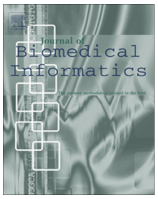
Journal of Biomedical Informatics 64 (2016) 44–54

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin



Methodological Review

Network inference from multimodal data: A review of approaches from infectious disease transmission

Bisakha Ray [a,*], Elodie Ghedin [b,d], Rumi Chunara [c,d]

[a] *Center for Health Informatics and Bioinformatics, New York University School of Medicine, USA*
[b] *Department of Biology, Center for Genomics & Systems Biology, USA*
[c] *Dept. of Computer Science and Engineering, Tandon School of Engineering, USA*
[d] *College of Global Public Health, New York University, USA*

ARTICLE INFO

Article history:
Received 28 March 2016
Revised 10 July 2016
Accepted 3 September 2016
Available online 6 September 2016

Keywords:
Network inference
Multimodal data
Bayesian inference
Infectious disease
Transmission

ABSTRACT

Networks inference problems are commonly found in multiple biomedical subfields such as genomics, metagenomics, neuroscience, and epidemiology. Networks are useful for representing a wide range of complex interactions ranging from those between molecular biomarkers, neurons, and microbial communities, to those found in human or animal populations. Recent technological advances have resulted in an increasing amount of healthcare data in multiple modalities, increasing the preponderance of network inference problems. Multi-domain data can now be used to improve the robustness and reliability of recovered networks from unimodal data. For infectious diseases in particular, there is a body of knowledge that has been focused on combining multiple pieces of linked information. Combining or analyzing disparate modalities in concert has demonstrated greater insight into disease transmission than could be obtained from any single modality in isolation. This has been particularly helpful in understanding incidence and transmission at early stages of infections that have pandemic potential. Novel pieces of linked information in the form of spatial, temporal, and other covariates including high-throughput sequence data, clinical visits, social network information, pharmaceutical prescriptions, and clinical symptoms (reported as free-text data) also encourage further investigation of these methods. The purpose of this review is to provide an in-depth analysis of multimodal infectious disease transmission network inference methods with a specific focus on Bayesian inference. We focus on analytical Bayesian inference-based methods as this enables recovering multiple parameters simultaneously, for example, not just the disease transmission network, but also parameters of epidemic dynamics. Our review studies their assumptions, key inference parameters and limitations, and ultimately provides insights about improving future network inference methods in multiple applications.

© 2016 Elsevier Inc. All rights reserved.

Contents

---

* Corresponding author at: Center for Health Informatics and Bioinformatics, New York University Langone Medical Center, 227 East 30th Street, 757 E, New York, NY 10016, USA.

E-mail address: bisakha.ray@nyumc.org (B. Ray).

http://dx.doi.org/10.1016/j.jbi.2016.09.004
1532-0464/© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Dynamical systems and their interactions are common across many areas of systems biology, neuroscience, healthcare, and medicine. Identifying these interactions is important because they can broaden our understanding of problems ranging from regulatory interactions in biomarkers, to functional connectivity in neurons, to how infectious agents transmit and cause disease in large populations. Several methods have been developed to reverse engineer or, identify cause and effect pathways of target variables in these interaction networks from observational data [1–3]. In genomics, regulatory interactions such as disease phenotype-genotype pairs can be identified by network reverse engineering [1,4]. Molecular biomarkers or key drivers identified can then be used as targets for therapeutic drugs and directly benefit patient outcomes. In microbiome studies, network inference is utilized to uncover associations amongst microbes and between microbes and ecosystems or hosts [2,5,6]. This can include insights about taxa associations, phylogeny, and evolution of ecosystems. In neuroscience, there is an effort towards recovering brain-connectivity networks from functional magnetic resonance imaging (FMRI) and calcium fluorescence time series data [3,7]. Identifying structural or functional neuronal pairs illuminates understanding of the structure of the brain, can help better understand animal and human intelligence, and inform treatment of neuronal diseases. Infectious disease transmission networks are widely studied in public health. Understanding disease transmission in large populations is an important modeling challenge because a better understanding of transmission can help predict who will be affected, and where or when they will be. Network interactions can be further refined by considering multiple circulating pathogenic strains in a population along with strain-specific interventions, such as during influenza and cold seasons. Thus, network interactions can be used to inform interventional measures in the form of antiviral drugs, vaccinations, quarantine, prophylactic drugs, and workplace or school closings to contain infections in affected areas [8–11]. Developing robust network inference methods to accurately and coherently map interactions is, therefore, fundamentally important and useful for several biomedical fields.

As summarized in Fig. 1, many methods have been used to identify pairwise interactions in genomics, neuroscience [12,13] and microbiome research [14] including correlation and information gain-based metrics for association, inverse covariance for conditional independence testing, and Granger causality for causation from temporal data. Further, multimodal data integration methods such as horizontal integration, model-based integration, kernel-based integration, and non-negative matrix factorization have been used to combine information from multiple modalities of 'omics' data such as gene expression, protein expression, somatic mutations, and DNA methylation with demographic, diagnoses, and phenotypical clinical data. Bayesian inference has been used to analyze changes in gene expression from microarray data as DNA measurements can have several unmeasured confounders and thereby incorporate noise and uncertainty [15]. Multi-modal integration can be used for classification tasks, to predict clinical phenotypes such as tumor stage or lymph node status, for clustering of patients into subgroups, and to identify important regulatory modules [16–20]. In neuroscience, not just data integration, but multimodal data fusion has been performed by various methods such as linear regression, structural equation modeling, independent component analysis, principal component analysis, and partial least squares [21]. Multiple modalities such as FMRI, electroencephalography, and diffusion tensor imaging (DTI) have been jointly analyzed to uncover more details than could be captured by a single imaging technique [21]. In metagenomics, network inference from microbial data has been performed using methods such as inverse covariance and correlation [2]. In evolutionary biology, the massive generation of molecular data has enabled Bayesian inference of phylogenetic trees using Markov Chain Monte Carlo chain (MCMC) techniques [22,23]. In infectious disease transmission network inference, Bayesian inference frameworks have been primarily used to integrate data such as dates of pathogen sample collection and symptom report date, pathogen genome sequences, and locations of patients [24–26]. This problem remains challenging as the data generative processes and scales of heterogeneous modalities may be widely different, transformations applied to separate modalities may not preserve the interactions between modalities, and separately integrated models may not capture interaction effects between modalities [27].

As evidence mounts regarding the complex combination of biological, environmental, and social factors behind disease, emphasis on the development of advanced modeling and inference methods that incorporate multimodal data into singular frameworks has increased. These methods are becoming more important to consider given that the types of healthcare data available for understanding disease pathology, evolution, and transmission are numerous and growing. For example, Internet and mobile connectivity has enabled mobile sensors, point-of-care diagnostics, web logs, and participatory social media data which can provide complementary health information to traditional sources [28,29]. In the era of precision medicine, it becomes especially important to combine clinical information with biomarker and environmental information to recover complex genotype-phenotype maps [30–33]. Infectious disease networks are one area where the need to bring together data types has long been recognized, specifically to better understand disease transmission. Data sources including high-throughput sequencing technologies have enabled genomic data to become more cost effective, offering support for studying

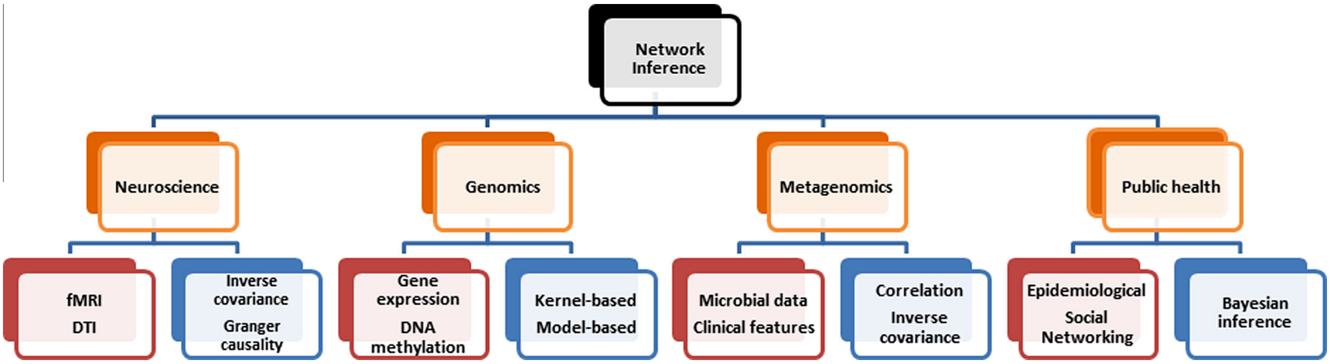


Fig. 1. Examples of multimodal network inference methods in different applications. Different modalities of data have been integrated in several applications for inferring specific networks. Most network inference methods focus on recovering network topology.

Download English Version:

https://daneshyari.com/en/article/4966957

Download Persian Version:

https://daneshyari.com/article/4966957

Daneshyari.com