010.0025
- Emmert-Streib, F., Glazko, G., Altay, G., and de Matos Simoes, R. (2012). Statistical inference and reverse engineering of gene regulatory networks from observational expression data. *Front. Genet.* 3:8. doi: 10.3389/fgene.2012.00008
- Erdős, P., and Rényi, A. (1959). On random graphs. I. *Publ. Math.* 6, 290–297.
- Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Comment. Acad. Sci. U. Petrop.* 8, 128–140.
- Gentleman, R., Whalen, E., Huber, W., and Falcon, S. (2010). *Graph: A Package to Handle Graph Data Structures*. R package version 1.46.0.
- Ideker, T., and Sharan, R. (2008). Protein networks in disease. *Genome Res.* 18, 644–652. doi: 10.1101/gr.071852.107
- Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467. doi: 10.1016/0022-5193(69)90015-0
- König, D. (1936). *Theorie der Endlichen und Unendlichen Graphen*. White River Junction, VT: Chelsea Green Publishing.
- Kotlyar, M., Fortney, K., and Jurisica, I. (2012). Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods* 57, 499–507. Protein-Protein Interactions. doi: 10.1016/j.jmeth.2012.06.003

- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. doi: 10.1126/science.298.5594.824
- Solomonoff, R., and Rapoport, A. (1951). Connectivity of random nets. *Bull. Math. Biophys.* 13, 107–117. doi: 10.1007/BF02478357
- Tukey, J. (1977). *Exploratory Data Analysis*. New York, NY: Addison-Wesley.
- Watts, D., and Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442. doi: 10.1038/30918
- Wollman, A. J. M., Nudd, R., Hedlund, E. G., and Leake, M. C. (2015). From animaculum to single molecules: 300 years of the light microscope. *Open Biol.* 5. doi: 10.1098/rsob.150019. Available online at: http://rsob.royalsocietypublishing.org/content/5/4/150019?utm_source=Adestra&utm_medium=email&utm_campaign=Open%20Biology%20content%20mailing%20-%20May%202015&utm_content=From%20Animaculum%20to%20single%20molecules%3A%20300%20years%20of%20the%20light%20microscope&utm_term=Open%20Biology%20content%20mailing%20-%20May%202015
- Yu, H., Kim, P., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* 3:e59. doi: 10.1371/journal.pcbi.0030059

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Emmert-Streib and Dehmer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.