

Preface

The past decade has seen an incredible growth of network methods following publications by Laszlo Barabasi and others. Excellent text books exist on general networks and graph theory, but these books typically describe unweighted networks. This book focuses on weighted networks. In weighted networks, the pairwise connection strength between two nodes is quantified by a real number between 0 and 1. It is worth emphasizing that **most of the material also applies to unweighted networks**. Further, unweighted networks can easily be constructed from weighted networks by dichotomizing the connection strengths between nodes. While unweighted networks permit graph-theoretic visualization techniques and algorithms, weighted networks can be advantageous for many reasons including the following:

1. They preserve the continuous nature of the underlying connectivity information. For example, weighted correlation networks that are constructed on the basis of correlations between numeric variables do not require the choice of a hard threshold (Chap. 5). Dichotomizing information and (hard)-thresholding may lead to information loss.
2. They often lead to highly robust results (Zhang and Horvath 2005). In contrast, results based on unweighted networks, constructed by thresholding a pairwise association measure, often strongly depend on the threshold.
3. They can sometimes be decomposed and approximated by simpler networks. For example, networks can sometimes be approximated by “factorizable” networks (Chap. 2). Such approximations are often difficult to achieve for sparse, unweighted networks.
4. They sometimes allow for a parsimonious parametrization (in terms of modules and conformities, see Sect. 2.3).
5. They often allow one to derive simple relationships between network concepts (statistics) (Sect. 3.8 and Chap. 6). In particular, weighted *correlation* networks facilitate a geometric interpretation based on the angular interpretation of the correlation (Sect. 6.7).
6. They can be used to enhance standard data-mining methods such as cluster analysis since (dis)-similarity measures can often be transformed into weighted networks (Sect. 7.7).

Many of the applied sections in this book present analysis techniques and strategies to the wider audience of applied researchers. The book assumes little mathematical and statistical knowledge, but some sections are rather abstract. To make the book self-contained, some sections review statistical and data-mining techniques. Since several technical sections and chapters are less relevant to applied researchers, they start out with the warning that they can be skipped. I also present abstract, theoretical material since it may be useful for quantitative researchers, who carry out methodological research. In my own experience, I have found that applied researchers can be expert users of network methods and software. Of course, domain-knowledge experts have often a superior intuition about how to arrive at a meaningful analysis of their data. Many weighted network methods arose from collaborations with cancer biologists, neuroscientists, mouse geneticists, and biologists (e.g., see the acknowledgement section and references).

Although the field of weighted network analysis only began a few years ago, it is already impossible to summarize it in one book. I have tried to cite as many articles as possible, but I apologize to my colleagues for failing to cite their work. Several people are mentioned throughout the book, see the index for names and page numbers. I am acutely aware that I leave unmentioned many important ideas and techniques. My only excuse for giving too much attention to my own work is that I understand it best.

While the methods are formulated in general terms, which facilitate their application to wide variety of data, most applications involve genes, proteins, and gene expression data. It has become clear that networks have important medical and biological applications. Gene co-expression networks bridge the gap from individual genes to clinically important, emergent phenotypes. Gene networks allow one to move beyond single-gene comparisons and systematically identify biologically meaningful relationships between gene products, pathways, and phenotypes. Weighted gene co-expression network analysis (WGCNA) has been used to identify candidate disease biomarkers, to annotate genes with regard to module membership, to study the relationships between co-expression modules, and to compare the network topology of different networks. Case studies show how WGCNA can be used to screen for genes, to understand the transcriptional architecture, and to relate modules in different mouse tissues. Integrating co-expression networks with genetic marker data facilitates systems genetic applications (Sects. 11.5 and 12.3), which make use of causal testing and network edge-orienting procedures.

Freely Available R Software

This book provides an in-depth description of the WGCNA R package (Langfelder and Horvath 2008), which provides functions for carrying out network analysis tasks. R is a freely available, open source language and environment for statistical computing and graphics, which has become a de-facto standard in data analysis (Ihaka and Gentleman 1996; Venables and Ripley 2002; Gentleman et al. 2004,

2005; Carey et al. 2005). The R environment integrates standard data analysis and visualization techniques with packages (libraries) implementing the latest advances in data mining, statistics, and machine learning. The WGCNA package is available from the Comprehensive R Archive Network (CRAN), the standard repository for R add-on packages. To install it, type the following command into the R session:

```
install.packages("WGCNA")
```

Most of the R code and data presented in the book chapters can be downloaded from the following webpage:

www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Book

Related scientific articles and presentations can be found at the following webpages: www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/.

Other relevant R packages mentioned throughout the book are freely available on the R CRAN package resource at

www.R-project.org

and/or on the Bioconductor webpage.

Carey VJ, Gentry J, Whalen E, Gentleman R (2005) Network structures and algorithms in bioconductor. *Bioinformatics* 21(1):135–136

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. *J Comput Graph Stat* 5(3):299–314

Gentleman RC, Carey VJ, Bates DJ, Bolstad BM, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth GK, Tierney L, Yang YH, Zhang J (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80

Gentleman R, Huber W, Carey V, Irizarry R, Dudoit S (2005) *Bioinformatics and computational biology solutions using R and bioconductor*. Springer, New York

Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform* 9(1):559

Venables WN, Ripley BD (2002) *Modern applied statistics with S*, 4th edn. Springer, New York

Zhang B, Horvath S (2005) General framework for weighted gene coexpression analysis. *Stat Appl Genet Mol Biol* 4:17

Los Angeles, USA

Steve Horvath



<http://www.springer.com/978-1-4419-8818-8>

Weighted Network Analysis
Applications in Genomics and Systems Biology
Horvath, S.
2011, XXIII, 421 p., Hardcover
ISBN: 978-1-4419-8818-8